universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Development of an Isobaric-Tag based Quantitative Cross-Linking Mass Spectrometry Method“

verfasst von / submitted by
Christian Stieger, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2018 / Vienna 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 862
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Chemie
Betreut von / Supervisor:	Univ.-Prof. Dr. Christopher Gerner
Mitbetreut von / Co-Supervisor:	

Acknowledgments

First of all, I want to thank Karl Mechtler for the great opportunity to work in his group at the inspiring environment of the Institute for Molecular Pathology; for the freedom of action, even if it was hard at the beginning and for his trust in me.

Furthermore, I want to express my gratitude to Prof. Christopher Gerner for arousing my interest in the field of mass spectrometry based proteomics and for supervising my thesis.

I gratefully thank Dr. Gerhard Dürnberger for providing the R scripts for normalization of my cross-link data and creating the volcano plots. Moreover, I would like to thank Dr. Elisabeth Roitinger for proofreading the thesis and all the good advice.

Special thanks go to all my colleagues and friends from the Mass Spectrometry groups of IMP and MFPL. To Gabi, Ines, Karel, Schutzi and Susanne, for maintaining the mass spectrometers and HPLC systems, but foremost for the comfortable environment and all the funny moments. Johannes for sharing his infinite wisdom on proteomics and my cross-linking colleagues Zsuzsi, Manuel and Rebecca for helpful discussion. Thanks to everyone I could not mention here, for the constructive but nonetheless relaxed working atmosphere.

Thank you, mum and dad, for supporting me, for raising me and for letting me become who I am today.

Last but not least, the most important person in my life. Thank you, Hanna, for enduring me during my desperate periods and for encouraging me all the time!

Abstract

Cross-linking mass spectrometry has become an important extension for the fields of proteomics. In comparison to classical proteomics, not only informations about the presence and quantity of proteins and peptides, but also structural insights into the organization of proteins can be obtained, making cross/linking mass spectrometry also useful for structural biologists. To gain a deeper insight into protein dynamics, there are several attempts to do quantitative XLMS. In this thesis a novel and simple quantitative cross-linking workflow employing isobaric tags for quantitation on the MS^2-level is described. In comparison to previously described quantitative methods, it allows multiplexing and does not require any new algorithms for quantitation.

For evaluation, the workflow was developed and tested on S. pyogenes Cas9 as model protein. Upon RNA binding, Cas9 undergoes a significant structural rearrangement, where both the free and the RNA bound protein is characterized by x-ray crystallography. This well characterized structural rearrangement represents an ideal biological system to study abundance changes of cross-linked peptides.

Using the distance restraints derived from quantitative cross-linking experiments, several structural models of the Cas9-riboprotein were generated. The best model reached a root mean square distance of <10 Å. These results demonstrate the developed approach to produce a meaningful outcome and highlights its potential application in future research projects.

Zusammenfassung

In den vergangenen Jahrzehnten hat sich Cross-Linking Massenspektrometrie als wichtige Ergänzung in der Proteom-Forschung etabliert. Im Vergleich zu klassischen Proteomics-Experimenten liefert diese Methode nicht nur Informationen über Präsenz und Quantität eines Proteins oder Peptids, sondern erlaubt darüberhinaus Rückschlüsse auf dessen dreidimensonale Struktur und ist damit auch ein nützliches Werkzeug für Strukturbiologen. Um tiefere Einblicke in die Dynamik von Proteinen zu erlangen, gibt es zahlreiche Ansätze um quantitative Methoden mit Cross-Linking Massenspektrometrie zu kombinieren. In dieser Arbeit wird eine neue und einfache Methode zur Quantifizierung von gecross-linkten Peptiden mit Hilfe von isobarischen Markierungsreagenzien zur Quantifizierung auf der MS^2-Ebene beschrieben. Im Vergleich zu bisher beschriebenen Methoden erlaubt sie „Multiplexing“ und benötigt keine neuen Algorithmen zur Quantifizierung.

Zur Evaluierung und Entwicklung wurde der Workflow an S. pyogenes Cas9 als Standardprotein getestet. Durch das Binden an gRNA vollzieht Cas9 eine drastische Konformationsänderung und sowohl das freie als auch das RNA-gebundene Protein ist röntgenkristallographisch charakterisiert. Diese gut charakterisierte Strukturänderung ist ein ideales Modell um die Änderung der Cross-Link Quantität auf einem biologischen System zu testen.

Mit Hilfe der Distanzinformationen, die aus den quantitativen Cross-Linking Experimenten gewonnen wurden, konnten einige Strukturen des Cas9-Riboproteins modelliert werden. Das beste Modell erreichte eine mittlere quadratische Abweichung von <10 Å im Vergleich zur Kristallstruktur. Diese Ergebnisse demonstrieren die potentiellen Anwendungs-möglichkeiten dieser Methode in zukünftigen Forschungsprojekten.

Table of Content

List of Abbreviations

PCR – Polymerase chain reaction
FAB – Fast atom bombardment
MALDI – Matrix assisted laser desorption/ionization
UV – Ultraviolet
IR - Infrared
ESI – Electrospray ionization
HPLC - High performance liquid chromatography
CE - Capillary electrophoresis
Q – Quadrupole
TOF – Time of flight
FT-ICR – Fourier transform ion cyclotron
PRM – Parallel reaction monitoring
HCD – High energy collisional dissociation
LFQ – Label free quantification
SRM – Single reaction monitoring
MRM – Multiple reaction monitoring
SILAC - Stable isotope labeling by amino acids in cell culture
XLMS – Cross-linking mass spectrometry
NMR - Nuclear magnetic resonance
SDS – Sodium dodecyl sulfate
CID - Collisional induced dissociation
ETD - Electron transfer dissociation
DSSO – Disuccinimidylsuberate sulfoxide
DSS – Disuccinimidyl suberate
BS^3 – Bis sulfosuccinimidyl suberate
NHS – N-hydroxysuccinimide
EDC – 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimid
DMTMM - 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride
DMSO - Dimethylsulfoxide
ABC-Buffer – Ammonium bicarbonate buffer
SEC – Size exclusion chromatography
SCX – Strong cation exchange
FDR – False discovery rate
RNP - Ribonucleoprotein
r.a.d. – reduced, alkylated and digested
MS – Mass spectrometry
MS/MS – Tandem mass spectrometry
PSM – Peptide spectrum match
MCX – Mixed mode cation exchange
QXLMS – Quantitative cross-linking mass spectrometry
DDA – Data dependent acquisition
NCE – Normalized collision energy
Cas9 – CRISPR-associated protein 9
crRNA – CRISPR RNA
tracrRNA – trans-activating crRNA
sgRNA – single guide RNA
PAM – Protospacer adjacent motif
FRET – Förster resonance energy transfer
gRNA – guide RNA (crRNA paired with tracrRNA)
ppm – parts per million
PD 2.2 – Proteome Discoverer 2.2
XL – Cross-link
CSM – Cross-link spectrum match
RMSD – root mean square distance
ID – Identification (cross-link)
rpm – Rotations per minute

1. Introduction

1.1 Mass Spectrometry based Proteomics

Protein biology was ever since a key area in biological research. While in the beginning mainly proteins in their entirety could be investigated, sequencing techniques as developed by Sanger or Edman, facilitated the determination of the exact amino acid sequence of a protein.[12] Anyhow, these methods are time consuming and only applicable to purified proteins. Omics- disciplines in general describe the investigation of the entirety of a certain biomolecule in a biological sample, like genomics does for genes with the four nucleobases as building blocks. Genomics owes its early breakthrough the availability of amplification methods like polymerase chain reaction (PCR) that allowed the analysis of low sample amounts. For proteins and/or peptides amplification is not possible. Therefore, highly sensitive instruments that allow high throughput analysis are required to study the proteome. Organic molecules and polymers of the size of peptides could already be analyzed with mass spectrometers in the mid-1900s. However, the analysis of biological macromolecules was impossible since the available ionization methods would have destroyed the analyte before measurement. With the development of soft ionization techniques, tandem mass spectrometers, diverse fragmentation methods and major improvements in chromatographic separation, proteomics finally became reality in the late 1990s.[3] By now, proteomics is a core discipline in biological and medicinal research. It allows both to identify and to quantify peptides and to draw conclusions about their biological role. Among others it helped to elucidate several cellular pathways, identify protein interaction partners and to discover the role of post translational modifications.[4] Nowadays proteomics is at the frontier of becoming a meaningful discipline in structural biology. Over the last decade, especially cross-linking mass spectrometry (see 1.2) evolved more and more to a real “omics” discipline, helping to solve structural questions concerning proteins, protein complexes and whole cells and organisms.[5]

1.1.1 Ionization Techniques in Proteomics

Generally, mass spectrometers analyze charged molecules in the gas phase by determining their mass to charge ratios. Therefore, the analyte needs to be ionized and transferred to the gas phase. Early known ionization techniques like electron ionization or chemical ionization require volatilization of the sample before ionization. Large biomolecules like proteins, nucleic acids and oligosaccharides will decompose rather than volatilize when heated, therefore these methods are not applicable.

The first step towards non-destructive biomolecule analysis was the development of fast atom bombardment (FAB) ionization. This involves the bombardment of an analyte dissolved in glycerol (matrix) with a high energy inert gas atom beam. FAB ionization was applied to ionize proteins and could generate mass spectra of trypsin (23kDa). Nevertheless, due to its lack of sensitivity it did not prevail.[6]

Further improvement of the sample matrix and the use of laser to desorb the molecules led to the development of matrix assisted laser desorption/ionization (MALDI). This method is based on the embedding of the analyte into an organic matrix that strongly absorbs UV- (e.g. 2,5-dihydroxybenzoic acid) or IR-light (e.g. succinic acid). This matrix and the embedded sample are irradiated with a laser beam, leading to a desorption and ionization of the analyte.[7] The main advantages of MALDI are the high sensitivity and the fact that mainly singly charged ions are formed, which simplifies spectra analysis. These strengths made MALDI the method of choice for proteomics experiments for a long time, and even nowadays it is routinely used by many laboratories.

The second ionization technique that is very important in modern proteome research is electrospray ionization (ESI). Although, it was already developed in the late 1960s for the analysis of synthetic polymers, it took till 1988 when John B. Fenn applied this method to proteins and peptides, leading to the Nobel Prize in Chemistry in 2002.[8] ESI can be easily interfaced with diverse separation techniques like high performance liquid chromatography (HPLC) or capillary electrophoresis (CE) and is therefore the method of choice for high throughput applications. In ESI a molecule dissolved in a volatile solvent is injected into the ion source through a needle, while high voltage is applied to the tip of the needle. This high voltage leads to the formation of highly charged solvent droplets containing the analyte. Depending on the applied voltage, positively or negatively charged droplets are formed. With solvent evaporation, the droplet size decreases, therefore the charge density on the droplet surface increases which subsequently leads to coulomb explosions. Finally, the charge(s) are transferred to the analyte molecule and the charged species is released in the gas phase where it is captured by the vacuum of the mass spectrometer. [9]

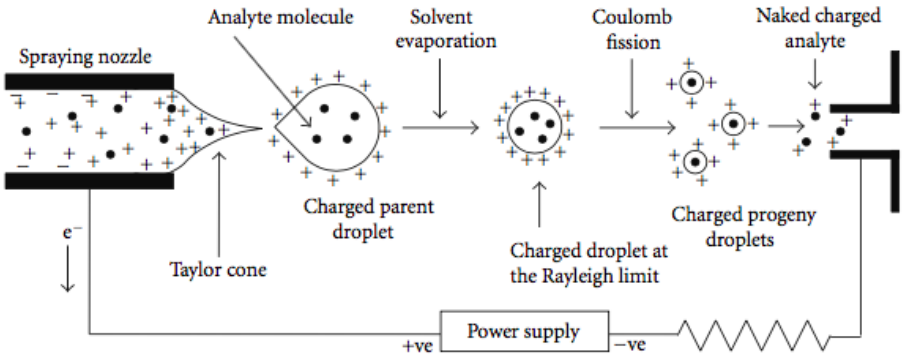


Figure 1: Schematic depiction of ESI (adapted from Banerjee et al.)[9]

Depending on size and structural properties of the molecule, several different charged species may be formed. In the case of peptides, the nature of the amino acid side chains as well as the pH of the solvent define the possible charge state.
The classical shotgun proteomics approach involves a tryptic digest of the protein sample. Trypsin selectively cleaves the amide bond C-terminally adjacent to arginine and lysine residues leading to peptides with at least two basic functional groups (N-terminal amine & C-terminal lysine or arginine side chain). If peptides are dissolved and separated under acidic conditions, the basic residues are protonated. Therefore, peptides are usually measured in positive ESI modus and are typically at least two times positively charged. Depending on the peptide size and side chain composition also higher charged species can be found.

1.1.2 Mass Analyzers Commonly Used in Proteomics

In general, there are five types of mass analyzers used in modern proteomics: quadrupole (Q), time-of-flight (TOF), ion trap, Fourier transform ion cyclotron (FT-ICR) and orbitrap. They can be either used as standalone instruments or with tandem-instrumentation, where the strengths of the individual analyzers are combined.
In the early days of proteomics, MALDI-TOF was one of the most popular strategies. In this approach, the mass of intact peptides is recorded by the mass spectrometer. After desorption and ionization, all analyte ions are accelerated to the same high kinetic energy. The ions are then separated in a drift region according to their velocity which depends on their m/z ratio. As no second mass spectrometric dimension is available, the sample needs to be of low complexity, so that the identified peptides can be matched onto one protein. Therefore, this approach was often combined with 2D-gel-electrophoreses. This method is often referred to as peptide-mass fingerprinting.
The introduction of fragmentation cells and tandem mass spectrometers made peptide and protein identification more clear-cut. After a full scan, i.e. identifying all available ions, a single ion is isolated and fragmented. This results in peptide backbone dissociation and allows amino acid sequencing of the selected peptide precursor. Due to its simplicity, high resolution, sensitivity and mass accuracy MALDI-MS/MS (Q-TOF, TOF-TOF) was and still is a popular method for peptide and protein identification.[10]
Moreover, so called “triple-quad” (QQQ) mass spectrometers, consisting of two quadrupole mass analyzers in series with a third quadrupole, that is used as a collision cell, between them, are popular in the proteomics community. QQQs are widely applied for targeted quantification.[4]
Nowadays they are often replaced by the latest generation of mass spectrometers, hybrid orbitrap mass spectrometers, allowing the latest targeted label free quantification strategy called parallel reaction monitoring (PRM).[11] The triumph of these modern machines started with the development of a high-resolution orbitrap mass analyzer by Alexander Makarov in 2000. Essentially, an orbitrap consist of two outer and one spindle-like central electrode.

This special structure allows it to act as analyzer and detector at the same time. After the ions are trapped in the Orbitrap by "electrodynamic squeezing", they harmonically oscillate around the central electrode. As every ion oscillates differently, they can be identified by measuring their frequency. Like mass spectrometers using ion-cyclotron resonance technology, orbitraps are Fourier Transformation based mass analyzers. Nevertheless, they are more compact and easier to operate and maintain.[12,13] One of the latest instrumentations employing orbitrap mass analyzers is the QExactive Orbitrap™ from Thermo Fisher Scientific. This instrument already exists in its fourth generation, where the third generation (QExactive HF) was used in this thesis and will be described below. It possesses an atmospheric pressure ion source followed by a stacked-ring radio frequency ion guide (RF-lens) focusing ions into a narrow beam. After focusing, the ion beam passes an injection flatpole and a bent flatpole with a large bore in the line of sight from the RF-lens, preventing solvent droplets and uncharged species from entering the mass analyzer. The injection flatpole of the third generation QExactive is capable to select ions and allows low-resolution removal of undesired ions prior to mass analysis. A segmented quadrupole mass filter serves as first mass spectrometric dimension. Either all ions can pass the quadrupole to generate a full scan (MS^1) or certain ions are filtered for MS^2 scans with an ion isolation width from 0.4 to 5600 Da. The quadrupole is followed by an exit lens and an octapole, transferring the ions into the C-trap. From the C-trap the ions can be ejected directly into the ultra-high-field orbitrap mass analyzer for MS^1 scans or into the higher-energy collisional dissociation (HCD) cell for fragmentation. After successful fragmentation, the total ion population is transferred via the C-trap into the orbitrap mass analyzer. This latest generation orbitrap used in the HF system allows measurements with a resolution of up to 500,000 at m/z 200 or a scan speed of up to 17 Hz.[10,14] This high resolution as well as the high scan speed allow deep coverage of the proteome, even within complex samples.

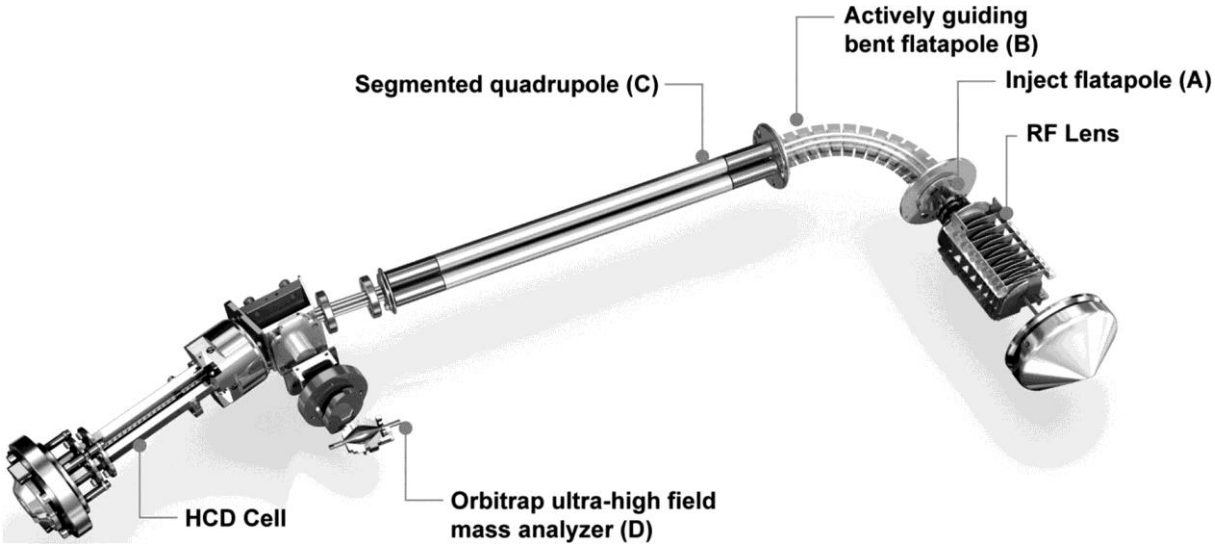


Figure 2: Schematic overview of a QExactive HF Orbitrap™
(adapted from Brochure: Q Exactive HF Orbitrap LC-MS/MS System)[101]

1.1.3 Quantitative Proteomics

In proteomics experiments not only qualitative, but especially quantitative experiments are of high importance. Quantitative information is essential, when e.g. protein expression, turnover and localization are investigated. Nowadays, there are many different methods that allow quantitation. The three main methodologies are metabolic labelling, chemical labeling and label-free quantification (LFQ). Apart from these proteome-wide quantitative methods, there are also targeted methods like single and multiple reaction monitoring (SRM and MRM) and PRM that allow the quantification of some selected peptides corresponding to the protein of interest.[15,16]

Metabolic labeling requires the introduction of defined stable-isotopes into the proteome of the investigated organism. The most popular method to do so is the "stable isotope labeling by amino acids in cell culture" (SILAC) introduced by the group of Matthias Mann in the early 2000s. In SILAC the cells are grown in a medium that contains stable isotope labeled arginine and lysine. These amino acids are subsequently incorporated into every ribosomal synthesized protein. With this method, the whole proteome can be accurately quantified and up to 3 different conditions can be compared within one experiment. SILAC however, requires immortal cell lines that can be labeled with heavy amino acids. Although recent studies showed, that also more complex organisms such as C. Elegans or mouse can be "SILAC-labeled", it remains very time consuming and expensive.[15] Moreover, SILAC and every other MS^1 based quantification strategy that involves the introduction of stable isotopes leads to an increased sample complexity, as every peptide occurs in its heavy and light version.

For chemical labeling, there are two different approaches. On the one hand, there are labels that can be used for MS^1 quantification (Dimethylation, tryptic digest in $H_2{}^{18}O$) on the other hand so called isobaric tags yielding reporter ions in the MS^2 scan are used. Chemical labeling for MS^1 quantification follows the principle of SILAC, with a few differences. In case of deuterium labeled reagents, there might be a significant shift in retention time, so that the light and heavy labeled molecules do not coelute perfectly, complicating quantification. Another drawback affecting all chemical labeling approaches, is the late stage of introduction. SILAC labeled cells can be mixed prior to cell lysis, so there is only one step, the cell culture, where a systematic error in sample preparation can occur. Chemical labeling is performed after protein digestion, leaving more space for potential errors.[15]

Compared to other quantitation methods, isobaric tags offer the great benefit of multiplexing, i.e. they allow comparison of up to ten different conditions at in one measurement. To date, there are two widely applied isobaric tags called iTRAQ (AB Sciex) and TMT (Thermo Fisher Scientific) allowing up to 8-plex and 11-plex comparison, respectively. These tags are supplied with reactive groups towards sulfhydryl groups (cysteine), aldehydes (per-iodate treated glycans) and amines (lysine and peptide N-terminus). The tags contain a labile bond that may be easily cleaved during fragmentation,

leading to the corresponding reporter ions. These reporter ions are rather small, ranging from 114-121 Da (iTRAQ) and 126-131 Da (TMT), so they do not interfere with other ions. Unfortunately, these quantification methods also involve some drawbacks. The first problem is the tag itself. As they are usually quite big and hydrophobic, the peptides shift towards higher retention times on reversed phase HPLC. This issue however can be compensated with a higher percentage of organic solvent within the HPLC gradient or increased temperature during elution. Apart from the afore mentioned shift in retention time, Pichler et al. reported that larger isobaric tags lead to a decreased number of identified peptides.[18] The second problem arises during quantification. Coeluting peptides with a similar mass as the selected precursor may be co-isolated for fragmentation hence falsifying the reporter ion ratio. This phenomenon is also referred to as ion suppression and may be overcome by quantification in MS^3, ideally with high resolution. This requires expensive equipment, moreover MS^3 in general elongates the duty cycle resulting in a lower number of recorded spectra and PSMs. Also, compared to MS^2 it lacks sensitivity.[19]

In LFQ, depending on the algorithm either every detected feature (e.g. every m/z species) or only those corresponding to a previously identified peptide are quantified. The quantitative information is summed up and the final peptide or protein abundance is compared between the replicates. LFQ utilizes the MS^1-level for quantification, is practically very simple, since special sample modification is necessary, but it requires complex algorithms for data analysis. The successful quantification depends on the consistent performance of the chromatographic separation system, the ion source and the mass spectrometer. Moreover, it is hard to combine with prefractionation techniques.[15,17]

Large scale quantitative workflows often identify candidate proteins that seem to be more abundant in one condition, but often the real ratio differs significantly from the reported result. To verify the quantitative outcomes, targeted strategies are commonly used. While in the past SRM or MRM using a triple-quadrupole instrument were the methods of choice, nowadays orbitrap instruments and PRM allow to quantify more transitions at once and become more and more popular.[11]

Taken together, each quantification method has its strengths and weaknesses. Depending on the problem and the available equipment, the ideal method needs to be selected carefully.

1.2 Cross-Linking Mass Spectrometry (XLMS)

Mass spectrometry has been used for decades to identify proteins and peptides by sequencing their primary structure. Unfortunately, not only the amino acid sequence of a protein is important for its function, but also its three-dimensional structure. In the past the 3D structure of proteins was solved using X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy. However, these techniques often fail for several reasons. X-ray structures can only be obtained if the protein of interest can be crystallized, which may be difficult for dynamic, disordered and extremely large proteins. NMR in contrast can be used to study protein dynamics, but is limited to rather small proteins. Over the last years, mass spectrometry became more and more popular to overcome the limitations of classical techniques in structural biology and is now on the edge of becoming an tool for whole cell protein structure/assembly elucidation.[20]

In XLMS a chemical reagent, consisting of two reactive moieties and a linker chain, is used two connect two amino acids within a native protein or protein complex. Depending on the cross-linker functionality, it can selectively connect lysines, cysteines and acid side chains. Additionally, photochemical cross-linking reagents are available, that allow unselective cross-linking with every amino acid side chain. After protein digestion, apart from normal peptides also cross-linked peptides are formed and can be identified by mass spectrometry. As the cross-linker and the two connected amino acid side chains have a defined length, it is possible to define the maximum between the residues to allow the cross-linking reaction to take place. This information about the local proximity of two amino acids can be used to computationally model the structure of a single protein or even large complexes. Nowadays XLMS is often used in combination with other low-resolution methods like small angle X-ray scattering or ion-mobility MS to solve the three-dimensional orientation and structure of complex biomolecules.[21,22]

1.2.1 History of Cross-Linking (Mass Spectrometry)

Chemical protein cross-linking was already used before proteomics had its great breakthrough. In the beginning, bifunctional aromatic fluorine-compounds like 1,5-difluoro-2,4-dinitrobenzene were used as so-called “bridging-reagent” to study the general assembly and properties of protein fibers and membrane proteins. This reaction is based on a nucleophilic aromatic substitution, where the two fluorine atoms are displaced with nucleophilic groups like the ε-amino groups of two lysines. Although the principle was already the same as in modern protein-crosslinking, due to a lack of high performance and throughput analysis methods, the conclusions drawn from these experiments were rather plain from a nowadays point of view.[23,24] In the 1970, crosslinking in combination with SDS-PAGE was used to identify the ribosomal protein-protein contacts using a homobifunctional, amine reactive crosslinker: dimethyl suberimidate.[25]

From the moment mass spectrometry started to be widely applied for protein identification and peptide sequencing, it was combined with chemical cross-linking. Since a cross-linked sample is way more complex than a "normal" tryptic digest of a protein, analysis was way more demanding and error-prone. However, in 2000 chemical cross-linking in combination with MALDI-TOF-MS and peptide-mass-fingerprinting was successfully applied to propose a topological model for the Nup84p complex.[26] In the same year it was shown, that the combination of HPLC separation with MALDI-TOF measurement and post-source decay can increase the number of assigned cross-links as well as the confidence, due to additional sequence information of the peptide. These approaches were the first to prove that XLMS can be used to gain low resolution three dimensional structures of proteins.[27]

Until the late 2000s, most studies using XLMS determined few (10-50) cross-links. With the improvements in shotgun mass spectrometry, it became possible to obtain more cross-links and therefore more structural information from one experiment. In 2010, Chen et al. used cross-linking in combination with mass spectrometry to analyze the 12-subunit 513kDa RNA Polymerase II. Back then, the X-ray structure of this large complex was already solved, so it was possible to evaluate the identified cross-links and to somehow conduct a quality control. This study revealed 106 XLs in a single LC-MS/MS experiment, of which 99 matched the crystal structure and the other 7 could be explained with protein flexibility, as they were within regions with high B-factors*.[28] This publication served as a benchmark study and paved the way to modern cross-linking mass spectrometry. Until now, cross-linking mass spectrometry helped to at least partially solve the structure of large biomolecular complexes like the proteasome, the ribosome and the cohesion loader Scc2.[29–31]

As a logical next step, the idea of expanding cross-linking to whole cells or cell compartments came up. Since this goal is magnitudes more complex than prior experiments, specially designed cross-linker were needed to facilitate data analysis. (see 1.2.6) These special reagents bear a MS (mostly collisional induced dissociation [CID] or electron transfer dissociation [ETD])) cleavable chemical group like sulfoxides or aspartate-proline bonds. [32,33] To allow selective cleavage of the cross-linker without fragmenting the whole peptide special chemical moieties are needed. The group of Lan Hunag developed several cross-linker containing the CID cleavable sulfoxide group. One of the most prominent candidates, DSSO (disuccinimidyl sulfoxide), was recently used to investigate the whole interactome of E. Coli lysate, HeLa cells and purified mouse heart mitochondria.[20,34,35] These studies impressively demonstrate the applicability of XLMS to investigate the three-dimensional localization, organization and structure of hundreds of proteins within whole cells and organelles, respectively. The only flaw is the lack of identified cross-links between low abundant proteins.[34] Nevertheless, with further improvements in MS instrumentation and the development of selectively enrichable cross-linking reagents also these borders will be crossed soon.

*B-factor: In case of protein structures a measure of local mobility within the macromolecule chain.

1.2.2 Cross-Linking Chemistry and Cross-linker Chemistry

Over the time, several cross-linker with different reactive groups, spacer length and additional features have been developed. The two amine-reactive crosslinker DSS and BS^3 are currently most widely applied within the cross-linking community. They only differ by a chemical modification of the reactive group. While DSS possesses two N-hydroxysuccinimide esters at each end of the suberic acid spacer (11.4 Å), BS^3 contains two N-hydroxy*sulfo*succinimide ester moieties. Both active-ester moieties react preferred with the primary amine residues of the protein N-terminus and the lysine side chains at pH 7-9. However, the two sulfonic acid groups change the hydrophobicity of the crosslinker, so that BS^3 primarily cross-links easily surface accessible lysines, while DSS is also able to react with lysines that are buried in hydrophobic regions. Moreover, DSS is cell membrane permeable whereas BS^3 can be used to cross-link uniquely cell surface proteins.[36,37] In addition to amines, also reactivity towards hydroxyl groups (Ser, Thr, Tyr) especially under slightly basic conditions was observed. Apart from nucleophiles in general, carboxyl- and thiol-residues can be selectively targeted.[37] To obtain reactivity towards cysteins, mostly haloacetyl- or maleimide-groups are used. Since cysteines are usually low abundant (5 times lower abundance compared to lysine) and most residues are involved in disulfide bonds, mostly heterofunctional thiol reactive cross-linking reagents (e.g. maleimide-NHS) are used. On the other hand, the low abundance of free cysteines offers an advantage: Site selective incorporation of a cysteine at an expected binding interface, and the use of such a heterofunctional cross-linker can help to proof hypothesis, and simplifies analysis.[38]
Targeting carboxylic acids is more challenging. The earliest approach was the use of activating agents like EDC, that has been used for antibody conjugate preparation. EDC forms an active ester with aspartate and glutamate side chains, that may easily react the terminal amines at lysines if they are close enough. As the optimal operation range of EDC is at pH 4-6 where the tertiary and quaternary structure of protein(complexes) may be perturbated, DMTMM which can be used at neutral pH is becoming more popular.[39,5] Since the formed active ester react traceless with lysines, EDC and DMTMM are termed “zero length” cross-linker, because they do not introduce any spacer. Therefore, the theoretical cross-linking distance is not 30 Å as for lysine-lysine DSS cross-links, but 15 Å.

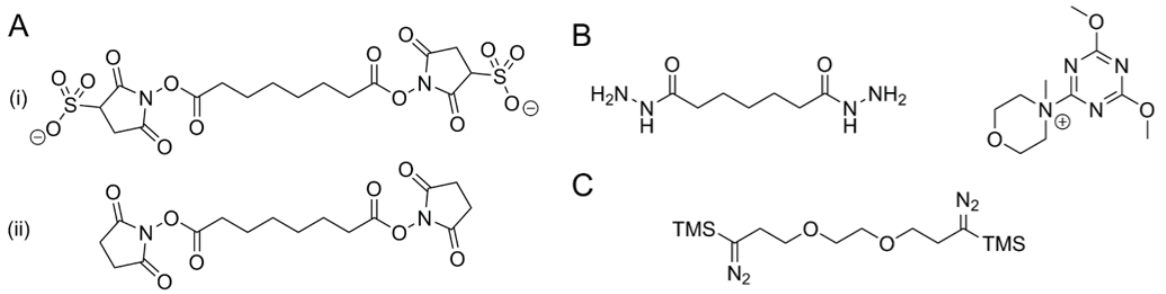


Figure 3: A - The two popular, amine reactive cross-linker BS^3 (i) and DSS (ii). B – The carboxylic acid targeting cross-linker PDH and the activating reagent DMTMM. C – Latest generation of acid reactive cross-linking reagents “Diazoker 1”.

Furthermore, it is possible to generate carboxyl-carboxyl cross-links. For this purpose, also activating reagents like EDC or DMTMM are used, but a cross-linker with two hydrazide groups (e.g. pimelic acid dihydrazide - PDH) is added. At a pH below 8.0, the hydrazide groups behave more reactive than primary amines, and are more likely to react with the active ester. Additionally to the acid-acid cross-links also zero-length cross-links are formed, which depending on the sample complexity and amount may be beneficial, as it results in a higher cross-link density.[39] Recently, a group of exclusively acid reactive cross-linker, that do not require an activating agent was described that uses trimethylsilyldiazo-groups to target carboxylic acids. Anyhow, to date, this reagent group has only been tested on a couple of standard proteins .[40]

Sometimes, the density of reactive amino acids within a protein, protein domain or protein-protein interface may be too low to cover it. For this purpose, photo-reactive cross-linking reagents have been developed, that react unspecifically with all amino acids. Mostly heterofunctional linker with an NHS- or maleimide functionalization on the one end and a diazirine or benzophenone on the other end are used. These photoreactive groups may also be introduced as non-canonical amino acids into recombinant expressed proteins. Upon UV irradiation, the photoreactive groups form a reactive carbene species that reacts with the next available amino acid. The highly heterogeneous mixture of cross-linked peptides with several different cross-linking sites that may coelute complicates subsequent data analysis. Recently a new feature of diazirine based cross-linking reagents was discovered. Apart from the carbene formation, the diazirine may also isomerize to a linear diazo group, that favorably reacts with carboxylic acids.[41,42]

Cross-linker diversity is also increased by the introduction of chemical groups within the spacer. Apart from chemical moieties that allow selective enrichment (see 1.2.3), several reagents carrying a chemical group or bond labile to MS fragmentation. The best established and most often used method for cross-linker cleavage is collisional induced dissociation (CID/HCD). Popular MS-cleavable groups are sulfoxides and urea structures and the aspartate-proline bond. Although all three are CID/HCD cleavable, their fragmentation behavior differs dramatically. Sulfoxides and D-P bonds are by magnitudes more labile compared to the peptide backbone, therefore fragmentation providing predominantly ions corresponding to the two peptides with one half of the linker. As the cleavage of the linker is asymmetrically, the resulting product ions occur as doublets of the peptide with the heavy part and the light part of the cleaved cross-linker. To further analyze the cross-link either the two cleavage products can be submitted to another MS level for separate sequencing, or a subsequent MS^2 scan with a complementary fragmentation method (e.g. ETD) is triggered.[20,34] The group of Jim Bruce developed a special case of these CID-cleavable cross-linking reagent called PIR-linker. Instead of one MS-labile group, this group of reagents bears two identical cleavable moieties. During fragmentation both bonds are cleaved simultaneously, yielding the ions of peptide A and peptide B plus the mass of the cross-linker stamp and a cross-link reporter ion with a defined mass. The ReACT algorithm calculates on the fly, weather the MS^2 is derived from a cross-link precursor and

which two ions correspond to the two distinct peptides. These two ions are subsequently sequenced in MS^3 like the afore mentioned doublets.[32]

In contrast, urea based, cleavable cross-linker like DSBU are designed to be cleaved together with the peptide backbone in one single MS^2 experiment. The urea group can be seen as a di-amide of carbonic acid therefore the bondage is roughly as labile as the amide bonds in the peptide backbone. Since this type of linker cleaves asymmetrically as well, characteristic doublet ions corresponding to the two cross-linked peptides are formed along with y- and b-ions in the same scan.

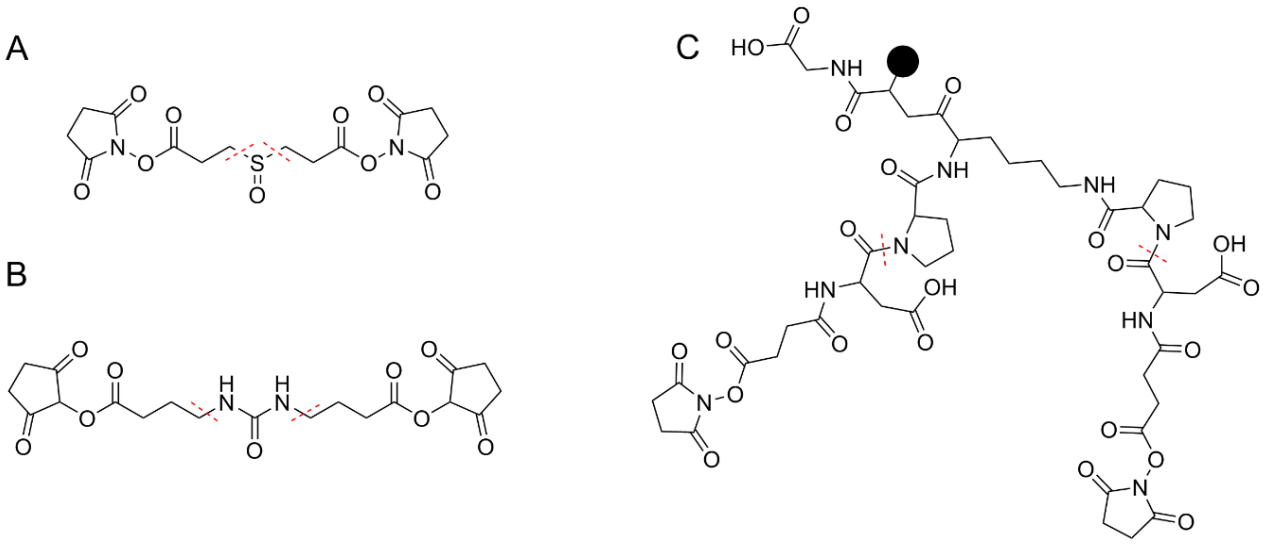


Figure 4: The commercially available cleavable cross-linker DSSO **(A)** and DSBU **(B)** and the peptide based cleavable and enrichable PIR-cross-linker BDP-NHP **(C)**. Cleavable bonds are marked with dashed red lines. The black dot represents a biotin functionality for affinity purification

Experiments showed that cleavable cross-linker tend to form less cross-links in contrast to their non-cleavable counterparts, However, the characteristic doublet ions and separate peptide sequencing in MS^3 facilitate subsequent data analysis and allow more confident identification of cross-linked peptides. (see 1.2.5)

The cross-linking reaction itself is typically carried out at close physiological pH in a buffer that is able to stabilize the protein(s) or the complex of interest. Optimal yield could be observed at protein concentrations above 1 mg/ml with a 6 to 100-fold excess of crosslinker. With a few exceptions (e.g. BS^3) crosslinker are not water soluble and therefore an adequate stock of the crosslinker dissolved in an anhydrous organic solvent (e.g. DMSO, DMF,…) is prepared. The crosslinking stock is added to the protein solution in a way, that the total organic solvent amount does not exceed 10 vol%, as more could affect the tertiary and quaternary structure of the analyte.[43] Highest cross-linking reaction yields are usually observed between room temperature and 37°C, but depending on the protein the reaction can also be carried out on ice or at 4°C in the fridge.[39,44] After a reaction time of some

minutes up to 2 hours the reaction is quenched by adding a primary amine (e.g. Tris) or ammonium (e.g. ABC-buffer) containing buffer. Quenching the cross-linking reaction is necessary to prevent the protein from over cross-linking, altering the three-dimensional structure. Afterwards, the cross-linked protein(s) are reduced, alkylated and digested using standard protocols, established for non-cross-linked proteins.[44,45]

1.2.3 Enrichment of Cross-Linked Peptides

As afore-mentioned, technical improvements in mass spectrometry and chromatography made it possible to analyze complex protein mixtures. Anyhow it remains challenging to investigate low abundant and often heterogenic protein features such as crosslinks with deep coverage. Treating a protein with a chemical cross-linking reagent enormously increases its complexity. Not only "real" crosslinks, also known as crosslinks type II, linking two independent peptides are formed, but also crosslinks of type I bridging two lysines without a cleavage site in-between or type 0 cross-links only modifying one lysine while the second NHS moiety is simply hydrolyzed are formed.[46] To successfully identify a higher number of cross-links and therefore gain more information about a protein(-complex) several methods to enrich type II crosslinks have been developed. At this point it needs to be mentioned, that modern instruments like the latest generation QExactive Orbitrap™ mass spectrometers do not require enrichment of cross-links, as long as the protein(complex) does not exceeds the size of 100 kDa, as suggested by Iacobucci et al. [47]
Cross-linked peptides differ from their linear counterparts mainly by two properties. First of all, a cross-linked peptide corresponds to two linear peptides and a linker therefore, they are in general larger and heavier. In the case of amine reactive cross-linker in combination with tryptic digest additional "missed cleavages" occur due to blocked lysine residues, leading to even longer peptides. Secondly, cross-linked peptides are more likely to carry higher charges in solution under acidic conditions compared to linear peptides. Especially in the case of a tryptic digest, as tryptic, cross-linked peptides have two N-terminal amines and two positively charged amino acid side chains on the C-terminus.[48]
To enrich cross-links according to these properties, there are two chromatographic methods that are widely applied: size exclusion chromatography (SEC) and strong cation exchange chromatography (SCX). SEC, also known as gel-filtration, is a well-known method to separate proteins and polymers based on their size and is routinely used to purify overexpressed proteins. In SEC, the column is filled with a porous resin. Depending on the size, more precisely the Stokes radius, the analyte molecule interacts differently with the stationary phase. This interaction is based on diffusion into the porous material. The larger the analyte molecule, the earlier it elutes. Single amino acids and buffer ingredients are retarded the most and large molecules (proteins, DNA/RNA oligos) elute first. Varying the pore size of the resin, columns can be designed for a certain size range. Special columns for peptides can be used to separate linear peptides form their cross-linked counterpart. [48]

Strong cation exchange chromatography is a well-known method for peptide fractionation upon MS-analysis. In SCX the resin is functionalized with negatively charged groups, often sulfonic acids, that interact with the loaded peptides. To ensure sufficient binding, the loaded peptides need to be highly acidic, so all basic amino acid residues are charged. The bound peptides are eluted using a salt gradient (e.g. 0mM-1M KCl), where higher charged peptides elute later, as they are bound more tightly. Cross-links are accumulated in the later fractions, as they bear additional charges compared to linear peptides. Using classical SCX-fractionation, cross-linked peptides are enriched in the later fractions, and the earlier fractions can be discarded. In addition to 1D-SCX, charge-based fractional diagonal chromatography (ChaFRADIC), a two-dimensional SCX enrichment strategy, has been developed. [49] In the first dimension, highly charged species are isolated from a Lys-C digest. They are subsequently digested with trypsin and the final digest is again subjected to SCX enrichment/fractionation. However, this approach is very time consuming and as every chromatographic dimension goes along with sample loss, this approach is only recommended if large amounts of protein are available. New protocols using MCX cartridges (combining the properties of a C-18 reversed phase and cation exchange stationary phase) , allow an efficient single-step enrichment for cross-linked peptides and therefore reduce the needed measurement time on the mass spectrometer. [45]
Beside of these general enrichment methods based on physico-chemical properties, also special chemical cross-linking reagents have been designed to allow the selective enrichment of cross-links. Most prominent, there are azide or alkine functionalized cross-linker that enable selective coupling to functionalized biotin molecules and further allow affinity enrichment using streptavidin beads.[50,51] Moreover, large cross-linker, carrying a biotin group have been used in several studies. These reagents have the benefit, that they do not need an additional click-reaction to introduce the biotin group. Though, the hydrophobicity and size of the biotin group as well as the general size of these reagents may influence the protein tertiary structure. [35,52]

1.2.4 Quantitative Cross-Linking Mass Spectrometry

Protein dynamics play a significant role in many biological functions. Therefore, it is a major goal of every structural biology related discipline to capture these dynamics. Quantitative cross-linking mass spectrometry (QXLMS) is a method, that may be used to monitor dynamic processes. Like in common proteomics studies, the use of stable isotopes is quite popular for quantification. An often-applied strategy is the use of differentially labeled cross-linking reagents like d_0/d_4-BS^3. In this case the two conditions that should be compared are reacted with either the heavy or light cross-linker and mixed equivalently. After reduction, alkylation, digestion and MS-analysis, the identified cross-linked peptides can be quantified at the MS^1-level by comparing the abundance of the heavy and light peaks. While at the beginning the quantification required manual integration of the MS^1 signal, software developed for quantitative proteomics experiments like MaxQuant or

pQuant have been adopted for cross-link quantification.[5,53] Moreover, the xTract software was specially designed for XL quantification and is also capable of LFQ without differentially labeled cross-linker. However, it can only be used in combination with xQuest/xProphet and requires a Linux interface.[54]

Recently a proof of principle study combining TMT isobaric labels and the MS-cleavable cross-linker DSSO was published. However, this strategy has not been applied to any biological question yet.[55]

An impressive example of a quantitative cross-linking experiment using metabolic labeling was published by Chavez et al. in 2015. In their study multidrug-resistant and drug-sensitive HeLa cells were compared using SILAC and the biotin tagged CID-cleavable PIR cross-linker "biotin aspartate proline N-hydroxyphthalimide" (BDP-NHP). The proteome-wide cross-link analysis revealed two proteins (histone H3 and TOP2A) with the same expression levels in heavy and light cells, but different abundances of certain cross-links. In the case of TOP2A, biochemical tests showed a correlation between decreased enzyme activity and the abundance of the XL involving lysine residues 489 and 798. This experiment impressively demonstrated, that a proteome-wide investigation of protein-protein interactions and protein structures is possible and may contribute to the understanding of biological processes.[56] To validate these results, the candidate cross-links were quantified once more using a PRM-method. To facilitate subsequent data analysis, a new software, XLink DB 2.0, was developed. It is a free online tool, that calculates all possible transitions for cross-linked peptides, independent of which cross-linking reagent has been used. This transition list can be easily imported into Skyline and used for targeted quantification.[57]

1.2.5 Data Acquisition in Cross-Linking Mass Spectrometry

Apart from physical enrichment methods, there is also a possibility to selectively acquire cross-link spectra. As already mentioned XL peptides are larger and more prone to be highly charged in solution. This also applies in the gas-phase. While the majority of linear peptides is two times positively charged in a common MS experiment, cross-linked peptides do not occur with a charge state below three. By excluding masses below 1300 Da and charge state 2+ in the DDA parameters, 59% of the linear peptides can be excluded from the measurement while only 2% of the cross-linked ones are not recorded. [58] For more complex samples, it may also be beneficial to exclude the 3+ charge state as this reduces data density.

Another aspect that should be considered is the lower abundance of cross-linked peptides. As a consequence, longer injection times may be considered as beneficial, although this elongates the duty cycle and leads to a lower number of acquired spectra. Apart from that, also the collision energy should be optimized for each cross-linker and instrument type. Kolbowski et al. optimized the fragmentation energy for the non-cleavable DSS/BS^3 cross-links on an Orbitrap Fusion™ Tribrid™ instrument (Thermo Fisher Scientific) and came up with a charge and m/z dependent decision tree. According to that, for most of the cross-

linked peptides HCD fragmentation using lower collision energies of NCE 26-30% works best. Only for ions with a high charge density (z = 5+, 6+ for m/z < 800 and z = 7+ for all m/z) electron transfer dissociation with supplementary HCD activation (EThcD) outperformed HCD. They also concluded, that this data dependent decision tree only offers a real benefit for more complex samples. Therefore, on common tandem mass spectrometers lower energy HCD should be the method of choice (for non-cleavable cross-linking reagents).[44]
Furthermore, the resolution at which the cross-linked samples are recorded is important. Usually, both the full scans and the MS/MS spectra are recorded at high resolution, as this is beneficial for subsequent data analysis.[34,35,44,48]

1.2.6 Identification of Cross-Linked Peptides and Data Analysis

The development of software capable to identify cross-links from large data sets was one of the biggest obstacles in the field. In proteomics, the recorded MS/MS spectra are usually searched against a theoretical spectra library of an in silico digested database. As a cross-link usually contains two different peptides, the search space increases exponentially compared to a linear increase in classical proteomics. This so called n^2 problem can be handled quite well for small databases (e.g. single proteins or small protein complexes), but impedes the analysis of large (e.g. whole proteome databases). [5] As it was pointed out by Yang et al., the search space for a cross-linked E. coli lysate is four magnitudes larger than a normal database search for linear peptides using the human proteome.[59] Moreover, a correct FDR estimation becomes difficult, as there are three possibilities for false positives (True-False, False-True and False-False). To tackle the afore mentioned challenges, until today over thirty algorithms for cross-link identification have been developed.[20,32,34,46,59–66] Most of the programs sink into oblivion after their development. However, there are some remarkable examples with a good reputation, that will be discussed below.
One of the most often used software is the Linux based xQuest/xProphet package. The standalone version of xQuest was already published in 2008 and expanded in 2012 by xProphet, a tool for an automated target decoy based FDR estimation.[65,67] If classical non-cleavable cross-linker are used, the software performs an exhaustive search, searching for all possible peptide-peptide combinations. In this modus, it is possible to use a database size of up to 100 proteins. To use larger databases, a mixture of differentially labeled cross-linker (e.g. DSS d0/d12) needs to be used, where the search space is reduced by searching for doublets in the MS^1 scan. In 2015 another extension called xTract was published.[54] This package supports both label-free and stable-isotope-based quantification of cross-linked peptides. It is utilized by several groups with both mass spectrometric and biological expertise and helped to answer several biological questions.[68–70] A major drawback however, is its Linux interface, complicating analysis for Apple and Windows user.
Another very popular software used by many groups is pLink. It supports every possible non-cleavable cross-linker, additionally the latest version can analyze DSSO-cross-linked when acquired as CID-ETD (MS^2-MS^2) experiment. To avoid the n^2 problem pLink treats

cross-links as linear peptides with an unknown modification. It identifies individual peptides and afterwards combines them to identify cross-links. pLink includes an algorithm for FDR estimation and according to the developer it can be used for databases as big as the human proteome.[59] The original program was a command line operated software, but in early 2018 pLink 2.0, a more user-friendly version including a GUI, was officially launched. The latest version already includes several stable-isotope labeling based quantification methods and it is planned to implement it into the pFind Studio pipeline. Just like xQuest it is used by many research groups and contributed to several recent publications. [71,72] Another popular software that uses a similar approach is Kojak. For FDR estimation however, it uses the Percolator algorithm that is widely applied in classical proteomics.[73–75]

At this point also StavroX needs to be mentioned. To minimize search space, the algorithm calculates all theoretically possible cross-link precursor masses and compares them to the precursor masses extracted from the input file. The positive hits are treated as potential candidates and the fragment ions of the MS^2 scan are compared to the theoretical spectrum calculated by StavroX. To estimate the FDR it uses a target/decoy approach. For decoy database generation it shuffles the assigned FASTA and finally reports a score cut-off corresponding to the respective FDR. [61] StavroX was used in several studies investigating single proteins or protein complexes.[76–78] Nevertheless, it is not suited for searching against large databases.

Recently, Gossenreiter et al. compared the four afore mentioned software packages for the number of IDs and the "real" FDR. Therefore, they cross-linked BSA with BS3 (d0/d4), analyzed it on a QExactive HF and searched the acquired data for the light cross-linker only, against a small contaminant database containing BSA, applying a target FDR of 5% on the spectrum level. The results were validated via manual inspection of the spectra. Cross-links were considered as true, if the heavy cross-linker isotope pattern was present in the full scan. Kojak yielded the highest number of identified cross-links (63) followed by pLink (37), xQuest (29) and StavroX (22). In terms of the observed FDR, Kojak (11.1%) and xQuest (6.3%) show higher values than specified. On the other side pLink (2.1%) and StavroX (0.9%) are significantly below the target value. These findings suggest, that StavroX has the most stringent FDR calculation and therefore misses many true positive hits. In contrast to that Kojak identifies way more cross-links but at a more than ten times higher FDR. Moreover, in this study the score distribution of false positive hits was investigated and proposed score cut-off for each of the software to minimize wrong identifications.[79]

Another method to facilitate cross-link identification is the use of cross-link reporter ions. Classical non-cleavable cross-linking regents like DSS were shown to produce ions corresponding to two lysines plus the mass of the cross-linker when fragmented with collision induced dissociation techniques.[63] SIM-XL is a software package that makes use of these characteristic ions, to minimize the number of spectra that need to be analyzed and to increase the identification reliability.[80] These reporter ions are more likely to be formed at higher collision energies. On the contrary, Kolbowski et al. have demonstrated a rather

gentile collision energy of 28% NCE to lead to the highest sequence coverage of cross-linked peptides and therefore the best identification rate.[44] That is one reason, why this type of cross-link reporter ions never prevailed.

A more promising reporter peak based approach makes use of MS-cleavable cross-linking reagents. During fragmentation the linker is cleaved, yielding ions corresponding to the respective peptides carrying a modification. As most of the cleavable cross-linking reagents have an asymmetric cleavage site, ion doublets with a characteristic Δ-mass emerge. Since each doublet corresponds to one of the cross-linked peptide, the number of possible candidate peptides can be significantly reduced. The spectrum containing a doublet is subsequently searched for fragments that could arise from a theoretical candidate and the best scored peptides are assigned. Depending on the cross-linker and the fragmentation method, either backbone fragmentation, cross-linker cleavage or a combination of both occurs. In the case, that predominantly the cross-link fragments are formed, another MS level is needed to sequence the two peptides. This can be achieved either by a second subsequent MS^2 scan using a complementary fragmentation method[20] or by subjecting the doublet ions to MS^3 for separate sequencing.[32,34] Fragmentation and subsequent MS^3-analysis of the cleaved peptides allows database search and FDR calculation like in standard proteomics experiments. In comparison to MS^2-based methods, MS^3 suffers from longer duty cycles and lower sensitivity.[5] To date there are only a few examples for search engines specifically designed for the analysis of cleavable cross-links. The most prominent examples are MeroX and XlinkX.[34,62] While MeroX can only analyze normal MS^2-experiments or previously merged spectra, XLinkX was designed to analyze both sequential MS^2 (e.g. CID-ETD) and MS^n (e.g. MS^2-MS^3) based experiments. Although, the characteristic doublet ions foster unambiguous cross-link identification, internal evaluation showed that the target FDR may still be exceeded in some cases.

In conclusio, there are dozens of software packages for cross-link identification. Each of them has a different strategy to deal with the problems arising in XLMS. Nevertheless, none of them is perfect, and all results have to be treated carefully. It is good practice to inspect all cross-link spectrum matches manually and to discard spectra that are not fulfilling certain quality criteria. Iacobucci et al. suggested five guidelines to avoid misassignments in cross-linking mass spectrometry.[81] This however is very time-consuming, and often score cut-offs may be more practical.

To bring cross-linking mass spectrometry to maturity, a general agreement on certain standards in sample preparation, data acquisition and data analysis are needed and will for sure help to increase the popularity of this application.

1.3 CRISPR-Associated Protein 9 (Cas9) from S. Pyogenes

Some sort of CRISPR-system is present in almost every known bacterium and archaea. These systems are part of the adaptive immune system and aim to protect the organism from invading nucleic acids (e.g. virus/phage DNA or RNA). There are several types of CRISPR systems that are divided into three major types. Type I systems use large multi-protein complexes to target the invading nucleic acids. In comparison to that, type II and III systems are more compact single protein enzymes.[82]

Probably best-known is CRISPR/Cas9 from S. pyogenes, a type II system that was heavily investigated by the groups of Emmanuelle Charpentier and Jennifer Dounda. Compared to the cascade (type I) and Csm/Cmr (type III) based CRISPR/Cas systems, Cas9 requires two RNA molecules, called CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA), to bind the target nucleic acid sequence. The crRNA can be divided into two segments. An approximately twenty nucleotides long sequence complementary to the target, also known as protospacer. Adjacent to that, the short palindromic repeat region that is used to interact with the tracrRNA is found. The tracrRNA has two crucial functions. On the one hand it is required for crRNA maturation, via recruiting RNAse III to the pre-crRNA. On the other hand, its secondary structure is necessary for correct complex formation with the Cas9 nuclease.

A major contribution to the breakthrough of CRISPR/Cas9 as the number one genome editing system was the development of a single chimeric RNA molecule that combines the features of both cr- an tracrRNA. This molecule was designed to have the target recognition sequence at the 5'-terminus followed by a long hairpin structure forming sequence, mimicking the tracrRNA. In biochemical tests the chimeric RNA (single guide RNA – sgRNA) showed the same target specificity and activity like Cas9 complexes formed with a cr/tracrRNA duplex.[83] For successful target binding, type II systems additionally require a so-called PAM region (protospacer adjacent motif) right after the targeted sequence. In case of the S. pyogenes Cas9 this sequence is 5'-NGG-3', where N is a random nucleotide.[82] Biochemical testing revealed that Cas9 has two different nuclease domains, each of them cleaving one strand of the target DNA. As these two domains are homologous to the readily described HNH and RuvC endonucleases, they were called HNH(-like)-domain and RuvC(-like)-domain, respectively. To identify the target/off-target strand specificity, inactivating point mutations were introduced within these domains. Mutation of D10 to alanine proofed to inactivate the RuvC domain. The obtained cleavage products of the single mutated nucleases indicate that the HNH domain cleaves the strand complementary to the crRNA, while the off-target strand is cleaved by the RuvC domain.[83] Moreover, these two mutations proved to be helpful in solving holoenzyme crystal structure, CRISPRi and for epigenetic applications.[84–86]

In 2014 Jinek et al. and Nishimasu et al. separately published crystal structures of the S. pyogenes Cas9 ternary complex with a sgRNA and target DNA. Jinek et al. additionally solved the structure of Cas9 in complex with sgRNA only and without any bound nucleic

acids (apo Cas9). These studies revealed a bilobed structure of Cas9. On the one side there is the recognition-lobe, that was later divided into the Rec-1, -2 & -3 domains, responsible for guide RNA binding and target recognition. On the other side there is the nuclease-lobe, responsible for PAM recognition and DNA cleavage. This lobe was divided into the already described nuclease domains HNH and RuvC and the PAM-interacting domain (PI-domain).

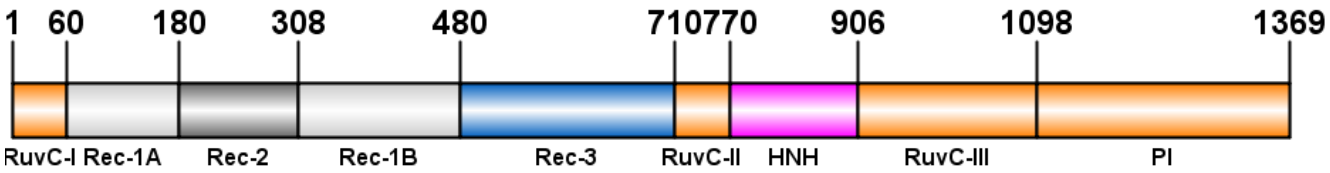


Figure 5: Illustration of S. pyogenes Cas9 domains.
The same color scheme was used for three dimensional structures in part 3 & 4

As no further direct contacts between the two lobes were found in the structures of the holo-complex, Nishimasu et al. suggested the two lobes Cas9 to behave rather flexible in the absence of nucleic acids.[87,88]

Apart from the overall structures, the major finding was the huge difference between the apo conformation and the sgRNA bound state. In the apo state the structure seems more disordered and the relative orientation of the two lobes differs dramatically from the nucleic acid bound conformations. Based on these findings, Jinek et al. suggested the apo state to adopt an auto inhibited conformation.[87]

Also, the localization of the distinct domains in the different states was surprising. In the crystal structure of the holo-protein, the HNH active site is located away from the anticipated DNA cleavage site 3nucleotides upstream of the PAM sequence. Therefore, it remained elusive how the HNH domain can adopt a conformation that allows cleavage of the complementary target DNA strand. Follow-up studies using bulk and single molecule Förster resonance energy transfer (FRET) as well as molecular simulations aimed to explain the detailed movements during nucleic acid recognition and cleavage. These studies indicate a third state (the active, docked state) of the HNH domain in addition to the two captured by x-ray crystallography. Apart from that, the single molecule based studies all suggest the apo state to be rather flexible than to adopt a stable auto inhibited conformation.[89–92] A very recent paper using high speed atomic force microscopy was able to monitor S. pyogenes Cas9 live in action. All three major conformations were investigated, and even the cleavage of a DNA strand by Cas9 can be seen. The recordings of the apo state clearly proof its disordered and flexible conformation, whereas the nucleic acid bound states are well structured.[93] The single molecule and electron microscopy studies were all published after the starting date of this master thesis, but support the herein presented findings. Nevertheless, cross-linking mass spectrometry offers unique possibilities and allows to capture dynamic changes of Cas9.

1.4 Aim of the Thesis

QXLMS is becoming more and more popular in the cross-linking mass spectrometry community. So far mainly MS^1-based quantification methods have been published for cross-linked peptides. In this thesis a simple and robust QXLMS workflow based on isobaric tags was developed. In comparison to previously described quantitative approaches, this one does not need any advanced equipment and does not require manual quantitation. Cross=link identification was done with pLink and the cross=links were quantified with Proteome Discoverer 2.2. To demonstrate the relevance of the workflow, it was developed on the basis of S. Pyogenes Cas9. Because of its well characterized structure and dynamics, and the existence of different conformations Cas9 is an ideal model protein for the development of a QXLMS method. Furthermore, the conformations rearrangement can be easily induced by addition of commercially available synthetic tracr/crRNA and DNA oligos. Moreover, several protocols for Cas9 expression and purification are well established and the protein is easily available.

2. Materials and Methods

2.1 Chemicals and Reagents

Table 1: List of used chemicals

Chemical (Purity)	Abbreviation	Supplier
Acetic Acid (≥99.5%)	AcOH	Fluka
Acetonitrile (≥99.9%, HPLC grade)	ACN	VWR Chemicals
Acrylamid/Bisacrylamide 29:1 (30% Solution)	Acrylamid	National Diagnostics
Ammonium Acetate (≥98%)	AmAc	Fluka
Ammonium Bicarbonat (≥99.5%)	ABC	Sigma Aldrich
Ammonium Persulfate	APS	Bio-Rad
Bromophenol Blue		Sigma Aldrich
Dimethylsulfoxide, anhydrous (≥99.9%)	$DMSO_{anh.}$	Sigma Aldrich
Disuccinimidyl suberate (NMR: 100%)	DSS	Thermo Fisher Scientific
1.4-Dithiothreitol (high purity)	DTT	Roche
Ethanol (absolute)	EtOH	Merck
Formic Acid (98%, MS grade)	FA	Fluka
Glycerol (87%)		AppliChem
HEPES (cell culture grade)	HEPES	AppliChem
Hydrochloric Acid (37%)	HCl	Sigma Aldrich
Hydroxylamine (50% wt. In H_2O)	NH_2OH	Sigma Aldrich
2-Iodoacetmide (>99%, crystalline)	IAA	Sigma Aldrich
Methanol (≥99.7%)	MeOH	Sigma Aldrich
Magnesium Chloride Hexahydrat (≥98.0%)	$MgCl_2$	Fluka
N,N,N′,N′-Tetramethylethane-1,2-diamine (~95%)	TEMED	Sigma Aldrich
Potassium Chloride (≥99.5%)	KCl	Fluka
Potassium Hydroxide (≥99.5%)	KOH	Fluka
Sodium Hydroxide (≥99.5%)	NaOH	Merck
TMTsixplex™ Isobaroic Label Reagent	TMT 6-plex	Thermo Fisher Scientific
Triethylammmonium bicarbonate (1M, pH 8.5)	TEAB	Sigma Aldrich
Trifluoroacetic acid (≥99%)	TFA	Sigma Aldrich
Tris(2-carboxyethyl)phosphine hydrochloride (≥98%)	TCEP	Sigma Aldrich
Tris(hydroxymethyl)-aminomethan (≥99.9%)	Tris-base	Sigma Aldrich
Tris(hydroxymethyl)-aminomethan –HCL (≥99.9%)	Tris-HCl	Sigma Aldrich
Trypsin Gold (Sequencing Grade)		Promega
Ultrapure Water (Milli-Q)	H_2O	Merck Millipore

Table 2: List of used buffers

Buffer	Composition (1x)	Supplier
NuPAGE LDS Sample Buffer	141 mM Tris-base 106 mM Tris-HCl 2% LDS 0.51 mM EDTA 0.22 mM SERVA™ Blue G-250 0.175 mM Phenol Red pH 8.5	Thermo Fisher Scientific
NuPAGE MOPS SDS Running Buffer	50 mM MOPS 50 mM Tris-base 0.1% SDS 1 mM EDTA pH 7.3	Thermo Fisher Scientific
Native Running Buffer	25 mM Tris-base 192 mM Glycine pH 8.3	
Native Sample Buffer	100 mM Tris-HCl 10% Glycerol 0.025% Bromophenol Blue pH 8.3	

Table 3: List of used protein gels

Gel	Recipe	Supplier
8% native Acrylamide Gel	For 2-3 Gels: 1.3 ml Tris-Acetate (3 M, pH 8.3) 5.3 ml Acrylamid Sol. 13.23 ml H_2O 84 µl APS 12.5 µl TEMED	
Bolt™ 4-12% Bis-Tris Plus Gels		Thermo Fisher Scientific

Table 4: Staining solutions

Staining solution	Supplier
InstantBlue Protein Staining	Expedeon
SafeBLUE Protein Stain	NBS Biologicals
SYBR™ Safe DNA Gel Stain	Thermo Fisher Scientific

Table 5: List of used nucleic acids. DNA target sequence is depicted as bold letters. PAM is highlighted in red

Nucleic Acid	Sequence
Synthego crRNA	GUGAUAAGUGGAAUUGCCAUG + Synthego EZ Linker
Synthego tracrRNA	Kept secret by the supplier
54-bp ssDNA/ duplex DNA target strand	5' – GCTCAATTTTGACAGCCA**CATGGCATTCCACTTATCACTGG**CATCC TTCCACTC – 3'
54-bp Duplex DNA non-target strand	3' – CGAGTTAAAACTGTCGGTGTACCGTAAGGTGAATAGTGACCGTAGG AAGGTGAG – 5'

Table 6: Sequence of the used S. Pyogenes Cas9 Halo-Tag© fusion protein. Red marked amino acids are mutations, required for S. Pyogenes Cas9 to be inactive. The Halo-Tag© is depicted in light blue with the bold black linker sequence.

Amino acid	Amino Acid Sequence
(1)	MHHHHHHVDKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
(61)	DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH
(121)	RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGD
(181)	LNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEK
(241)	KNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
(301)	KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFD
(361)	QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
(421)	IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
(481)	TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEG
(541)	MRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT
(601)	YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKR
(661)	RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQV
(721)	SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
(781)	QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL
(841)	SDYDVD**A**IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLIT
(901)	QRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV
(961)	KVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY
(1021)	KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV
(1081)	WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG
(1141)	FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
(1200)	LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNE
(1261)	QKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLF
(1321)	TLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD**GGSR**S
(1381)	AEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLHGNPTSSYVWRNIIPHVAPTHRC
(1441)	IAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVLVIHDWGSALGFHWAKRNPE
(1501)	RVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGTLPMGVVRPL
(1561)	TEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPKLLFWG
(1621)	TPGVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLSTLEISG

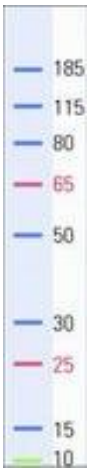


Figure 6: Illustration of the PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa (Thermo Fisher Scientific) on a 4-12% Bis-Tris gel using MOPS running buffer

2.2 Sample Preparation

2.2.1 Reconstitution of ApoCas9, gRNA-Cas9 and gRNA-targetDNA-Cas9

The inactive S. Pyogenes Cas9 variant D10A, H847A with a fused Halo-Tag was expressed as described by Deng et al.[85] After purification, the protein was buffer exchanged to HEPES buffer (100 mM, pH 7.4, 250 mM KCl) using Bio-Rad P-6 Micro Bio-Spin columns according to the manufacturer protocol. The apo protein samples were diluted 1:1 using 1 M KCl and H_2O to achieve a final protein concentration of 10 µM. The samples for ribo- and holo-protein formation were diluted with water in the way, that after the nucleic acid addition the final protein concentration was 10µM.

In case of the purchased synthetic cr/tracrRNA (Synthego, CA), 5nmol of the corresponding tracerRNA was dissolved in 20 µl H_2O. For gRNA formation, the previously prepared solution was added to the flask containing crRNA. The two oligos were annealed by heating to 78°C for 10 minutes on a thermocycler subsequent 30 minutes in the incubator (37°C) and final cooling on the benchtop. Synthetic DNA oligos were dissolved in nuclease free water to a final concentration of 1 mM. All nucleic acids were stored up to one month at -20 °C and defrosted immediately before usage.

For riboprotein formation, previously prepared Cas9 and annealed gRNA were mixed in a molar ratio of 1:2 and incubated for 15 minutes at room temperature. To form the Holo-protein synthetic ssDNA (5-fold excess) was added to the sample and incubated for 30 minutes at room temperature. In every case, complex formation was checked on an 8% native acrylamide gel using a Bio-Rad Mini-PROTEAN® vertical electrophoresis chamber. The gel was run for 45 minutes at 4°C. Nucleic acids were stained with SYBR™ Safe DNA Gel Stain and proteins were stained using SafeBLUE Protein Stain. After successful complex formation, the samples were immediately cross-linked and processed further.

2.2.2 Cross-linking of the different Cas9 Complexes

Cross-linker and protein concentration as well as cross-linking time, temperature and buffer composition were optimized to obtain optimal results (see Results 3.1). In the final protocol, ApoCas9 as well as the reconstituted Cas9 complexes were crosslinked at 4°C, using the homobifunctional amine reactive crosslinker DSS at a concentration of 250 mM in a 50 mM HEPES buffer containing 500 mM KCl with pH 7.4. The crosslinker was dissolved in a minimum amount of anhydrous DMSO to give a maximum organic solvent concentration of 10 vol% in the final sample. Cross-linking reactions were quenched with 100 mM Tris-buffer (pH 7.5) for 30 minutes at 4°C. After quenching, samples were checked on a SDS-Page for successful cross-linking. Therefore, roughly 1 µg protein was mixed with 2.5 µl 4x NuPAGE LDS Sample Buffer and diluted to a final volume of 10 µl using H_2O. The sample was heated to 78°C for 15 minutes just before applying it onto a precast 4-12% Bis-Tris acrylamid gel (Thermo Fisher Scientific). The gel was run for 35 minutes at 200V using NuPAGE MOPS SDS Running Buffer in a XCell SureLock® Mini-Cell (both Thermo Fisher Scientific). To estimate the size of the protein bands, 3 µl of PageRuler Plus were run in a separate lane on each gel. For protein staining the gel was incubated in 15-20 ml InstantBlue protein staining solution for 1-24 hours followed by destaining with water.

2.2.3 Protein Digestion

For Protein digestion, optimized in-house protocols were used. All reagents were either freshly prepared or defrosted immediately before use.

2.2.3.1 In-Solution Digest

The Protein sample was reduced using 10 mM DTT at 37°C for 30 minutes under permanent shaking. To alkylate the reduced cysteine residues IAA was added to a final concentration of 20 mM and the reaction mixture was incubated at room temperature in the dark for 30 minutes. Afterwards Trypsin was added at an enzyme-to-protein ratio of 1:30 and the sample was digested at 37°C over night. On the next day, the digestion was stopped by acidifying the sample with 10% TFA to a final concentration of 1% TFA. Samples were frozen and stored at -20°C until analysis.

2.2.3.2 In-Gel Digest

After cutting out the band of interest, it was diced into pieces of roughly 1x1 mm. To destain the pieces, they were washed for 10 minutes on the shaker altering with 200 µl 100 mM ABC-buffer and 200 µl 50 mM ABC-buffer/50% ACN. After each step, the supernatant was discarded, and this procedure was repeated until the pieces were completely colorless.
In the next step, 100 µl 100% ACN were added and the pieces were incubated till they turned milky white. The organic solvent was taken off and the gel pieces were rehydrated in 100 µl 6 mM DTT solution in 100 mM ABC-buffer to reduce the cysteines. The reduction

was carried out at 60°C for half an hour. For alkylation 100 µl of an aqueous 27 mM IAA solution were added and incubated in the dark for 30 minutes. The alkylation reagent was withdrawn, and the gel pieces were washed twice as described for destaining. Again 100 µl 100% ACN were added till pieces turned milky white indicating dehydration of the gel. The pieces were rehydrated with the required amount of Trypsin solution (12.5 ng/µl, 100 mM ABC-buffer) and the same amount of 100 mM ABC-Buffer was added on top. The digestion was carried out over night at 37°C. On the next day 10 µl 10% FA were added to stop the digest. Subsequently, the supernatant was transferred to a 200 µl PCR tube. To further extract peptides, two times 20 µl 5% FA/50% ACN were added on top of the gel slices and the sample was sonicated for ten minutes each time. All fractions were pooled and evaporated with a vacuum concentrator (Eppendorf Concentrator plus) until dryness. Each sample was reconstituted in 50 µl 0.1% TFA and stored in the freezer until analysis.

2.2.4 TMT-Labeling of Tryptic Digest

20 µg peptide of the digested sample was desalted using C-18 stage tips. Therefore, tree plugs of C-18 material were put into a 200 µl tip. The tip was washed with 100 µl 100% MeOH and 100 µl 80% ACN/0.1% FA before equilibrating the material with 0.1% FA. After loading the sample, it was washed once with 0.1% FA prior to eluting it with two times 30 µl 80% ACN/0.1% FA. The sample was evaporated and reconstituted in 20 µl 50 mM TEAB. A 0.8 mg equivalent of the TMT-reagent was equilibrated to room temperature just before use and was dissolved in 41 µl anhydrous ACN. 25% of the labeling reagent were mixed with the samples and incubated for 1h at room temperature. The reaction was quenched by adding 2.5 µl 5% hydroxylamine and incubated for 15 minutes. To remove the organic solvent, the sample was acidified and evaporated to roughly one third. To ensure complete labeling each sample was measured separately on a QExactive HF instrument (detail see 2.3). If labeling efficiency was above 98% differently labeled samples were mixed in an equimolar ratio, desalted using C-18 stage tips, dissolved in 20 µl 0.1% TFA and stored at -20°C until further use.

2.2.5 Size-Exclusion-Chromatography

To enrich cross-linked peptides SEC was performed on a TSKgel SuperSW2000 column (300 mm × 1 mm × 4 µm, or 300 mm × 4.5 mm × 4 µm, Tosoh Bioscience). 2-10 µl (1 µg/µl) of the digested and labeled peptides were injected and eluted isocratically with 30% ACN and 0.1% TFA in water and a flowrate of 10 µl/min on the small column. The column with 4.5 mm in diameter was eluted with a flow rate of 200 µl/min. A Dionex UltiMate 3000 HPLC system was used to monitor the enrichment runs at 214 and 254 nm. Fractions were collected peakwise (maximum 1.5 minutes), evaporated, resuspended in 50 µl 0.1% TFA and stored at -80°C until measurement.

2.3 LC-MS/MS Analysis

The following settings and methods can be seen as optimized operating conditions and were used for all experiments if not stated differently.

2.3.1 Quality Control

Before every injection into the mass spectrometer, the sample was analyzed on a Dionex UltiMate 3000 HPLC RSLC nanosystem equipped with a PepSwift Monolithic® RSLC column (0.2 x 5 mm, Thermo Fisher Scientific) for two reasons. First of all, to ensure the purity of the sample and that the tryptic digest is complete, which may be challenging for cross-linked proteins. The second reason is to estimate the amount of peptides, that may be subsequently injected into the mass spectrometer. To do so every ten runs 500 fmol of a tryptic bovine serum albumin digest is analyzed as a reference. Prior to the quality control, the samples were acidified and diluted with 0.1% TFA to an appropriate concentration.
After injection, the samples are concentrated and desalted on a PepSwift Monolithic® Trap (0.2 x 5 mm, Thermo Fisher Scientific). Peptides are eluted using a 20 minutes gradient ranging from 2-90% buffer B. The gradient is followed by a 5 minutes washing step at 90% buffer B, before the concentration of buffer B is changed to 2% within two minutes. Finally, the column is equilibrated with 2% buffer B for 13 minutes. The standard column oven temperature of the monolith is set to 60°C.

2.3.2 High Performance Liquid Chromatography

Prior to MS analysis, the digested peptides were separated using a Dionex UltiMate 3000 HPLC RSLC nanosystem coupled to an QExactive HF Orbitrap mass spectrometer (Thermo Fisher Scientific) via a Nanospray Flex ion source (Thermo Fisher Scientific).
For sample concentration, washing and desalting, the peptides were trapped on an Acclaim PepMap C-18 precolumn (0.3x5mm, Thermo Fisher Scientific), using a flowrate of 25 µl/min and 100% buffer A (99.9% H_2O, 0.1% TFA). The separation was performed using an Acclaim PepMap C-18 column (50 cm x 75 µm, Thermo Fisher Scientific) at a flowrate of 230 nl/min and 30°C. For separation, a solvent gradient ranging from 2-35% buffer B (80% ACN, 19.92% H_2O, 0.08% TFA) was applied. The time over which the gradient was applied varied from 30 minutes to 5 hours, depending on the sample complexity. In general, every HPLC method was built up as follows: 0-10 min 2% B, 10-X min 2-35% B (where X depends on the gradient length), followed by a 5 min gradient from 35-95% B, 5 min 95% B, 2 min 90-2% B and finally re-equilibrating the column with 2% B over 17 minutes.

2.3.3 MS Data Acquisition

2.3.3.1 Cross-Link Samples

All samples were recorded an a Orbitrap QExactive HF mass spectrometer (Thermo Fisher Scientific; San Jose, CA). The mass spectrometer started the acquisition 20 minutes after the injection into the HPLC and was operated in data-dependent-acquisition (DDA) mode. MS^1 scans were acquired over the m/z range from 350-2000 m/z at a resolution of 60,000 (@200 m/z). Up to 10 MS^1 precursor (most intense) were selected for fragmentation using HCD with 28% normalized collision energy. The isolation window for precursor ions was set to 1.6 m/z with a maximum injection time of 150 ms. Fragment ions were recorded in the Orbitrap using the dynamic first mass option with a resolution of 30,000 (@200 m/z). The precursor (±10 ppm) was excluded from MS^2 acquisition by a dynamic exclusion of 30 seconds.

2.3.3.2 TMT-Labeled Cross-Link Samples

The mass spectrometer was operated like for the XL-only samples with small modifications. Resolution for MS^1 was set to 120,000 and for MS^2 to 60,000. MS^1 spectra were recorded from 350-1650 m/z and fragmentation was performed using HCD with a stepped normalized collision energy of 33-35-37%. MS^2 scans were recorded with a fixed first mass of 115 Da.

2.4 Data Analysis

2.4.1 Identification of cross-linked peptides

At first, Thermo Xcalibur raw files were converted into mgf files using Proteome Discoverer (version 2.2, Thermo Fisher Scientific). To identify cross-linked peptides, the generated mgf files were analyzed with pLink (Version 1.23). For cross-link identification the following search parameters were used:

Fragmentation Method	HCD
Enzyme	Trypsin
Enzyme Specificity	C-terminally to R and K, if not N-terminally to P
Max. Missed Cleavages	2 (by default)
Peptide Length	4-60 amino acids (by default)
Peptide Mass	400-6000 Da (by Default)
Fixed Modifications	Carbamidomethylation (+57.021 Da, @ C)
Dynamic Modifications	Oxidation (+15.995 Da, @ M)

Cross-Linker	DSS (reactivity: K, N-terminus; cross-link and loop-link +138.068 Da, mono-link +156.079 Da)
Database	Sequence of S. Pyogenes Cas9 with fused HaloTag®
Decoy Database	Reversed sequence of S. Pyogenes Cas9 with fused HaloTag® (by default)
FDR	1% (on the PSM level)
E-Value cut off	1
MS^1 Tolerance	10 ppm
MS^2 Tolerance	20 ppm

Table 7: Search parameters for XL-identification using pLink

For the TMT labeled samples, TMT 6-plex (+ 229.163 Da) was added as dynamic modification on lysines and as a static modification at the peptide N-termini. pLink automatically produces several Excel files containing lists of cross-linked peptides with additional information, a graphical output as .png and an additional pLabel output. Both were used for manual cross-link validation.

2.4.2 Mapping of Cross-Links onto the Crystal Structures

To ensure, that the cross-linking experiment gives meaningful information, the identified amino acid linkages were plotted onto the corresponding crystal structure to obtain distance information. As the solved crystal structure of apo Cas9 contains many regions that could not be captured using crystallography, a complete homology model was generated with the help of the I-TASSER server.[94] (see 2.4.4)
To visualize the published crystal structures and the homology model, UCSF Chimera version 1.12, a python based freeware, was used.[95] Chimera was operated using its command line function. To calculate the Euclidean distance (ED) between the Cα-atoms of the identified cross-linked lysines the following command was used:

distance#1:AA1.a@CA#2:AA2.a@CA

where “AA1” and “AA2” represent the positions of the linked amino acids and “.a” selects the chain defined in the PDB-file. “CA” defines, that the distance is measured from the Cα atom. In certain cases, to gain a deeper insight into the orientation of the lysines, also the Cε (e.g. the last carbon atom in the lysine side chain) was used.

2.4.3 Quantification of Cross-Linked Peptides

For cross-link quantification, reporter ion intensities of TMTsixplex™ Isobaroic Labels were calculated using Proteome Discoverer 2.2. Reporter ions were normalized on the total peptide amount. The "Quan Spectra" category was exported to an excel file. This file contains every recorded MS^2 scan and the corresponding reporter ion intensity recorded for every TMT channel. Identified cross-link spectrum matches were grouped for their unique cross-linking site and the reporter intensities were summed up. Only spectra with a co-isolation below 30% were used for quantification.

For volcano plot analysis, the procedure was the same, but TMT-reporter intensities were generated in P.D. 2.2 without normalization. Before the plot was generated with an in-house R-script, the cross-links were normalized on the total cross-link amount.

2.4.4 Modeling of Cas9 conformations

To obtain structural models derived from the identified cross-links, the I-TASSER web server was used. For homology model generation, the sequence of S. pyogenes Cas9 without Halo-Tag©, linker region and poly-histidine at the N-terminus (Table 6) was submitted. The partially solved crystal structure (4cmp:A) was specified as template without alignment. Moreover, all known crystal structures and EM structures of Cas9 in complex with any other biomolecule were excluded from the template library. I-Tasser automatically renks the generated models according to their calculated C-score*. The best ranked homology model was used for subsequent cross-link mapping. Additionally, to the predicted models, also the top ten PDB structures closest to the best model are reported. These hits included structures for S. aureus Cas9, and F. novicida Cas9 both in complex with sgRNA and target dsDNA. Due to the high structural similarity and comparable size range, these two proteins were used as template for later models.

Cross-link derived models for gRNA bound Cas9 were generated with I-TASSER using either the apo structure (4cmp:A) or the structure of the two homologous Cas9 variants from S. aureus (5axw:A) and F. novicida (5b2o:A) as template without alignment. In addition, cross-linked residues were assigned as distance restraints in the following format:

DIST Res1 CA Res2 CA 30

DIST indicates that the two atoms are a certain distance apart from each other. Res1 and Res2 indicate the two cross-linked lysine residues and CA defines that the distance should be measured from the alpha-carbon atom of the indicated residue. Finally, 30 specifies the distance between the two alpha atoms in Angstrom.

Each of the five generated models was evaluated based on the following criteria. First, the Euclidian distance of the assigned restraints was measured. Second, the structure was checked for general coherence (e.g. are secondary structure elements present, how elongated or disordered is the structure, etc.). Moreover, all models were directly compared with the solved crystal structure (4zt0). Comparison was conducted in UCSF Chimera 1.12 using the MatchMaker tool, the Needleman-Wunsch alignment algorithm and the BLOSUM-62 Matrix.

3. Results

3.1 S. Pyogenes Cas9 cross-linking

3.1.1 Optimizing the Reaction Conditions for Cas9 Cross-Linking

In order to develop a robust cross-linking workflow for Cas9, several conditions were tested to obtain an ideal result. According to previous findings with other proteins, 50 mM HEPES at pH 7.5 seemed as an adequate starting point for cross-linking experiments. To determine the ideal cross-linker concentration, the protein was aliquoted and reacted with 500, 1000 and 2000 μM of DSS for 60 minutes at 37°C. Subsequent SDS-gel analysis showed a high degree of multimer-formation (Fig. 7), indicating over cross-linking. To check weather these multimers significantly influence the results or if they may be neglected, the sample cross-linked with 500 mM DSS was digested in solution and analyzed on the mass spectrometer. The analysis of the obtained cross-links showed that many of them do not fit the published crystal structure of apo Cas9. In addition to the in-solution digest, the monomer band on the gel was in-gel digested and analyzed in the same way. Unfortunately, also this experiment yielded several cross-links incompatible with the known crystal structure. This can be explained by destruction of the tertiary structure by too many cross-links. Moreover, the loss of positive charges by blockage of lysine side chains may contribute to this phenomenon.

The following experiment was carried out using a lower cross-linker concentration (250 mM), shorter reaction time (30 minutes) and lower reaction temperature (4°C). As slight improvements in terms of multimer-formation could be observed for the lower DSS concentration at 4°C on the gel, another experiment with lower protein concentration but the same reaction conditions was carried out. Unfortunately, this experiment did not lead to any improvement (data not shown), so the protein concentration was kept at 10μM. As a next step, the buffer system was changed to see if this influences the cross-linking reaction in a positive way. Facing the problem, that primary amine containing buffers like glycine and Tris are not suitable for cross-linking with NHS-chemistry, phosphate buffer was chosen, even if this could influence the following LC-MS experiment. 50 mM phosphate buffer containing 15 mM NaCl did only slightly improve the reaction outcome, whereas in 200 mM phosphate buffer with 200 mM NaCl, the multimerization could be completely inhibited. (Fig. 7) To analyze the sample on the mass spectrometer, an aliquot containing roughly 0.5 μg protein was diluted 1:20 with 0.1% TFA, so the final phosphate concentration was only 10 mM. To minimize the influence of the buffer, the sample was washed for 10 extra minutes with Buffer A after trapping it on the pre-column of the HPLC system. As the analysis results show, the number of cross-links incompatible with the crystal structure significantly decreased. The majority of the remaining overlength cross-

links could be explained with flexible and unstructured regions within the protein. Afterwards it was hypothesized, that the in general high ion strength rather than the phosphate buffer led to the desired result. Therefore, 50 mM HEPES buffer containing 500 mM NaCl was tested as cross-linking buffer. The obtained results were comparable with those of the 200 mM phosphate buffer sample., hence this condition was used in all following experiments.

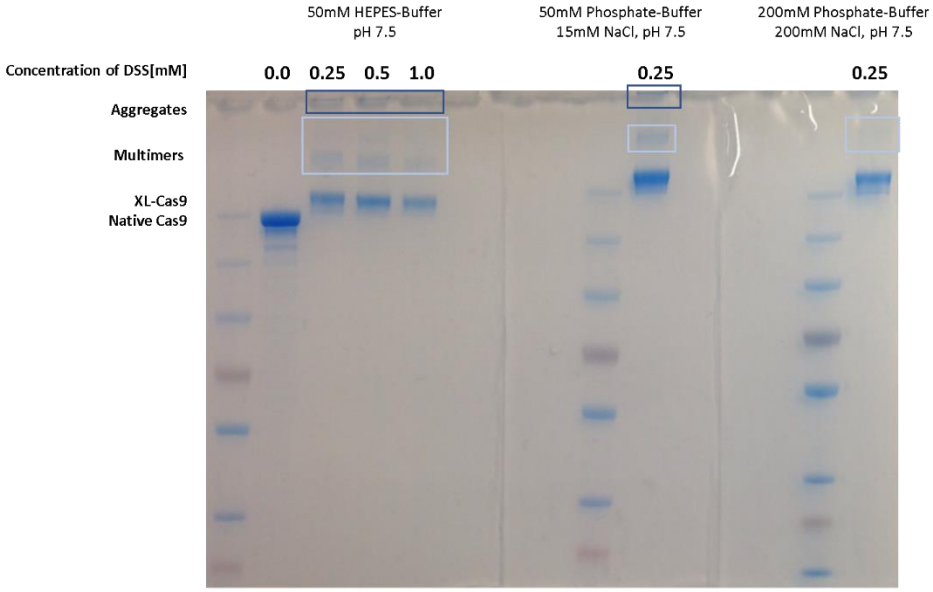


Figure 7: SDS-Gel experiments towards the optimal cross-linking conditions

3.1.2 Formation of the different Cas9 complexes

Initial RNA-binding and complex formation experiments were conducted using in vitro transcribed sgRNAs, as a gift from an in-house group. The sgRNAs were obtained at very low concentrations, ranging from 84-300 ng/µl. To concentrate the nucleotides, they were precipitated by the addition of 0.1 volumes 7.5 M ammonium acetate and 2.5 volumes 100% ethanol. The mixture was incubated at -20 °C overnight, to ensure the complete precipitation. On the next day, the RNA was pelleted at 20000 rpm for 15 minutes in a micro centrifuge. The pellet was washed, dried and finally dissolved in 10 µl RNAse free H_2O at a concentration of 300 µM. For secondary structure formation the RNA was heated to 96 °C for 5 minutes. After it has cooled down to room temperature it was used to titrate the readily prepared apoCas9 solution with 1 to 8-fold molar excess of RNA. To monitor the complex formation, the samples were checked on an 8% native Tris-Acetat gel. As Cas9 with the fused HaloTag® has an estimated isoelectric point of 8.3 which roughly equals the pH of the native gel, it is expected to move slower than the sgRNA:Cas9 complex containing negatively charged phosphate groups from the nucleic acid. After staining, the gel showed two bands for the lanes containing RNA and only a smear at the very top for the lane containing Cas9 only. It was assumed, that the upper band corresponds to unbound Cas9, whereas the lower one represents the formed complex. (Fig. 8 A) As the different sgRNAs

were designed to target several genes, they can be seen as a heterogeneous mixture. Therefore, not one sharp band but a large smear with multiple distinct bands was obtained on the gel. Interestingly, even with a large excess of RNA the band corresponding to apo Cas9 did not disappear completely.

Immediately after they were loaded onto the gel, the remaining samples were cross-linked using optimized conditions and applied onto a SDS-PAGE gel. It is noteworthy that at low sgRNA excess (1.5 & 3-fold) a significant amount of multimer could be observed, while starting from 4.5-fold excess, predominantly the monomer band appeared. This observation indicates an increase in structural stability of the protein, is in accordance with other studies published within the last year.[91–93,96] Additionally to gel electrophoresis, the samples with 4.5-fold excess and 8-fold nucleic acid excess were digested in solution and analyzed on the mass spectrometer. Brief inspection of the results showed considerable changes in the amount and position of the detected cross-links, hence the experiment was brought to the next level by using synthetic gRNA.

A small amount of Cas9 was again titrated with the afore prepared crRNA:tracrRNA complex and analyzed via native gel electrophoresis. It is emphasized, that the freshly synthesized RNA was of better quality compared to the in vitro transcribed sgRNA, thus only 1.5, 2 and 3-fold excess was used. Already at a molar ratio of 1:2 (protein to nucleic acid), the band linked to the apo conformation fully disappeared indicating a complete complex formation.

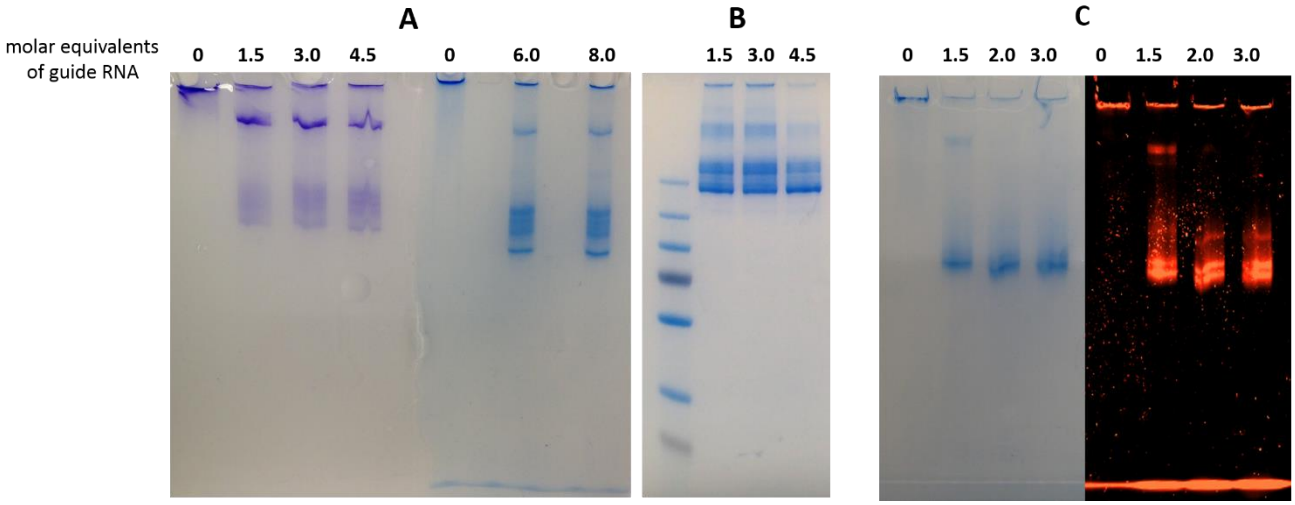


Figure 8: Native gel analysis of Cas9 titrated with different amounts of in vitro transcribed sgRNA (A) and SDS-PAGE analysis of the cross-linked complexes (B). Native gel showing Cas9 titrated with synthetic guide RNA stained with Coomassie (left) and SYBR™ Safe (right), respectively. (C)

In a next step, the formation of the holo-complex, consisting of Cas9, gRNA and target DNA, was investigated. Therefore, the preformed riboprotein was mixed with a five-fold excess of ssDNA complementary to the synthetic RNA. Comparative native gel analysis of the two complexes revealed no shift-changes. However, staining for nucleic acids showed an increased intensity in the holo-protein lane, attributable to the additional nucleic acid. (Sup. Fig. 1) The following reaction with DSS resulted in a very low cross-linking yield. It was

anticipated, that the high concentration of negative charges arising from both RNA and DNA backbones negatively influences the cross-linking reaction. Therefore, only apo Cas9 and the riboprotein were investigated by comparative and quantitative cross-linking experiments.

3.2 Cross-Linking Mass Spectrometry Experiments

3.2.1 Comparative Analysis of apo Cas9 and gRNA-Cas9

After the optimal cross-linking conditions and the conditions for complex formation were elucidated, the two different conformations were compared on the cross-link level. Therefore, the two different states (RNA bound and RNA free) were reconstituted, cross-linked, reduced, alkylated and digested in parallel. The digested samples were measured in duplicate on the same instrument to ensure the best reproducibility. Only cross-links that were identified at least twice in the two replicate runs were considered for comparison (Suppl. Tab. 1). After "one-hit-wonder" exclusion, 162 and 93 unique cross-links remained for the apo- and ribo-protein, respectively. This is in accordance with the observations in previous gel electrophoresis experiments, where RNA containing samples showed less band broadening and/or multimerization. A first comparison of the identified cross-links revealed that only a small subset (41 XLs, 19%) was identified in both samples while 56% were identified in the apo sample exclusively. The small number of shared cross-links indicated a large structural rearrangement, as described in the crystal structures. As a next step, the linked sites were assigned to their corresponding domain in the Cas9 protein and inspected group wise according to their identification.

At first, those contained in both conformational states were plotted onto the homology model of apo Cas9 and the crystal structure of sgRNA bound Cas9 (PDB 4zt0, chain A). Most of these linkages were formed within the same domain, and only 33% were so called inter-domain cross-links. Already published structural studies investigating Cas9 did not report severe structural changes within the separate domains. The intra-domain cross-links found in both samples did not exceed the maximum cross-linking distance and therefore were not investigated any further. Taking a closer look at the linkages between different domains showed, that almost 70% are located at the RuvC-HNH-PI interface. This interface seems to be conserved among the two conformations, which corresponds to the measured distances as well as the relative orientation. Additionally, two cross-links connecting lysine 233 (Rec2) with lysine 772 & 775 (HNH) were in common. Both linkages would exceed the max. XL-distance in the RNA bound state by a few Å. Nevertheless, recent FRET studies and molecular simulations suggest that the HNH domain may adopt conformations, that allow these cross-links to form.[91,92]

In contrast with the afore mentioned, the cross-link connecting lysines 673 and 948 (Rec3 and RuvC) is not compatible with the apo crystal structure but found multiple times in every

MS run. Moreover, there is one cross-link (439-1118) that does not fit any of the reported crystal structures but that is supported by several CSMs. Since the region of amino acids 1100-1200 represents a flexible loop in both structures it may still be possible, that this overlength linkage is formed.

Next, the cross-links occurring exclusively in the riboprotein samples were inspected more precisely. In contrast to the other groups, only the minority (35%) of the links are within the same domain. Like before they perfectly fit both structures and do not show significant distance differences, hence they were not investigated into detail. In case of the inter-domain cross-links, again many of them (12 XLs, 36%) were located at the RuvC-HNH-PI interface. Except one, they did not show any changes in the Euclidean distance measured on the two structures. In addition, 3 cross-links at the Rec2-HNH interface and one between lysine 377 and 218 (Rec1-Rec2) were identified, that neither show significant distance changes. However, over 50% of the cross-links uniquely identified in the riboprotein sample were found to connect the Rec3 with the RuvC and the HNH domain, respectively. Rec3 is the domain that is affected the most by the structural rearrangement upon guide RNA binding therefore, it is not surprising, that these links are not found in the apo protein sample. The theoretical distances these links would span in the apo conformation range from 61 to more than 100 Å. Furthermore, 4 cross-links that fit neither of the two conformations and one only compatible with the apo state were identified.

In the apo-only group there are three intra domain links formed between lysine residues that are torn apart by guide RNA binding and therefore unlikely to be formed in the RNP-state. The other intra-links do not show severe changes in XL-distance, as in the previously investigated groups. The linkages connecting different domains give a different picture compared to the common and RNP-only cross-links. 38 of the 63 inter-links exceed the maximal cross-linking distance in both structures by up to more than 30Å. These overlength cross-links are formed by lysine residues from every domain, but the majority of them includes amino acids 1100-1200. Moreover, there are 9 links that are impossible to be formed in the riboprotein only. Seven of them contain the Rec1 domain, that is shifted away from RuvC- and PI-domains upon RNA binding. The remaining cross-links do not show any differences between the two structures and are likely to be a product of the higher cross-linking rate of the apo protein.

Overall, it is striking, that apo Cas9 yield a by far higher number of identified cross-links compared to the gRNA:Cas9 complex. Additionally, the cross-links found only in the apo sample show a significantly increased number of overlength cross-links. These two findings support previous reports, indicating a highly flexible conformation of the apo protein. In addition to the more rigid conformation of the riboprotein, also the steric hindrance of the gRNA may contribute to the overall lower number of cross-links. Moreover, the comparative cross-linking experiments highlights the regions undergoing the largest structural rearrangement. However, also many cross-links that are in agreement with both crystal structures are exclusively found in one of the two samples. This may be due to the dynamic range problem, facing in XLMS. On the one hand the riboprotein only yields a low

cross-link amount, therefore some of them may be disregarded during acquisition. On the other hand, the highly flexible apo state forms loads of different cross-links, leading to a highly complex chromatogram. Due to the complexity, some cross-links may be overlooked. Further, two lysine-lysine connections that should only be formed in the RNP-conformation, are found in both samples or even exclusively in the apo sample. To achieve a more comprehensive comparison of the two conformations, a quantitative approach would be necessary to be independent of the right precursor selection during DDA.
Until now, several methods aiming for quantitative cross-linking using SILAC, LFQ and differently labeled cross-linker have been reported. Recently even a paper using TMT-labels to quantify cross-links has been published, but without applying it to a biological problem. This publication used an advanced instrumentation and MS^3 for quantitation. To make TMT based quantification more widely available and to proof the concept on a known structural rearrangement, we aimed to develop a simple and robust method to quantify and identify cross-links in a single MS^2 experiment.

3.2.2 Developing a Multiplexed QXLMS Method on the MS^2-Level

To investigate weather MS^2 based quantification using tandem mass tags yields meaningful results, apo Cas9 and the gRNA-Cas9 complex were cross-linked under the same conditions in parallel. Cross-linking reactions were stopped by the addition of Tris-buffer (pH 8) to a final concentration of 100 mM. This represents the first challenge in this methodology. TMT-reagents are designed to react with free amine groups on the peptides or proteins via a NHS-ester, the same chemistry most common cross-linking reagents use. This reaction is inhibited by any primary amines contained in the quenching buffer, but quenching is required to prevent over cross-linking. Therefore, proteins were reduced, alkylated and digested as usual, but additionally the final digest was desalted using C-18 stage tips, to get rid of the Tris molecules. Further the eluted and dried peptides were dissolved in 50 mM TEAB buffer, as recommended for TMT-labeling. Peptides labeling was conducted as advised by the supplier (see 2.2.4). For preliminary experiments, only two TMT-channels were used, 127 for the apo state and 128 for the riboprotein sample. After labeling, the samples were mixed equimolar, according to the UV chromatogram of the monolith that was carried out as quality control after the digest. Subsequently, the mixed samples were measured using the standard settings for cross-linked peptides, but with a lower first mass of 115 Da, to capture all reporter ions. Analysis with Proteome Discoverer 2.2 revealed an almost ten-fold higher intensity for the reporter ions derived from the gRNA-Cas9 sample. Additionally, many peptides with no or insufficient TMT-labeling were identified. As insufficient and unequal labeling leads to meaningless results, the leftover of the unmixed samples was reacted with the TMT-reagent again. To ensure comparability, both samples were again desalted with C-18 stage tips, dried, and resuspended in 50 mM TEAB buffer prior to the reaction. After quenching, both samples were measured separately with a 30 minutes HPLC gradient and the same instrument settings as before, to check for complete

labeling of both samples. As this time both samples only contained <1% (false discovery rate was set to 1%) of peptides that were incompletely labeled, the samples were again mixed in a 1:1 ratio and measured on a QExactive HF. The quantitative values for linear peptides as obtained in PD 2.2 were comparable. Therefore, the Thermo Xcalibur raw file was converted into a mgf file and analyzed with pLink. This search yielded less than 50 identified unique cross-links compared to over 300 in the standard samples. Since the quality control of the instrument was not conspicuous, the recorded spectra were inspected manually to search for anomalies. Prima facie, it was obvious, that almost exclusively the MS^1 precursors were present in the MS^2 spectra. (Fig. 9 A) This is of course limiting the identification rate, as unambiguous assignment of the recorded spectra is not possible. In comparison to that, the same cross-linked peptide pair without TMT-labels showed good fragmentation behavior under the same analysis conditions. (Fig. 9 B)

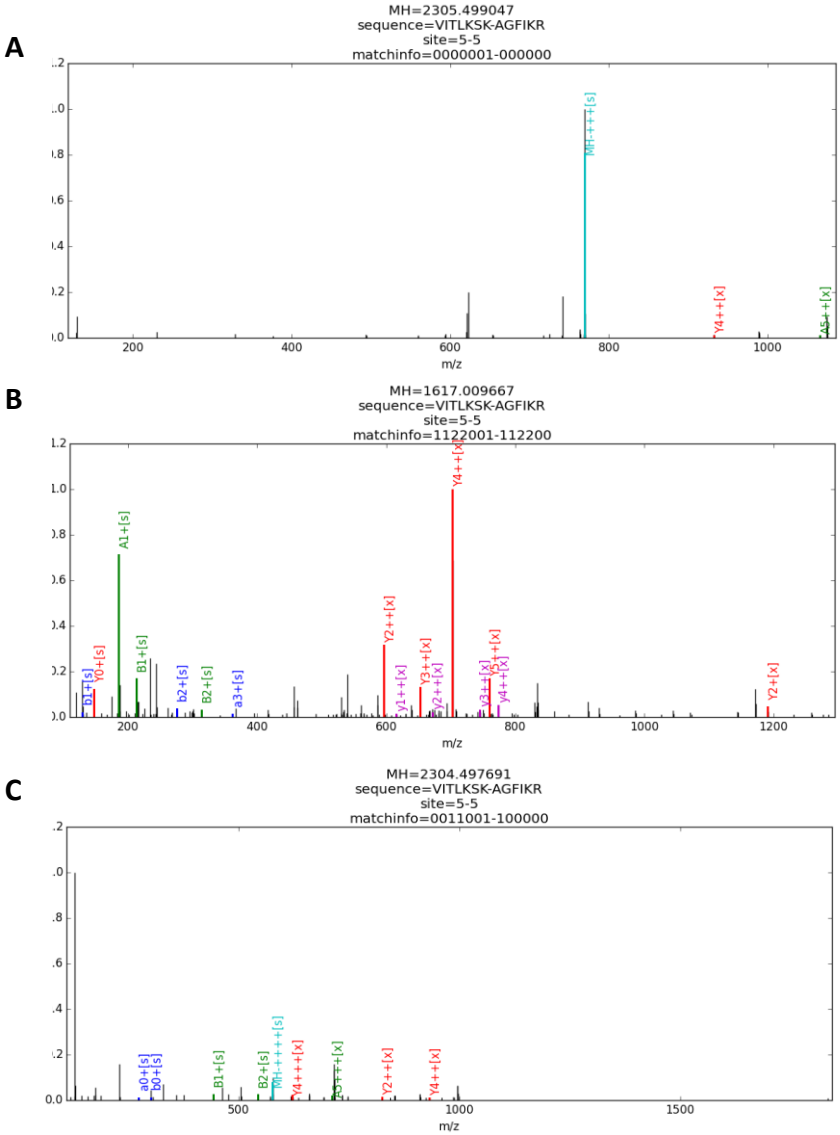


Figure 9: Graphical pLink output for the same cross-linked peptide **A:** labeled with TMT-6plex, fragmented with HCD at a NCE of 28% **B:** unlabeled peptide fragmented with HCD at a NCE of 28% **C:** labeled with TMT-6plex, fragmented with HCD at a NCE of 35%

From this finding, it was anticipated, that an increased NCEs could be beneficial for the fragmentation of cross-linked and TMT-labeled peptides.

To proof this hypothesis, the TMT-labeled apo Cas9 sample was recorded several times on a QExactive HF instrument using different NCEs ranging from 28-42% in steps of 2%. To save time each NCE was tested on a one-hour HPLC gradient (2-35% buffer B), where each measurement was followed by a 30 minutes wash to reduce carry-over.

The outcome supports the initial theory, as the number of identified cross-links significantly increased in all samples recorded with NCEs above 33%. Manual evaluation of the recorded CSMs showed a stepwise reduction of the precursor ion intensity compared to the fragment and reporter ions, with an almost complete loss of the precursor at energies above an NCE of 35% (Fig. 9 C). However, the graphical output of pLink for TMT-labeled samples is not representative, as only ions with a relative intensity of 5% of the most abundant peak are assigned. As cross-linked peptides have at minimum 2-4 primary amine groups that carry a TMT-label, the reporter signal is many times more intense than all the other signals, therefore most fragments have an intensity below this limit. Nevertheless, this is not affecting cross-link identification, since the search engine does not use this cut-off. Selected spectra were manually inspected, and almost the complete series of y- and b-ions could be assigned. The measurements applying NCEs from 33-42% yielded similar numbers (±10%) of CSMs but partially complementary unique cross-linking sites. (Fig. 10) To identify the highest number of XLs in further experiments, a stepped NCE ranging from 33-37% considered to be advantageous as it combines the fragments obtained from three different fragmentation events.

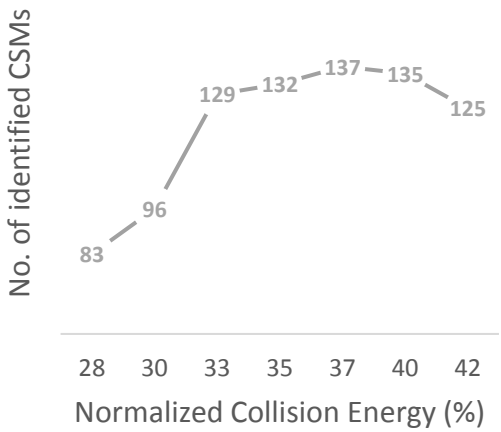


Figure 10: Number of identified CSMs at the respective NCE

Using the optimized collision energy, the sample was measured once more using a 3h HPLC gradient. This measurement resulted in a total of 100 unique cross-link sites with 176 corresponding cross-link spectrum matches. Compared to the unlabeled samples, this number is still much lower. Anyhow, this may be due to the increased complexity and the dilution effect, emerging through mixing of the different samples. Moreover, a certain percentage of identifications is lost due to the isobaric tag, as reported elsewhere. [18]

As the available sample amount was relatively low, only a few methods to increase the identification rate were available. The first strategy involved off-line enrichment/ fractionation using a miniaturized SEC column that was shown to work with sample amounts as low as 1 µg.[97] (see 2.2.5) The three fractions, that were anticipated to contain most of the cross-links (F2, F3, F4) were analyzed on the mass spectrometer. While in the first analyzed fraction no cross-linked peptide pair could be identified, fraction F3 and F4

contained 170 and 12 cross-links, respectively. Every XL obtained from fraction 4 was also found in fraction 3. Hence, only fraction 3 was used for further data analysis. The second strategy that was chosen to obtain a higher number of cross-links was the elongation of the chromatography gradient to 5 hours. The longer gradient allows to inject more material into the HPLC which can be essential for cross-link identification. Additionally, at least in theory, the peak capacity may be increased, if the HPLC system is operated under perfect conditions.[98]

The 5h gradient yielded a total of 177 unique cross-links which is roughly the same as obtained with SEC enrichment. However, for SEC enrichment an almost four-time higher amount of digested peptides was used compared to the amount injected for the 5h gradient. A closer look at the results derived from the two different methods shows that 80-83 percent of the cross-links identified in the first method are also captured in at least one of the two superior methods, while 72% are shared by all of them. The 5h gradient and the SEC enriched sample show an overlap of roughly two thirds. Overall it could be shown, that the number of cross-links IDs can be significantly increased by either enriching the sample using off-line SEC fractionation, or by elongation of the HPLC gradient. The outcome of the two experiments is comparable and therefore both methods may be used, depending on the available material amount or measurement time, respectively. However, if enough time and sample are available, both methods may be used, as they yield partially complementary cross-linked residues.

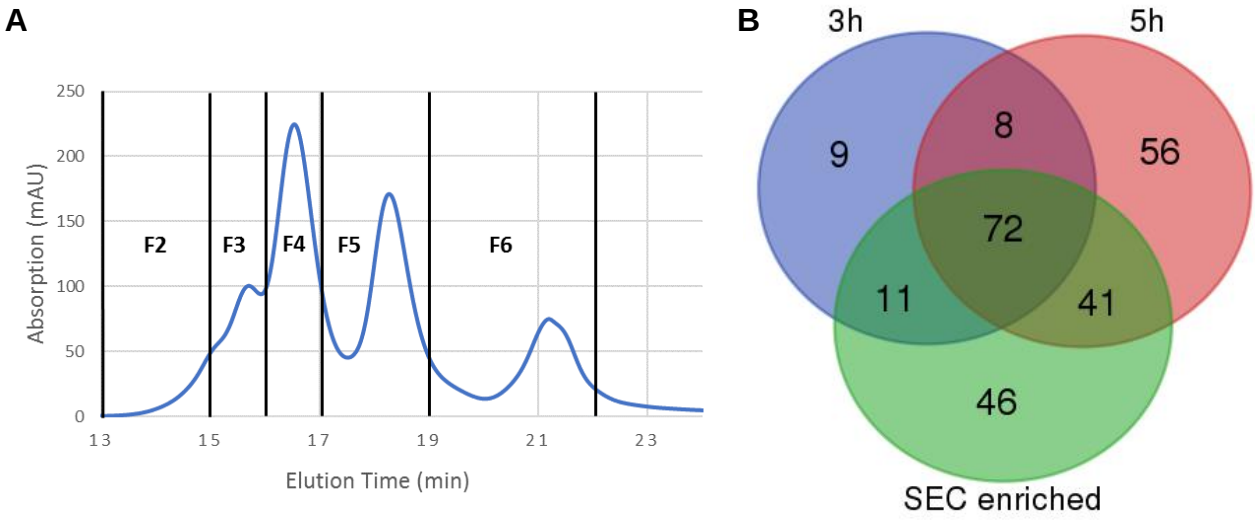


Figure 11: SEC UV/Vis-Chromatogram of TMT labeled apo & RNP Cas9 mixture **(A)** Overlap of unique cross-link sites between the three different analysis methods **(B)**

3.3 TMT-Based Monitoring of the Apo to RNP rearrangement

Both, SEC fraction 3 and the 5h HPLC-gradient run were quantified separately as described in 2.4.3. The calculated ratios for the apo and the RNP channel showed, a tendency towards the apo sample for almost three quarters of all identified cross-links in both samples. This findings are not surprising, since observations made on the SDS gel and in the comparative cross-linking study already indicated a higher flexibility and therefore reactivity of the apo conformation.

Nevertheless, 18 different linkage sites could be highlighted that are associated with the RNA bound Cas9 structure. Among them 9 have been identified in both experiments. This double identification with similar reporter ion ratios, boosts the confidence for these cross-links to be characteristic for the riboprotein. Evaluation of these cross-links on the known structures revealed the majority of them to be located at the RuvC-Rec3-HNH interface that is formed upon gRNA binding. With one exception, they perfectly fit the reported crystal structure, whereas most of them are not in agreement with the apo Cas9 homology model or are connecting lysine residues significantly more far apart in the apo state. (Fig. 12, Suppl. Tab. 2) The one cross-link violating the RNP-crystal structure (marked with a black arrow in Fig. 12) connects two solvent exposed lysines on the HNH- and Rec2-domain, respectively (233-772). As already discussed in 3.2.1, it is very likely, that this cross-link is formed in the gRNA bound state, since the HNH domain was observed to be quite mobile. Cross-links with an increased TMT-reporter signal for the apo sample proofed to be more difficult to evaluate. Not only the Euclidian distance between the two lysines, but also other factors, such as steric hindrance through gRNA, may contribute to a different cross-link amount. However, a substantial portion had a cross-linking distance above 30 Å in both

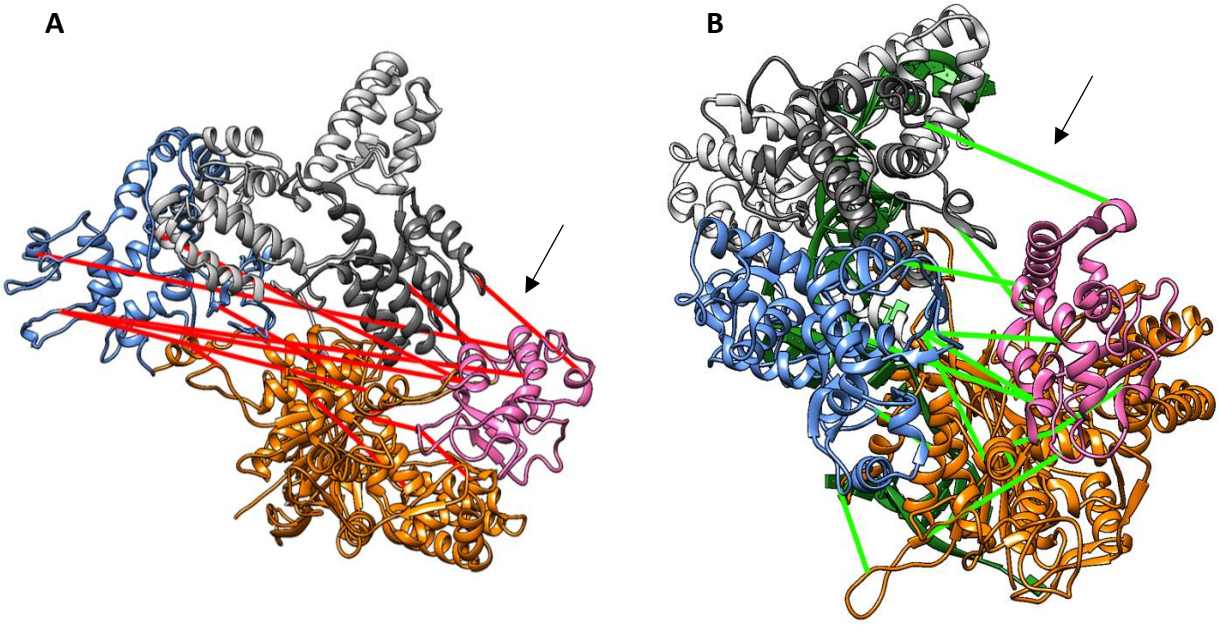


Figure 12: Identified XLs with an at least 2-fold increased reporter ion signal in the channel corresponding to the RNP sample plotted onto the apo Cas9 homology model **(A)** and the RNP crystal structure 4zt0 **(B)**, respectively.

possible conformations, again indicating increased flexibility in the nucleic acid free state. Moreover, the higher cross-linking rate of the apo protein may influence this result.

Since identified cross-links can not only be plotted onto known crystal structures, the potential of using the distance restraints derived from the cross-links referred to the gRNA bound conformation for molecular modeling was investigated. To do so, the webserver based modeling platform I-TASSER was used. The crystal structure of apo Cas9 (4cmp:A) was assigned as a template and every cross-link with an at least 2-fold increased reporter signal for the riboprotein was specified as a restraint. The generated models were compared with the RNP crystal structure (4zt0:A). While the best ranked I-TASSER model showed a high similarity to the apo crystal structure, the two lowest scoring models were highly disordered. In contrast, the models ranked second and third had a certain resemblance to the RNP structure. (Fig. 13 A/C)

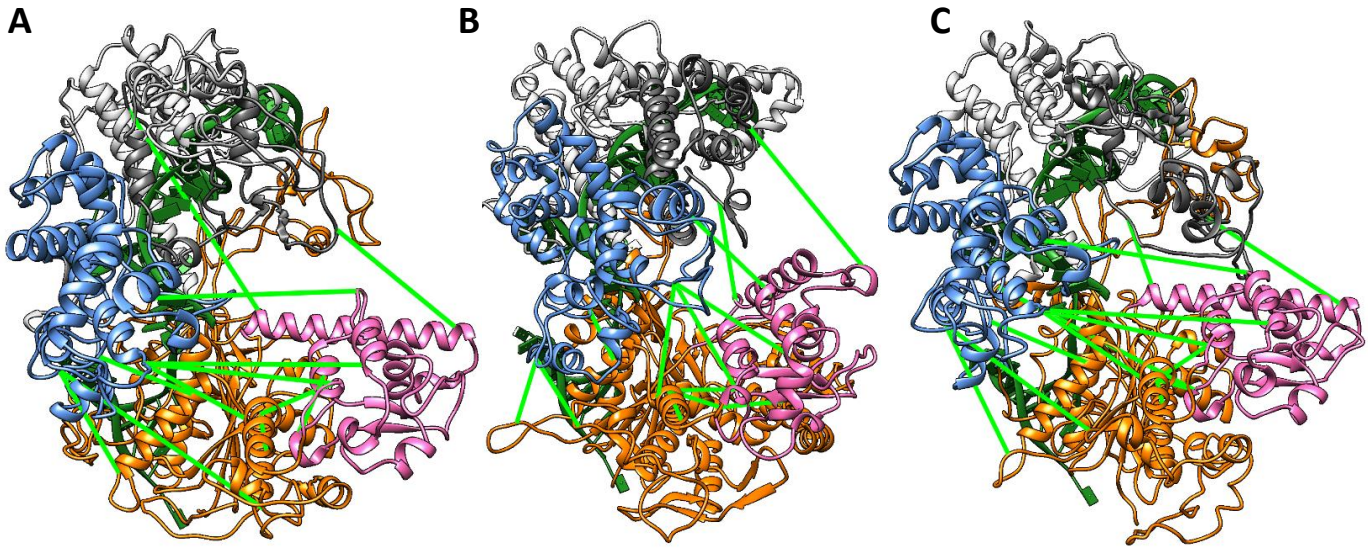


Figure 13: I-Tasser Models that are in good agreement, using 4cmp as template and 2-fold XLs from 5h experiment **(A,C)**, XLs on 4zt0 respectively **(B)**

MatchMaker calculated a RMSD of 24.6 and 21.5 Angstroms for model A and C, respectively. Measurement of the Euclidian distance between the lysine residues defined as distance restrains, revealed that most of them would represent over-length cross-links. Anyway, the C_α-C_α distance of these links was reduced at least by the factor two compared to the apo structure. Closer evaluation of the structural model however showed, that the distinct domains show a high similarity to those of the apo structure.

Taken together, this approach allowed the correct spatial arrangement of the six Cas9 domains, although local structures were partially completely different. Still, since a quite simple modeling workflow was used, these results are very remarkable. In comparison, a recent study using a more sophisticated modelling approach (Integrative Modeling Platform -IMP) achieved an average RMSD of 15 Å for a protein of comparable size.[99]

To obtain a more confident result, and to proof that this experiment is reproducible as a full TMTsixplex™ experiment, 3 biological replicates of apo Cas9 and of sgRNA bound Cas9 were reconstituted, cross-linked and prepared as described in section 2. Each replicate was labeled with another isobaric reagent and was checked individually for labeling efficiency. After quenching, the samples were mixed according to their UV-spectra on the monolith. In a first run, the TMT-labeled peptide mixture was analyzed on the mass spectrometer applying the same settings as before and a 3h HPLC gradient. From this experiment a total of 115 cross-linked spectrum matches corresponding to 57 unique cross-linked sites could be identified. The even lower number of identified cross-links compared to the first experiment, where only two experiments without replicates were compared, may be explained by the higher complexity. Nevertheless, 115 spectra represent only 3.4% of all identified spectra (cross-links, loop-links, mono-links, regular PSMs).
After normalization, the fold-change and p-value of the identified cross-links was calculated, and they were illustrated in a volcano plot (Suppl. Fig. 2, Suppl. Tab. 3). In this plot several linked residues show a significant fold-change for the RNP conformation. The most confident of them highlight once more the Rec3-HNH-RuvC interface to be newly formed after gRNA binding. The cross-links that were associated with the apo conformation, mostly involve the Rec-1 domain. This is in accordance with the observations in the apo crystal structure, where this domain is located closer to the RuvC and HNH domains compared to the riboprotein structure. Comparison of the cross-linking distance and structure fitting shows, that they are in better agreement with the apo structure. (see Sup. Tab. 1 and Suppl. Fig. 3 & 4) However, for a protein like Cas9 there are not enough cross-linked residues to seriously predict structural changes or as input for molecular modeling. To increase the number of IDs, SEC enrichment was performed, and the most promising fractions were analyzed on the mass spectrometer. (Fractions F2 & F3, Figure 14).

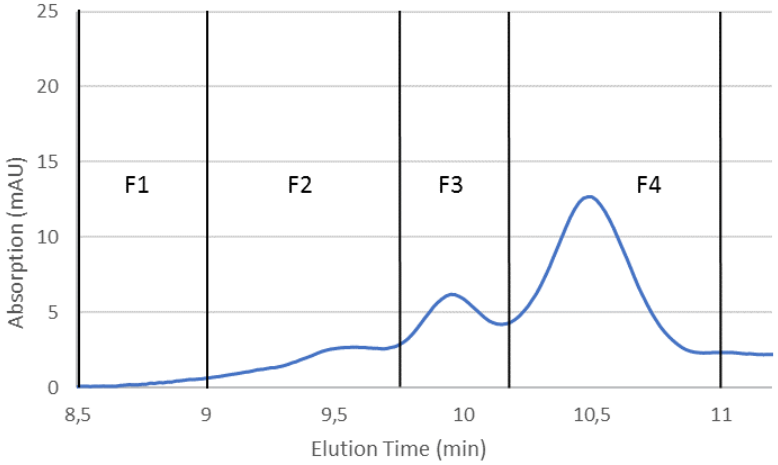


Figure 14: SEC UV/Vis-Chromatogram of the TMT-sixplex experiment

Overall 235 unique cross-linked peptide pairs corresponding to 633 cross-link spectrum matches could be identified in fraction F2 (see Sup. Tab. 4), while F3 contained only regular

peptides, loop- and mono-links. The Vulcano plot of the identified and quantified cross-links emphasizes many links that are preferentially formed in the gRNA bound state. (Fig. 15) Again, the cross-links were evaluated on the known crystal structure and showed that

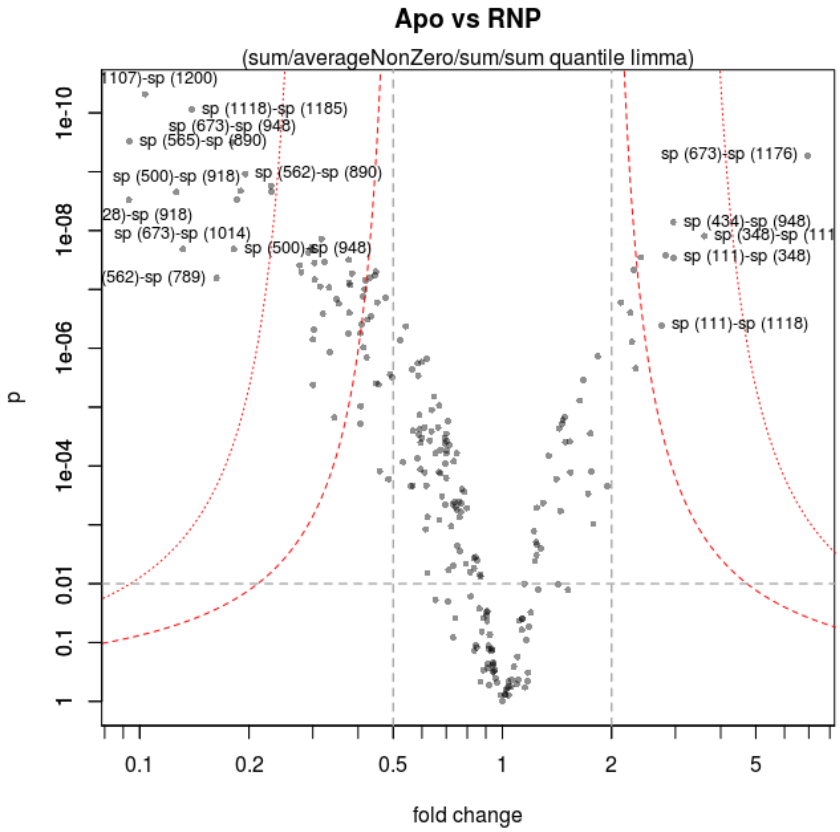


Figure 15: Vulcano Plot of the cross-links identified and quantified from the SEC enriched 6-plex experiment

10 among them are formed between the HNH/RuvC and Rec3 domains. Euclidian distance measurements showed good correlation between fold change and distance difference in the different conformations. Some links were slightly "over-length" but could be explained with known dynamics of Cas9, as elaborated before.[92,96]

The cross-links associated with the riboprotein-structure (min. 3-fold reporter signal change and. p-value ≤ $1e^{-6}$) were used to create another model with I-TASSER. In addition to the apo-structure, also the crystal structures of the two homologous proteins S. aureus Cas9 and F. noviciade Cas9 were used for modeling. Out of the 15 created models (Suppl. Fig. 5-7), 3 had less than 5 overlength cross-links (4 cross-links are overlength on the crystal structure itself) and showed a coherent structure. The model created with F. novicida as template that was scored the fifth best by I-TASSER had only two overlength connections of 31 Å and 33 Å, respectively. Based on the defined selection criteria this model ("final model") was chosen as final structure. In comparison with the S. pyogenes Cas9 X-ray structure, this model had a RMSD of 7.9 Å. Apart from the overall structure, also the distinct domains had a good overlap with the crystal structure and showed RMSDs from 5.7 Å to 8.1 Å. Moreover, the structure would even allow correct nucleic acid binding, as it contains

channels for both gRNA and the target DNA. In Figure 16 the final model and the crystal structure are compared next to each other. Apart from some structural elements, the two structures seem very similar, proving a correct arrangement of the different domains. This outcome highlights the potential of the quantitative cross-linking mass spectrometry workflow presented herein.

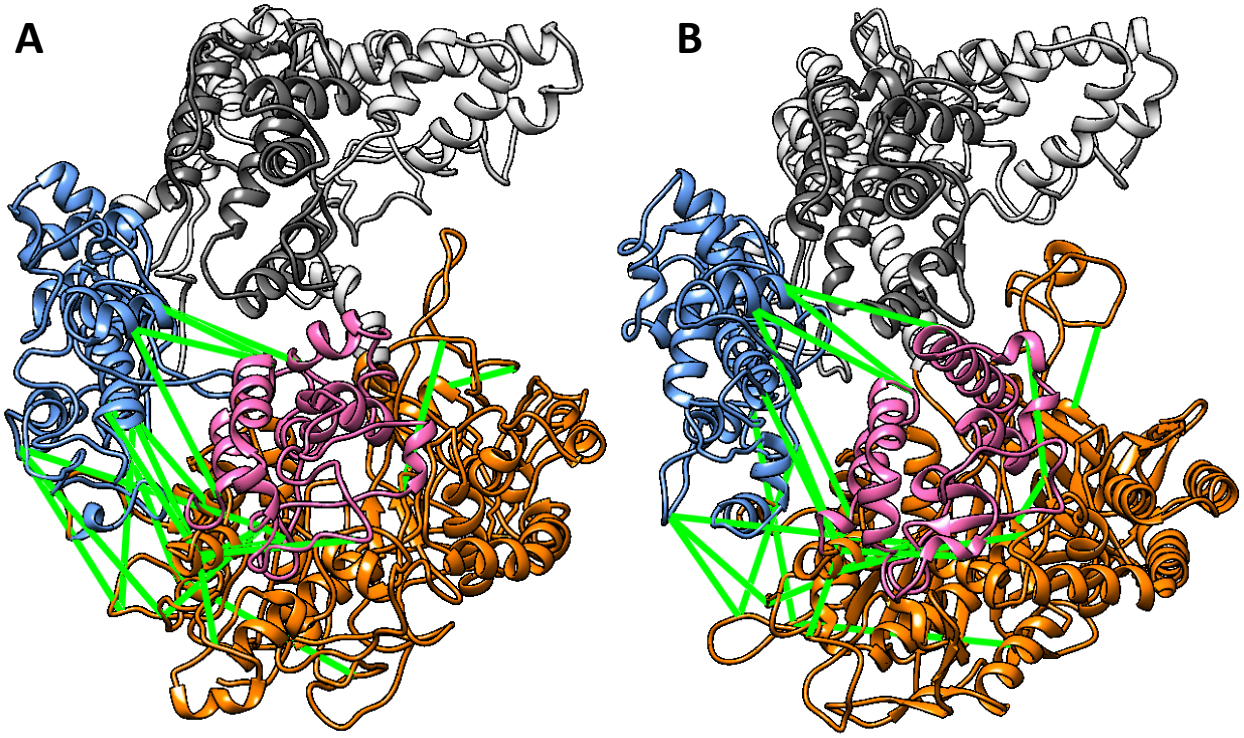


Figure 16: Best structural model generated wit I-TASSER (A) and riboprotein crystal structure (B). Cross-links used for modelling are depicted in green.

4. Discussion and Outlook

Over the last decade, XLMS has become a powerful technique in structural biology, but is not yet grown up. There are large controversies about how to conduct cross-linking, how to measure samples and how to analyze and interpret them.
While several pioneers of cross-linking mass spectrometry stick to well established non-cleavable cross-linker, other groups prefer MS-cleavable cross-linker as their unique properties allow a less ambiguous identification of cross-linked peptides. The fraction preferring MS-cleavable cross-linker is further divided into proponents of multistep MS workflows (e.g. MS^2[CID]-MS^2[ETD] or MS^2[CID]-MS^3[CID]) and supporters of simple MS^2 based methods. Every group promotes their own strategies and until now there are not enough objective and independent studies that systematically compare different methods. Search engines and their algorithms represent an even greater challenge. Some software are developed to identify as many cross-links as possible and thereby produce many false positives. Recent studies highlighted that the real FDR for cross-linked spectrum matched and especially for unique cross-links is way higher than expected.[79,100] Hence, results from cross-linking experiments may be less trustworthy than other proteomics experiments and require manual validation which is very time consuming. To date the field is missing a gold-standard, like MaxQuant or Mascot are for classical proteomics.
Nevertheless, cross-linking mass spectrometry is a valuable technique for gaining structural information about proteins, when other approaches fail. Moreover, it is a powerful extension for interactome studies, as cross-linker can stabilize transient interactions. Additionally, details about the binding interface may be discovered whereas other methods may only identify the binding partners.

In this thesis, a novel and simple quantitative approach for cross-linking mass spectrometry was developed and tested on a well described protein. The obtained results support previous findings, proofing this method to produce a both reliable and meaningful outcome. Since the quantitative information dramatically reduces the number of cross-links associated with a certain protein conformation, subsequent modeling and scoring of the obtained models is simplified. However, some protein regions may not be captured by the cross-links that show a certain fold-change. Therefore, this strategy is best suited for proteins that undergo significant structural rearrangements upon stimulation and specially to identify the regions that are affected the most. TMT-based quantification offers several advantages like the possibility to compare up to 11 conditions in one experiment and to use prefractionation techniques. But it always comes along with some drawbacks. Apart from the significantly decreased number of IDs, also the quantitative values do not need to be exact. Especially, cross-links are often low abundant and co-elute with other isobaric peptides leading to ion suppression and inaccurate ratios.

Nonetheless, the results identified the Rec3 to be the most affected from structural rearrangement upon gRNA binding. Using the XL-derived distance restraints and a simple web-based modeling workflow allowed to create an appropriate three-dimensional model. Apart from the appropriate structural model of the riboprotein, the highly dynamic and flexible conformation of the apo-protein could be confirmed.

In the future multiplexed quantitative cross-linking mass spectrometry will help solving many questions concerning the dynamic behavior of large protein machines like the proteasome or spliceosome.

5. References

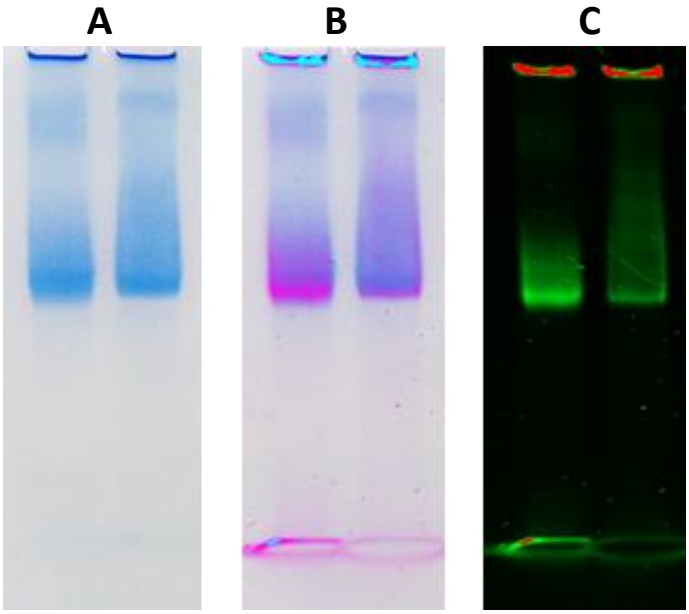
(1) Sanger, F. *Biochem. J.* **1949**, *45* (5), 563.
(2) Edman, P.; Begg, G. *Eur. J. Biochem.* **1967**, *1* (1), 80.
(3) Rehm, H.; Letzel, T. *Der Experimentator: Proteinbiochemie/Proteomics*; 2010.
(4) Aebersold, R.; Mann, M. *Nature* **2003**, *422* (6928), 198.
(5) Yu, C.; Huang, L. *Anal. Chem.* **2017**, acs. analchem.7b04431.
(6) Lovrić, J. *Introducing proteomics : from concepts to sample separation, mass spectrometry and data analysis*; Wiley-Blackwell, 2011.
(7) *MALDI MS: A Practical Guide to Instrumentation, Methods and Applications*; Hillenkamp, F., Peter-Katalinic, J., Eds.; Wiley-VCH: Weinheim, 2007.
(8) Greenberg, A. *Twentieth-Century Science. Chemistry: Decade by Decade*, 1 edition.; Cannon, W. J., Ed.; Facts on File: New York, 2007.
(9) Banerjee, S.; Mazumdar, S. *Int. J. Anal. Chem.* **2012**, *2012*, 1.
(10) Michalski, A.; Damoc, E.; Hauschild, J.-P.; Lange, O.; Wieghaus, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S. *Mol. Cell. Proteomics* **2011**, *10* (9), M111.011015.
(11) Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J. *Mol. Cell. Proteomics* **2012**, *11* (11), 1475.
(12) Makarov, A. *Anal. Chem.* **2000**, *72*, 1156.
(13) Zubarev, R. A.; Makarov, A. *Anal. Chem.* **2013**, *85* (11), 5288.
(14) Scheltema, R. A.; Hauschild, J.-P.; Lange, O.; Hornburg, D.; Denisov, E.; Damoc, E.; Kuehn, A.; Makarov, A.; Mann, M. *Mol. Cell. Proteomics* **2014**, *13* (12), 3698.
(15) Bantscheff, M.; Lemeer, S.; Savitski, M. M.; Kuster, B. *Anal. Bioanal. Chem.* **2012**, *404* (4), 939.
(16) Vidova, V.; Spacil, Z. *Anal. Chim. Acta* **2017**, *964*, 7.
(17) Schubert, O. T.; Röst, H. L.; Collins, B. C.; Rosenberger, G.; Aebersold, R. *Nat. Protoc.* **2017**, *12* (7), 1289.
(18) Pichler, P.; Köcher, T.; Holzmann, J.; Mazanek, M.; Taus, T.; Ammerer, G.; Mechtler, K. *Anal. Chem.* **2010**, *82* (15), 6549.
(19) Rauniyar, N.; Yates, J. R. *J. Proteome Res.* **2014**, *13* (12), 5293.
(20) Liu, F.; Rijkers, D. T. S.; Post, H.; Heck, A. J. R. *Nat. Methods* **2015**, *12* (12), 1179.
(21) Sinz, A.; Arlt, C.; Chorev, D.; Sharon, M. *Protein Sci.* **2015**, *24* (8), 1193.
(22) Rutsdottir, G.; I. Rasmussen, M.; Hojrup, P.; Bernfur, K.; Emanuelsson, C.; Söderberg, C. A. G. *Proteins Struct. Funct. Bioinforma.* **2018**, *86* (1), 110.
(23) Zahn, H. *Angew. Chemie* **1955**, *67* (19–20), 561.
(24) Berg, H. C.; Diamond, J. M.; Marfey, P. S. *Science (80-.).* **1965**, *150* (3692).
(25) Clegg, C.; Hayes, D. *Eur. J. Biochem.* **1974**, *42* (1), 21.
(26) Rappsilber, J.; Siniossoglou, S.; Hurt, E. C.; Mann, M. *Anal. Chem.* **2000**, *72* (2), 267.
(27) Young, M. M.; Tang, N.; Hempel, J. C.; Oshiro, C. M.; Taylor, E. W.; Kuntz, I. D.; Gibson, B. W.; Dollinger, G. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (11), 5802.
(28) 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. *EMBO J.* **2010**, *29* (4), 717.
(29) Guerrero, C.; Milenkovic, T.; Przulj, N.; Kaiser, P.; Huang, L. *Proc. Natl. Acad. Sci. U.*

S. A. **2008**, *105* (36), 13333.
(30) Lauber, M. A.; Reilly, J. P. *J. Proteome Res.* **2011**, *10* (8), 3604.
(31) Chao, W. C. H.; Murayama, Y.; Muñoz, S.; Jones, A. W.; Wade, B. O.; Purkiss, A. G.; Hu, X.-W.; Borg, A.; Snijders, A. P.; Uhlmann, F.; Singleton, M. R. *Nat. Commun.* **2017**, *8*.
(32) Tang, X.; Bruce, J. E. *Mol. Biosyst.* **2010**, *6* (6), 939.
(33) Kandur, W. V; Kao, A.; Vellucci, D.; Huang, L.; Rychnovsky, S. D. *Org. Biomol. Chem.* **2015**, *13* (38), 9793.
(34) Liu, F.; Lössl, P.; Scheltema, R.; Viner, R.; Heck, A. J. R. *Nat. Commun.* **2017**, *8*, 15473.
(35) Schweppe, D. K.; Chavez, J. D.; Lee, C. F.; Caudal, A.; Kruse, S. E.; Stuppard, R.; Marcinek, D. J.; Shadel, G. S.; Tian, R.; Bruce, J. E. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (7), 1732.
(36) Rappsilber, J. *J. Struct. Biol.* **2011**, *173* (3), 530.
(37) *Thermo Scientific Crosslinking Technical Handbook*; 2012.
(38) Chen, Z.; Stokes, D. L.; Rice, W. J.; Jones, L. R. *J. Biol. Chem.* **2003**, *278* (48), 48348.
(39) Leitner, A.; Joachimiak, L. A.; Unverdorben, P.; Walzthoeni, T.; Frydman, J.; Förster, F.; Aebersold, R. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (26), 9455.
(40) Zhang, X.; Wang, J.-H.; Tan, D.; Li, Q.; Li, M.; Gong, Z.; Tang, C.; Liu, Z.; Dong, M.-Q.; Lei, X. *Anal. Chem.* **2018**, *90* (2), 1195.
(41) Lössl, P.; Sinz, A. Humana Press, New York, NY, 2016; pp 109–127.
(42) Belsom, A.; Mudd, G.; Giese, S.; Auer, M.; Rappsilber, J. *Anal. Chem.* **2017**, *89* (10), 5319.
(43) Bhattacharjya, S.; Balaram, P. *Proteins* **1997**, *29* (4), 492.
(44) Kolbowski, L.; Mendes, M. L.; Rappsilber, J. *Anal. Chem.* **2017**, *89* (10), 5311.
(45) Schmidt, R.; Sinz, A. *Anal. Bioanal. Chem.* **2017**.
(46) Schilling, B.; Row, R. H.; Gibson, B. W.; Guo, X.; Young, M. M. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 834.
(47) Iacobucci, C.; Götze, M.; Hage, C.; Arlt, C.; Ihling, C. H.; Piotrowski, C.; Sinz, A. *under Rev.* **2018**.
(48) Leitner, A.; Reischl, R.; Walzthoeni, T.; Herzog, F.; Bohn, S.; Förster, F.; Aebersold, R. *Mol. Cell. Proteomics* **2012**, *11* (3), M111.014126.
(49) Tinnefeld, V.; Venne, A. S.; Sickmann, A.; Zahedi, R. P. *J. Proteome Res.* **2017**, *16* (2), 459.
(50) Vellucci, D.; Kao, A.; Kaake, R. M.; Rychnovsky, S. D.; Huang, L. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (8), 1432.
(51) Burke, A. M.; Kandur, W.; Novitsky, E. J.; Kaake, R. M.; Yu, C.; Kao, A.; Vellucci, D.; Huang, L.; Rychnovsky, S. D. *Org. Biomol. Chem.* **2015**, *13* (17), 5030.
(52) Petrotchenko, E. V; Serpa, J. J.; Borchers, C. H. *Mol. Cell. Proteomics* **2011**, *10* (2), M110.001420.
(53) Chen, Z. A.; Fischer, L.; Cox, J.; Rappsilber, J. *Mol. Cell. Proteomics* **2016**, *15* (8), 2769.
(54) Walzthoeni, T.; Joachimiak, L. A.; Rosenberger, G.; Röst, H. L.; Malmström, L.; Leitner, A.; Frydman, J.; Aebersold, R. *Nat. Methods* **2015**, *12*, 1185.
(55) Yu, C.; Huszagh, A.; Viner, R.; Novitsky, E. J.; Rychnovsky, S. D.; Huang, L. *Anal. Chem.* **2016**, *88* (20), 10301.
(56) Chavez, J. D.; Schweppe, D. K.; Eng, J. K.; Zheng, C.; Taipale, A.; Zhang, Y.; Takara, K.;

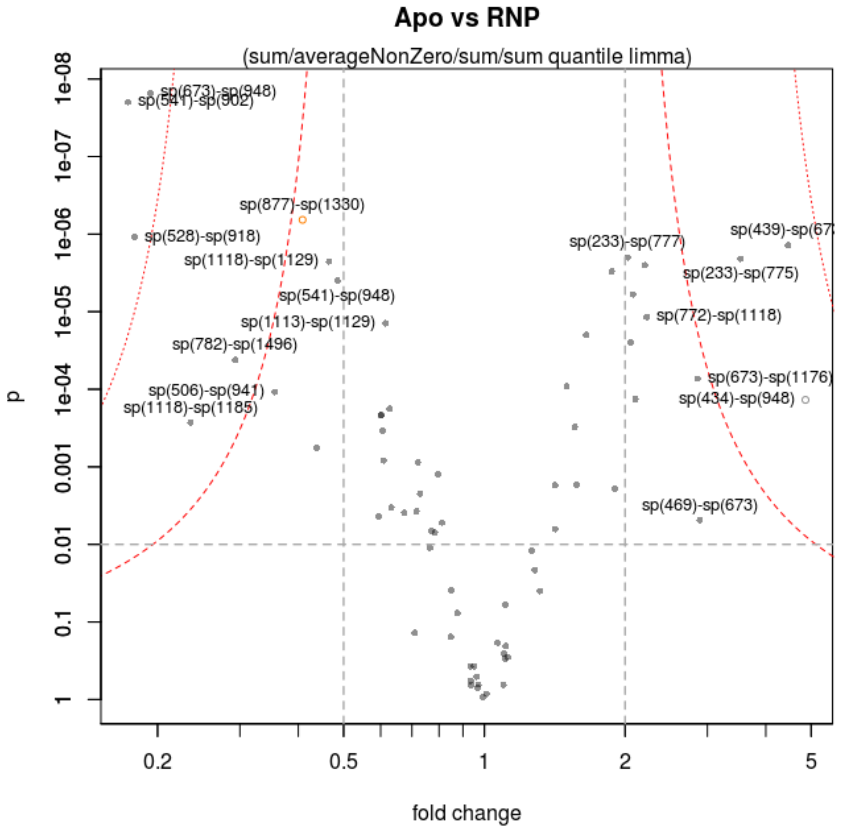
Bruce, J. E. *Nat. Commun.* **2015**, *6*, 7928.
(57) Chavez, J. D.; Eng, J. K.; Schweppe, D. K.; Cilia, M.; Rivera, K.; Zhong, X.; Wu, X.; Allen, T.; Khurgel, M.; Kumar, A.; Lampropoulos, A.; Larsson, M.; Maity, S.; Morozov, Y.; Pathmasiri, W.; Perez-Neut, M.; Pineyro-Ruiz, C.; Polina, E.; Post, S.; Rider, M.; Tokmina-Roszyk, D.; Tyson, K.; Vieira Parrine Sant'Ana, D.; Bruce, J. E. *PLoS One* **2016**, *11* (12), e0167547.
(58) Giese, S. H.; Fischer, L.; Rappsilber, J. *Mol. Cell. Proteomics* **2016**, *15* (3), 1094.
(59) Yang, B.; Wu, Y.-J.; Zhu, M.; Fan, S.-B.; Lin, J.; Zhang, K.; Li, S.; Chi, H.; Li, Y.-X.; Chen, H.-F.; Luo, S.-K.; Ding, Y.-H.; Wang, L.-H.; Hao, Z.; Xiu, L.-Y.; Chen, S.; Ye, K.; He, S.-M.; Dong, M.-Q. .
(60) McIlwain, S.; Tamura, K.; Kertesz-Farkas, A.; Grant, C. E.; Diament, B.; Frewen, B.; Howbert, J. J.; Hoopmann, M. R.; Käll, L.; Eng, J. K.; MacCoss, M. J.; Noble, W. S. *J. Proteome Res.* **2014**, *13* (10), 4488.
(61) Götze, M.; Pettelkau, J.; Schaks, S.; Bosse, K.; Ihling, C. H.; Krauth, F.; Fritzsche, R.; Kühn, U.; Sinz, A. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (1), 76.
(62) Götze, M.; Pettelkau, J.; Fritzsche, R.; Ihling, C. H.; Schäfer, M.; Sinz, A. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (1), 83.
(63) Iglesias, A. H.; Santos, L. F. A.; Gozzo, F. C. *Anal. Chem.* **2010**, *82* (3), 909.
(64) Trnka, M. J.; Baker, P. R.; Robinson, P. J. J.; Burlingame, A. L.; Chalkley, R. J. *Mol. Cell. Proteomics* **2014**, *13* (2), 420.
(65) Rinner, O.; Seebacher, J.; Walzthoeni, T.; Mueller, L. N.; Beck, M.; Schmidt, A.; Mueller, M.; Aebersold, R. *Nat. Methods* **2008**, *5*, 315.
(66) Mayne, S. L. N.; Patterton, H.-G. *Brief. Bioinform.* **2011**, *12* (6), 660.
(67) Walzthoeni, T.; Claassen, M.; Leitner, A.; Herzog, F.; Bohn, S.; Förster, F.; Beck, M.; Aebersold, R. *Nat. Methods* **2012**, *9*, 901.
(68) Wang, Z.; Schey, K. L. *Exp. Eye Res.* **2017**, *159*, 23.
(69) Armony, G.; Jacob, E.; Moran, T.; Levin, Y.; Mehlman, T.; Levy, Y.; Fass, D. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (47), 13384.
(70) Heuck, A.; Schitter-Sollner, S.; Suskiewicz, M. J.; Kurzbauer, R.; Kley, J.; Schleiffer, A.; Rombaut, P.; Herzog, F.; Clausen, T. *Elife* **2016**, *5*, e21516.
(71) Bouguenina, H.; Salaun, D.; Mangon, A.; Muller, L.; Baudelet, E.; Camoin, L.; Tachibana, T.; Cianférani, S.; Audebert, S.; Verdier-Pinard, P.; Badache, A. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (50), E10687.
(72) Schilbach, S.; Hantsche, M.; Tegunov, D.; Dienemann, C.; Wigge, C.; Urlaub, H.; Cramer, P. *Nature* **2017**, *551* (7679), 204.
(73) Hoopmann, M. R.; Zelter, A.; Johnson, R. S.; Riffle, M.; MacCoss, M. J.; Davis, T. N.; Moritz, R. L. *J. Proteome Res.* **2015**, *14* (5), 2190.
(74) Berger, S.; Procko, E.; Margineantu, D.; Lee, E. F.; Shen, B. W.; Zelter, A.; Silva, D.-A.; Chawla, K.; Herold, M. J.; Garnier, J.-M.; Johnson, R.; MacCoss, M. J.; Lessene, G.; Davis, T. N.; Stayton, P. S.; Stoddard, B. L.; Fairlie, W. D.; Hockenbery, D. M.; Baker, D. *Elife* **2016**, *5*, e20352.
(75) Kim, J. ook; Zelter, A.; Umbreit, N. T.; Bollozos, A.; Riffle, M.; Johnson, R.; MacCoss, M. J.; Asbury, C. L.; Davis, T. N. *Elife* **2017**, *6*, e21069.
(76) Ren, Z.; Ranganathan, S.; Zinnel, N. F.; Russell, W. K.; Russell, D. H.; Raushel, F. M. *Biochemistry* **2015**, *54* (21), 3400.
(77) Schacherl, M.; Gompert, M.; Pardon, E.; Lamkemeyer, T.; Steyaert, J.; Baumann, U. *Biochim. Biophys. Acta - Biomembr.* **2017**, *1859* (10), 1859.

(78) Halder, S.; Parrell, D.; Whitten, D.; Feig, M.; Kroos, L. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (50), E10677.
(79) Gossenreiter, T.; Hartl, M. In *7th Symposium on Structural Proteomics*; 2017.
(80) Lima, D. B.; de Lima, T. B.; Balbuena, T. S.; Neves-Ferreira, A. G. C.; Barbosa, V. C.; Gozzo, F. C.; Carvalho, P. C. *J. Proteomics* **2015**, *129*, 51.
(81) Iacobucci, C.; Sinz, A. *Anal. Chem.* **2017**, *89* (15), 7832.
(82) Koonin, E. V; Makarova, K. S. *RNA Biology*. Taylor & Francis May 2013, pp 679–686.
(83) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. *Science (80-.).* **2012**, *337* (6096).
(84) Larson, M. H.; Gilbert, L. A.; Wang, X.; Lim, W. A.; Weissman, J. S.; Qi, L. S. *Nat. Protoc.* **2013**, *8* (11), 2180.
(85) Deng, W.; Shi, X.; Tjian, R.; Lionnet, T.; Singer, R. H. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (38), 11870.
(86) Liszczak, G. P.; Brown, Z. Z.; Kim, S. H.; Oslund, R. C.; David, Y.; Muir, T. W. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (4), 681.
(87) Jinek, M.; Jiang, F.; Taylor, D. W.; Sternberg, S. H.; Kaya, E.; Ma, E.; Anders, C.; Hauer, M.; Zhou, K.; Lin, S.; Kaplan, M.; Iavarone, A. T.; Charpentier, E.; Nogales, E.; Doudna, J. A. *Science (80-.).* **2014**, *343* (6176).
(88) Nishimasu, H.; Ran, F. A.; Hsu, P. D.; Konermann, S.; Shehata, S. I.; Dohmae, N.; Ishitani, R.; Zhang, F.; Nureki, O. *Cell* **2014**, *156* (5), 935.
(89) Sternberg, S. H.; LaFrance, B.; Kaplan, M.; Doudna, J. A. *Nature* **2015**, *527* (7576), 110.
(90) Dagdas, Y. S.; Chen, J. S.; Sternberg, S. H.; Doudna, J. A.; Yildiz, A. *Sci. Adv.* **2017**, *3* (8), eaao0027.
(91) Zuo, Z.; Liu, J. *Sci. Rep.* **2017**, *7* (1), 17271.
(92) Yang, M.; Peng, S.; Sun, R.; Lin, J.; Wang, N.; Correspondence, C. C. *CellReports* **2018**, *22*, 372.
(93) Shibata, M.; Nishimasu, H.; Kodera, N.; Hirano, S.; Ando, T.; Uchihashi, T.; Nureki, O. *Nat. Commun.* **2017**, *8* (1), 1430.
(94) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. *Nat. Methods* **2015**, *12* (1), 7.
(95) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25* (13), 1605.
(96) Huai, C.; Li, G.; Yao, R.; Zhang, Y.; Cao, M.; Kong, L.; Jia, C.; Yuan, H.; Chen, H.; Lu, D.; Huang, Q. *Nat. Commun.* **2017**, *8* (1), 1375.
(97) Rampler, E.; Stranzl, T.; Orban-Nemeth, Z.; Hollenstein, D. M.; Hudecz, O.; Schlögelhofer, P.; Mechtler, K. *J. Proteome Res.* **2015**, *14* (12), 5048.
(98) Köcher, T.; Pichler, P.; Swart, R.; Mechtler, K. *Nat. Protoc.* **2012**, *7* (5), 882.
(99) Chen, Z. A.; Pellarin, R.; Fischer, L.; Sali, A.; Nilges, M.; Barlow, P. N.; Rappsilber, J. *Mol. Cell. Proteomics* **2016**, *15* (8), 2730.
(100) Fischer, L.; Rappsilber, J. *Anal. Chem.* **2017**, *89* (7), 3829.
(101) Thermo Fisher Scientific Inc. Higher-Quality Data, Faster Than Ever Thermo Scientific Q Exactive HF Orbitrap LC-MS/MS System https://assets.thermofisher.com/TFS-Assets/CMD/brochures/BR-64052-LC-MS-Q-Exactive-HF-Orbitrap-BR64052-EN.pdf (accessed Mar 7, 2018).

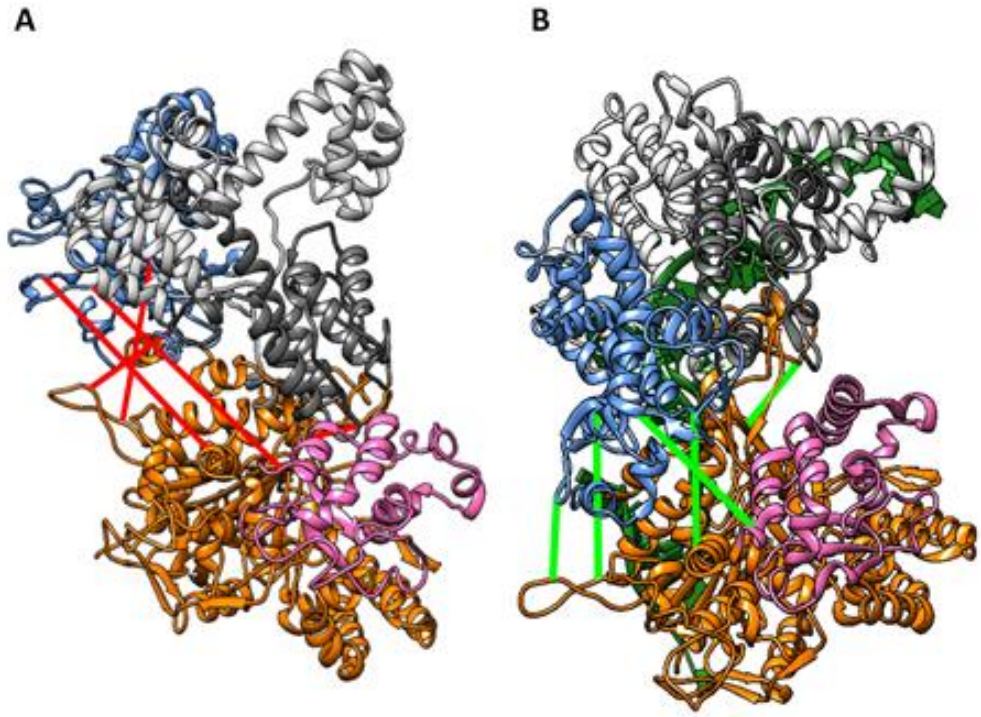
Supplementary Information



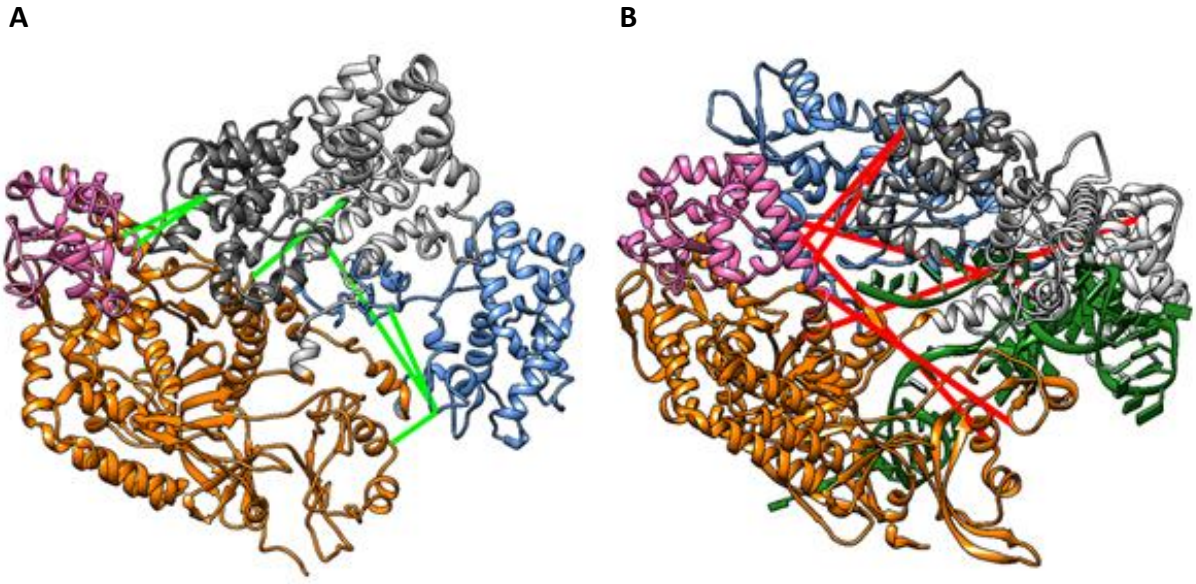
Supplementary Figure 1: 8% native PAGE stained with Coomassie (A), SybrSafe (C) and the overlay (B). The right lane corresponds to Cas9 in complex with gRNA, the left lane represents the holo-complex.



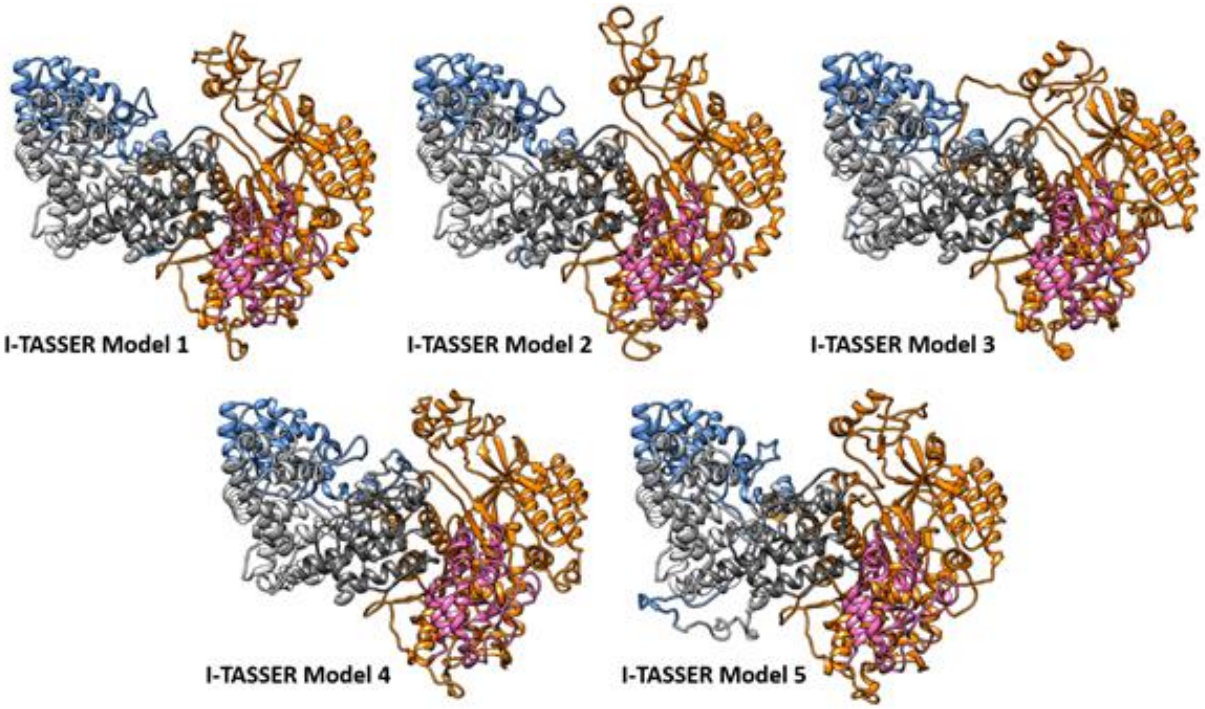
Supplementary Figure 2: Vulcano Plot of the identified cross-links from the 6-plex experiment, recorded with a 3h HPLC gradient



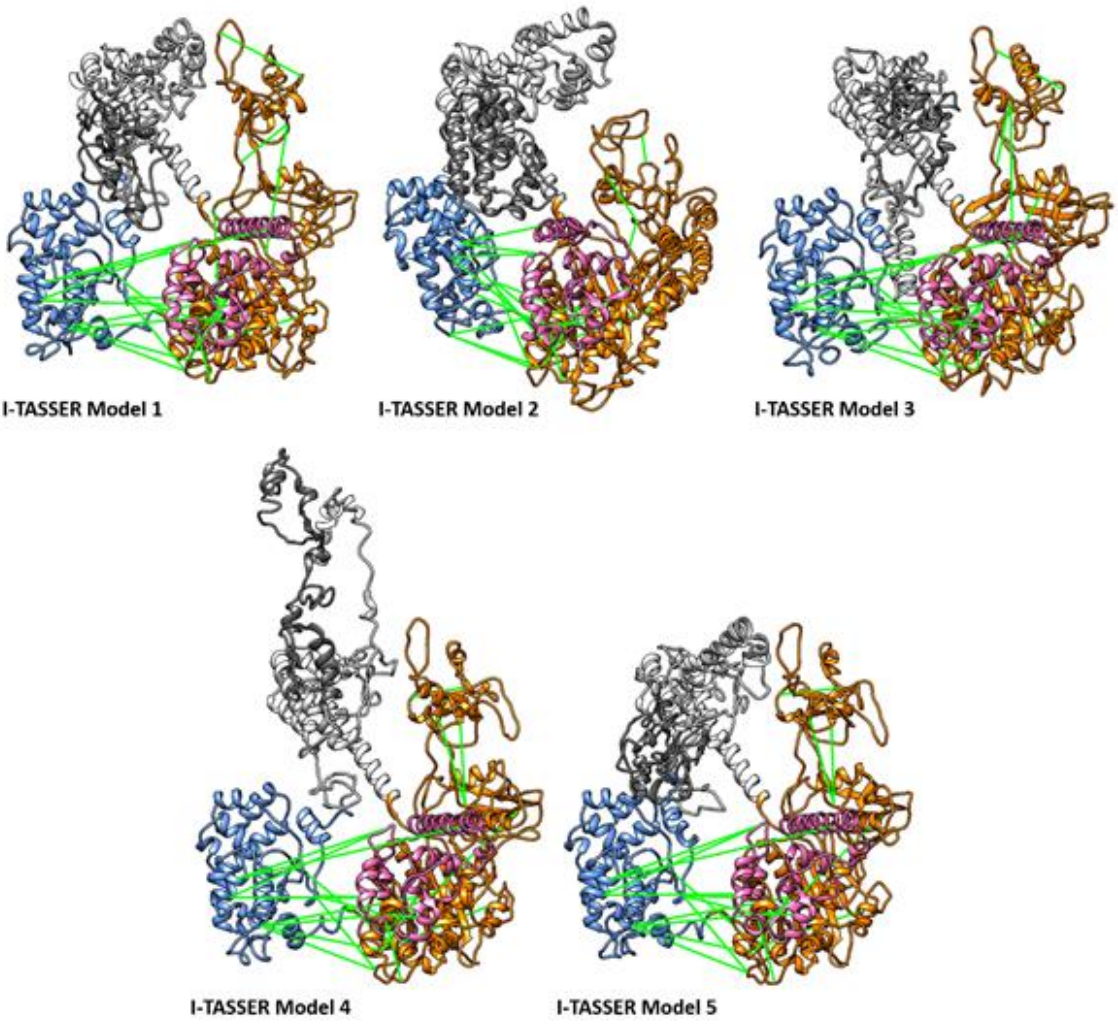
Supplementary Figure 3: Identified XLs with a more than 3-fold increased abundance in the RNP sample plotted onto the apo Cas9 homology model **(A)** and the RNP crystal structure 4zt0 **(B)**, respectively.



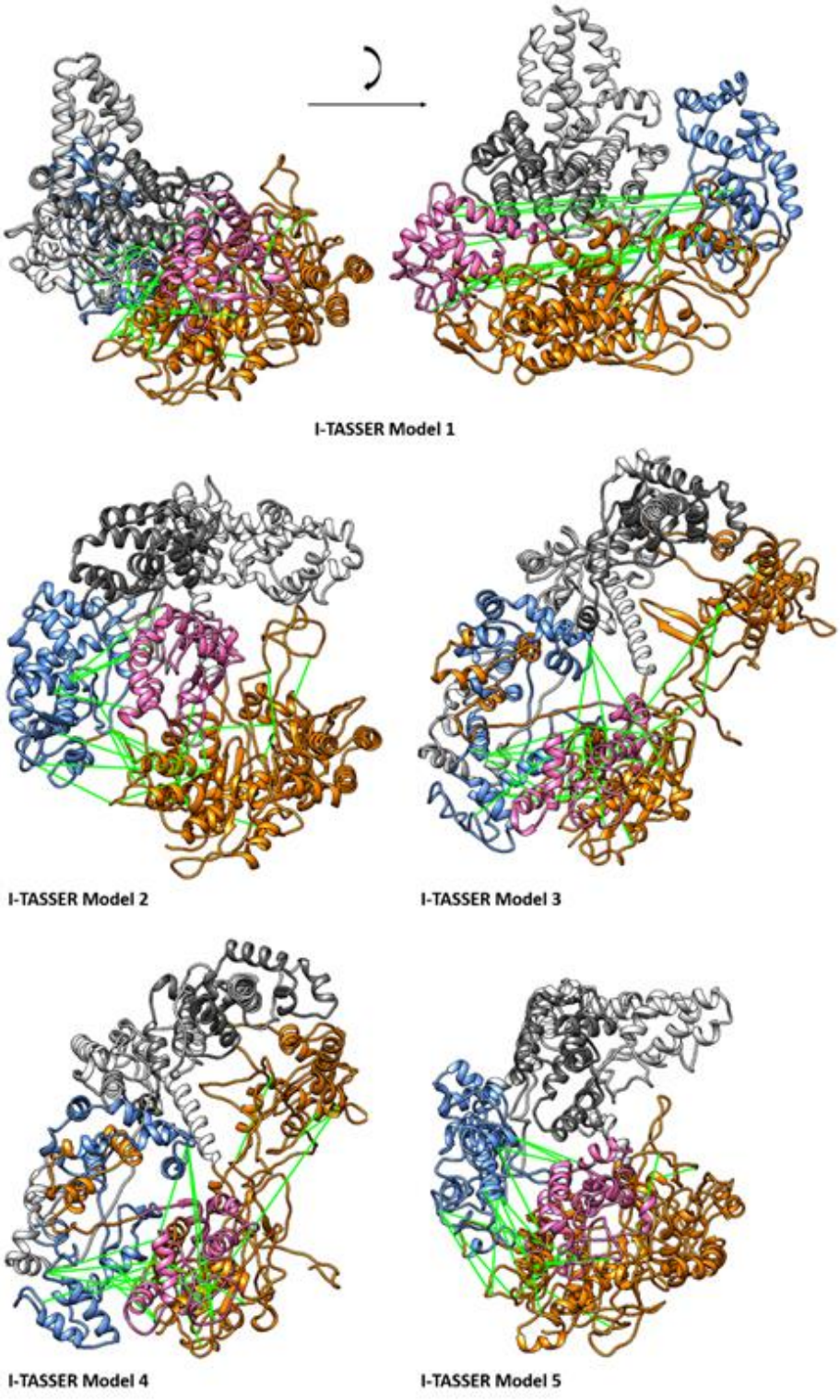
Supplementary Figure 4: Identified XLs with a more than 3-fold increased abundance in the Apo sample plotted onto the apo Cas9 homology model **(A)** and the RNP crystal structure 4zt0 **(B)**, respectively.



Supplementary Figure 5: Top 5 models, predicted by I-TASSER using the apo crystal structure of S. pyogenes Cas9 as homology template.



Supplementary Figure 6: Top 5 models, predicted by I-TASSER using S. aureus Cas9 as homology template. The cross-links used as distance restraints are depicted as green lines.



Supplementary Figure 7: Top 5 models, predicted by I-TASSER using F. novicida Cas9 as homology template. The cross-links used as distance restraints are depicted as green lines.

Supplementary Table 1: Cross-links identified by at least 2 CSMs in the comparative XL experiment.

Fund in sample	XL Position 1	XL Position 2	Apo CSMs	RNP CSMs
Apo	33	44	4	
Shared	33	30	8	3
Apo	33	442	2	
Apo	33	673	4	
Apo	45	33	6	
RNP	45	1097		3
Apo	111	348	4	
Apo	111	665	6	
Apo	111	1113	17	
Apo	111	1118	9	
Apo	112	348	2	
Apo	112	665	7	
Apo	131	141	10	
Apo	131	148	2	
Apo	148	536	6	
Shared	233	890	11	2
Shared	233	772	7	3
Apo	233	348	4	
RNP	263	775		4
Apo	263	377	6	
Apo	263	1113	22	
Apo	263	1118	50	
Apo	314	141	6	
Apo	323	218	5	
Apo	336	112	6	
Apo	348	112	7	
Apo	348	558	2	
Apo	348	1118	3	
RNP	377	218		4
Shared	382	346	7	18
Apo	434	439	9	
Apo	442	948	7	
Apo	468	314	10	
Apo	468	500	5	
Apo	468	665	3	
Apo	468	673	4	
Apo	468	33	3	
Apo	468	742	9	
Apo	468	749	4	
Apo	484	439	3	
Apo	484	442	17	
Apo	484	500	2	
Apo	484	536	3	
Apo	484	665	106	
RNP	500	1020		2
Apo	500	439	3	
Apo	500	942	3	
Apo	500	1192	3	
Shared	506	665	8	6
RNP	506	442		3
Apo	506	1118	3	
Apo	506	1192	2	
Apo	510	500	15	
Apo	510	558	15	
RNP	528	880		5
RNP	528	1020		2
Shared	528	536	12	
RNP	536	880		15
RNP	536	902		3
RNP	536	1020		9
Apo	536	439	10	
Shared	536	528		68

Apo	536	558	9	
Apo	536	33	5	
Apo	536	1118	5	
Apo	545	528	4	
Shared	570	562	4	7
Apo	571	500	4	
Apo	571	536	13	
Apo	571	562	4	
Apo	571	565	3	
Apo	599	646	2	
Apo	599	649	7	
Shared	599	558	4	6
Apo	637	649	25	
Shared	646	602	4	4
RNP	665	673		15
Apo	665	112	4	
Apo	665	1129	7	
Apo	665	1192	8	
Apo	677	536	7	
Apo	677	546	2	
Apo	684	665	5	
Shared	684	673	14	5
Apo	684	1118	4	
Apo	710	442	12	
Shared	710	665	3	19
Shared	710	673	12	4
Apo	710	742	8	
Apo	710	749	3	
Apo	710	1107	12	
Apo	742	1097	5	
Apo	742	1197	4	
RNP	749	33		3
Apo	749	665	4	
Shared	**772**	**959**	**24**	**2**
Shared	**772**	**961**	**4**	**4**
Shared	**772**	**968**	**9**	**7**
Apo	772	1113	9	
Apo	772	1118	6	
Apo	772	1129	10	
RNP	775	558		4
RNP	789	218		5
RNP	789	233		3
Shared	789	890	4	3
RNP	797	377		4
RNP	797	233		4
Apo	810	1255	4	
RNP	848	1246		2
Shared	**848**	**1255**	**24**	**7**
Apo	848	866	9	
Apo	848	902	4	
Apo	848	913	3	
Apo	848	968	6	
RNP	855	968		3
RNP	862	218		2
RNP	866	913		5
Apo	866	855	4	
RNP	866	862		17
Apo	866	348	5	
RNP	877	536		8
Shared	877	880	2	6
RNP	890	565		4
Apo	896	918	2	
RNP	902	536		8
RNP	902	918		7
RNP	902	959		2
Apo	902	1020	21	
RNP	913	902		5
RNP	913	959		2

Apo	913	1024	12	
RNP	929	500		8
RNP	929	959		3
Apo	942	442	18	
Shared	948	673	15	7
RNP	948	954		4
Apo	948	1020	12	
Apo	948	442	9	
RNP	959	775		4
RNP	959	918		8
Shared	**968**	**775**	**11**	**4**
RNP	968	959		2
Shared	1000	1031	2	4
Shared	1003	1020	24	13
Shared	1003	1031	3	8
RNP	1014	673		4
Apo	1014	918	2	
RNP	1014	1020		14
Apo	1014	1031	9	
RNP	1024	536		4
RNP	1031	1000		13
RNP	1031	1020		7
Shared	1059	1031	17	7
Apo	1059	1255	9	
Shared	1085	44	16	20
Shared	**1097**	**1197**	**13**	**7**
Apo	1097	1113	8	
Apo	1097	1118	20	
Apo	1107	1176	2	
Shared	1107	1118	4	32
Shared	1107	1124/1120	5	18
Apo	1107	665	5	
Apo	1107	1096	5	
Apo	1113	1118	10	
Shared	1113	1129	4	5
Apo	1113	665	9	
RNP	1118	439		4
Shared	1118	439	6	4
Apo	1118	1158	2	
Apo	1118	1188	4	
RNP	1118	1129		16
RNP	1118	1151		5
Shared	1118	1113	8	5
Shared	1118	1124	4	6
Apo	1118	558	4	
Apo	1118	665	2	
Apo	1118	673	13	
Apo	1124	1113	5	
Apo	1130	1107	5	
Apo	1130	1118	3	
Apo	1130	1123	9	
Apo	1130	1176	13	
Apo	1130	1200	4	
Apo	1153	1158	9	
Shared	1156	1151	4	4
RNP	1158	1153		10
Shared	1158	1151	16	13
Apo	1176	1197	5	
RNP	1176	1123		6
Shared	1176	1118	11	5
Shared	1176	1192	2	28
Apo	1176	348	7	
Apo	1176	665	4	
Apo	1176	677	10	
Shared	1185	1113/1118	10	15
RNP	1188	1151		2
RNP	1192	1188		6
Apo	1197	1118	7	

Apo	1197	1188	5	
Apo	1197	665	16	
Apo	1200	1113	3	
Apo	1200	1118	4	
RNP	1231	1085		6
Apo	1231	1255	6	
Shared	1231	1246	8	4
RNP	1246	775		3
RNP	1246	862		2
Shared	**1246**	**959/961**	**4**	**3**
RNP	1255	772		2
RNP	1255	968		7
Shared	**1255**	**968**	**6**	**2**
Apo	1255	1244	4	
Apo	1263	1246	9	
Apo	1340	1158	9	
Shared	**1244/1246**	**772**	**4**	**3**

Supplementary Table 2: Identified and quantified cross-links from 2-plex experiment obtained from the 5h HPLC gradient experiment

Cross-link	Ratio (Apo/RNP)
1014-673	0.139469257
528-880	0.144849778
565-890	0.160546182
948-673	0.18425726
789-218	0.196660089
1185-1118	0.246641147
902-959	0.343082756
1107-1200	0.347586775
263-772	0.361684873
797-233	0.390889267
528-918	0.418998497
913-959	0.465517311
902-528	0.472245754
848-1246	0.495115116
500-929	0.504131655
1129-1118	0.518961193
848-959	0.595591508
1118-1151	0.59744909
1246-862	0.687806106
1118-1153	0.799267864
1107-1118	0.875239716
1031-918	0.968221403
1200-772	1.003021016
1003-954	1.02873729
377-218	1.061047807
649-602	1.064923988
866-913	1.070639661
646-602	1.106732178
1014-1031	1.112803989
1024-918	1.113983741
1158-1153	1.120567354
500-948	1.121503286
1246-961	1.137221123
1148-1191	1.141088819
1255-1244	1.171933535
442-652	1.180935631
1118-1129	1.205957935
810-1246	1.229045787
1014-954	1.255544086
442-1118	1.278099991
1014-1024	1.282265993
500-677	1.328575276
913-902	1.376471265
929-959	1.379613119
1246-775	1.408084594
1158-1191	1.43416765
382-346	1.505115347
1129-1113	1.506094672
382-749	1.506406149
772-961	1.518339452
1200-1191	1.591127861
1148-1188	1.609394055
1003-1031	1.653572307
570-562	1.670203489
1096-1200	1.671942951
1176-1124	1.789222128
902-918	1.836213165
710-562	1.842950761
1246-959	1.901503039
810-855	1.934593858
1113-775	1.942231517
948-954	1.951964745
1158-1151	1.979409868
974-959	1.987018105
33-44	1.987339609
599-1158	2.043201143
1188-1113	2.081463808
866-862	2.124799572
1246-772	2.170623313
1113-76	2.180927431
1014-918	2.201591687
500-442	2.215791763
1118-652	2.218589289
890-775	2.328730755
111-649	2.379259956
506-1197	2.380318823
646-652	2.411400872
637-652	2.441561648
484-439	2.456820384
1244-772	2.474188844
1200-1129	2.534429955
506-439	2.546576018
637-649	2.554633471
772-974	2.626775552
1107-1096	2.653290024
1096-1113	2.655441481
673-1113	2.677728742

1129-1151	2.681569439
1296-1300	2.693801459
1014-948	2.694138199
44-30	2.772283582
673-1191	2.824526547
500-33	2.831595672
710-673	2.907921009
705-673	3.01297427
1107-772	3.024841197
500-942	3.041365167
755-749	3.093059757
1197-1191	3.09563755
1113-1151	3.19083978
528-33	3.223337883
772-1129	3.277908505
1340-1151	3.312029252
877-880	3.373589235
742-500	3.39466382
442-948	3.422607362
1176-1188	3.424003226
1340-1158	3.471767642
112-336	3.498661403
233-789	3.637041554
929-918	3.848283749
902-880	3.880033614
506-33	4.05805586
33-30	4.128329522
33-673	4.157275888
959-775	4.234925695
1197-1118	4.378857607
1107-742	4.379011707
942-749	4.42444034
772-959	4.510525904
1197-673	4.577405952
506-749	4.801141231
1096-1118	4.80458869
1107-1197	4.868002173
263-1113	4.967711963
500-1197	5.079199123
233-775	5.083553436
112-1113	5.130513944
500-749	5.155355289
789-890	5.31067595
528-439	5.416835651
1176-500	5.537867374
1107-673	5.584690334
637-1118	5.634323088
233-772	5.697217921
263-377	5.703241157
442-673	5.761595621
233-890	5.843441157
1176-506	5.993761173
1176-1129	6.090490587
749-673	6.351916749
1107-1124	6.354844767
336-111	6.390064517
112-348	6.605734334
348-1118	6.702922353
942-673	6.917916625
1118-1158	7.750871129
263-1118	7.796050002
1176-1197	7.925685431
1176-1113	8.25477904
599-649	8.399279934
1118-673	9.153333278
1107-1176	9.231541198
111-348	9.290862822
1124-1113	10.33224867
1118-1113	10.3350644
749-33	10.62286701
948-442	12.03792044
772-1118	12.49347382
959-918	12.57179596
1176-1118	12.76992652
1124-1118	15.10408116
918-959	15.91360925
323-218	18.88183889
336-348	27.06012387
1176-673	51.52741422

Supplementary Table 3: Identified and quantified cross-links from 6-plex experiment without enrichment

Cross - Linked Residues	Fold-Change	p-Value
673 - 948	0.19	6.817E-07
541 - 902	0.17	6.817E-07
877 - 1330	0.41	1.506E-05
528 - 918	0.18	1.877E-05
439 - 673	4.46	1.926E-05
233 - 777	2.03	1.931E-05
233 - 775	3.53	1.931E-05
1118 - 1129	0.47	1.931E-05
673 - 942	2.21	1.931E-05
918 - 959	1.87	2.081E-05
541 - 948	0.48	2.486E-05
673 - 1629	2.08	3.441E-05
772 - 1118	2.23	6.219E-05
1113 - 1129	0.61	6.960E-05
673 - 1197	1.65	9.175E-05
1031 - 1065	2.06	1.070E-04
782 - 1496	0.29	1.697E-04
673 - 1176	2.86	2.796E-04
1113 - 1124	1.50	3.336E-04
506 - 941	0.36	3.740E-04
439 - 948	2.10	4.300E-04
434 - 948	4.86	4.300E-04
775 - 1246	0.63	5.335E-04
562 - 568	0.60	5.957E-04
775 - 890	0.60	5.957E-04
1118 - 1185	0.24	7.122E-04
673 - 1118	1.56	7.814E-04
565 - 1292	0.61	8.490E-04
541 - 1031	0.44	1.350E-03
1296 - 1300	0.61	1.906E-03
1107 - 1118	0.72	1.951E-03
439 - 506	0.80	2.664E-03
1118 - 1176	1.58	3.515E-03
673 - 1113	1.42	3.515E-03
541 - 1629	1.90	3.798E-03
866 - 913	0.73	4.271E-03
1116 - 1129	0.63	6.216E-03
918 - 961	0.72	6.780E-03
867 - 913	0.67	6.954E-03

772 - 918	0.59	7.465E-03	**772 - 959**	1.07	2.313E-01
469 - 673	2.89	8.216E-03	**442 - 948**	1.11	2.511E-01
772 - 1246	0.81	8.581E-03	**775 - 1113**	1.10	3.084E-01
775 - 1118	1.42	1.017E-02	**918 - 929**	1.12	3.387E-01
33 - 541	0.77	1.054E-02	**33 - 506**	1.11	3.488E-01
1113 - 1176	0.78	1.077E-02	**772 - 961**	0.95	4.251E-01
772 - 902	0.76	1.655E-02	**959 - 1248**	0.93	4.251E-01
772 - 1113	1.26	1.775E-02	**439 - 500**	0.96	5.647E-01
1113 - 1118	1.28	3.074E-02	**772 - 1255**	0.94	6.273E-01
33 - 673	0.85	5.507E-02	**862 - 1246**	0.97	6.888E-01
862 - 866	1.32	5.507E-02	**673 - 710**	1.10	6.888E-01
902 - 918	1.11	8.128E-02	**901 - 918**	0.94	6.888E-01
562 - 570	0.88	1.022E-01	**959 - 1246**	0.97	7.294E-01
1106 - 1129	0.71	1.791E-01	**1098 - 1197**	1.01	8.552E-01
1246 - 1300	0.85	1.986E-01	**775 - 959**	0.99	0.9309586

Supplementary Table 4**:** Identified and quantified cross-links from 6-plex experiment after SEC enrichment

Cross - linked Residues	Fold-Change	p-Value			
528 - 918	0.09	5.856E-08	**772 - 918**	0.42	5.762E-07
565 - 890	0.09	1.839E-08	**775 - 1246**	0.42	4.578E-07
1107 - 1200	0.10	1.036E-08	**442 - 1113**	0.42	5.185E-06
500 - 918	0.13	5.200E-08	**439 - 1176**	0.42	1.474E-06
673 - 1014	0.13	2.609E-07	**959 - 1003**	0.43	4.403E-07
1118 - 1185	0.14	1.036E-08	**442 - 1118**	0.43	1.313E-06
562 - 789	0.16	4.403E-07	**148 - 442**	0.44	4.195E-07
673 - 948	0.18	1.839E-08	**742 - 1113**	0.45	1.236E-05
500 - 948	0.18	2.609E-07	**1014 - 1024**	0.45	3.958E-07
862 - 918	0.18	5.856E-08	**1118 - 1129**	0.45	8.515E-07
439 - 1118	0.19	5.200E-08	**954 - 1014**	0.46	1.273E-05
562 - 890	0.20	4.275E-08	**1113 - 1296**	0.46	2.432E-04
848 - 959	0.23	5.200E-08	**775 - 961**	0.48	7.463E-07
913 - 959	0.23	5.200E-08	**33 - 565**	0.49	3.104E-04
848 - 913	0.28	3.387E-07	**1014 - 1031**	0.49	9.211E-06
673 - 959	0.28	3.958E-07	**30 - 65**	0.50	9.922E-06
33 - 942	0.29	2.779E-07	**961 - 1246**	0.52	2.831E-06
959 - 1014	0.30	2.609E-07	**866 - 1031**	0.53	1.720E-04
565 - 902	0.30	2.609E-07	**775 - 890**	0.54	1.810E-06
848 - 1246	0.30	2.815E-06	**434 - 506**	0.56	3.983E-04
1096 - 1200	0.30	1.287E-05	**1113 - 1200**	0.56	7.637E-06
506 - 942	0.30	2.004E-06	**1096 - 1107**	0.57	3.937E-04
918 - 1024	0.30	3.210E-07	**1118 - 1296**	0.57	6.307E-05
500 - 929	0.30	4.526E-07	**673 - 1188**	0.58	1.507E-04
562 - 902	0.31	5.375E-07	**439 - 506**	0.58	6.230E-06
565 - 959	0.32	2.180E-07	**1107 - 1118**	0.59	9.602E-06
500 - 942	0.32	1.213E-06	**439 - 562**	0.59	7.113E-05
442 - 506	0.32	3.210E-07	**1113 - 1289**	0.59	6.019E-05
442 - 1176	0.33	5.474E-07	**673 - 1296**	0.59	3.848E-05
1148 - 1188	0.33	4.333E-06	**848 - 862**	0.59	2.271E-04
848 - 961	0.34	4.129E-05	**855 - 1246**	0.60	8.931E-05
959 - 1031	0.35	7.732E-07	**442 - 500**	0.60	5.990E-06
954 - 1003	0.35	8.670E-07	**749 - 1124**	0.60	7.923E-05
565 - 772	0.38	2.297E-06	**918 - 1118**	0.61	5.754E-05
148 - 918	0.38	3.083E-07	**772 - 1255**	0.61	2.455E-04
148 - 673	0.38	4.914E-07	**1289 - 1629**	0.61	1.827E-03
346 - 382	0.38	1.194E-06	**902 - 1113**	0.62	3.937E-04
599 - 649	0.38	5.225E-07	**562 - 570**	0.62	5.381E-06
565 - 775	0.38	4.071E-07	**1113 - 1129**	0.62	9.100E-03
902 - 959	0.41	2.277E-06	**948 - 1024**	0.62	1.168E-03
918 - 961	0.41	2.793E-05	**918 - 954**	0.63	8.584E-05
866 - 1024	0.41	5.057E-05	**148 - 1176**	0.64	6.436E-05
848 - 918	0.41	1.747E-06	**772 - 1246**	0.64	3.200E-05
918 - 1031	0.41	7.427E-07	**866 - 913**	0.65	1.996E-05
1113 - 1185	0.41	3.668E-06	**848 - 866**	0.65	2.503E-02
			742 - 1107	0.65	1.286E-04
			772 - 848	0.67	5.754E-05

775 - 855	0.67	2.455E-04	**233 - 789**	0.95	3.420E-01
902 - 913	0.67	2.755E-05	**862 - 1176**	0.95	3.233E-01
1296 - 1300	0.67	1.294E-03	**506 - 749**	0.96	4.332E-01
33 - 500	0.67	1.156E-04	**348 - 918**	0.97	5.035E-01
148 - 348	0.67	2.537E-04	**772 - 855**	0.99	7.856E-01
434 - 1113	0.68	5.727E-04	**314 - 775**	1.00	9.786E-01
346 - 356	0.69	9.201E-05	**1118 - 1197**	1.02	7.604E-01
652 - 673	0.69	7.923E-05	**346 - 862**	1.02	7.794E-01
1096 - 1113	0.69	1.819E-04	**673 - 1031**	1.02	6.535E-01
913 - 1024	0.70	7.590E-04	**948 - 1031**	1.03	6.106E-01
673 - 1300	0.70	3.018E-04	**506 - 1176**	1.04	4.885E-01
434 - 442	0.70	1.294E-04	**637 - 1118**	1.04	5.797E-01
929 - 959	0.70	6.862E-05	**918 - 1153**	1.04	7.917E-01
439 - 1188	0.70	1.157E-04	**862 - 902**	1.06	5.797E-01
772 - 890	0.70	8.591E-05	**439 - 772**	1.06	4.560E-01
772 - 961	0.71	4.600E-05	**775 - 1118**	1.07	2.861E-01
772 - 902	0.71	2.632E-02	**111 - 673**	1.09	5.354E-01
1113 - 1197	0.71	8.150E-03	**772 - 1113**	1.10	1.999E-01
506 - 1113	0.71	9.781E-05	**866 - 902**	1.10	4.602E-01
637 - 649	0.72	1.623E-03	**1113 - 1118**	1.11	5.278E-02
1151 - 1340	0.73	9.859E-02	**348 - 948**	1.13	7.363E-02
918 - 929	0.73	7.555E-04	**434 - 673**	1.14	5.001E-02
33 - 1629	0.73	7.029E-03	**755 - 942**	1.14	4.939E-02
1118 - 1289	0.73	1.692E-04	**1107 - 1176**	1.15	1.369E-02
1289 - 1300	0.74	7.054E-04	**336 - 348**	1.15	5.925E-01
1176 - 1188	0.74	4.864E-02	**1031 - 1176**	1.16	1.082E-01
439 - 1296	0.74	8.211E-04	**673 - 1197**	1.17	4.858E-01
1107 - 1113	0.75	1.286E-04	**649 - 948**	1.18	3.525E-01
500 - 1118	0.75	8.891E-04	**506 - 1107**	1.19	6.639E-02
218 - 348	0.75	3.380E-03	**772 - 1107**	1.20	3.945E-02
652 - 1151	0.76	6.952E-04	**1031 - 1118**	1.22	1.959E-03
1118 - 1200	0.76	1.174E-03	**1176 - 1629**	1.23	5.894E-03
439 - 442	0.77	4.136E-03	**348 - 649**	1.24	2.974E-03
855 - 866	0.77	9.528E-04	**862 - 866**	1.24	4.792E-03
866 - 918	0.77	7.250E-04	**862 - 1118**	1.24	3.209E-03
506 - 1118	0.77	4.374E-04	**33 - 506**	1.24	8.466E-04
1003 - 1031	0.78	4.863E-04	**442 - 942**	1.25	1.694E-02
442 - 772	0.79	6.502E-03	**111 - 439**	1.28	3.759E-03
1113 - 1188	0.80	8.466E-04	**348 - 439**	1.30	7.250E-04
673 - 673	0.81	8.879E-03	**772 - 1118**	1.34	1.408E-04
233 - 1118	0.82	2.204E-02	**111 - 1113**	1.41	3.104E-04
33 - 439	0.83	7.707E-03	**500 - 1197**	1.43	1.390E-02
439 - 500	0.84	5.312E-03	**233 - 772**	1.43	5.904E-05
500 - 1107	0.84	1.586E-01	**500 - 1176**	1.45	9.383E-04
33 - 673	0.84	5.136E-03	**1118 - 1176**	1.46	5.057E-05
789 - 1191	0.84	1.307E-01	**918 - 959**	1.47	4.567E-05
637 - 652	0.85	1.438E-01	**1031 - 1629**	1.48	4.129E-05
855 - 862	0.85	5.766E-03	**673 - 942**	1.49	8.776E-05
565 - 862	0.86	3.431E-02	**959 - 1176**	1.52	1.708E-02
948 - 1014	0.87	9.686E-03	**1176 - 1197**	1.53	8.776E-05
673 - 1113	0.87	4.925E-01	**263 - 1118**	1.54	2.455E-04
789 - 890	0.87	1.041E-02	**263 - 1113**	1.64	2.295E-05
954 - 961	0.88	7.919E-02	**439 - 948**	1.67	1.115E-05
918 - 1014	0.89	4.864E-02	**673 - 1629**	1.72	5.159E-04
673 - 866	0.89	3.233E-01	**233 - 890**	1.75	6.844E-05
772 - 974	0.90	3.913E-02	**439 - 942**	1.76	2.443E-04
948 - 1118	0.90	4.416E-02	**782 - 1176**	1.78	1.497E-03
673 - 684	0.91	5.350E-02	**33 - 749**	1.83	4.997E-06
1113 - 1176	0.91	3.826E-02	**348 - 562**	1.95	3.959E-04
775 - 1113	0.91	2.607E-01	**673 - 1107**	2.12	8.515E-07
348 - 434	0.91	3.015E-01	**348 - 1113**	2.25	1.194E-06
772 - 959	0.92	1.341E-01	**348 - 1031**	2.28	2.997E-06
434 - 439	0.92	1.505E-01	**442 - 673**	2.31	3.929E-07
749 - 755	0.92	5.563E-01	**233 - 775**	2.33	7.469E-06
948 - 954	0.92	8.915E-02	**442 - 948**	2.41	2.959E-07
1031 - 1031	0.93	1.620E-01	**111 - 1118**	2.75	1.792E-06
959 - 1246	0.93	1.487E-01	**439 - 673**	2.82	2.947E-07
902 - 918	0.94	1.554E-01	**434 - 948**	2.96	1.301E-07
500 - 742	0.94	3.436E-01	**111 - 348**	2.97	2.959E-07
346 - 673	0.94	2.792E-01	**348 - 1118**	3.60	2.081E-07
565 - 1118	0.94	2.496E-01	**673 - 1176**	6.94	2.516E-08
775 - 959	0.94	2.644E-01			