Cross-linking Mass Spectrometry for Investigating Intrinsically Disordered Proteins

Dissertation

To obtain the doctoral degree in Natural Sciences (Dr. rer. nat)

Presented to the Faculty

of Natural Sciences I

Martin-Luther-Universität Halle-Wittenberg

From Alessio Di Ianni

Reviewers:	1. Prof. Dr. Andrea Sinz
	2. Prof. Dr. Stefan Hüttelmaier
	3. Prof. Dr. Mathias Schäfer
Public date of defense:	16.09.2025

Table of contents

Index of Figures

Appendix Figures

Index of Tables

Abbreviations

ABC	Ammonium bicarbonate
ACN	Acetonitrile
AcOH	Acetic acid
APS	Ammonium persulfate
ATII	Angiotensin II
BS^2G	Bis(sulfosuccinimidyl)glutarate
CE	Collision energy
CID	Collision induced dissociation
CD	Circular dichroism
CEM	Chain ejection model
CRM	Charged residue model
CTD	C-terminal regulatory domain of p53
CuAAC	Cu-catalyzed azide-alkyne cycloaddition
DBD	DNA binding domain of p53
DNA	Deoxyribonucleic acid
DMSO	Dimethyl sulfoxide
DMTMM	4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride
DSAB	Disulfide azide agarose beads
DSBSO	Disuccinimidyl bis-sulfoxide
DSBU	Disuccinimidyl dibutyric urea
DSSI	Disuccinimidyl disuccinic imide
DSSO	Disuccinimidyl sulfoxide
DSPP	Disuccinimidyl phenyl phosphonic acid (PhoX)

DTT	Dithiothreitol
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EIC	Extracted ion chromatogram
EM	Electron microscopy
ESI	Electrospray ionization
ETD	Electron transfer dissociation
EThcD	Electron transfer with supplemental higher-energy collisional dissociation
ETciD	Electron transfer with supplemental collision induced dissociation
EtOH	Ethanol
FA	Formic acid
Fast ISD	Fast in-solution digestion workflow
FDR	False discovery rate
FPLC	Fast protein liquid chromatography
FRET	Förster resonance energy transfer
FTICR	Fourier-transform ion cyclotron resonance
HCD	Higher-energy collision induced dissociation
HDX	Hydrogen-deuterium exchange
HEPES	4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid
HLT	Histidine-lipoyl-TEV tag
HPLC	High-performance liquid chromatography
IAA	Iodoacetamide
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
IEM	Ion evaporation model

IMAC	Immobilized ion metal affinity chromatography
IM-MS	Ion mobility-mass spectrometry
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
LB	Lysogeny broth
LC	Liquid chromatography
LTQ	Linear ion trap (ThermoFisher marketing name)
LIT	Linear ion trap
MALDI	Matrix-assisted laser desorption ionization
MD	Molecular dynamics
MDM	Mouse double minute
MeOH	Methanol
MW	Molecular weight
MWCO	Molecular weight cut-off
MS	Mass spectrometry, mass spectrometer
MS/MS	Tandem mass spectrometry
m/z	Mass to charge ratio
NLS	Nuclear localization signal-containing region of p53
NMR	Nuclear magnetic resonance
NHS	N-hydroxy succinimide
PAC4	Piperidinium-alkyne-C4
PAGE	Polyacrylamide gel electrophoresis
PASEF	Parallel accumulation serial fragmentation
PCAB	Photo-cleavable azide agarose beads
PDB	Protein data bank

PONDR	Predictor of natural disordered regions
ppm	Parts per million
PRR	Proline-rich region
RE-DNA	Response element-DNA
RF	Radiofrequency
RISE	Reporter ion scan event
SAXS	Small angle X-ray scattering
SCX	Strong cation exchange chromatography
SEC	Size exclusion chromatography
SDA	Succinimidyl diazirine
SDS	Sodium dodecyl sulfate
SOC	Super optimal broth with catabolite repression medium
SPAAC	Strain-promoted azide-alkyne cycloaddition
S/N	Signal to noise ratio
TAD	Transactivation domain of p53
TCEP	Tris(2-carboxyethyl)phosphine
TD	Tetramerization domain of p53
TEMED	N,N,N′,N′-Tetramethyl ethylenediamine
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
THPTA	Tris(3-hydroxypropyltriazolylmethyl)amine
TIC	Total ion current
TOF	Time of flight
TP1	Test peptide 1

Tris	Tris(hydroxymethyl)aminomethane
u	Atomic mass unit
v/v	Volume per volume
w/w	Weight per weight
w/v	Weight per volume
XL-MS	Cross-linking mass spectrometry
YT	Yeast tryptone
z	Charge

Proteinogenic amino acids

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophane	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Summary

Until the end of the last century, the classical sequence → structure → function paradigm and related 'lock-and-key' as well as 'induced fit' models were the major theoretical basis for understanding the molecular mechanisms of protein function [1]. However, the situation started to change at the turn of this century. In fact, the potential functional importance of proteins and protein regions without defined structure is increasingly being recognized. These are now known as intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs) [2]. Regardless of their lack of unique 3D structures, IDPs and IDRs are biologically active and possess a multitude of specific functions. IDPs/IDRs are found in all proteomes characterized so far (archaea, bacteria, eukaryotes and viruses). Dysfunction of IDPs/IDRs is commonly related to the pathogenesis of various human diseases [3]. The structural and functional complexity of protein disorder is further increased by the fact that sites of various posttranslational modifications (PTMs) are commonly located within IDPs/IDRs. One of the most prominent representatives of this protein class is the tumor suppressor p53, which is often referred to as the 'guardian of the genome'. p53 plays an important role in DNA repair, cell cycle control, and apoptosis in human cells [4, 5]. Human p53 is biologically active as a homotetramer formed by the association of initial dimers. More than 40 years after the discovery of p53, the structure and spatial arrangement of the tetrameric p53 complex remain elusive. Structural proteomics techniques, namely chemical cross-linking in combination with high-resolution mass spectrometry (XL-MS), native mass spectrometry (Native MS), hydrogen/deuterium exchange MS (HDX-MS), and protein footprinting-MS, represent alternative approaches for 3D-protein structure elucidation and protein/protein-complexes topology analysis [6]. Chemical cross-linkers are typically bifunctional reagents that covalently connect proteins and protein complexes and act as 'molecular rulers', enabling measurement of distances between the connected amino acids. Native MS is a technique used for intact protein analysis and is especially important to investigate the stoichiometry of protein assemblies and protein-(ligand) complexes. Finally, HDX-MS and protein footprinting-MS have been extensively adopted in recent years to study in-solution protein dynamics and surface topology. An integrated structural biology approach combining structural MS with computational modeling to derive 3D structures of proteins and protein assemblies offers an appealing opportunity to answer important biological questions that would otherwise remain unaddressed. Human full-length wild-type p53 was expressed in *Escherichia coli* as a soluble fusion protein with an N-terminal lipoyl domain, derived from the dihydrolipoyl transacetylase of the *Bacillus stearothermophilus* pyruvate dehydrogenase complex according to an existing protocol [7]. During p53 purification, the

fusion protein was cleaved via its TEV (tobacco etch virus) cleavage site, allowing retrieval of the full-length tag-less protein. Formation of the p53 tetramer as a functional unit *in vivo*, as well as the specific binding of response element (RE) DNA were used as controls for confirming the activity of purified p53. Quantitative XL-MS and protein-footprinting data obtained with the homobifunctional amine-reactive bis(sulfosuccinimidyl)glutarate (BS^2G) were in agreement with previously published disuccinimidyl dibutyric urea (DSBU) data. The data were used to quantitatively probe possible conformational changes in the C-terminal domain of p53 upon binding of a 26-mer-RE DNA. It could be shown that the structure of the C-terminal region of full-length p53 tetramer is more compact than perceived by previous models. Moreover, even the DNA-free form of p53 is more compact and remains conformationally unchanged upon specific binding of a 26-mer RE-DNA. Structural data from the isolated p53 *in vitro* might not completely reflect the *in vivo* situation and/or the cellular context of p53. Furthermore, p53 functions *in vivo* are extensively fine-tuned by a large number of PTMs, some of them even present at the same time, making a thorough evaluation impossible to achieve. For this reason, future studies are needed using cellular systems to investigate p53 functions as well as p53 protein-protein interaction networks.

XL-MS has now matured as a powerful tool to investigate intricate and entangled protein networks in cells, organisms and tissues [8, 9]. This has been achieved by the use of extensive prefractionation, such as size exclusion or strong cation exchange chromatography of cross-linked peptide mixtures prior to MS analysis. An alternative route consists in adding an enrichment handle to the cross-linker molecule. These reagents are known as ‘trifunctional cross-linkers’ and allow to separate cross-links from linear peptides via highly specific affinity chromatography [10, 11]. In this work, two trifunctional cross-linkers have been evaluated for future *in cellulo* studies on IDPs. The first one, called PAC4, makes use of an alkyne handle for copper-catalyzed azide alkyne cyloaddition (CuAAC) enrichment of cross-linked peptides. The second cross-linker, named DSSI, possesses an ethylphosphonate tag, which allows cross-links to be enriched via highly selective immobilized ion metal affinity chromatography (IMAC). Their CID-MS/MS fragmentation behavior has been characterized in detail using synthetic peptides as model systems. Finally, the performance of both trifunctional linkers has been tested on model proteins and more complex systems in perspective for future proteome-wide cross-linking studies on IDPs.

1 Introduction

1.1 The ‘protein folding’ problem

A fundamental challenge in structural biology is to understand the principles, which govern how protein fold in solution, from a random-coil (or unfolded state, U) to a functional folded state (also named as native state, N). The so called ‘protein folding problem’ has been tackled by a large number of theoretical as well as experimental studies. At the end of the 1960s, Christian B. Anfinsen, reinforced by his experiments on ribonuclease A, postulated that under physiological conditions every protein has a unique, stable, and kinetically accessible minimum of its free energy. This is encoded in its amino acid sequence, which ensures the functionality of this protein [12, 13]. This concept is known as ‘Anfisen’s dogma’ or ‘thermodynamic hypothesis’. Concurrently, Levinthal’s paradox dominated ideas about protein folding. In his 1969 paper, ‘How to fold graciously’ [14], he noted that, because of the very large number of degrees of freedom in an unfolded polypeptide chain, the molecule has an astronomical number of possible conformations to sample. Martin Karplus in 1997 [15] stressed that the concept introduced by Levinthal is that the appropriate point of reference for protein folding is a random search problem. This means that all conformations of the polypeptide chain, except its native state, are equally probable, so that the native state can be found only by an unbiased random search. This has been referred to as the ‘golf course’ model of the protein potential energy surface [16]. For such a surface, the time to find the native state is given by the number of configurations of the polypeptide chain (10^{70} for a 100-residue protein) multiplied by the time required to find one configuration (10^{-11} seconds). This leads to an enormously long folding time ($\sim 10^{59}$ seconds or about 10^{52} years). Since proteins generally fold in timeframes in the order of milliseconds to seconds, there was indeed a paradox. Therefore, ‘if a protein were to attain its correctly folded configuration by sequentially sampling all the possible conformations, it would require a time longer than the age of the universe to arrive at its correct native conformation’. Levinthal’s solution to the protein folding problem was that there were well-defined pathways to the native state, so that protein folding was under ‘kinetic’ control. This statement was questioned and refined by Baldwin in 1994, whose theoretical studies pointed to the fact that there is no unique pathway for folding and no unique transition state. On each of these pathways there is a high-energy intermediate that is close in structure to the native form. The new view expressed by Baldwin and supported by Ken A. Dill and Hue Sun Chan [17] replaced the Levinthal’s folding pathways solution and highlighted the idea of rugged, funnel-shaped energy landscapes with multiple intermediate states leading to the protein’s native form.

1.2 Intrinsically disordered proteins and their relevance in biological processes

All of the previous concepts have been the basis for the sequence→structure→function paradigm. This paradigm applied very well, since almost everything the community learned about proteins by that time was in line with these concepts. Enzymes historically have been the primary targets of scientific research, and their catalytic activities were critically dependent on the presence of unique 3D-structures, at least within the active center of an enzyme. Therefore, the validity of such a 'rigid' perspective on protein functionality was beyond any doubt. Around the turn of the millennium, some conceptual studies were published. The results of these investigations disclosed many proteins contain unstructured regions or are completely 'intrinsically disordered' under native conditions. These studies laid the foundation for a new area of research [2, 18, 19]. A multitude of observations established the field of 'unstructured' biology, which today is an integral part of structural biology and biochemistry [18, 20]. The paradigm shift was supported by a series of experimental observations of 'disorder' achieved within a few proteins. Structural biology at the end of the 90s was based on around 18,000 structures stored in the Protein Data Bank (PDB), while the PDB now includes 224,572 structures (9th September 2024). It has now been found that thousands of these proteins also contain unstructured or disordered regions which can transition onto an ordered structure when attached to certain interaction partners. This explains why many proteins can only be crystallized with binding partners and many regions are often structurally unclear [2]. The investigation of these structurally unexplained protein regions with advanced structure elucidation methods such as circular dichroism (CD), nuclear magnetic resonance spectroscopy (NMR) enabled the creation of a database for intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs), the so called DisProt database, which now includes 3024 protein entries [21-23]. Additional databases have been annotated as well, such as the Mobi Database [24], the Database of Disordered Protein Prediction (D^2P^2, [25]), and the DescribeProt Database [26]. Another boost for the advancement of the IDP field were newly developed algorithms for secondary structure prediction from amino acid sequences. The first predictor, developed in 2002, is called PONDR (Predictor of Natural Disorder [27]), whereas today many more are available as well. These algorithms are based on different principles [28], but they can all together be used to assess overall protein disorder content. In 2021 DeepMind, a British artificial intelligence (AI) company acquired by Google, has released AlphaFold [29], that is an AI algorithm that predicts a protein's 3D structure from its amino acid sequence. Since it has been released, it has revolutionized the field of structural biology as it has given access to all the scientific community with ~214,683,829 structures up to date, including 48 complete proteomes available for bulk download. It is now even possible to predict the structure of oligomeric proteins as well as

protein complexes, with the release of AlphaFold Multimer [30], and with a glance, identify regions which are predicted with low confidence which are typically IDRs. With the help of these prediction methods it has been proved that far more proteins than previously assumed belong to the class of IDPs or have IDRs, some of them even included in Nobel Prize's discoveries [31], and have been found to be quite abundant in all the different domains of life (archaea, bacteria, eukaryotes) and in viruses [32]. IDPs play significant roles in very important biological processes [33] (differentiation, transcription, DNA condensation, cell cycle, mRNA processing and splicing, mitosis, apoptosis and membrane trafficking [34]) and their functions are regulated by PTMs, mainly located in the IDRs [35, 36]. Prominent examples of proteins belonging to the class of IDPs are Cdk inhibitors (p27), Cadherins, RIP1/RIP3-Kinase-complex and transcription factors such as p53 [37].

1.3 A prominent example: p53 – 'The guardian of the genome'

1.3.1. Historical overview

In the 1970's, the attention of 'modern' cancer researchers was mostly focused on cancer-promoting viruses. In particular, it became known that such viruses carried oncogenes. Over the next fifteen years a long list of oncogenes was identified and it became clear that oncogenes were the cause of cancers in animals. But how did these viral oncogenes act to transform cells and generate tumors in animals? It was proposed that the tumorigenic virus oncogenes encode viral proteins that lead indirectly to the excessive induction of putative cellular oncoproteins. It was on this fertile theoretical concept that the tumor suppressor p53 was discovered for the first time. Experimentally-induced tumors in animals by small DNA tumor viruses, such as the *Simian virus* (SV40), typically express few viral encoded proteins. These are recognized by the immune system of the host, leading to the production of antibodies against these proteins. Initially, these proteins were named viral tumor antigens. Subsequent genetic analysis revealed that the genes encoding these viral tumor antigens were often the ones responsible for the transformation of the cell, namely the viral oncoproteins. In the case of SV40, the two viral proteins identified in this manner were called large T-antigen and small t-antigen, respectively. It was while studying these SV40-derived tumor antigens that several groups independently stumbled on p53. This happened in 1979: David Lane and Lionel Crawford discovered that when sera from animals bearing SV40-induced tumors were employed to immunoprecipitate SV40 large T-antigen, a non-viral protein with an apparent molecular mass of about 53 kDa was also co-isolated [38]. Further analysis revealed that this cellular protein was physically complexed with SV40 large T-antigen. Thus the viral protein, previously shown to be the

main responsible for the transforming and tumorigenic activity of the SV40 virus, had selected this so far unknown cellular protein as its partner. At the same time, Daniel Linzer and Arnold Levine [39] applied a similar experimental approach to SV40 transformed cells, and came up with essentially the same observations. Three other groups, those of Alan Smith in the UK, Robert Carroll in New York and Pierre May in France, simultaneously made very similar findings, all published in 1979 [40, 41]. In the 1980s, new interactions of viral proteins with p53 were observed, such as complex formation with an oncoprotein from adenovirus and human papillomavirus [42-44]. This incorrectly prompted at a possible role of p53 as oncogene. By analyzing the cDNAs of transformed cell cultures, it was noticed that many of p53 cDNA clones contained mutations and that the wild-type p53 prevented transformation of cell cultures. Most importantly, these mutations were found in both alleles of the p53 gene, which suddenly caused a shift of p53's role from an oncoprotein to a tumor suppressor protein. These findings were confirmed by four observations in the 1990s:

1) Patients with Li-Fraumeni syndrome who have mutations in the p53 gene have a very high probability of developing cancer, with the formation of different independent tumors.
2) Mice with a loss of function mutation in the p53 gene get early tumors.
3) More than 50% of all human tumors are caused by mutation of the p53 gene.
4) Unusually high expression levels of p53 are observed in many tumor-containing tissues.

These observations demonstrated that p53 plays a crucial role in preventing cancer development. Additionally, a good number of publications showed that p53 binds to specific DNA sequences (response elements; RE) and thus enhances/reduces the transcription of various target genes [45, 46]. It became then clear that p53 has also the function of a transcription factor.

1.3.2. p53 structure-function relationship

p53 is a tumor suppressor protein and has to fulfill a single but fairly important task: being a transcription factor [46]. To behave as such, the protein needs to bind a specific DNA sequence, also known as p53-RE DNA, associated with a target gene promoter. In this way, p53 specifically recognizes genes which contain half-site regions and regulate their expression levels (by either inducing or decreasing a target gene's transcription), hence their functions. p53 controls a large number of genes which are involved in the control of crucial regulatory processes of the cell. These processes include apoptosis, senescence, anti-angiogenesis, and cell-cycle arrest [4, 5, 47-49]. Another important signalling pathway activated by p53 that is not necessarily linked to programmed

cell death and/or cell stasis is the activation of different DNA repair mechanisms [50] which are trigged in response to DNA damage. Under physiological conditions, it is logical to think the p53 levels have to be quite low, since no genotoxic or cellular stress are happening. For this reason, the cell has to adopt a complex regulation network to inactivate p53 function when this is not needed (Figure 1.1). In non-stress conditions for the cell, the concentration of p53 is in the nanomolar range [51]. This is achieved by negative regulators such as mouse double minute proteins MDM2 and MDM4 (also termed MDMX), which belong to family of E3 ubiquitin ligases. These proteins are responsible for the ubiquitination [52, 53] and consequent degradation of p53 by the human 26S proteasome degradation system.

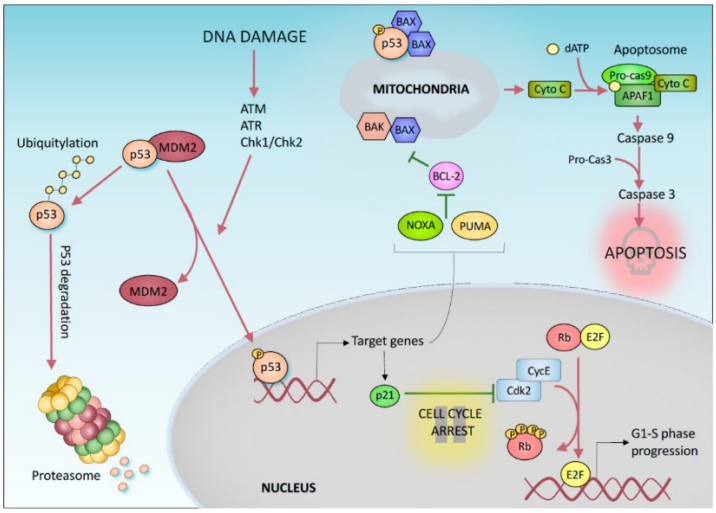


Figure 1.1: Simplified representation of the cellular regulation network of p53.

In response to stress signals, p53 tumor suppressor protein is activated through a complex mechanism involving for example, protein kinases. In combination with coactivators such as p300/CBP, p53 is able to regulate the activation of different target genes which then determine the cellular response. In an unstressed state, the levels of p53 are maintained quite low through its ubiquitination and consequent proteasomal degradation through the action of Ubiquitin E3 ligases negative regulators, e.g. MDM2 and MDM4. Stress-inducing signals lead to phosphorylation of p53 N-terminus by different protein kinases (e.g. ChK1/2), so that MDMX proteins are not able to bind p53. Coactivators like p300/CBP compete for the N-terminal binding of p53 and further activate it to exert its function as transcription factor. p53 can also translocate to the cytoplasm and in mitochondria, to trigger signaling pathways responsible for cell apoptosis. Figure is adapted from [49].

In case of cellular stress conditions, e.g. activation of an oncogene, hypoxia, DNA damage, p53 is post-translationally regulated [54] (phosphorylated by protein kinases and/or deubiquitinated by deubiquitinases). In this way, on one hand p53 is not able to bind MDMX negative regulators anymore and on the other hand, additional coactivators, such as p300/CBP (CREB binding protein) can bind p53. This leads to the accumulation and increase of the p53's levels in the cell, so that the transcriptionally active form of the protein, the homotetramer, can be formed. The type of gene target which is thereby activated depends on the stress-causing signal and on the different post-translational

modifications (PTMs) the protein carries [54]. A large variety of PTMs are well-known for p53, such as phosphorylation, sumoylation, acetylation, methylation, neddylation, ubiquitination, which act as regulators for deciding which target gene will be activated/repressed [55]. To further complicate the picture, p53 can also translocate into the cytoplasm and reach mitochondria, where it is able to recruit other protein interactors (e.g. Bcl-2–associated X protein, BAX) and trigger apoptosis in stress conditions. p53 has a modular structure composed of five different domains plus a linker region, as shown in Figure 1.2. All those domains are involved either in important protein-protein as well as in protein-DNA interactions.

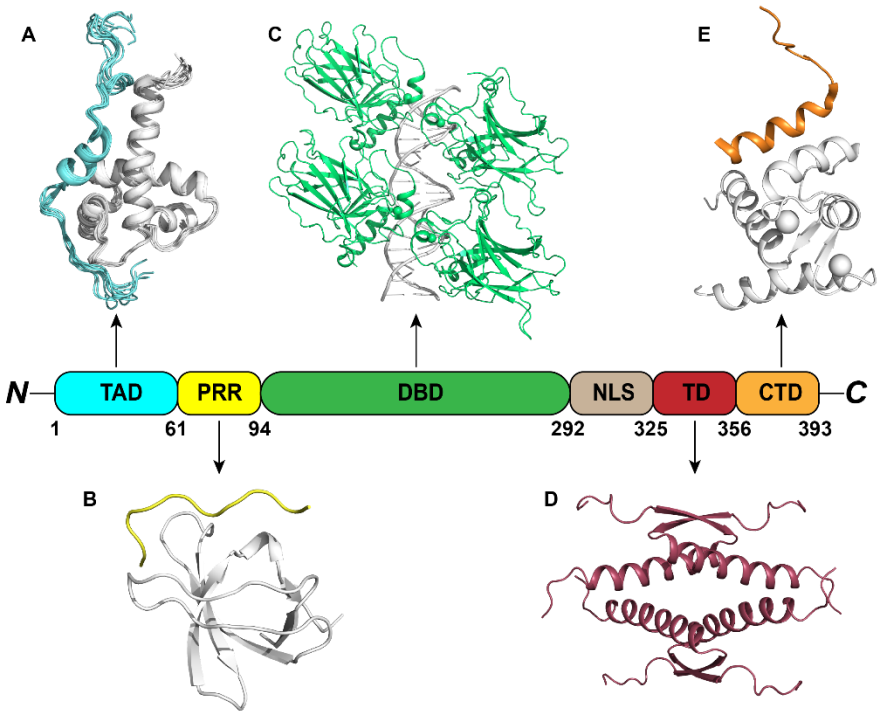


Figure 1.2: Structural domain organization of p53.

The tumor suppressor protein p53 can be divided in five domains and a linker region: the transactivation domain (TAD, cyan), the proline-rich region (PRR, yellow), the DNA binding domain (DBD, green), the nuclear localization signal-containing region (NLS, gray) or linker region, the tetramerization domain (TD, red) and the regulatory C-terminal domain (CTD, orange). All these domains are involved in important interactions with different partners, e.g. A) TAD complex with TAZ-domain of p300; B) Polyproline-helix complex with an SH3 domain; C) DBD complex with RE-DNA; D) TD as tetrameric complex; E) CTD complex with S100β. Depending on the interaction partner, this region can undergo disorder-to-order transition or remains disordered. p53 domains are colored according to the scheme, the interactions partners are colored in white.

Specifically, p53 has a modular domain organization with two ordered domains, the DNA binding domain (DBD) and tetramerization domain (TD), and contains approximately 40% of IDRs

comprising the N-terminal transactivation domain (TAD) and proline-rich region (PRR) as well as the C-terminal regulatory domain (CTD) [4]. The transactivation domain is further subdivided in TAD-1 and TAD-2. These subdomains are responsible for the regulation of the transcriptional activity of p53. Both negative regulators (MDM2 and MDM4) [56] as well as positive regulators (p300/CBP) [57] bind to this domain. This domain has intrinsic disorder as feature to be able to bind the highest amount of protein partners and act as one of the main regulatory points of p53 [58, 59]. In those IDRs, IDPs usually contain the so called *molecular recognition features (MORFs)* [60, 61], which are small (10-70 residues) IDRs in proteins that undergo a disorder-to-order transition upon binding to their partners. Examples of MORFs are the structured α-helix formed upon binding of MDM2/4 [62] or the formation of an amphipathic α-helix upon binding of Replication A protein or Tfb1 [63], a subunit from the TFIIH complex from *Saccharomyces cerevisiae*. The binding sites for different interaction partners often overlap [64]. This behavior can only be achieved by the tremendous flexibility present in IDRs. For this reason, the binding of different effector proteins is strongly influenced from PTMs in those regions. For the N-terminal domain of p53, phosphorylation of lysines and serines [65, 66] is the main responsible for the shift in binding affinity towards the different interaction partners. The TAD is followed by the proline-rich region (PRR) [67]. This region is characterized by a repeating motif PxxP (with x being a variable amino acid). It is known that PRR domains can mediate interactions with Src Homology 3 (SH3) domains of proteins [68]. The high proline content typical of this domain is known to make the protein more rigid [69]. PxxP motifs are able to adopt a poly-proline helix-like structure [70]. The function of this domain is yet not completely understood. However, the presence of a rigid linker with a defined length between the TAD and the other domains of the protein is definitely needed for its function. The first well-folded as well as the largest domain of p53 is the DNA binding domain (DBD) or core domain [71-73]. DBD adopts an immunoglobulin-like β-sandwich fold that provides a basic and ideal scaffold for the DNA-binding surface. It can be further divided into two structural motifs that bind to the minor groove and major groove of target DNA, respectively [4]. The loop-sheet-helix motif, which docks to the DNA major groove, includes loop L1, β-strands S2 and S2', parts of the extended β-strand S10, and the C-terminal helix. The other half of the DNA-binding surface is formed by two large loops (L2 and L3), which are stabilized by a zinc ion and bind to the DNA's minor groove. The zinc ion is tetrahedrally coordinated by a histidine and three cysteine side chains (H179, C176, C238, and C242). Loss of zinc causes a significant decrease in thermodynamic stability of the protein, enhances aggregation tendencies, and is followed by structural fluctuations in neighboring loops that cause loss of DNA binding specificity [74, 75]. p53's DBD is marginally stable and have a melting temperature slightly above the body temperature

(44°– 45°C) [76, 77], which dictates the stability of the full-length protein. In contrast, the DBD of the paralog p63, which, on the basis of homology modeling, has a similar structure, is surprisingly stable. The explanation for this lower stability is probably due to the presence of buried polar amino acids that are not involved in any hydrogen bonds, which may be one of the key factors contributing to p53's inherent instability [4]. Substitution of such amino acids with the corresponding hydrophobic residues found in the paralogs p63 and p73 increases the stability of p53's DBD by 1.6 kcal/mol [71, 78, 79]. As phylogenetic analyses and functional studies suggest that the genes of p63 and p73 are more ancient than that of p53, it seems that instability of p53 has developed evolutionary to allow for more rapid cycling between folded (active) and unfolded (inactive) states and hence a tighter control of cellular protein levels. Furthermore, a more flexible core of the β-sandwich may provide the structural plasticity required for optimal function. The main task of the DBD is the specific binding of RE-DNA. These DNA consist of two palindromic decameric motifs of general sequence 5'-RRRCWWGYY Y-3' (R = A, G; W = A, T; Y = C, T), which can be separated by 0-13 base pairs (bp) [45, 80, 81]. Thus, each half-site of the p53 consensus sequence can be described as an inverted repeat of two pentameric quarter sites. Four p53 DBD domains bind to these response elements in a highly cooperative manner to give a complex with 4:1 stoichiometry, resulting in bending and twisting of the DNA, which is increased upon binding of the full-length protein compared to isolated core domains. However, how exactly the DNA binding mode works is still a matter of debate. The most accepted and common hypothesis is that p53 dimers bind each one a RE-half site and thus a complete RE-DNA is bound by a p53 tetramer [82]. The DBD ends with an unstructured linker region, termed the nuclear localization signal-containing region (NLS), which contains a nuclear-shuttling signal sequence responsible for p53 translocation into the nucleus [83]. This linker region connects the DBD with the second ordered domain of the protein, the tetramerization domain (TD). The structure of the TD has been elucidated both by X-ray crystallography and NMR spectroscopy and consists of a short β-strand, followed by a tight turn and an α-helix [84, 85]. Two of these domains form an initial dimer through the formation of an intermolecular, antiparallel β-sheet and an antiparallel α-helix stack. Two associated dimers associated via their helices to form a tetrameric α-helix bundle with D2 symmetry [4]. The tetramerization process of the TD is mainly driven by hydrophobic interactions. Mutation of some of some crucial amino acids (e.g. L344) has shown to affect the oligomerization state of the TD so that is not able to form tetramers, but only dimers (e.g. mutant L344A) or monomers (e.g. mutant L344P) [51, 86, 87]. Some of these TD mutants have also been found in some p53-related diseases, such as Li-Fraumeni syndrome [88]. The regulatory C-terminal domain (CTD) is also intrinsically disordered and exhibits local disorder-to-order transition

behavior when bound to different interactors. Such a high number of binding partners is only possible because the CTD, which is intrinsically disordered in solution in the unbound state, can mimic various secondary structure elements, such as an α-helix after binding of S100β, a β-sheet after binding of the deacetylase sirtuin or different loop conformations after binding of CBP [89] and cyclinA2. CTD plays an important role in assisting DBD-specific DNA binding. This is due to the presence of a highly-positive electrostatic surface with a rich density of lysines, which has evolutionary developed to bind negatively charged DNA [90, 91]. This makes it able to slide along double-stranded DNA sequences and help the DBD in finding the specific RE-sites and regulate those genes. Some studies have shown that the capacity of the CTD sliding onto the DNA in the so called 'Search mode, S' which anticipates the specific 'Recognition mode, R' [92, 93] is influenced by varying concentrations of intracellular cations [93]. Non-specific DNA binding of CTD can be reduced or even completely prevented by acetylation of some lysines. Moreover, both specific DNA binding and transcriptional activity are lost when removing the last C-terminal 30 amino acids of the protein. Based on these observations, the CTD has two main roles: binding different interactors which regulate the protein activity (e.g. through PTMs at the CTD) and sliding along the target DNA until a RE-DNA is found. Then the DBD can bind the RE and the CTD is simultaneously released from the DNA. Whether this is achieved through a parallel acetylation *in vivo* has not yet been clarified, but evidence suggests this might be one possibility for p53 function [94].

1.3.3. p53 tetramer: current status

More than 40 years after the discovery of the tumor suppressor protein p53, the structure and spatial arrangement of the tetrameric p53 complex remain elusive. The ordered domains, DBD and TD, have already been independently structurally elucidated. The structure of the TD has been elucidated by both X-ray crystallography and in-solution NMR [84, 85]. The structures of the DBD alone (and some of its mutants) as well as the DBD in complex with RE-DNA have been determined by X-ray crystallography [72, 85]. For the IDRs of the proteins, such as TAD and CTD, only small peptides in complex with different binding partners have been either crystallized or solved by NMR [56, 57]. For the p53 full-length protein in its biologically active oligomeric form, the homotetramer, several models have been proposed, with some of them showing quite striking differences. The first model, proposed in 2006 [95], is solely based on cryo-EM data of murine p53 and shows a hollow-skewed cube-like structure (Figure 1.3 B). The molecule has eight vertices/nodes of two types (large and small nodes) connected by linkers. The N-terminus of one monomer flanks the C-terminus of the

partner in the dimer, forming so called 'N/C nodes'. DNA binding induces only a small conformational change involving a slight rotation of the DBD, in which only two out of the four DBD bind one half of the RE-DNA (Figure 1.3 D). This model has been refined and re-discussed over the years [96]. The second model, published only one year later, is a result of an integrative structural approach combining NMR, small angle X-ray scattering (SAXS), negative staining-EM and ab-initio molecular modeling studies [97]. This resulted in two models, respectively for the DNA-free and DNA-bound p53 states (Figures 1.3 A, C). In the DNA-free form, the p53 tetramer displays a cross-shaped structure composed of loosely interacting dimers. After DNA binding, the tetramer becomes a more compact structure in which the DBD are clamping around the target RE-DNA. The four DBD bind each one to a RE-DNA quarter site. The third model, revealed in 2017 [98], is a result of a combination of EM data [99] and molecular dynamics (MD) simulations (Figure 1.3 E). In this model, each DBD of the tetramer is binding to a quarter site as well and interestingly the C-terminal domain has been found to interact via electrostatic interactions with the dsDNA by means of MD. Moreover, the conformation of the C-terminus is more compact than the one of the DNA-free cross-shaped structure in which the N- and C- termini are completely extended. Finally in 2022, a cryo-EM structure of p53 with the nucleosome, H2A, H2B, H3.1, and H4 plus DNA fragment containing the p53 binding site was proposed [100], but the low resolution obtained for p53 did not allow resolving its structure in the cryo-EM map (Figure 1.3 F). All the generated models show a highly different organization of the tetrameric complex and lack experimental information regarding the IDR regions of p53. To probe which ones are actually true, further structural studies need to be carried out and the role of the C-terminal IDR in DNA binding has been investigated in this work.

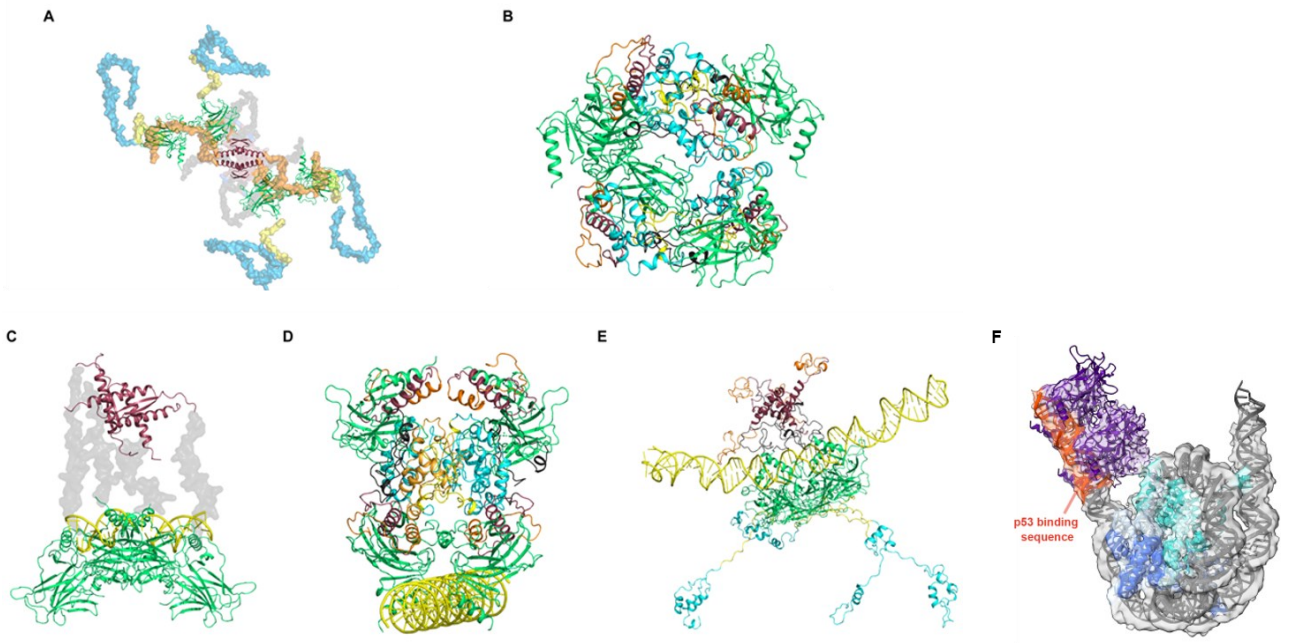


Figure 1.3: Structural models of full-length tetrameric p53 in absence (A, B) and presence of DNA (C, D, E, F).

Models A and C [97] are based on a combination of different structural techniques. Models B [95], D [96] and F [100] are based on cryo-EM experiments. Model E [98] results from the combination of negative staining EM [99] and MD simulations.

1.4. Structural biology toolbox to study structure and functions of IDPs

With the discovery that a large fraction of the proteomes from different organisms contains IDPs or IDRs [31, 32], it became clear that disorder is involved in a large number of protein functions and has an impact on protein dynamics and structural stability [33, 58]. Consequently, the structural biology field has shown great interest in understanding protein disorder from a structure/function perspective by means of different structural techniques. CD spectroscopy has been extensively used to study the impact of IDRs on the overall protein structure [101, 102]. NMR spectroscopy is one of the most frequently used approaches for studying IDP dynamics [103] although it is challenging due to the fast interconversion between transient conformational states, leading to a severe spectral overlap. Performing experiments at low temperatures can partially mitigate this issue. Alternatively, the detection of other NMR-active nuclei (particularly ^{13}C and ^{15}N in combination with ^{1}H, e.g. ^{1}H-^{15}N HSQC, or ^{13}C-^{15}N CON) can also overcome the problem of amide proton exchange, and has the benefit of increased chemical shift dispersion compared to ^{1}H-NMR [104, 105]. Indeed, IDPs typically show much higher dispersion (hence high NMR resolution) on the $^{15}N/^{13}C$ scale. This has

resulted in several studies, in which the dynamics of IDPs has been probed with or without a ligand/binding partner [94]. Alternatively, fluorescence techniques, such as Förster resonance energy transfer (FRET), fluorescence polarization assay (also known as fluorescence anisotropy) are also commonly used to probe IDP dynamics in solution [106, 107]. The main advantages of these techniques include their exceptional sensitivity (requiring only low nM to pM protein concentrations) and the extensive data they provide, such as distances between interacting domains, radius of gyration, diffusion coefficients, molecular volumes, and affinity constants. However, a significant drawback is the requirement for highly hydrophobic dyes, typically maleimide derivatives for cysteine labeling. These dyes can significantly affect protein solubility and may cause aggregation or distortion of protein conformation if labeling conditions are not meticulously controlled. Small angle X-ray scattering (SAXS) [97, 108] is another useful and complementary technique which gives details about important protein features, such as overall hydrodynamic radius and protein's size and shape. Cryo-electron microscopy (cryo-EM) has become in the last ten years the main high-resolution technique for probing structures of huge molecular machineries [109, 110]. With advancements in sample preparation protocols (e.g. GraFix [111]) and software tools for processing cryo-EM data (from particle classification to 3D shape reconstruction [112-114]) it is now possible to achieve high structural resolution, particularly for well-folded regions. However, IDRs often appear as fuzzy areas of electron density [115]. This technique requires lower sample concentration/purity compared to X-ray crystallography/NMR. It is mainly suited to resolve high molecular weight complexes (typically around 200 kDa as lower limit) and it is complementary to NMR (which generally have 35 kDa as upper limit). Finally, structural proteomics approaches have gained remarkable popularity in the last twenty years to investigate protein structures and protein-protein interactions [116, 117]. Native MS, alone or in combination with ion-mobility mass spectrometry (IM-MS) [118-120], has been used to characterize IDPs conformational ensembles, investigating IDPs interactions with binding partners as well as getting detailed information about their rotationally-averaged collisional cross-section (CCS), a parameter which reflects protein's size, shape and molecular conformation. HDX-MS displays protein dynamics and changes in a time-dependent manner [121-123]. In HDX, backbone amide protons of a polypeptide chain are limited in their H/D exchange if they are involved in hydrogen bonding networks. This technique has received great attention and further developed to study IDP dynamics in the milli-second time range [124] since within the usual kinetics time points adopted in HDX (seconds to hours), IDPs are fully exchanged. Last but not least, XL-MS has evolved as a promising technique to investigate the 3D structures of proteins and protein complexes in their native environment [6, 125]. The development of cross-linking reagents with different features [10,

11, 126-145] as well as the outstanding development in MS instrumentation and MS workflows [8, 146-150], have pushed the boundaries of XL-MS, making this technique even capable of handling very complex mixtures (from cell lysates up to whole tissues) [9, 151]. The synergy of cryo-EM with structural MS techniques (mainly HDX-MS and XL-MS) has clearly proven in the last years to be an amazing combination in the structural biologist's toolbox for studying systems, in which classical high-resolution approaches have failed [152-154]. The structural constraints obtained via XL-MS can be used as input for computational protein modeling, which at the moment is still struggling in predicting IDRs with acceptable accuracy. The technique has now reached a fairly mature standpoint and is now being adopted in combination with AlphaFold to get exciting insights from both *in vitro* as well as *in cellulo* data [155, 156].

1.5. XL-MS goes into cells/tissues: proteome-wide XL-MS to study protein interaction networks

As mentioned earlier, disorder is a crucial feature to regulate protein function as well as allowing a higher number of possible interactors, therefore the ability of those proteins to modulate intricate protein interaction networks and signaling cascade pathways. The tumor suppressor p53 greatly displays this behavior, since it is one of the proteins involved in the highest amount of protein-protein interactions, both nuclear and cytoplasmic. Structural studies on such isolated proteins, pursued as well in this work, can give a glimpse about how the protein mediates specific tasks and how the IDRs contribute to regulate those functions. On the other hand, drawing biological conclusions could be extremely risky. The reason is that *in vitro* investigations are lacking the simple but fine feedback regulation mechanisms present in a cellular environment, which get even more complex at the tissue/organ's level. Secondly, PTMs which are crucial in regulating proteins and protein-protein interaction networks are not taken into account. For this reason, studying the interaction networks of IDPs in cellular systems gives the opportunity to obtain biologically relevant insights. XL-MS has evolved over the years as a powerful approach to study protein structures and protein-protein interaction networks in complex matrices. In the very beginning, the technique was confined in studying single proteins up to five-six proteins at a time. Nowadays, it is possible to obtain insights from so called 'proteome-wide' XL-MS experiments, which are carried out on cell lysates, intact cells, or tissue sections [9]. This has been made possible through the development of cross-linking reagents with different characteristics as well as advanced sample preparation workflows aiming to enrich cross-links peptides, which represent the minority in a complex sample. On top of that, the

advancements in MS instrumentation from what concerns speed/sensitivity of the analysis as well as dedicated MS methods developed for the detection of cross-linked species has moved the technique forward, enabling its use for proteome-wide studies. Two novel cross-linking reagents were characterized and evaluated in the present work for their potential application in future IDPs proteome-wide XL-MS analyses.

1.6. Mass spectrometry: basic concepts

As mentioned in section 1.4, mass spectrometry (MS) is an important tool for the structural elucidations of proteins. The basic principle of MS is the identification of a compound by the molecular or atomic mass of its components. Different substances (from inorganic to organic compounds, peptides or proteins) are ionized generating charged species (positive or negative depending on the type of acquisition mode) into the gas phase. These ions are then showing up in a mass spectrum with a defined mass to charge ratio (*m/z*), which is unique for a defined species, allowing to discriminate the different compounds present in the sample. The ionization process takes place under atmospheric pressure or in a vacuum. On the other hand, detection always takes place in high vacuum. In order to efficiently use MS to analyze biomacromolecules, gentle ionization methods are required. These ensure that peptides, proteins and/or protein complexes can be transferred intact into the gas phase. Besides matrix-assisted laser desorption/ionization (MALDI) [157, 158], electrospray ionization (ESI) [159] is the most commonly used soft ionization method. In detail, ESI has become the golden-standard ionization method to transfer proteins and protein complexes from their native in-solution state to the gas-phase as well as for a broad spectrum of '-omics' applications. The most common way to identify proteins by MS is by the conventional 'bottom-up' approach, in which proteins are proteolytically digested into peptides prior to MS detection. To unambiguously identify a peptide belonging to a defined protein, especially in large scale proteomics experiments in which there's a higher occurrence of 'isobaric' peptides, tandem mass spectrometry (MS/MS) is used. By using tandem MS, the exact amino acid sequence of a peptide can be determined. MS sequencing is achieved by selecting a precursor ion in the mass spectrometer and cleaving it into fragment ions using an appropriate fragmentation technique e.g. collision-induced dissociation (CID), electron-transfer dissociation (ETD). The fragments are then recorded in a 'product-ion' or 'fragment ion' mass spectrum [160]. To determine the *m/z* values of both precursor and fragment ions, different mass analyzers are employed. The most common ones are quadrupoles, time-of-flight (TOF), ion cyclotron

resonance (FTICR), ion trap, and orbitrap analyzers. A basic description of the ESI process as well as selected fragmentation techniques and mass analyzers used in this work are also presented.

1.6.1 Electrospray Ionization (ESI)

The most frequently used method for the ionization of biomolecules is electrospray ionization (ESI) in combination with mass spectrometric analysis [161]. This is one of the few methods, in which ionization takes place at atmospheric pressure. The conceptual development of ESI was described by the Dole group in 1968 [162] and in the late 1980s coupled with MS by John Fenn [159] who was awarded the Nobel Prize in Chemistry in 2002. The schematic structure of an ESI source is shown in Fig 1.4. In ESI, a solution is sprayed in the presence of an electric field. This result in the formation of a Taylor cone, from which a mist of micrometer-sized charged droplets is generated. The electrostatically charged aerosol is continuously passed into the mass spectrometer. This continuous character makes ESI optimally suited for coupling with liquid chromatography (LC). The aerosol is sampled into the first vacuum stage of a mass spectrometer through a capillary carrying a potential difference of several kV, which can be heated to aid solvent evaporation from the charged droplets. These droplets reduce their volume through evaporation processes until the critical charge density, the so called 'Rayleigh limit', is exceeded [163]. At this point, the droplet distorts because the electrostatic repulsion of charges of the same polarity becomes more powerful than the surface tension holding the droplet together. The droplet undergoes Coulomb fission where the original droplet explodes creating many smaller, more stable droplets. The new droplets undergo desolvation and subsequently further Coulomb fissions until ions are released in the gas phase.

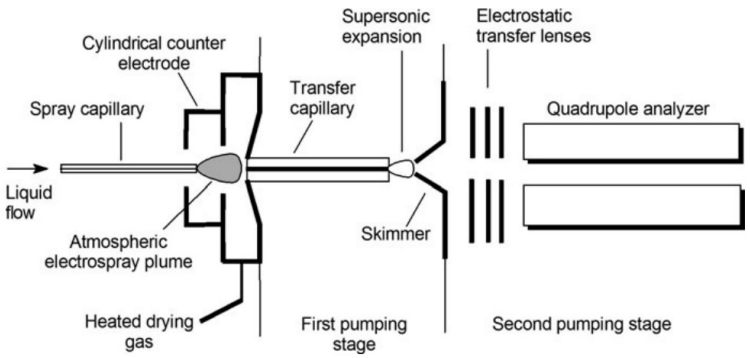


Figure 1.4: Schematic representation of an ESI source.

The analyte solution is applied into a capillary at the entrance of the mass spectrometer. By applying an electric voltage, a spray is created that leads to the desolvation of the ESI droplets and generate charged ions which enter into the mass spectrometer. Figure is adapted from [160].

The formation of these ions in the ESI process can be explained by using two different models: 1) the Ion Evaporation Model (IEM) by Iribarne and Thomson [164] and 2) the Charged Residue Model (CRM) by Dole [162]. The IEM model assumes that single, desolvated, highly charged ions evaporate directly from the surface of the droplets. The CRM model instead postulates that electrospray droplets undergo a series of evaporation-fission cycles, eventually leading to droplets that contain on average one analyte ion. The gas-phase ions form after the remaining solvent evaporation, with the analyte taking all the charges that the droplet carried. Both models are neither proven nor disproved and there are observations which can only be explained by one model or the other [165, 166]. Low molecular weight (MW) species that exist as preformed solution-phase ions are thought to be transferred into the gas phase via the IEM whereas the CRM applies to large multicharged species, such as globular proteins. A chain ejection model (CEM) [163] has also been proposed for disordered polymers and explains dissociation as well as unfolding experiments on proteins in the gas phase.

1.6.2 Mass analyzers

1.6.2.1 Quadrupole mass analyzer

Quadrupoles (Q) are one of the most prominent mass analyzers that are components of triple quadrupole (QqQ) instruments as well as hybrid mass spectrometers [167]. They can be used as pure ion optics for the directed transfer of ions or as mass filters. A quadrupole mass analyzer consists of four hyperbolically or cylindrically shaped rod electrodes extending in the z-direction and mounted in a square configuration (xy-plane, Figure 1.5 A). The pairs of opposite rods are held at the same potential which is composed of a DC (Direct Current, U) and an AC (Alternate Current, V) component. As an ion travels along the z-axis into the quadrupole, it is subjected to an attractive force from one of the rods, which carries a charge opposite to that of the ion. If the rods are subjected to a periodic and alternating voltage, ions oscillate along the x and y axes. The equations describing the quadrupolar field are called Mathieu's equations, from the French mathematician who developed them. The DC and AC voltages and the frequency determine which ions with a particular *m/z* ratio or *m/z* range will travel through the quadrupole in the z-direction on a stable trajectory (region I in the stability diagram, Figure 1.5 B). Ions that are not in this stabilized *m/z* range will not describe stable orbits and move radially out of the quadrupole or parallel and hit the quadrupole's rodes. Without the DC component, quadrupoles operate in the so called 'RF-only mode', meaning they behave as pure ion guides transferring a wide *m/z* range.

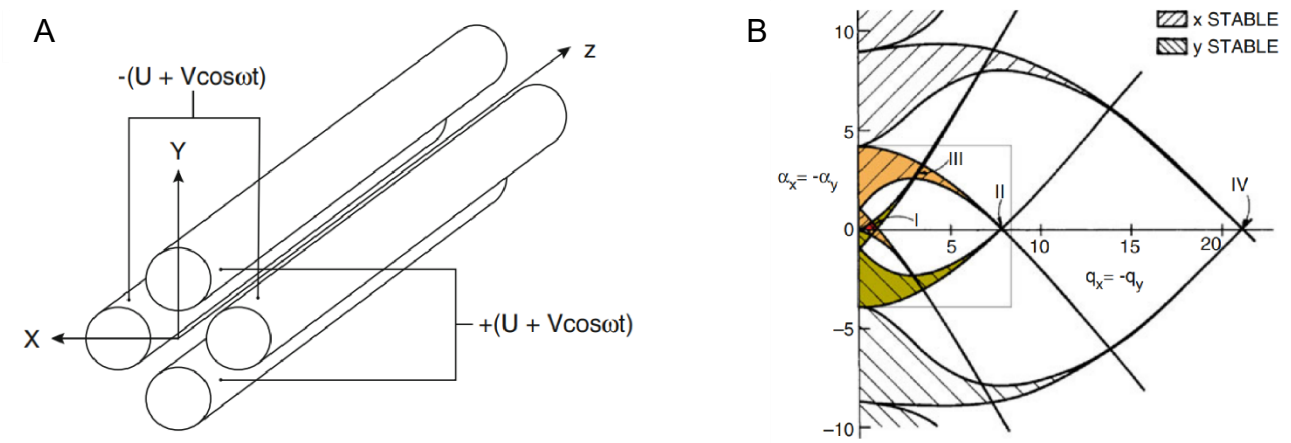


Figure 1.5: Schematic representation of a quadrupole mass filter.

A) Schematic view of a quadrupole with above mentioned axes (U=DC voltage, V=AC voltage). B) Stability diagram for a linear quadrupole analyzer showing four stability regions (I–IV) for x- and y-motion. Figure is adapted from [160].

1.6.2.2 Time-of-flight mass analyzer

The first time-of-flight (TOF) analyzer was developed in 1946 by W. E. Stephens [168]. The principle of TOF is relatively simple: ions of different *m/z* are dispersed in time during their flight along a field-free drift path of defined length [169]. Given that all the ions start their journey at the same time or at least within a short time window, the smaller ions will arrive earlier at the detector than the heavier ones. This requires that they emerge from a pulsed ion source which can be obtained by pulsing ion packages out of a continuous beam or by employing a true pulsed ionization method (e.g. MALDI) [170]. The main advantages of TOF instruments are:

1) In theory, the *m/z* range for a TOF analyzer is unlimited
2) Mass spectra are obtained from ions created during a short (nanoseconds) ionizing event, e.g., from a laser shot in MALDI
3) TOF analyzers offer high ion transmission (and thus high sensitivity)
4) The mass spectral acquisition rate of TOFs is very high, > 10^3 Hz (or mass spectra/s)
5) TOF instrument design/building is comparatively simple
6) Modern TOF instruments allow accurate mass measurements and tandem MS experiments

In a TOF, a charged particle in an electric field has a potential energy, which is converted into kinetic energy, with the following relationships applying:

$$E_{el} = ezU = \frac{1}{2} m_i v^2 = E_{kin} \quad (Eq.\,1.1)$$

$$v = \frac{s}{t} \quad (Eq.\,1.2)$$

$$v = \sqrt{\frac{2ezU}{m_i}} \quad (Eq.\,1.3)$$

where E_{el} is the potential energy, e is the charge of the elctron, z is the charge of the ion, m is its mass and v its velocity. Consequently, v is given by the ratio of the ion path s divided by the time needed for the ion to run that trajectory. From equation (1.3), velocity v is inversely proportional to the square root of *m/z*, implying ions with lower *m/z* values will have a higher velocity and will reach the detector earlier than ions with larger *m/z* values. After acceleration in the electric field, all ions should have identical kinetic energy and in the field-free region they will reach the detector at different times according to their *m/z.* Increasing the distance s the ions travel in the flight tube, higher resolving power is obtained. To achieve the same without increasing the length of the flight tube, a so called 'reflectron' [171] was developed. The reflectron is composed of a series of ring-shaped electrodes at increasing potential. It acts as an 'ion mirror', which substantially diminishes the spread of flight times of the ions with identical *m/z* values caused by differences in the initial kinetic energy of these ions at the exit of the ion source. To further gain resolving power, modern TOF instruments additionally implement 'delayed extraction' [172], which is another way to focus ion packets of the same *m/z* before detection.

1.6.2.3 Orbitrap mass analyzer

The first step that led to the invention of the orbitrap mass analyzer was the implementation of orbital trapping as method for ion trapping, which can itself be used for mass analysis. The Kingdon trap, named after its inventor, was invented in 1923 [173] and utilizes a purely electrostatic field for ion trapping, using neither magnetic (like in FTICR or magnetic sector field instruments) nor dynamic (RF) electric fields (like in quadrupoles or 3D-ion traps). It consists of a thin-wire central electrode, an electrically isolated coaxial outer cylindrical electrode and two endcap electrodes. Knight further developed the principle by changing the shape of the outer electrode so that its radius decreases from the center to the outer edges [174]. This produces an axial quadrupole term in addition to the radial logarithmic term in the electric potential, which allows ions to oscillate axially in the z-direction. Based on this, the orbitrap was invented by Alexander Makarov (Figure 1.6). It can be considered as

a modified 'Knight-style' ideal Kingdon trap with specially shaped inner and outer electrodes [175]. Ions are introduced tangentially into the orbitrap using the C-trap (curved trap) and follow stable oscillations around the central electrode. The quadro-logarithmic static electric field causes the ions, in addition to their radial motion, to move periodically along the z-axis of the central electrode. This movement induces an ion image current, which in analogy to FTICR, is followed by fast Fourier Transform algorithm to convert the recorded frequency-domain signal into a mass spectrum. The axial oscillation frequency (ω_z) of the trapped ions is indirectly proportional to the square root of the *m/z* value, with the following relationships:

$$\omega_z = \sqrt{k\left(\frac{z}{m_i}\right)} \qquad (Eq.\ 2)$$

$$R = \frac{m}{\Delta m} = \frac{1}{2\Delta w_z}\sqrt{k\left(\frac{z}{m}\right)} \qquad (Eq.\ 2.1)$$

While the axial oscillation frequency ω_z in an orbitrap is proportional to $\sqrt{(z/m)}$, the cyclotron frequency ω_c in FTICR is proportional to z/m. Accordingly, FTICR offers higher resolving power for ions with low *m/z* values, while the orbitrap may outperform it at high *m/z* values, since there is no linear decay with increasing *m/z* value.

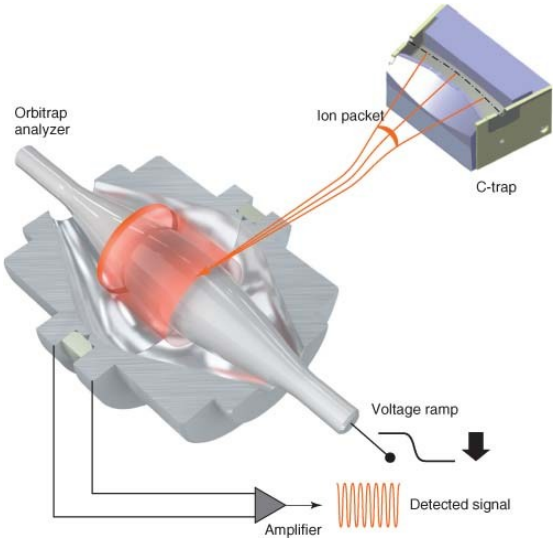


Figure 1.6: Principle of operation of the orbitrap mass analyzer.

Ions are injected as coherent packages from the C-trap tangentially into the orbitrap and follow stable axial oscillations around the central electrode. This generates an image current in the outer electrodes, which is amplified and Fourier transformed to generate mass spectra. Figure is adapted from [160].

1.6.3 MS fragmentation methods

In this work, a variety of different MS fragmentation approaches have been applied. Collision-induced dissociation (CID) is one of the most commonly used techniques. Here, precursor ions are accelerated with the help of an electric potential and undergo collisions with neutral molecules of an inert gas (e.g. nitrogen, helium or argon). Through collisions, a bond cleavage/break is promoted from the precursor ions, generating different types of fragment ions [176]. Usually, the weakest bond is cleaved, which in the case of peptides and proteins represents the peptide bond (HN-C=O bond). Following fragmentation, a series of b- and y-type ions are generated according to where the positive charge localizes. This is explained best by the 'mobile proton theory' [177-180]. In the case of b-type ions, after bond cleavage the charge remains on the N-terminal fragment, while in the case of y-type ions it localizes on the C-terminal fragment (Figure 1.7) [181]. Neutral losses are also common, such as NH_3, H_2O and CO losses. The latter results on the formation of a-type ions from b-type ions. A similar, more informative fragmentation approach is the so called 'higher energy collision-induced dissociation (HCD)', commonly adopted in Orbitrap instruments. With this method, a bond cleavage is also induced between the carbonyl group and the adjacent nitrogen atom of the peptide bond. The difference in this case, is that HCD is a beam-type fragmentation technique, resulting in a more effective fragmentation of primarily precursor ions and the formation of secondary fragments from the already generated product ions, resulting in internal fragments in peptides. In the beginning, the C-trap was used as a collision cell simply by raising the RF voltage from eV to keV [182]. Nitrogen was typically used as collision gas. After several collisions, most of the precursor ions have undergone multiple dissociation events and all ions accumulate in the C-trap. This type of fragmentation has been named 'higher-energy C-trap dissociation'. An increase in RF voltage of the C-trap unfortunately causes a decrease in the *m/z* range that can be stored concurrently in the C-trap. These disadvantages are avoided by attaching an RF-only octopole collision cell to the back end of the C-trap (now known as Ion Routing Multipole, IRM). This octopole collision cell provides fragment trapping over a wide *m/z* range. The fragments generated are then filled back into the C-trap, thermalized, and transferred into the orbitrap for final *m/z* analysis. With the orbitrap fusion tribrid mass spectrometer from Thermo Fisher Scientific [183], the HCD-generated fragment ions can be analyzed both in the dual-pressure linear ion trap or in the orbitrap. Moreover, electron-transfer dissociation (ETD) and variations thereof are also available. ETD [184] is a non-ergodic process, in which low-energy electrons are transferred from an electron-donating radical anionic species to the precursor ions, such as fluoranthene, which leads to their dissociation. In a reaction chamber in the vicinity of the S-lens of the instrument, an electron from the supplied nitrogen is transferred to

fluoranthene by applying a strong electric field. The fluoranthene radical anion is then transferred to the linear ion trap where the selected precursor ions are stored. The electron transfer from the fluoranthene radical to the analyte ion leads to the formation of an unstable charge-reduced radical cation and its fragment ions [184]. Fragmentation takes place along the peptide backbone at the HN-Cα bond, resulting in c- and z-type fragment ions. The instrument is also capable of combing collisional activation schemes with ETD fragmentation, and thus generate as many possible and complementary fragment ions in one single spectrum. As such, it is possible to perform so called 'ETciD' and 'EthcD' experiments, in which the precursor is first subjected to ETD followed by collisional activation by CID or HCD. All the activation approaches described herein have been tested for the gas-phase characterization of the permanent positively charged cross-linker PAC4.

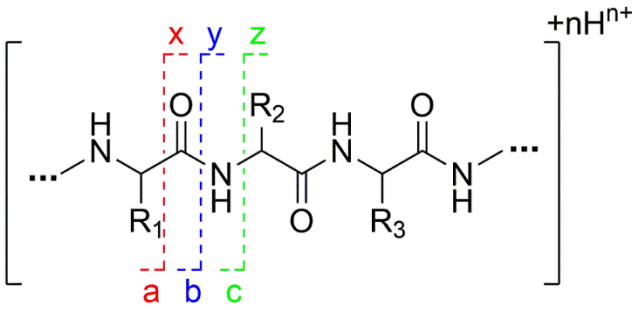


Figure 1.7: Roepstorff-Fohlmann-Biemann nomenclature.

Depending on the fragmentation technique, different bonds are preferentially cleaved generating a- and x-, b- and y-, or c-and z-type ion series.

1.6.4 Native mass spectrometry (Native MS)

Proteins interact with their binding partners mainly through non-covalent electrostatic interactions, Van der Waals and hydrophobic interactions as well as hydrogen bonds. These partners can include proteins, nucleic acids, lipids, sugars and small molecules. As bio-molecular interactions play a fundamental role in biology, their structural investigation is definitely necessary. One of the techniques that has been growingly adopted in the last years is native MS [185-187]. This technique is able to preserve non-covalent interactions between biomolecules, allowing the investigation of the stoichiometry of those intact complexes in the gas-phase, which should reflect the situation arising from the two partners being in solution. Fundamental papers confirmed that transfer from solution to the gas-phase does not alter macromolecules in terms of their structure and function, within the time of analysis [165]. Even organization of protein assemblies can be determined by mixing different subunits of a multiprotein molecular machinery, such as ribosomal proteins. Protein solutions are

generally sprayed from highly volatile salt solutions, such as ammonium/triethylammonium acetate or triethylammonium bicarbonate. These salts guarantee the necessary ionic strength and pH values close to the physiological conditions (pH ~7.4). When large proteins and their complexes are ionized, ions with high *m/z* values (> 3000) need to be analyzed in suitable instruments. These are usually mass spectrometers, which have been modified to increase the transmission of high molecular weight species (up to *m/z* 80,000). The high-mass Q-TOF II (Waters/MS Vision) used in this work and the orbitrap EMR/UHMR series (Extended Mass Range/Ultra High Mass Resolution) are examples of commercially available mass spectrometers. Transmission of ions with high *m/z* vlaues is usually achieved by reducing the RF frequencies in the quadrupole (Q-TOF II) or reduction of the frequency of RF voltages applied to the injection. Also, bent flatapoles, quadrupole, transfer multipole, C-trap and HCD cell, with adjustment of the voltage ramp rate on the central orbitrap electrode, significantly improve transmission of ions with high *m/z* values (orbitrap UHMR) [187]. Native MS has many advantages for studying the topology and stoichiometry of macromolecular complexes. It can be applied to samples that drastically change in mass, flexibility, symmetry and polydispersity. Moreover, multiple oligomeric states can be observed and studied in a single experiment without the need to separately accumulate individual species. In this way, dynamics of the quaternary structure of complexes can be investigated in real time. Furthermore, native MS does not require complex preparation of the samples to be analyzed; typically, a buffer exchange step in a compatible and volatile buffer is required. Despite having many benefits, this technique presents also some disadvantages. First, the analysis of intact protein complexes is performed in the gas-phase; this results in discrepancies in the relative abundances of the detected species. Ion abundances depend on ionization efficiency, transmission and detectability of each individual complex, making it not applicable for rigorous quantification purposes. Also, the strength of non-covalent interactions might change in the gas-phase. As the aqueous solvent is evaporated, hydrophobic interactions become much weaker, while electrostatic interactions as well as hydrogen bonds become stronger. This leads to the fact that complexes that are mainly based on hydrophobic interactions, are very difficult to analyze. Furthermore, protein interactors, which individually have their own oligomeric state, might be difficult to probe for binding if their K_d for homo-oligomerization is lower than the K_d for complex formation. Here, a mild fixation of the complex by a cross-linking reagent prior to analysis could solve this issue. Native MS can also be coupled with other techniques, such as ion mobility spectrometry (IMS) or chemical cross-linking [188]. On one hand, IMS provides information about protein conformation via measurement of its CCS. On the other hand, chemical cross-linking can be

used to fix weak and/or transient interactions, obtain low-resolution structural data and determine the interaction surfaces of proteins in complex.

1.6.5 Hydrogen/deuterium exchange mass spectrometry (HDX-MS)

Probing dynamics of proteins in solution is not trivial. Folded proteins, such as globular proteins as well as enzymes, experience defined conformational states in solution or even partially unfolded states within time scales usually in a range from hours to several days [189]. In the context of IDPs, the situation changes dramatically as these proteins sample a huge and flat surface composed of an abundance of different isoenergetic conformers, as well as unfolded populations within a much faster time scale (from nanoseconds to milliseconds, depending on the specific protein) [190]. This makes these kind of proteins quite difficult to probe in terms of structural dynamics. NMR has been extensively used in the last years to study fast dynamic processes in IDPs. Recently, hydrogen-deuterium exchange mass spectrometry (HDX-MS) [121] has become quite a popular approach to tackle this issue. Different papers have been published, in which milli-second HDX-MS using stop-flow apparatus is able to capture and study dynamics of IDPs such as α-synuclein . HDX-MS displays protein dynamics and changes in a time-dependent manner by giving information on hydrogen bonding networks and complements techniques, such as XL-MS and protein footprinting. It measures changes in mass associated with the incorporation of deuterium into backbone amide hydrogens when a protein is diluted into a deuterated solution (usually D_2O). Side chains of amino acids can exchange deuterium as well, but typically undergo fast back-exchange within the time of analysis and are not detected. The rate of H/D exchange in a polypeptide chain, provided that pH, temperature, and ionic strength are constant, depends on the folded state of the protein and its dynamics. Specifically, the stability of hydrogen bond networks can be investigated, with backbone amide protons involved in stable hydrogen bonds displaying limited H/D exchange when compared to amide protons exposed to the solvent or involved in weak hydrogen bonding. First HDX measurements date back to the 1950s. NMR spectroscopy was used as method of detection in 1970-1980s. Mass spectrometry became the method of choice for detecting HDX since the 1990s; this is because it is able to analyze large proteins, accepts low-concentrated samples, and tolerates complex sample matrices. Advances in LC-MS technology as well as different software pipelines have increased the speed of data analysis by several orders of magnitude in the last ten years. Nowadays, it is possible to analyze large multi-subunit proteins within hours, while in the past the whole analysis could take up to weeks or months. However, caution should be taken when performing such experiments. For example, a sample quality

assessment of the protein of interest should precede the HDX experiments. As such, native MS analysis or size exclusion chromatography to investigate the oligomeric state of the protein as well as functional assays should be performed beforehand. Additionally, the buffer used in the labeling reaction should have enough buffering capacity to ensure constant pH values and the protein solution has to be pre-equilibrated at the temperature used for the exchange reaction and kept stable during D_2O labeling. Finally, in a rigorous HDX experiment, the level of back exchange of the LC-MS system should be defined to correct the % of deuterium uptake for each peptide in the analysis.

1.6.6 Cross-linking mass spectrometry (XL-MS)

1.6.6.1 Principle of the technique and historical timeline

Classical high-resolution approaches for protein 3D-structure analysis, such as X-ray crystallography and NMR spectroscopy, have been able to solve the structures of a multitude of proteins that are deposited in the PDB database (https://www.rcsb.org). However, their structure determination power reached their limits with IDPs or membrane proteins. For IDPs, this is due to their large degree of flexibility and disorder which makes them unsuitable for structural resolution, requiring alternative techniques. XL-MS has emerged as a powerful approach to characterize also, but not only, IDPs and IDRs. XL-MS is a low-resolution structural method; the very first application of cross-linking reagents, such as glutaraldehyde to fix cells, dates back to the 1960s [191]. Chemical cross-linking has been coupled with MS detection in the early 2000s [192, 193]. In the beginning, the technique was used to characterize only single proteins. Nowadays, it has reached a mature standpoint, thanks to the outstanding advances in LC-MS instrumentation as well as the development of several cross-linking reagents [6]. In fact, it is now possible to characterize protein complexes as well as protein assemblies/molecular machineries composed of several proteins up to the investigation of protein interaction networks in cell lysates, intact cells, organelles and tissues. These are known as 'proteome-wide' XL-MS studies which are becoming more and more important to understand and characterize dynamic signaling pathways as well as identifying transient interactions in a more '*in vivo* like' scenario on a system-wide scale [9, 194]. The principle of the technique is straightforward: a bifunctional organic reagent, termed cross-linker, is able to bridge two amino acid side chains in a protein or protein complex. This defines a cross-linker as a sort of 'molecular ruler', which can be literally used to measure distances within a protein. A great number of cross-linkers have been developed so far; the selection of a specific reagent depends on the scientific question to address [117]. In case low-resolution restraints for protein structural analysis are needed, short as well as

'zero-length' reagents can be used, with the last ones providing the highest spatial resolution. If an interaction network of a protein is to be characterized, medium to longer reagents (12 Å or longer) are more suitable for that purpose.

1.6.6.2 Anatomy of a cross-linker

1.6.6.2.1 Cross-linking chemistries

A chemical cross-linker is characterized by four main features (Figure 1.8). The first one is the amino acid reactivity, encoded in the reactive groups present on each side of the molecule. So far, different classes of amino acids can be targeted. A cross-linker can be defined as homobifunctional if both reactive groups possess the same amino acid specificity, otherwise they are called heterobifunctional [195]. In the latter case, different combinations of reactive groups are employed, making these reagents attractive to study proteins which lack the amino acid targeted by one single reactive group. However, some of them commonly require a two-step labeling protocol. Here, a protein with the first modifiable reactive site is labeled. After labeling, the excess of cross-linker is usually removed by filters with a certain molecular weight cut-off (MWCO). Then, the protein interaction partner is added and the second cross-linker reactivity is induced. A classic example is the heterobifunctional amine/acidic reactive succinimidyl diazirine (SDA) [137, 196]. Lysine is the most commonly targeted amino acid and a very good nucleophile, usually reacted with the classical N-hydroxy succinimide (NHS) ester chemistry [197] (subchapter 1.6.6.2.2).

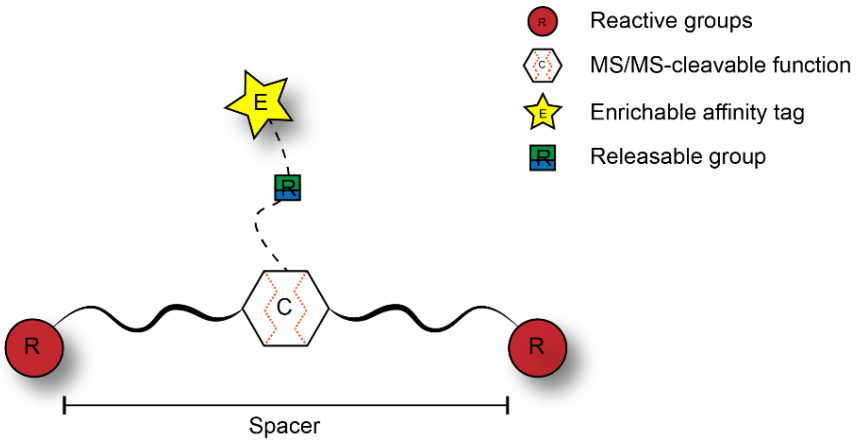
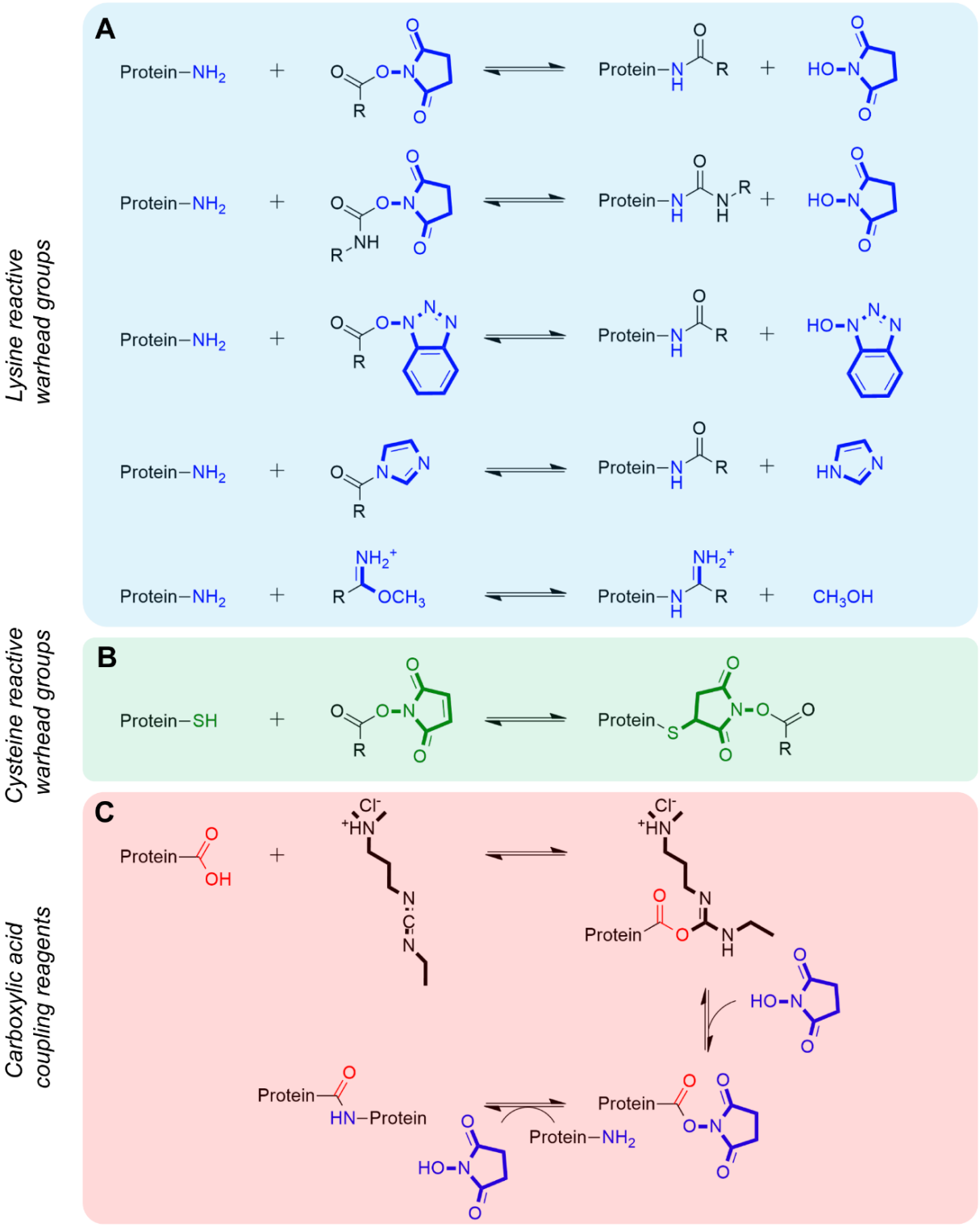


Figure 1.8: Illustration of the main features of a cross-linker.

A cross-linker has two reactive groups with specific amino acid reactivities (red spheres). It can optionally contain a MS/MS-cleavable function (white hexagon), an enrichable tag (yellow star) and a releasable group (green/blue square) to selectively release cross-linked peptides. The spacer length of the cross-linker (hence

the distance it can measure) can vary so to have zero length/short (0-5 Å), medium (10-15 Å) and long (> 15 Å) cross-linking reagents.

Alternative chemistries are also available, such as carbamates, benzotriazoles, imidazoles and imidoesters, but generally less used (Figure 1.9A) [135, 198-200]. Cross-linking reactions are normally carried at pH 7.4 so to mimic a physiological environment, at which lysine residues are almost completely protonated and hence, not reactive. Nevertheless, this is still enough to guarantee acceptable cross-linking efficiency. It has to be noted that the reaction would be far more efficient at pH 10, but with consequent denaturation of the protein under investigation. NHS reactivity has been extensively characterized: Not only lysine residues, but also protein N-termini and hydroxyl-containing amino acids can react forming an amide bond or an ester bond [197]. Cysteines are also well-known soft nucleophiles, which can be targeted with a large number of reactive warheads, with the maleimide chemistry being the most used, proceeding via Michael addition (Figure 1.9 B) [136, 141]. On the contrary, the acidic side chains of aspartate and glutamate residues are quite weak nucleophiles, but quite abundant in proteins. In order to react, carboxylic acids usually need to be activated with coupling reagents such as 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) coupled with NHS chemistry (Figure 1.9 C) [201] as well as 4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) (Figure 1.9 D) [202] alone or coupled with hydrazide cross-linking (Figure 1.9 E). Additionally, diazirine chemistry ((Figure 1.9 F) has been also used in photo-cross-linking to target carboxylic groups in proteins via formation of linear diazo-compounds [137]. On the other hand, diazirines target all 20 amino acids via formation of a short-lived reactive carbene [203]. Other reactive groups targeting all 20 proteinogenic amino acids by UV-A irradiation are aryl azides and benzophenones (Figure 1.9 G, H) [204]. Photo-reactive cross-linking principles can also be introduced in the form of unnatural amino acids, such as photo-metionine or photo-leucine [205]) directly via a genetic engineering approach developed by Schultz [206]. Last but not least, arginine can be targeted via glyoxal-based chemistry, such as the so called ‘ArGO/KarGO’ reagents [207], and finally also tyrosine and tryptophane/phenylalanine residues (Figure 1.9 I, L) [208, 209].


A
Lysine reactive warhead groups
Protein–NH2
Protein–NH
R
HO–N
B
Cysteine reactive warhead groups
Protein–SH
Protein
S
C
Carboxylic acid coupling reagents
Protein
OH
Cl-
+HN
N
C
HO–N
HN–Protein
Protein–NH2
CH3OH

(continued).

Carboxylic acid coupling reagents

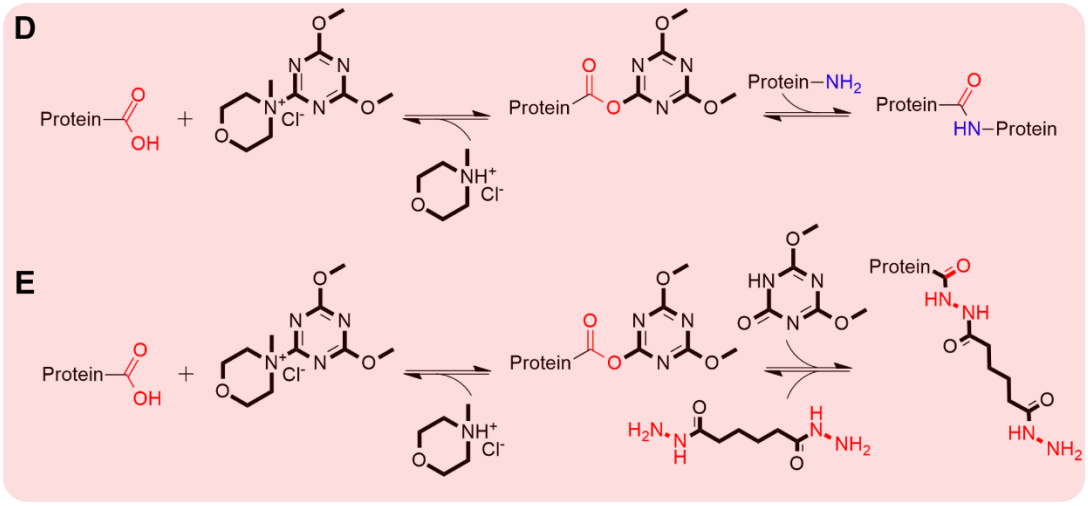
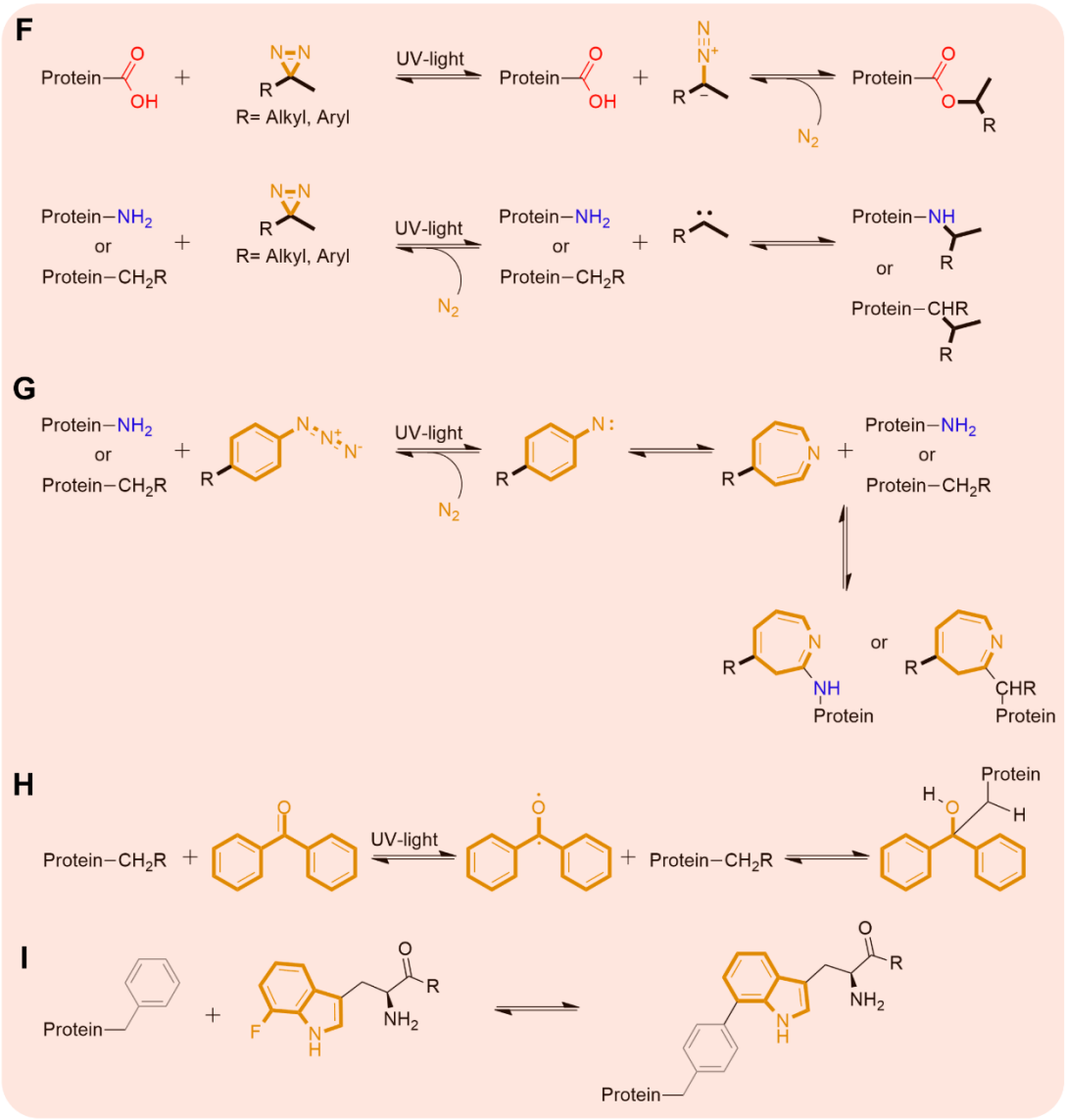


Photo-cross-linking principles



Arginine reactive warhead groups

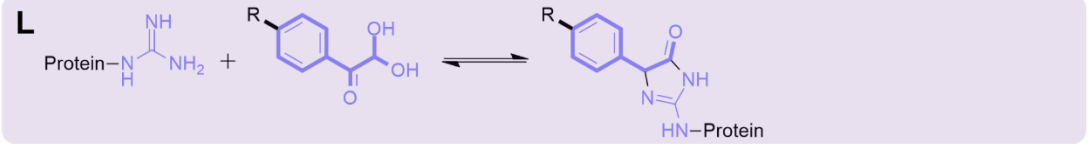


Figure 1.9: Summary of the main cross-linking chemistries.

A) Reactive ("warhead") groups targeting lysine residues (as well as protein N-termini and hydroxy group-containing amino acids, such as serine, threonine, and tyrosine residues). B) Maleimide warhead group for targeting cysteine residues. C) EDC carboxylic acid coupling to amines. D) DMTMM carboxylic acid coupling to amines. E) DMTMM carboxylic acid coupling with hydrazide-based cross-linkers. F) Photo-cross-linking with diazirines proceeds according to two major pathways: isomerization to linear diazo compounds with subsequent functionalization of carboxylic acids (upper panel); formation of a short-lived carbene intermediate that can insert into a C-H or heteroatom-H bond (lower panel). G) Photo-cross-linking with aryl azides forming short-lived nitrenes that rearrange and react with nucleophiles or insert into C-H bonds. H) Photo-cross-linking with benzophenones creates a biradical species upon UV irradiation. The oxygen radical then abstracts a hydrogen radical. The alkyl radical generated reacts by forming a new C–C bond between the photophore and the receptor protein. I) Photo-cross-linking with 7-F-tryptophane, leading to the formation of a new C-C bond with phenylalanine residues. L) Glyoxal-based chemistry for targeting arginine side chains in proteins.

1.6.6.2.2 N-hydroxy succinimide ester cross-linkers

For the most commonly used NHS-ester based cross-linkers, various cross-linking products are identified. These are intra-peptide cross-links (also known as 'loop-links' or type I cross-links), in which the cross-linker is bridging two amino acids within the same peptide, 'dead-ends' (or type 0 cross-links), in which one side of the cross-linker has reacted with a peptide and the second group has been hydrolyzed by water/ammonia, as well as inter-peptide cross-links (type II cross-links) which are the most interesting for the analysis (Figure 1.10). The analysis of XL-MS experiments is generally performed using a 'bottom-up' approach, in which the cross-linked proteins are digested into peptides with the aid of a suitable protease. This is usually done in-solution right after the cross-linking reaction or after electrophoretic gel separation (SDS-PAGE) to enrich for the cross-linked species of interest. The resulting peptide mixtures are separated via liquid chromatography and then analyzed by mass spectrometry. The data are then processed with a suitable software able to search for cross-linked products. Here, the experimentally obtained fragment ion spectra are compared with *in silico* theoretically predicted spectra and the cross-linked spectral matches (CSMs) are reported.

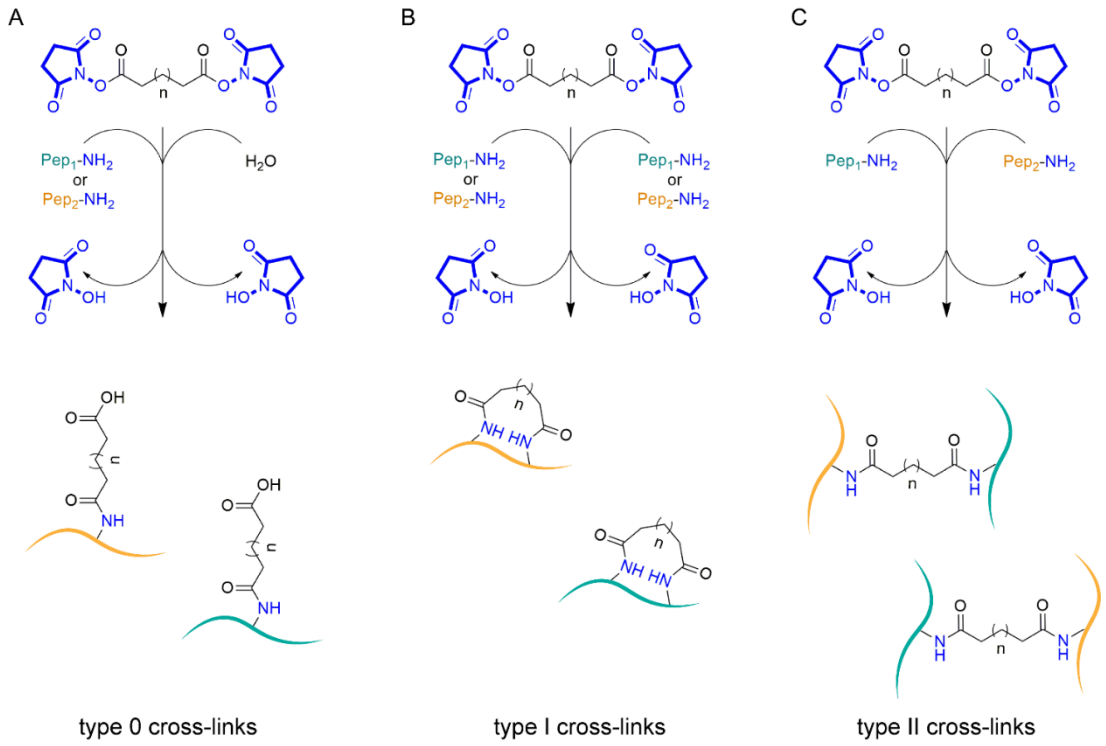


Figure 1.10: Nomenclature of cross-linked products.

When the XL-MS reaction takes place, three major products are formed: 'loop-links' (or type I cross-links), in which the cross-linker is bridging two amino acids (lysines shown here) within the same peptides, 'dead-ends' (or type 0 cross-links), in which one side of the cross-linker has reacted with the peptide while the other one is hydrolyzed by water, and finally inter-peptide cross-links (or type II cross-links), in which two different peptides are connected. The corresponding nomenclature is used accordingly to Schilling et al. [210].

1.6.6.2.3 MS/MS-Cleavable cross-linkers

In addition to their reactivities, cross-linkers can be grouped into cleavable and non-cleavable reagents [211]. MS/MS cleavability is achieved by introduction of a labile functional group (generally in the middle of the spacer), which can be fragmented in the mass spectrometer generating diagnostic signature ions for each peptide involved in the XL-MS reaction (Figure 1.8). Examples are the cross-linkers disuccinimidyl dibutyric urea (DSBU) and disuccinimidyl sulfoxide (DSSO), which respectively possess a urea and a sulfoxide as cleavable groups [128, 129]. The greatest benefit of MS-cleavable cross-linkers is their application to complex matrices, such as cells or tissues. This is due to the reduction of the so called quadratic ('n^2') search space problem to a linear (2n) search [212]. As the software has to search for each linear peptide modified with a cross-linker fragment, the search space becomes linear and does not increase quadratically with the number of peptide combinations. Some software pipelines are able to analyze both cleavable and non-cleavable cross-linkers (such as MeroX, XlinkX, Kojak, XiSearch, MaxLynx [213-217]), while some others are

developed for analyzing only cleavable (such as MS Annika [218]) or non-cleavable reagents (such as pLink 2 [219]).

1.6.6.2.4 Enrichment of cross-linked species

It is important to note that when cross-linking complex mixtures, the number of cross-linked lysine residues is usually very low (estimated between 1-5%). This means that in XL-MS experiments, cross-linked peptides have much lower abundance than linear peptides. To alleviate this situation, several analytical strategies have been developed. In the first place, extensive pre-fractionation of cross-linked samples has been applied by performing size exclusion (SEC) and strong cation-exchange (SCX) chromatography [149]. These two approaches exploit properties, such as size or charge, which are generally increased in cross-linked peptides. This usually results in large amounts of samples, each one still containing a background of unmodified linear peptides. A second workaround consists in installing an enrichment handle on the cross-linker (Figure 1.8). These reagents are known as 'trifunctional cross-linkers' and allow to separate cross-links from linear peptides via highly specific affinity chromatography. So far, two generations of trifunctional cross-linkers have been designed. The first one makes use of a biotin affinity handle directly installed as third function in the spacer region of the cross-linker [220]. The second generation relies on the introduction of smaller chemical functionalities than a biotin group in the cross-linker's spacer to perform selective enrichment of cross-links. These functions include azides/alkynes for biorthogonal click-chemistry enrichment (e.g. SPAAC, CuAAC) [11, 221, 222] and, more recently, a phosphonate handle to perform immobilized ion metal affinity chromatography (IMAC) enrichment of cross-linked peptides. Two examples are the enrichable cross-linkers DSBSO (disuccinimidyl bissulfoxide) [132] and DSPP (disuccinimidyl phenyl phosphonic acid, commonly named as PhoX) [10]. The latter linker has been developed by the groups of Heck and Scheltema, inspired by the groundbreaking advancements in phosphoproteomics, where IMAC has reached high levels of enrichment specificity. In the PhoX linker that contains a phosphonate group, the P-O bond of the phosphate is replaced by a P-C bond. This substitution solves the problem of phosphate group's lability during sample handling and MS analysis. Additionally, since the P-C bond is not amenable to hydrolysis, phosphopeptides can be removed by treatment with a phosphatase prior to the enrichment step. Finally, the cross-linker can have a releasable group to detach cross-linked peptides after enrichment on beads. For example, DSBSO contains an acidic cleavable acetal group to release 'clicked' cross-linked peptides on-beads.

In most cases, the beads contain cleavable groups, either by UV light or by chemical reduction, to selectively release the cross-links from the beads after enrichment [223].

1.6.6.2.5 Cross-linkers used in this work

In this work, the non-cleavable cross-linker bis-sulfosuccinimidyl glutarate (BS^2G) was used to investigate structural changes in the tumor suppressor protein p53 (see chapter 4.1.2). Additionally, two trifunctional cross-linkers, named PAC4 [138] and DSSI, were tested (see chapters 4.2 and 4.3). DSSI was also compared to DSBU for proteome-wide XL-MS applications.

2 Aim of the work

Despite the enormous effort research has made in the last 40 years, the structure of the tumor suppressor p53 protein has not yet been elucidated. As p53 is an intrinsically disordered protein (IDP), traditional methods for protein 3D-structure elucidation showed their limitations in unveiling the disordered regions (IDRs) of p53. To date, different models have been proposed for the active form of p53, which act as a homotetramer *in vivo*, both in the DNA-free and DNA-bound state. Those models were generated using both experimental as well as computational approaches and differ substantially in their domain organization (Figure 1.3). To further investigate the structural properties of p53, with a special focus on the understudied disordered C-terminal region, an integrative structural biology approach has been pursued. Specifically, the topology of p53 in the absence and presence of DNA in solution was investigated. In contrast to the majority of structural data published for p53 where IDRs are missing completely or where IDR-derived peptides are used as substitutes, full-length wild-type human p53 was recombinantly expressed following established protocols [7, 120] and used for this study. The combination of complementary structural MS methods, structure prediction with AlphaFold2, data-driven protein modeling, and in-depth scoring and validation of derived models allowed getting novel insights into the C-terminal IDR of p53 and evaluate potential structural changes thereof upon DNA binding.

In order to investigate p53 from a structural point of view in a more '*in vivo*-like' perspective, XL-MS studies should preferably be performed in cellular systems. Trifunctional MS-cleavable reagents are needed to define entangled protein-protein interaction networks in their native environment. Two trifunctional CID-MS/MS cleavable cross-linkers were evaluated in this work with the aim of developing an integrated XL-MS platform enabling the study of IDPs in the cellular milieu.

3 Materials and methods

3.1 Materials

3.1.1 Chemicals

Acetic acid	CH_3COOH	Roth
Acetonitrile (LC-MS grade, HiPerSolv)	CH_3CN	VWR
Acrylamide/Bisacrylamide solution (37.5:1); 40% (w/v)		Roth
Agar-Agar		Merck
Agarose (SeaKem LE)		Lonza
Ammonium acetate	NH_4CH_3COO	Sigma-Aldrich
Ammonium bicarbonate	NH_4HCO_3	Sigma-Aldrich
Ammonium persulfate	$(NH_4)_2S_2O_8$	Sigma-Aldrich
Coomassie Brilliant Blue G250		Sigma-Aldrich
Coomassie Brilliant Blue R250		Sigma-Aldrich
Cesium Iodide	CsI	Sigma-Aldrich
cOmplete Protease Inhibitor (EDTA-free)		Roche
D-Glucose	$C_6H_{12}O_6$	Sigma-Aldrich
Dimethylsulfoxide	DMSO	Thermo Fisher Scientific
Dithiothreitol	DTT	AppliChem
DNA ladder (1kb, 100 bp)		Thermo Fisher Scientific
DNA sample buffer (6x)		Thermo Fisher Scientific
Ethanol	C_2H_5OH	Merck
Ethylen-diamino-tetraacetic-acid	EDTA	Sigma-Aldrich
Formic acid	FA	Roth
Glycerol		Merck
Hydrochloric acid (0.1 N, endotoxin-free)	HCl	Sigma-Aldrich
4-(2-Hydroxyethyl)-1-piperazinethan-sulfonic acid	HEPES	Roth

Iodoacetamide	IAA	Sigma-Aldrich
Imidazole		Merck
Isopropanol	i-PrOH	Merck
Isopropyl-β-D-1-thiogalactopyranoside	IPTG	Roth
LaemmLi sample buffer (2x)		Bio-Rad
Magnesium chloride	$MgCl_2$	Thermo Fisher Scientific
Methanol	CH_3OH	VWR
2-Mercaptoetanol		Roth
Nickel chloride	$NiCl_2$	Merck
PageRuler Prestained Protein Ladder		Thermo Fisher Scientific
PageRuler Plus Prestained Protein Ladder		Thermo Fisher Scientific
Potassium chloride	KCl	Roth
Rotiphorese 10x SDS-PAGE		Roth
Sodium chloride	NaCl	Roth
Sodium dodecyl sulfate	SDS	Roth
Sodium hydrogenphosphate	Na_2HPO_4	Sigma-Aldrich
Sodium hydroxide	NaOH	Sigma-Aldrich
N, N, N', N'-Tetramethylethylendiamine	TEMED	Bio-Rad
Trifluoroacetic acid	TFA	Merch
Tris(2-carboxyehtyl)phosphine hydrochloride	TCEP	Sigma-Aldrich
Tris(hydroxymethyl)aminomethane	TRIS	Sigma-Aldrich
Trypton/Pepton from casein		Roth
Urea		Roth
Yeast extract		Roth
Zinc chloride	$ZnCl_2$	Sigma-Aldrich

3.1.2 Equipment

3.1.2.1 Chromatographic systems

UHPLC system:	Agilent Technologies 1200 Series	Agilent
Nano-HPLC system:	Ultimate 3000	Dionex
Precolumn:	Acclaim PepMap C18 300 µm x 5 mm, 5 µm, 100 Å	
Analytical columns:	uPAC pillar array 50 cm, 5 µm i.d., 2.5 µm interpillar distance;	Thermo Fisher Scientific
	self-packed Picofrit nanospray emitter (360 µm od, 75 µm id× 150 mm length, 15 µm tip id) with C18-stationary phase (3.0 µm, 120 Å)	New Objective Dr. Maisch GmbH
FPLC systems:	ÄKTA FLPC Pure 25	GE Healthcare
	ÄKTA FLPC Prime Plus with Fraction collector Frac-920	GE Healthcare
Columns:	HisTrap FF, 1 mL	GE Healthcare
	HisTrap FF, 5 mL	GE Healthcare
	HiLoad Superdex 200 pg 16/600	GE Healthcare
	HiLoad Superdex 200 pg 26/600	GE Healthcare
	Superdex Peptide 3.2/300	GE Healthcare

3.1.2.2 Mass spectrometers

Orbitrap LTQ-XL mass spectrometer	Thermo Fisher Scientific
Orbitrap Fusion Tribrid mass spectrometer	Thermo Fisher Scientific
High-mass Q-TOF II mass spectrometer	Micromass/MS Vision
timsTOF Pro mass spectrometer	Bruker Daltonik

All mass spectrometers used in this work were coupled to nano-HPLC systems via nano-ESI sources.

3.1.2.3 Laboratory equipment

Balance LE225D	Sartorius
Autoclave V75	Systec
DNA electrophoresis chamber Mini-Sub-Cell GT	Bio-Rad
Electrophoresis system Mini-Protean Tetra Cell	Bio-Rad
Gel documentation system Gel Doc XR	Bio-Rad
Incubation shaker innova 44	New Brunswick Scientific
Incubation shaker Thermoshake	Gerhardt
Laminar flow hood HERAsafe KS	Thermo Fisher Scientific
Magnetic stirrer R3T	mLw
Magnetic stirrer M21/1	Framo
Milli-Q H_2O dispenser	Thermo Electron LED
NanoDrop^{TM} spectrophotometer	Thermo Fisher Scientific
pH meter PH211	Hanna Instruments
Pipettes (0.1-5000 µL)	Eppendorf and VWR
Platform shaker Titramax 101	Heidolph
Voltage transmitter PowerPac 300	Bio-Rad
Thermocycler TPersonal	Biometra
Thermocycler	Eppendorf
Ultrasonic bath Ultrasonic cleaner	VWR

Ultrasonic homogenizer Vibra-Cell 75185	Sonics & Materials
Ultrasonic homogenizer	Thermo Fisher Scientific
Vacuum concentrator Savant SPD1010	Thermo Fisher Scientific
Vortex mixer 7-2020	neoLab
Vortex mixer	VWR
Centrifuge 5414R	Eppendorf
Centrifuge 5804R	Eppendorf
LED photo-cross-linking device	Home-built (Center for Structural MS, MLU Halle-Wittenberg)

3.1.2.4 Laboratory consumables

Amicon Ultra 15 mL centrifugation tubes (Molecular weight cut-off: 10 kDa, 30 kDa)	Millipore
Amicon Ultra 0.5 mL centrifugation unit filters (Molecular weight cut-off: 10 kDa, 30 kDa)	Millipore
Centrifugation tubes (15 mL, 50 mL)	Corning
DNA LoBind tubes	Eppendorf
Plastic cuvettes	Sarstedt
Protein LoBind tubes	Eppendorf
Safe Lock reaction tubes	Eppendorf
Sterilization filters (Filtropur S 0.2)	Sarstedt
Ultracentrifugation tubes	Beckman

3.1.3 Plasmid

HLT_p53FL_pET28a	Bastian Bräuning, Weizmann Institute, Rehovot, Israel

3.1.4 Oligonucleotide

The oligonucleotide sequences used in this work were ordered from Microsynth (Balgach, Switzerland). After synthesis, the oligonucleotides were purified via HPLC and used directly according to manufacturer's instructions without any further cleaning steps.

REP_TS	5'-CGCGGACATGTCCGGACATGTCCCGC-3'
REP_BS	5'-GCGGGACATGTCCGGACATGTCCGCG-3'

3.1.5 Model peptides

Angiotensin II (ATII)	DRVYIHPF
Test peptide 1 (TP1)	Ac-TRTESTDIKRASSREADYLINKER

3.1.6 Model proteins

Bovine serum albumin (BSA)	Sigma-Aldrich
Carbonic anhydrase (CA2)	Sigma-Aldrich

3.1.7 Bacterial strains

BL21 (DE3)	*E.coli, F⁻ompT gal dcm Ion hsdS$_B$ ($r^-_Bm^-_B$) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*
DH5α	*E.coli, F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF) U169, hsdR17 (rK- mK+), λ-*

3.1.8 Media and antibiotics

Antibiotics:

Kanamycin	30 mg/mL in H_2O, final concentration; 30 µg/mL, Roth

Media:

2 x YT-Medium	16 g/l Trypton, 10 g/l Yeast extract, 5 g/l NaCl, pH 7.0
Agar plates	15 g/l Agar-Agar in LB-Medium
LB medium	10 g/l Tryptone, 5 g/l Yeast extract, 10 g/l NaCl, pH 7.2
SOC medium	20 g/l Tryptone, 5 g/l Yeast extract, 0.5 g/l NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM Glucose, pH 7.0

3.1.9 Buffers

All the buffers were prepared using Milli-Q H_2O deionized and of the highest purity, if not stated otherwise.

TAE buffer (50x)	100 mM EDTA (0.5 M, pH 8.0), 5.71% (v/v) CH_3COOH, 2 M Tris
IMAC binding buffer	50 mM Na_2HPO_4 pH 8.0, 300 mM NaCl, 2.5 mM TCEP 20 mM imidazole
IMAC elution buffer	50 mM Na_2HPO_4 pH 8.0, 300 mM NaCl, 2.5 mM TCEP 500 mM imidazole
SEC buffer	50 mM HEPES pH 7.2, 300 mM NaCl, 2.5 mM TCEP, 10 % (v/v) Glycerol
Phusion buffer	5 x HF-Buffer
Peptide cross-linking buffer	50 mM HEPES, pH 8.5
Protein cross-linking buffer	50 mM HEPES, pH 7.4
HEK293T lysate cross-linking buffer	20 mM HEPES, 150 mM KCl, 10 mM $MgCl_2$, pH 7.4

SDS-PAGE solutions:

PAGE staining solution:	1% (w/v) Coomassie Brilliant Blue R250 in 40% (v/v) methanol, 10% (v/v) acetic acid
PAGE destaining solution:	10% (v/v) acetic acid, 40% (v/v) methanol

HPLC phases:

Loading pump solution:	0.1% (v/v) TFA in H_2O (LC/MS grade, VWR)
Nano pump solution A:	0.1% (v/v) FA in H_2O (LC/MS grade, VWR)
Nano pump solution B:	0.08% (v/v) FA in acetonitrile (LC-MS grade, VWR)

3.1.10 Cross-linking reagents

Bis(sulfosuccinimidyl) glutarate-D_0	BS^2G-D_0	Thermo Fisher Scientific
Bis(sulfosuccinimidyl) glutarate-D_4	BS^2G-D_4	Thermo Fisher Scientific
Disuccinimidyl dibutyric urea	BuUrBu/DSBU	CF Plus Chemicals/Bruker Daltonik
1,1-bis(4-((2,5-dioxopyrrolidin-1-yl)oxy) -4-oxobutyl)-4-ethynylpiperidin-1-ium	PAC4	CF Plus Chemicals
(2-(4-((2,5-dioxopyrrolidin-1-yl)oxy)-N-(4-((2,5-dioxopyrrolidin-1-yl)oxy)-4-oxobutanoyl) -4-oxobutanamido)ethyl)phosphonic acid	DSSI	CF Plus Chemicals

3.1.11 Enzymes

AspN	New England Biolabs
Benzonase nuclease	Promega
ProTEV Plus	Promega
Trypsin (porcine, sequencing grade)	Promega

3.1.12 Commercially available kits

GeneJET Plasmid Miniprep-Kit	Thermo Fisher Scientific
Bio-Rad Protein Bradford Assay	Bio-Rad

3.1.13 Software

Chromeleon (7.0)	Control of the HPLC-systems of the mass spectrometers (Thermo Fisher Scientific)
MassLynx (4.1)	Acquisition and processing of data from high-mass Q-TOF II mass spectrometer (Waters)
MeroX (2.0.1.7)	Identification and evaluation of XL-MS products from non-cleavable and MS/MS-cleavable cross-linkers (Michael Götze, MLU-Halle-Wittenberg, www.StavroX.com)
Proteome Discoverer (2.4)	Analysis of MS, MS/MS, and LC-MS data (Thermo Fisher Scientific)
Peptide Cutter/Peptide Mass	*In-silico* digestion of proteins with single and/or proteases combination (Expasy, www.expasy.org)
ProtParam	Calculation of pI and extinction coefficient of proteins (Expasy, www.expasy.org)
Protein Prospector	*In silico* generation of fragment ion spectra of peptides (https://prospector.ucsf.edu/prospector/mshome.htm)
pyCirclize	Python package to display circos plots from cross-linking data (https://moshi4.github.io/pyCirclize/)
Qual Browser (4.3.73.11)	Visualization and processing of LC-MS data (Thermo Fisher Scientific)
Quantity One (4.6.2)	Visualization and processing of SDS-PAGE and Agarose gels from Gel Doc XR (Bio-Rad)
SnapGene (2.8.1)	Visualization of DNA sequences (www.snapgene.com)

Unicorn (7.0)	Control of the ÄKTA-FLPC Pure 25 System (GE Healthcare)
Unicorn (5.11)	Control of the ÄKTA-FLPC Prime System (GE Healthcare)
Xcalibur (4.0)	Analysis of data from orbitrap mass spectrometers (Thermo Fisher Scientific)
XiNET	Representation of cross-linking data on protein sequences (https://crosslinkviewer.org/)
PyXLinkViewer	PyMOL plug-in for XL-MS data mapping into high-resolution protein structures

3.2 Methods

Unless stated otherwise, all the experiments have been performed with MilliQ-H_2O.

3.2.1 Molecular biology methods

3.2.1.1 Preparation of chemically competent cells

The preparation of chemically competent cells was carried out according to the protocol of $CaCl_2$ method from Sambrook and Russell [224]. 200 mL of LB medium were inoculated with 5 mL of an overnight culture of *E. coli* DH5α or BL21 (DE3) cells and cultured at 37 °C up to an optical density OD_{600} of ~0.6. The cells were then centrifuged for 15 minutes at 4,000 g, 4 °C). The cell pellet was resuspended in 50 mL of a solution of 0.1 M $CaCl_2$ (4 °C and sterile-filtered) and incubated on ice for 90-120 minutes. After a second sedimentation step (15 minutes, 4,000 g, 4 °C), the cell pellet was resuspended in 2 mL of a 0.1 M $CaCl_2$ solution and incubated for 90-120 minutes. Finally, the suspension was mixed with glycerol (final concentration 10% (v/v)) and shock-frozen in aliquots of 50 µL in liquid nitrogen and stored at -80 °C.

3.2.1.2 Transformation of chemically competent cells

An aliquot of 50 µL of chemically competent *E. coli* cells was thawed on ice and mixed with 1-5 µL of plasmid DNA. After 30 minutes of incubation on ice, a heat shock at 42 °C was performed. For DH5α and BL21 (DE3) cells, the duration of the heat shock was respectively 90 and 45 seconds. Cells were incubated again at 4°C for 5 minutes. Then, 400 µL of SOC medium, preheated to 37 °C, were added to the cells and these were incubated for 1 hour at 37°C with shaking. Subsequently, the transformed cells were plated out on a selection medium and cultured overnight at 37 °C.

3.2.1.3 Isolation of plasmid DNA from *E. coli*

Plasmid DNA was isolated using a commercially available kit (GeneJET Plasmid Miniprep Kit, Thermo Fisher Scientific) from overnight cultures of *E. coli* DH5α cells according to the manufacturer's instructions.

3.2.1.4 Hybridization of complementary DNA single strands

Double stranded RE-P DNA was prepared from two complementary oligonucleotides (see section 3.1.4) by means of hybridization. Equimolar amounts of each top and bottom strand (40 µM) were mixed in a batch volume of 25 µL in Phusion HF buffer. The hybridization mixture was heated using a PCR thermocycler at 98 °C for 5 min and gradually cooled down to 4 °C. The success of the hybridization step was checked via agarose gel electrophoresis and the resulting product used for further experiments.

3.2.1.5 Agarose gel electrophoresis

The DNA was separated on the basis of its molecular size by means of electrophoretic separation methods using agarose gel electrophoresis. 4.0 % (w/v) agarose gels were prepared for the hybridized p53 RE-DNA in TAE buffer according to the manufacturer's instructions. The DNA was mixed with DNA sample buffer and DNA dye and then pipetted into the gel's wells. The electrophoretic separation was performed at 200 V. Detection of DNA in agarose gel was carried out with the Quantity One (4.6.2) software (Gel Dock XR).

3.2.1.6 DNA sequencing

DNA constructs used or produced in this work were sent to Seqlab (Sequence Laboratories Göttingen GmbH) for sequencing and the results were then verified using the SnapGene software.

3.2.2 Cell biology methods

3.2.2.1 Expression of recombinant, full-length human p53 in *E. coli*

Expression of recombinant, full-length human p53 in *E. coli* was carried out according to an established protocol developed by Dr. Christian Arlt, MLU Halle-Wittenberg [7, 120]. The final expression plasmid (HLT_p53FL_pET28a, see chapter 3.1.3) was used to transform chemically competent *E. coli* BL21(DE3) cells. Kanamycin-LB agar selection medium was used to pick successfully transformed clones. From these clones, corresponding cultures were used to make cryo-stocks and used for protein expression. Specifically, 500 mL of 2 x YT medium supplemented with kanamycin (final concentration 30 µg/mL) and inoculated with a 5 mL preculture, was cultivated at

37 °C to an OD_{600} of 0.5-0.6. Then an acclimatization step was performed, by gradually cooling down the temperature of the expression culture around 18 °C at an OD_{600} of 0.9. Induction was carried out with 1 mM D-isopropylthiogalactose (IPTG) right after this step or at latest at an OD_{600} of 1.2. Parallel to the induction, $ZnCl_2$ was added to the expression mixture (final concentration 0.1 mM) to provide sufficient amount of Zn^{2+} ions for stabilizing the DNA binding domain (DBD) of the full-length p53 protein, ensuring successful protein expression and avoiding protein aggregation. Expression has been performed at 18 °C overnight.

3.2.2.2 Cell disruption by sonication

Cells were sedimented by centrifugation (4,000 x g, 30 minutes, 4 °C). The cell pellet was then resuspended in IMAC binding buffer (5 mL/1g of biowet mass) with the addition of protease cocktail inhibitor (Roche Complete, EDTA-free) at 4 °C. This was followed by cell disruption with ultrasound sonication at an amplitude of 35%. 6 disruption cycles of 30 seconds were carried out with pause intervals of 30 seconds in an ice bath, for a total duration of 6 minutes. 1 µL of benzonase was added to the mixture to cleave DNA/RNA, for 45 minutes on ice. Subsequently, an ultracentrifugation step was performed (35000 x g, 45 minutes, 4 °C). Cell debris and other insoluble components from cell disruption were separated from the soluble fraction. The decanted supernatant was used for the purification of recombinant full-length p53.

3.2.3. Protein biochemistry methods

3.2.3.1. Buffer exchange and concentration of protein solutions

Buffer exchange and concentration of protein solutions was performed at 4 °C. The determination of protein concentration was carried out by UV absorption measurements at 280 nm or using the Bio-Rad protein Bradford assay method according to manufacturer's instructions [225]. Centrifugation filters were used to change the buffer system from p53 solutions (Amicon Ultra 15 mL or 0.5 mL, 3.2.1.4) with a molecular weight cut-off (MWCO) of 30 kDa. First, the filters were conditioned twice with the maximum filling volume of the appropriate final buffer (5 minutes, 4000 x g). Subsequently, p53 solutions were transferred onto the centrifugation unit and concentrated at 4000 x g to about 1/10 of the maximum volume. This step was repeated five to six times to achieve a $1:10^5$-10^6 dilution of

the starting buffer, which represents an efficient buffer exchange. Concentration of protein solutions to the desired volume or final concentration was carried out by centrifugation at 4,000 x g.

3.2.3.2. Purification of full-length human wild-type p53

The purification of human full-length wild-type p53 was carried out using a four-step purification strategy consisting of an affinity chromatography step and a subsequent size exclusion chromatographic step, HLT tag cleavage using ProTEV protease and a further gel filtration step according to a procedure that had previously been established in the Sinz lab [7, 120]. The recombinantly expressed protein obtained from the plasmid HLT_p53FL_pET28a is a fusion protein consisting of an octahistidine (8xHis) tag followed by a lipoyl domain, a TEV cleavage site and the sequence of full-length wild-type p53.

Immobilized metal ion affinity chromatography (IMAC)

IMAC was performed as the first affinity chromatographic step. This is based on complexation or chemical coordination of the poly-histidine tag with the nickel ions of the IMAC column. Purification was performed using an ÄKTA FPLC system (GE Healthcare) using a nickel-NTA column (HisTrap FF, 1 mL). Column was equilibrated with IMAC binding buffer (10 column volumes) before the soluble fraction of the *E. coli* lysate was added at a flow rate of 1 mL/min. After sample application, a washing step using 10% (v/v) of IMAC elution buffer was applied to wash unspecifically bound proteins until the UV absorption signal at 280 min dropped down to baseline level. p53 was eluted with 40% (v/v) of IMAC elution buffer (corresponding to a final imidazole concentration of ~200 mM). Then, the column was washed with 100% (v/v) of the same buffer at a flow rate of 1 mL/min. The elution fractions (2 mL) were collected in a fraction collector by means of 96-well plates containing Protein LoBind tubes. The course of the purification was checked via SDS-polyacrylamide gel electrophoresis (SDS-PAGE) analysis (chapter 3.2.3.3).

Concentration of eluates

p53-containing fractions were pooled, filtered, and buffer-exchanged through centrifugal filtration units in the SEC buffer or directly loaded into a sample loop after washing of the loop with SEC

buffer prior to the first gel filtration step (2D mode using the versatile valve of the ÄKTA FPLC system Pure 25).

Size exclusion chromatography 1 (SEC)

Using SEC, proteins are separated from each other for the difference in their size or, more appropriately, their hydrodynamic radius. Thus, the HLT-p53 fusion protein can be separated from other protein contaminants. A SEC column with a separation range of 10,000-600,000 Da (HiLoad Superdex 16/600 200 pg) was used. An isocratic separation was performed with SEC buffer at a flow rate of 1.5 mL/min. Fractions of 2 mL were collected and analyzed by SDS-PAGE and peptide fragment fingerprint analysis was performed via LC-MS/MS. The protein was stored at -80 °C in aliquots of 500 µL at a concentration of ~10 µM in SEC buffer after shock-freezing with liquid nitrogen.

Proteolytic cleavage with ProTEV Plus protease

HLT-p53 aliquots were subjected to proteolytic cleavage of the tag with TEV protease (ProTEV Plus, Promega, 15 units) performed accordingly to the manufacturer's information. Proteolysis was carried out overnight at 4 °C for 16 hours and the success of the tag cleavage was verified by SDS-PAGE analysis.

Size exclusion chromatography 2 (SEC)

Cleaved HLT tag and TEV protease were separated using SEC using identical conditions as described above.

3.2.3.3. SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were separated with regard to their apparent molecular weights (more precisely, their hydrodynamic/Stokes radii) using SDS-PAGE. An example of the general composition of stacking and resolving gels is shown in Table 3.1. Volumes were adjusted accordingly for the preparation of 8% and 10% gels.

Table 3.1: Composition of resolving and stacking gels for SDS-PAGE analysis.

	Resolving Gel (12%)	Stacking Gel (5%)
Acrylamide/Bisacrylamide 40% (w/v)	3000 µL	650 µL
1.5 M Tris-HCl, pH 8.8	2500 µL	-
0.5 M Tris-HCl, pH 6.8	-	1250 µL
MilliQ-H_2O	4340 µL	3000 µL
10% (w/v) SDS	100 µL	50 µL
10% (w/v) APS	50 µL	25 µL
TEMED	10 µL	10 µL

Protein samples were mixed in a 1:1 (v/v) ratio with Laemmli sample buffer prior to electrophoresis and subjected to denaturation (95 °C, 10 minutes). Gel electrophoresis was carried out at 100-130 V for 10 min, followed by approximately 50-70 minutes at 200 V. Protein signals were subsequently analyzed and made visible by using colloidal Coomassie Brilliant Blue solution.

Samples were subjected to conventional Coomassie staining [226]:

Staining solution: 0.1% (w/v) Coomassie Brilliant Blue R250, 10% (v/v) acetic acid, 40% (v/v) ethanol, 50% (v/v) H_2O

Destaining solution: 10% (v/v) acetic acid, 25% (v/v) methanol, 65% (v/v) H_2O

Gels were destained by changing the destaining solution several times until the desired degree of decoloration was obtained.

3.2.3.4. Enzymatic in-solution digestion

The desired input amounts of protein solution were concentrated almost to dryness in a Speedvac concentrator (3.1.2.3.). Then, 25 µL of an 8M urea solution in 400 mM ammonium bicarbonate (ABC) were added. This was followed by reduction with 5 µL of a 45 mM Dithiotreitol (DTT) solution for 30 minutes at 56 °C. The reduced samples were alkylated using 10 µL of a 100 mM Iodoacetamide (IAA) solution for 30 minutes at room temperature in the dark. Excess of IAA was quenched for 15 minutes with 10 µL of a 45 mM DTT solution. Before addition of the protease, the

reaction mixture was diluted with water up to a 0.8 M final urea concentration. In the case of single proteolytic cleavage, trypsin was used as endoprotease and added to an enzyme:protein ratio of 1:20 (w/w). Samples were incubated at 37 °C for 16 hours. For a double proteolytic cleavage, the enzyme AspN was first added (enzyme:protein ratio 1:20) and incubated at 37 °C for 16 hours. Then, trypsin was added (enzyme:protein ratio 1:20) and the samples were incubated for 4 hours at 37 °C. The reaction was stopped by adding 10% (v/v) TFA and the proteolytic mixture's volume was reduced in a vacuum concentrator (40-50 µL) prior to mass spectrometric analysis.

3.2.3.5. Enzymatic in-gel digestion

The following solutions were freshly prepared for in-gel proteolysis:

Reduction solution:	10 mM DTT in 100 mM NH_4HCO_3
Alkylation solution:	55 mM IAA in 100 mM NH_4HCO_3
Destaining solution:	ACN:100 mM ABC, 1:1 (v/v)
Enzyme solution:	2 µL of AspN and/or trypsin stock solution, 78 µL of 25 mM NH_4HCO_3
Extraction solution:	ACN:5% (v/v) TFA, 2:1 (v/v)

Protein bands were excised from the SDS-PAGE gel and diced with a volume of approximately 1 mm^3. These cubes were transferred in 1.5 mL Protein LoBind tubes and dehydrated with 500 µL of ACN. After the supernatant was discarded, 50 µL of reduction solution were added and if needed, 100 mM ABC solution was used to cover the gel pieces completely. Samples were incubated for 30 minutes at 56 °C under constant shaking, then 500 µL of ACN were added again. The supernatant was discarded and the dried gel pieces were mixed with 50 µL of alkylation solution and incubated for 30 minutes at room temperature in the dark. Then, the gel pieces were washed for 10 minutes with 500 µL of ACN, washed for 10-30 minutes with 100 µL of destaining solution and dried again in ACN for 10 minutes. This process was repeated until efficient destaining of the gel pieces was achieved. After destaining, the proteolysis step was performed. In the case of single proteolytic cleavage, 20 µL of trypsin solution (enzyme:protein ratio 1:20) were added and eventually 100 mM ABC solution was used to completely cover the gel cubes. Samples were kept for 1 hour at 4 °C, then the proteolysis step was carried out at 37 °C for 16 hours. For a double proteolytic cleavage, the samples were primarily proteolytically digested with the enzyme AspN (enzyme:protein ratio 1:20)

for 16 hours as for trypsin cleavage. Then, trypsin solution (enzyme:protein ratio 1:20) was added to the samples and incubated for 4 hours at 37 °C. Afterwards, two extraction steps with 80 µL of extraction solution each were performed, which also stopped the proteolysis step. The final extracted peptide mixture was concentrated and reconstituted with a volume of 40-50 µL using a vacuum concentrator prior to MS analysis.

3.2.4. Cross-linking

3.2.4.1. Quantitative p53 cross-linking

Cross-linking reactions were carried out in 50 mM HEPES buffer, pH 7.4. The concentration of DMSO was not exceeding 2% (v/v) to not affect the protein structure. For the reactions in the presence of RE-DNA, hybridized double-stranded DNA was added to a final concentration of 2.5 µM (corresponding to a 1:1 ratio tetrameric p53:DNA). The outcome of the reaction was checked by SDS-PAGE. Solutions containing 10 µM wild-type p53 in 50 mM HEPES, 300 mM NaCl, and 2.5 mM TCEP, 10% (v/v) glycerol, pH 7.2, were incubated overnight with or without RE-DNA and cross-linked at 4 °C for 1 h with 0.5 mM $BS^2G\text{-}D_0$ and $BS^2G\text{-}D_4$, respectively. Cross-linkers were dissolved in neat dimethyl sulfoxide (DMSO) immediately before adding it to the protein solutions. The cross-linking reactions were performed in triplicates for each condition and repeated with reversed deuterium labeling. The reactions were quenched by adding ammonium bicarbonate to a final concentration of 20 mM. For XL-MS, samples with and without RE were mixed (1:1) and separated by SDS-PAGE (10% resolving/5% stacking gel). The monomeric p53 band was excised from the gel and in-gel digested with AspN (37 °C, overnight) and trypsin (37 °C, 4 hours). Peptides were recovered using an extraction buffer (5% (v/v) TFA/acetonitrile, 1:2 (v/v)) and the solution was subjected to LC-MS/MS analysis.

3.2.4.2. Peptide cross-linking

For the cross-linking experiments with synthetic model peptides, Angiotensin II (ATII) and Test peptide 1 (TP1) were used. For ATII, the peptide was dissolved in water at a concentration of 5 mM. To 4 µL of ATII (final concentration 2 mM), 2.8 µL of 50 mM PAC4 (final concentration 14 mM) and 3.2 µL of 50 mM HEPES buffer, pH 8.5 were added. The cross-linking reaction was performed

for 1 hour at room temperature. The reaction was quenched using ABC to a final concentration of 20 mM, for 15 minutes.

For TP1, an aliquot (2.8 µg, 1 nmol) was reconstituted in 20 µL of 50 mM HEPES, pH 8.5 (final concentration 50 µM). To the peptide solution, 1 µL of 20 mM of PAC4/DSSI cross-linker was added for 1 hour at room temperature (final concentration 1 mM). The reaction was quenched by adding ABC to a final concentration of 20 mM, for 15 minutes. Cross-linked peptides were desalted using C18 Zip Tip columns (Millipore) according to manufacturer's instructions prior to MS analysis.

3.2.4.3. Protein cross-linking

All cross-linking reaction were performed using 50 mM HEPES, pH 7.4. For the PAC4 cross-linking experiments with model proteins, Bovine serum albumin (BSA, 66.4 kDa) and Carbonic anhydrase from *bos taurus* (CA2, 29 kDa) were used. BSA solutions (5 µM, ~0.33 µg/µL) were cross-linked with increasing concentrations of PAC4 (10-, 25-, 50-, 100-, 200- fold excess) for 15, 30 and 60 minutes at room temperature. CA2 solutions (5 µM, ~0.15 µg/µL) were cross-linked with increasing concentrations of PAC4 (10-, 25-, 50-, 100- fold excess) for 30 minutes at room temperature. For the DSSI cross-linking experiments with model proteins, α-synuclein (α-syn, 14.5 kDa) was used. A solution of α-syn (10 µM, ~0.14 µg/µL), was cross-linked with DSSI (1 mM final concentration). The reactions were quenched using ABC to a final concentration of 20 mM for 15 minutes. For PAC4, the cross-linked proteins (after 60 minutes reaction) were checked for the number of modifications using Native MS.

3.2.4.4. Cross-linking of HEK293T lysates

HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS). After washing with phosphate buffered saline (PBS) three times, cell pellets were resuspended in the cross-linking buffer (20 mM HEPES, 10 mM $MgCl_2$, 150 mM KCl, 1%v/v DDM, pH 7.4). The cell debris was removed by centrifugation at 13,000 rpm for 15 min at 4°C. The supernatant was collected and adjusted to a final concentration of 1 mg/mL before carrying out the cross-linking reaction. DSSI/DSBU was added for 1h at room temperature (2 mM final concentration). The cross-linker was dissolved in neat DMSO immediately before adding it to the cell lysate. The reaction was quenched by adding ABC to a final concentration of 20 mM.

3.2.4.5. Cross-linked peptides enrichment

3.2.4.4.1 Cu-catalyzed azide-alkyne cycloaddition (CuAAC) reaction/enrichment with TP1

TP1 was cross-linked with PAC4 as previously described (chapter 3.2.4.2.). 10 µL of photo-cleavable azide agarose beads (PCAB) or disulfide azide agarose beads (DSAB) were mixed with the TP1 samples, THPTA (25:1), copper sulfate (5:1) and sodium ascorbate (50:1), where the ratios are given in regard to PAC4 input quantity. The reaction mixture was incubated for 2 hours at room temperature, with shaking. Supernatant was removed and collected in separate tubes. PCAB/DSAB beads were washed four times with HEPES and twice with 0.1% formic acid to remove linear peptides unspecifically bound to the beads. Cross-linked peptides were eluted by using 10 seconds UV irradiation via a LED photo-cross-linking device (PCAB beads) or 45 mM DTT in H_2O at 56 °C for 30 minutes (DSAB beads).

3.2.4.4.2 CuAAC enrichment of cross-linked peptides

BSA was cross-linked with PAC4 as previously described (chapter 3.2.4.3.). Samples were buffer-exchanged with the same cross-linking buffer prior to CuAAC enrichment to remove excess of unreacted cross-linker. 10 µL DSAB were mixed with the BSA samples, THPTA (25:1), $Cu(CH_3CN)_4BF_4$ (12.5:1) and sodium ascorbate (50:1), where the ratios are given in regard to PAC4 input quantity. Freshly prepared THPTA, $Cu(CH_3CN)_4BF_4$ and sodium ascorbate were preincubated for 10 minutes and added to the BSA cross-linked samples. The reaction mixture was incubated for 2 hours at 40 °C, with vigorous shaking (1400-1600 rpm). Supernatant was removed and collected in separate tubes. DSAB beads were washed four times with HEPES and twice with 0.1% formic acid to remove linear peptides unspecifically bound to the beads. Cross-linked peptides were eluted by using 45 mM TCEP in H_2O at 56 °C for 30 minutes.

3.2.4.4.3 Size exclusion chromatography of cross-linked HEK293T samples

All three replicates of digested HEK293T lysates were subjected to size exclusion chromatography (SEC) on an ÄKTA FPLC system (equipped with UPC-900, P-920 pump, and Frac-950 fraction collector, Amersham Biosciences) for pre-fractionation before LC-MS/MS analysis. Peptides were separated on a Superdex peptide 3.2/300 column (GE Healthcare) using a 2% (v/v) ACN, 0.1% (v/v)

TFA solution at flow rate of 50 µl/min. 2 min fractions were collected over the whole run; fractions A12 to B9, corresponding to elution volumes from 1.1 to 2.1 ml, were selected for subsequent LC-MS/MS analyses.

3.2.5. Mass spectrometry

3.2.5.1. Native MS

p53 samples were also analyzed via native MS with a high-mass Q-TOF II mass spectrometer (Micromass/MS Vision). Protein solutions were adjusted to a final concentration of 10 µM. For the functionality assay, RE-DNA was added to a final concentration of 2.5 µM and incubated for 16 hours at 4 °C. The buffer was manually exchanged in ammonium acetate (500 mM, pH 6.8) using centrifugal filtration units. BSA and CA2 (final concentration 5 µM) cross-linked with PAC4 were analyzed using a home-built online-buffer exchange system to check for the number of cross-linker modifications per molecule. This system is composed of a SEC column (3,000-9,000 Da MW range) used for desalting the sample which was collected into a sample loop (the fraction of the void volume containing high molecular weight proteins being desalted) and injected into the mass spectrometer [227]. Samples were buffer exchanged in ammonium acetate (250 mM, pH 6.8) and analyzed. Samples were ionized by nano-ESI using in-house, gold-coated borosilicate glass capillaries for manually buffer-exchanged samples or commercially available nano-ESI emitters connected to a self-packed SEC column (3,000-6,000 MW range) for online buffer-exchanged samples. Sample volumes used for all the native MS experiments were 2.5-5 µL. For manually buffer-exchanged samples, capillary voltage was set to 1.2-1.4 kV and the cone voltage to 90-120 V. For online buffer-exchanged samples, capillary voltage was set to 2.2-2.4 kV. The pressure in the source area of the MS (p1) was set to 10 mbar and the pressure in the collision cell (p3) was set to $1x10^{-2}$ mbar. Cesium iodide (CsI) measurements were carried out for the recalibration of the data and proper verification of MS signals.

3.2.5.2. HDX-MS

H/D exchange experiments with p53 were performed at room temperature. 1 µL of 10 µM p53 in 50 mM HEPES, 300 mM NaCl, 2.5 mM TCEP, 10 % (v/v) glycerol, pH 7.2 was diluted with 9 µL of 10 % (v/v) glycerol in 90 % (v/v) D_2O for 30 seconds, 1 or 2 minutes. Samples were injected to a home-built HDX apparatus, consisting of an HP1200 (Agilent, Santa Clara, USA) HPLC system,

cooled to 4 °C in a DB951 showcase (Polar Refrigeration, Bristol, UK), and a column setup comprising an Enzymate protein pepsin column, XBridge BEH C18 with VanGuard cartridge 5 mm × 2.1 mm, 2.5 µm, 300 Å and an XBridge Peptide BEH C18 column 100 mm × 1 mm, 3.5 µm, 300 Å (Waters, Milford, USA). Digestion was conducted for 10 min at 25 µL/min 0.1% (v/v) formic acid in water; separation was performed within 6 min at a flow rate of 80 µL/min using a gradient of 3–50 % (v/v) acetonitrile in water with 0.1 % (v/v) formic acid. Peptic peptides were analyzed with a timsTOF Pro mass spectrometer (Bruker Daltonik, Bremen, Germany) with ESI source at a temperature of 50 °C, nebulizer gas pressure of 3 bar, dry gas flow of 8 L/min, and capillary voltage of 3,500 V. Mass spectra were recorded in the *m/z* range 100–1700 using the standard proteomics settings of the timsTOF Pro instrument in the COMPASS software (Bruker Daltonik, Bremen, Germany). The HDExaminer 3.3.0 software (Sierra Analytics, Modesto, USA) was used to extract the deuterium content of peptic peptides. Data were corrected for back-exchange using data from six independent 10 µL injections of p53 stored in buffer containing 80 % (v/v) D_2O, 10 % (v/v) glycerol, and 10 % (v/v) H_2O.

3.2.5.3. Direct infusion orbitrap mass spectrometry

Peptide samples were desalted with C18 ZipTip columns (Millipore, 3050 Spruce St. Louis Missouri 63304. USA). Samples were measured immediately after sample preparation. Offline Nano-ESI-MS/MS analyses were performed on an orbitrap fusion tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA) with nano-ESI source (Nanospray Flex Ion Source; Thermo Fisher Scientific). Samples were loaded into gold-coated capillaries (4', 1.2 mm/0.68 mm o.d./i.d., World Precision Instruments, Sarasota, FL) that were prepared in-house (Model P1000 Flaming/Brown Micropipette Puller; Sutter Instruments, Novato, CA). The voltage was set to 1.3-1.6 kV in positive ionization mode and the source temperature was held at 275 °C. MS data were collected in the *m/z* range 300–1700 (R = 120,000 at *m/z* 200). For MS/MS measurements, ions were isolated in the quadrupole (isolation window 2 Da), fragmented by CID and HCD (NCE: 0 to 40 %, steps of 5 %) and fragment ions were analyzed in the orbitrap mass analyzer (R = 30,000 at *m/z* 200). External calibration was carried out prior to the MS measurements.

3.2.5.4. Nano-HPLC/nano-ESI orbitrap mass spectrometry

Peptide mixtures were analyzed by LC-MS/MS on an UltiMate 3000 RSLC nano-HPLC system (Thermo Fisher Scientific) coupled to an orbitrap fusion tribrid mass spectrometer equipped with an EASY Spray ion source (Thermo Fisher Scientific). Peptides were trapped on a C18 precolumn (Acclaim PepMap 100, 300 µm × 5 mm, 5 µm, 100 Å, Thermo Fisher Scientific) and separated on a µPAC C18, 50 cm column (Thermo Fisher Scientific). After trapping, peptides were eluted by a concave 90-min water-acetonitrile gradient from 3 to 30 % B (solvent B: acetonitrile, 0.1% [v/v] formic acid) at a flow rate of 300 nL/min. Data were acquired in data-dependent MS/MS mode using stepped higher-energy collision-induced dissociation (HCD) at normalized collision energies of 27 %, 30 %, and 33 %. High-resolution full scans (*m/z* 300–1700, R = 120,000 at *m/z* 200) were followed by high-resolution product ion scans (R = 15,000) for 5 s, starting with the most intense signal in the full-scan mass spectrum. Precursor ions with charge states > 2 + and < 8 + were selected for fragmentation; the isolation window was set to 2 Th. Dynamic exclusion of 60 s (window 2 ppm) was enabled to allow the detection of less abundant ions. Data acquisition was controlled via XCalibur 4.3 (Thermo Fisher Scientific). TP1 cross-linked products were analyzed by LC-MS/MS with the same configuration, using a concave 30-min water-acetonitrile gradient from 3 to 30 % B at a flow rate of 300 nL/min. An inclusion list of target precursors was provided to the instrument. If detected, these precursors were subjected to sequential CID- and HCD-MS/MS with energy ramps increasing from 10 to 40% NCE in increments of 5% NCE.

3.2.5.5. Nano-HPLC/nano-ESI TIMS-Q-TOF mass spectrometry

Digested peptide mixtures of BSA and α-syn were analyzed in three technical replicates, HEK293T lysate samples in three biological replicates by LC-MS/MS on an UltiMate 3000 RSLC nano-HPLC system (Thermo Fisher Scientific) that was coupled to a timsTOF Pro mass spectrometer [228-230] equipped with CaptiveSpray source (Bruker Daltonik). Peptides were trapped on a C18 precolumn (Acclaim PepMap 100, 300 µm × 5 mm, 5 µm, 100 Å) (Thermo Fisher Scientific) and separated on a self-packed Picofrit (New Objective) nanospray emitter (360 µm id × 75 µm id x 150 mm length, 15 µm Tip id) with C18 stationary phase (3.0 µm, 120 Å, Dr. Maisch GmbH). After trapping, peptides were eluted by a linear 90-min (for BSA and α-syn) or 180-min (for HEK293T lysates) water-acetonitrile (ACN) gradient from 3 % (v/v) to 35 % (v/v) ACN. The column was washed at a flow

rate of 300 nL/min with the following gradient: 35% (v/v) to 85% (v/v) ACN (5 minutes), 85% (v/v) ACN (5 minutes), 85% (v/v) to 3% (v/v) ACN (5 min). All steps were performed at RT. For the timsTOF Pro MS settings, the following parameters were adapted, starting from the PASEF method for standard proteomics. Mobility-dependent collision energy ramping was set to 95 eV at an inversed reduced mobility ($1/k_0$) of 1.6 V*s/cm^2 and 23 eV at 0.73 V*s/cm^2. Collision energies were linearly interpolated between these two $1/k_0$ values and kept constant above or below. No merging of TIMS scans was performed. Target intensity per individual PASEF precursor was set to 20,000. The scan range was set between 0.6 and 1.6 V*s/cm^2 with a ramp time of 166 ms. 14 PASEF MS/MS scans were triggered per cycle (2.57 s) with a maximum of seven precursors per mobilogram. Precursor ions in an *m/z* range between 100 and 1,700 with charge states ≥ 3+ and ≤ 8+ were selected for fragmentation. Active exclusion was enabled for 0.4 min (mass width 0.015 Th, $1/k_0$ width 0.015 V*s/cm^2).

3.2.6. Identification of cross-linked products

3.2.6.1. BS^{2}G MeroX cross-linking analysis

For the analysis of BS^{2}G cross-linking experiments, the following settings were applied: proteolytic cleavage C-terminally at Lys and Arg (up to 3 missed cleavages were allowed) and N-terminally at Asp and Glu (up to 3 missed cleavages were allowed); peptide length: 5–30 amino acids; modifications: alkylation of Cys by iodoacetamide (fixed), oxidation of Met (variable); cross-linker specificity: Lys, Ser, Thr, Tyr, N-terminus; search algorithm: quadratic mode; precursor mass accuracy: 10 ppm; fragment ion mass accuracy: 20 ppm; signal-to-noise ratio > 1.5; precursor mass correction enabled; false discovery rate (FDR) cut-off: 1%, and minimum score cut-off: 20.

3.2.6.2. PAC4 MeroX cross-linking analysis

For the analysis of PAC4 cross-linking experiments, the following settings were applied: proteolytic cleavage C-terminally at Lys and Arg (up to 3 missed cleavages were allowed); peptide length: 5–30 amino acids; modifications: alkylation of Cys by iodoacetamide (fixed), oxidation of Met (variable); cross-linker specificity: Lys, Ser, Thr, Tyr, N-terminus; search algorithm: RISEUP, max. number of missing ions 3; precursor mass accuracy: 10 ppm (orbitrap), 15 ppm (timsTOF); fragment ion mass

accuracy: 20 ppm (orbitrap), 25 ppm (timsTOF); signal-to-noise ratio > 1.5; precursor mass correction enabled; false discovery rate (FDR) cut-off: 1 %, and minimum score cut-off: 20.

3.2.6.3. DSSI/DSBU MeroX cross-linking analysis

For the analysis of DSSI cross-linking experiments, the following settings were applied: proteolytic cleavage C-terminally at Lys and Arg, up to 5 missed cleavages (α-syn) or 3 missed cleavages (HEK293T samples) were allowed; peptide length: 5–30 amino acids; modifications: alkylation of Cys by iodoacetamide (fixed, only for HEK293T lysates), oxidation of Met (variable); cross-linker specificity: Lys (α-syn) or Lys, Ser, N-terminus (HEK293T lysates); search algorithm: RISEUP mode (α-syn) with 3 maximum missing ions or Proteome-Wide mode (HEK293T samples) with a minimum peptide score of 10; precursor mass accuracy: 15 ppm; fragment ion mass accuracy: 25 ppm; signal-to-noise ratio > 1.5; precursor mass correction enabled (α-syn) and disabled (HEK293T lysates); false discovery rate (FDR) cut-off: 1%, and minimum score cut-off: 20. For α-syn samples, the fasta sequence of the protein was used for the MeroX search. For HEK293T lysates samples, 1 µg of DSSI and DSBU cross-linked samples were injected independently into the timsTOF and acquired with a 2+ to 8+ DDA-PASEF method. The identified proteins in Proteome Discoverer were used to build a cross-linker specific fasta database for carrying out the MeroX searches for all the proteome-wide XL-MS experiments which were performed.

3.2.7. Quantification of BS^2G cross-linked products

Cross-links were manually quantified with XCalibur 4.0. Extracted ion chromatograms (XICs) of the precursor masses for BS^2G-D_0 and BS^2G-D_4 cross-linked peptides were extracted and BS^2G-D_4/D_0 ratios were calculated using the relative intensities of the precursor ions' isotope distributions. In case of identifying a cross-linked peptide at different charge states or in case it contained an oxidized methionine, ratios were calculated for all possible peptide states. Identification of cross-links and 'dead-end' cross-links was performed with MeroX 2.0.1.7 [213]. 'Dead-end' cross-links were analyzed to extract quantitative footprinting data using Mass Spec Studio 2.0 (https://www.msstudio.ca) [231]. The following settings were applied: protein states, plusDNA, minusDNA; reagents, $BS^2G\_OH\_D_0$ (composition $C_5H_6O_3$, monoisotopic mass 114.0317), $BS^2G\_OH\_D_4$ (composition $C_5D_4H_2O_3$, monoisotopic mass 118.0568); amino acid modification,

methionine oxidation (variable), cysteine carbamidomethylation (fixed); acquisition mode, DDA; fragmentation, HCD; precursor tolerance, 10 ppm; number of tolerable termini, 2; peptide charge, 2–6; peptide length, 5–40; max. modifications per peptide, 3; enzyme, trypsin; XIC *m/z* selector tolerance, 6 ppm; XIC peak to background ratio, 2; fragment mass selector, 20 ppm. Absolute residue labeling yields were normalized to the unbound state.

3.2.8. Data visualization

All plots and data visualization have been performed using Python or R scripts if not stated otherwise.

4 Results

4.1. Plasmid construct design for recombinant expression of p53

In order to recombinantly express human p53 from *E. coli*, an expression vector was employed that had been designed by Dr. Christian Arlt in the Sinz lab [7, 120, 232] to have an N-terminal HLT (His-Lipoyl-TEV) tag (Figure 4.1). This tag is featured by an octa-histidine chain (8xHis), followed by the first lipoyl domain of the dihydrolipoyl transacetylase E2p of the pyruvate dehydrogenase complex of *Bacillus stearothermophilus* and a TEV cleavage site [ENLYFQ(G/S)]. The lipoyl domain contributes to enhance protein solubility, while the TEV site is important to get the tag-less native p53 sequence with an additional serine at the N-terminus after cleavage.

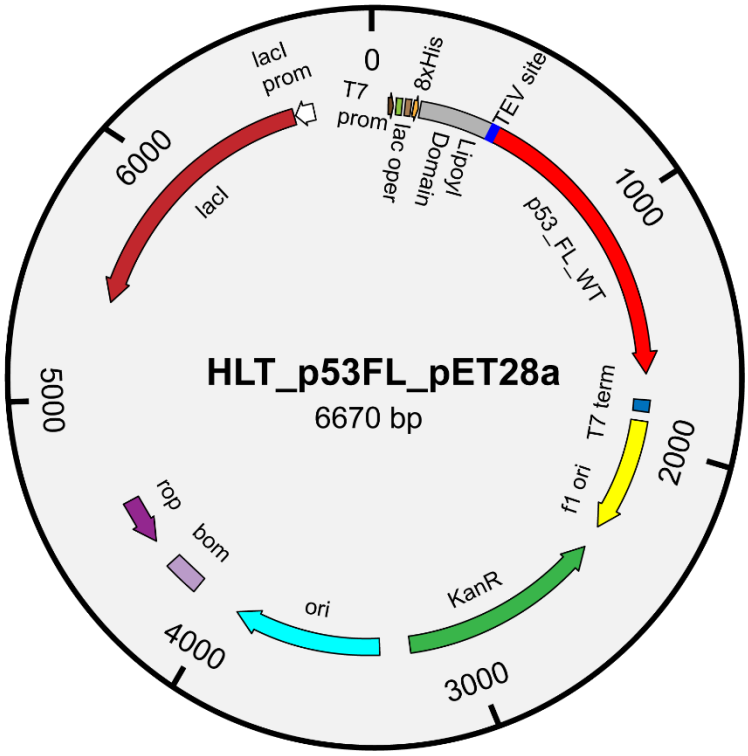


Figure 4.1: Plasmid map for the recombinant expression of full-length wild-type human p53.

The plasmid was used to transform chemically competent *E. coli* DH5α and BL21(DE3) cells (chapter 3.1.7 and 3.2.1.2). Colonies from the selection medium were picked and their DNA were sent for sequencing (chapter 3.2.1.6). After confirmation of the desired sequence, chemocompetent *E. coli* DH5α and BL21(DE3) transformed successfully were cryo-cultured and stored at -80 °C. DH5α cells were used for storage and if needed further DNA extraction, BL21(DE3) cells were used for the purification of recombinant p53.

4.1.1 Expression and purification of full-length wild-type human p53

Expression of the HLT-p53 was performed as described in chapter 3.2.2.1. As medium for the expression, 500 mL 2 x YT were prepared (chapter 3.18). For the production of HLT-p53, 6-10 g of wet biomass could be harvested. Expression tests were performed prior to the purification of the bigger batch, confirming the presence of the tagged target protein in the soluble fraction which could be produced in sufficient amount for the next experiments. The workflow for expressing and purifying full-length HLT-p53 had already been established by Dr. Christian Arlt [7, 120, 232] in the Sinz lab. The purification protocol comprised of IMAC (Figure 4.2) followed by a SEC (Figure 4.3) that were carried out at 4 °C due of the thermal instability of the protein (chapter 3.2.3.2). The total lysate (1:20 dilution), the soluble fraction (1:20 dilution), the flow-through (FT) and the elution fractions from the IMAC were analyzed via SDS-PAGE (Figure 4.2 B). As it can be seen from the SDS gel, the first IMAC step enabled to get high protein yield and overall good purity before the second chromatographic step. The HLT-p53 protein signal has an apparent molecular weight of approximately 70 kDa, which is higher than the one expected for the fusion protein (~56.2 kDa). Also the untagged p53 protein appears with an apparent molecular weight of 53 kDa which is higher than the actual protein molecular weight (43.7 kDa). This can be explained due to the fact that a large fraction of the protein sequence is intrinsically disordered, so that its hydrodynamic radius and actual electrophoretic mobility are different from a standard globular protein, causing a shift to a higher molecular weight. The flow-through showed that a large number of *E. coli* proteins did not bind to the nickel column during IMAC (Figure 4.2B). Also the fusion protein HLT-p53 did not completely bind to the nickel column. IMAC elution fractions between 109 and 112 mL were combined for a second SEC step (Figure 4.3; see also chapter 3.2.3.2). Two elution signals (E_1 and E_2) were detected at elution volumes of 49.6 and 55.8 mL (Figure 4.3) and analyzed by SDS-PAGE (Figure 4.3 B).

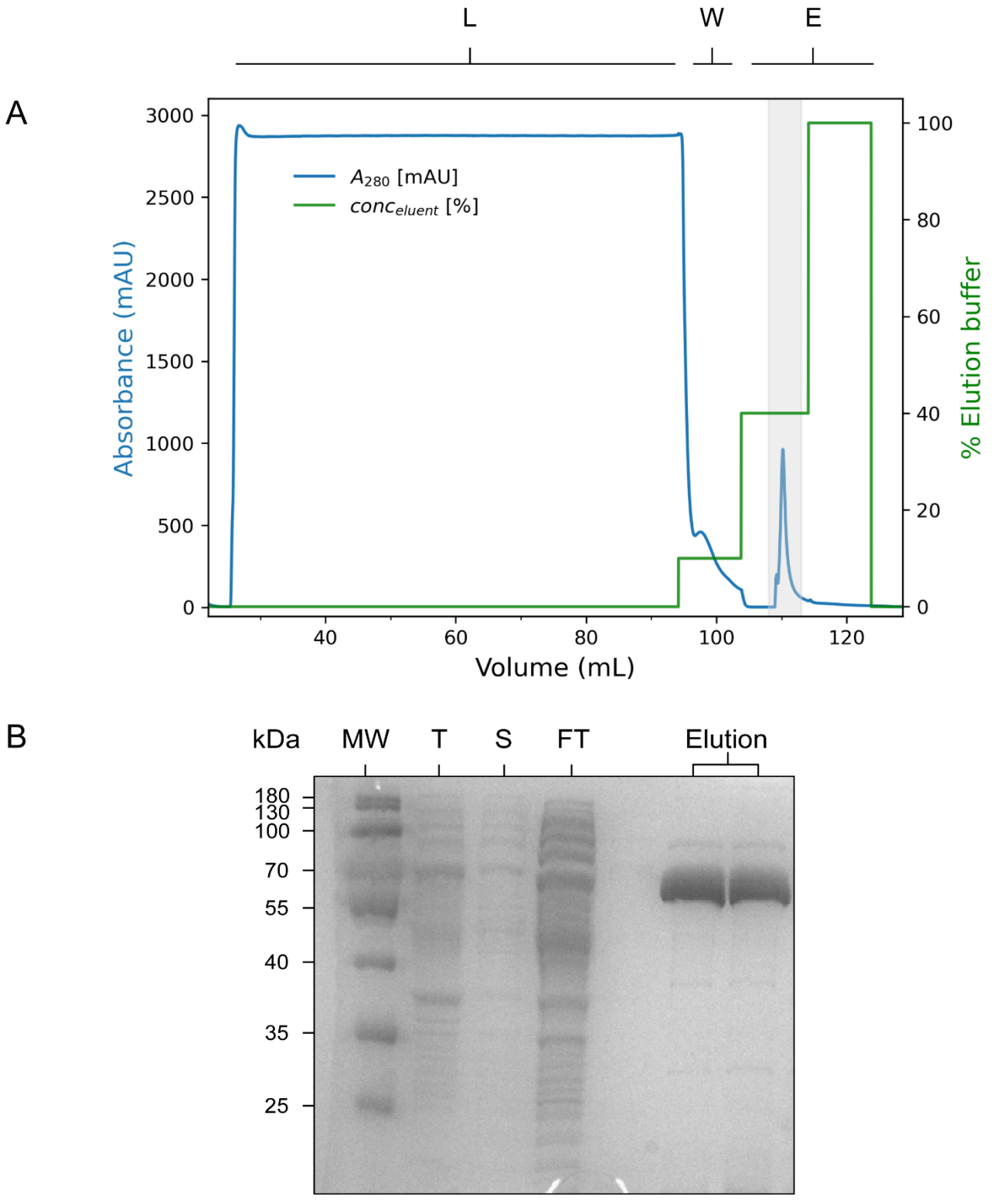


Figure 4.2: IMAC purification of HLT-p53.

A) Sample loading (L), column washing (W) and elution steps (E) are indicated above. The elution fractions analyzed via SDS-PAGE are highlighted in gray. UV absorbance at 280 nm (blue) and the percentage of elution buffer (chapter 3.2.3.2) (green) are plotted. B) SDS-PAGE of the IMAC purification step. MW - molecular weight ladder, T - total lysate, S - soluble lysate, FT - flow-through and elution fractions are indicated on the gel.

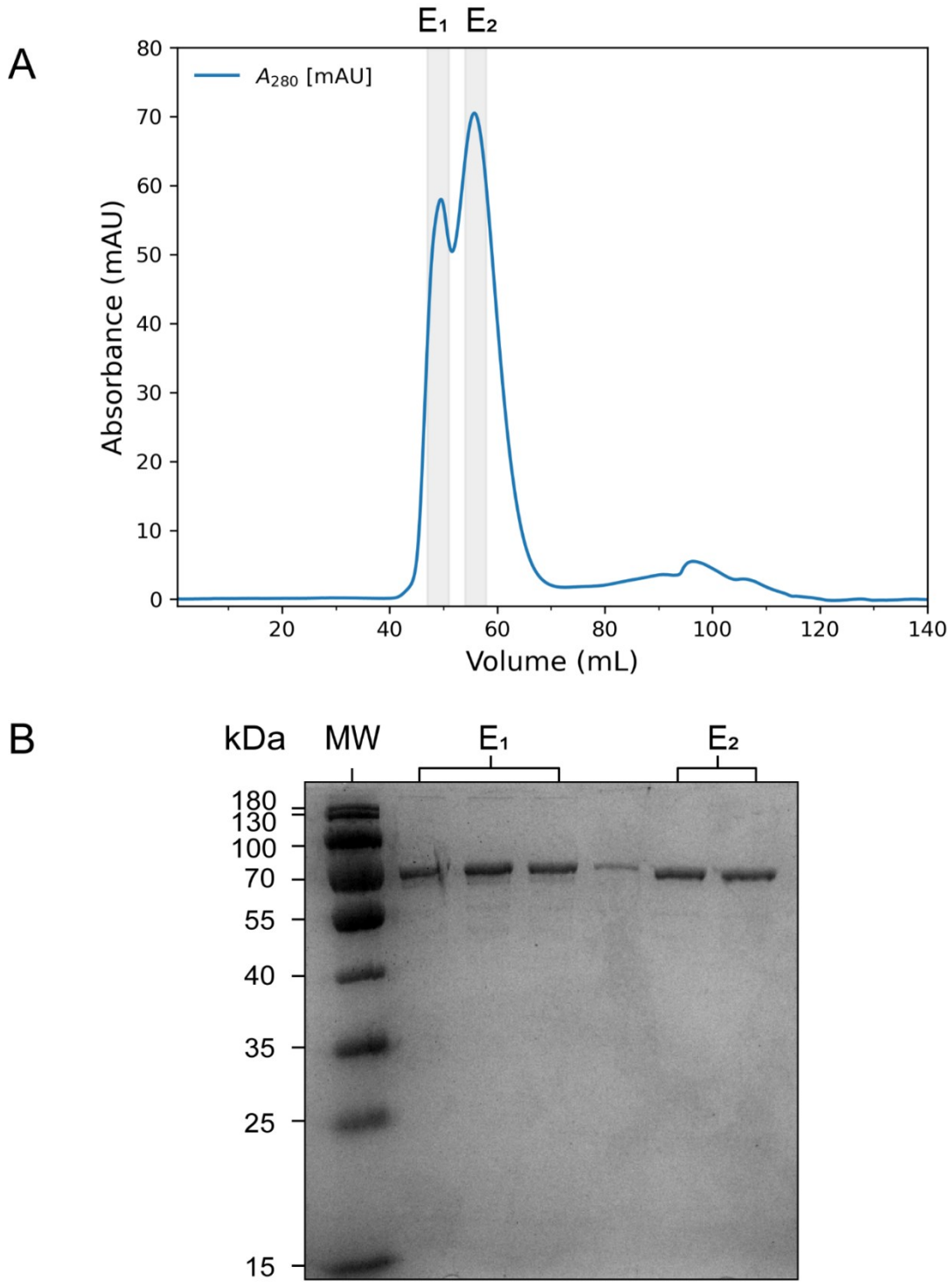


Figure 4.3: SEC purification of HLT-p53.

A) SEC chromatogram for the purification of HLT-p53. The elution fractions analyzed via SDS-PAGE are highlighted in gray. UV absorbance at 280 nm wavelength is plotted. B) SDS-PAGE analysis of the SEC purification step. MW - molecular weight ladder, E_1 and E_2 elution fractions are indicated on the gel.

SEC elution fractions exhibited one protein band with an apparent molecular weight of ~70 kDa, corresponding to HLT-p53. The elution time of the HLT-p53 fusion protein did not correlate with the expected molecular weight. The species in E_1 eluted very close to the void volume (~500-550 kDa), while E_2 eluted later (corresponding to a molecular weight of ~300-400 kDa). As the *in vitro* K_D value for the p53 tetramer-dimer equilibrium is ~55 nM, at a p53 concentration of 10-18 µM that is usually obtained from the first IMAC step (Figure A1), prevalently tetramer should be present and only a single elution peak is expected.

The presence of an additional peak very close to the void volume suggests the presence of higher order oligomeric states of p53. The Stokes radius of the protein during SEC seems to be larger than expected for tetrameric p53, due to the large IDR content. Fractions corresponding to the second elution peak were pooled, concentrated and used for HLT cleavage using the ProTEV protease (3.2.3.2). To separate the HLT tag, remaining impurities, and the TEV protease from p53, a second SEC step was performed using identical conditions (Figure 4.4 A, 3.2.3.2). The peak present in the previous SEC run centered around 50 mL, corresponding to the higher-order oligomers in the preparation, is now almost absent and two additional signals could be detected. A SEC elution peak at ~58.7 mL presumably corresponds to the tag-free tetrameric p53. The peak at ~93 mL probably corresponds to the HLT tag, appearing at a molecular weight of ~15 kDa in the gel (Figure 4.4 B). The E_2 fractions were combined. A fraction of the sample (5 µg) was used for protein footprinting by means of proteolytic cleavage followed by nano-HPLC/nano-ESI-MS/MS analysis to confirm the identity of p53. The outcome of TEV cleavage was monitored by SDS-PAGE (Figure 4.4 B, left hand side). Under the selected conditions, a nearly complete cleavage of the HLT tag cleavage was observed. One cleavage product migrated at an apparent molecular weight of ~53 kDa (presumably, corresponding to tag-less p53), while the smaller cleavage product migrated at an apparent molecular weight of ~15 kDa. The TEV protease used for the cleavage reaction has a molecular weight of 48 kDa and is not visible in the gel below the p53 signal.

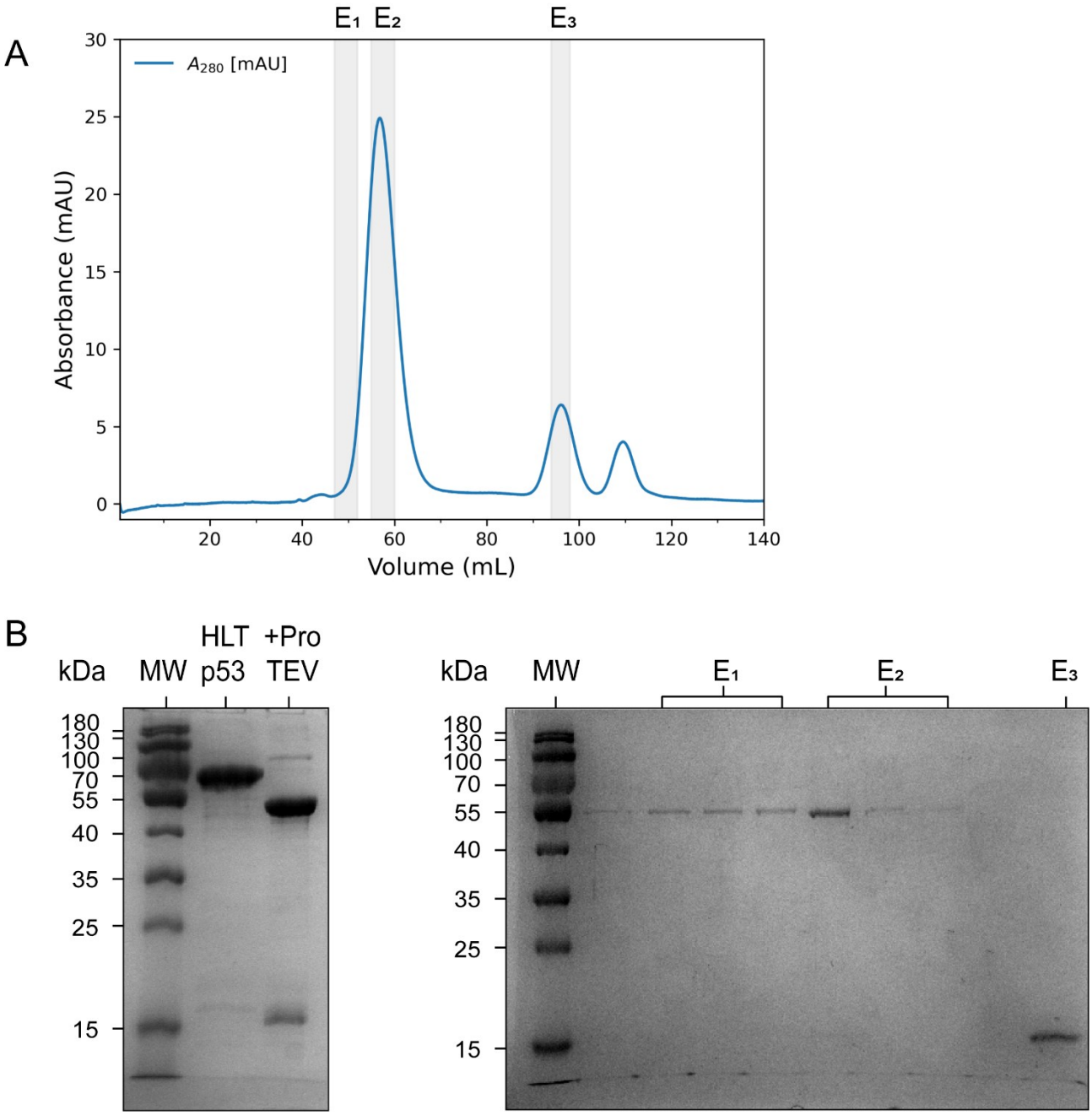


Figure 4.4: SEC after cleavage of the HLT tag.

A) SEC after cleavage of the HLT tag by the ProTEV Plus protease. UV absorbance at 280 nm is plotted. The elution fractions analyzed via SDS-PAGE are highlighted in gray. B) SDS-PAGE analysis of the SEC purification step. MW - molecular weight ladder, E_1, E_2 and E_3 elution fractions are indicated on the gel.

4.1.2 Functional characterization of purified p53 by MS

After purification, p53 was characterized regarding its DNA binding and tetramerization propensity. ESI-MS analysis at denaturing conditions revealed the presence of a wide p53 protein distribution (charge states 28+ to 57+) centered around the 46+ charge state (Figure 4.5). This is in accordance with the number of basic amino acids (47 arginine and lysine residues) in p53.

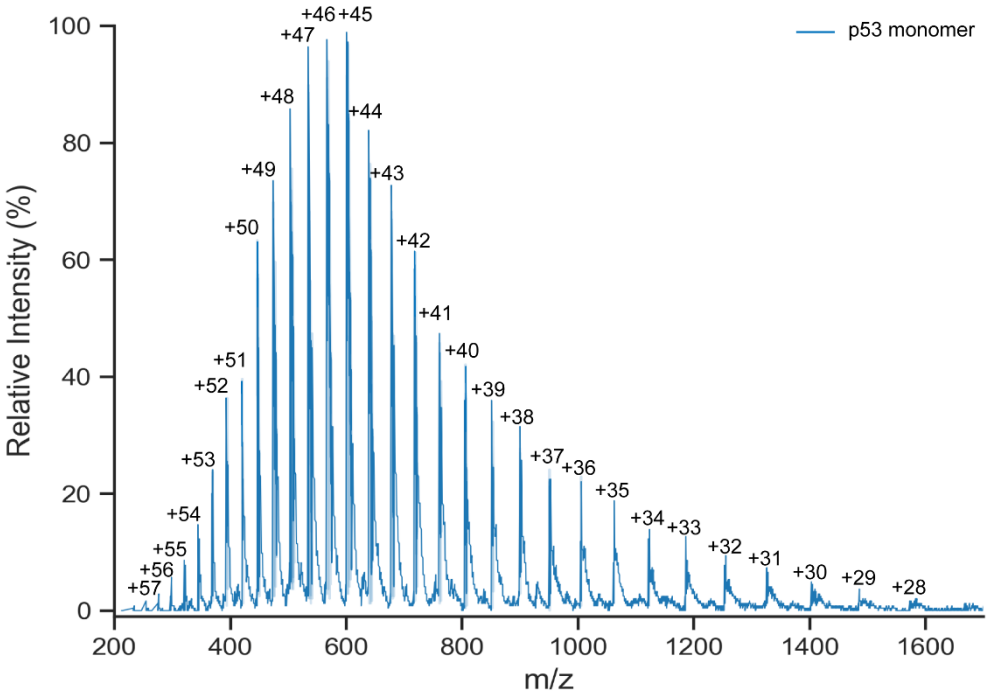


Figure 4.5: ESI-MS analysis of p53 under denaturing conditions (50% ACN/0.1% FA).

DNA binding and tetramerization propensity of p53 was assessed by two main criteria:

1) Ability of the protein to tetramerize to verify the correct folding and function of the tetramerization domain.
2) Capability of specific DNA binding to check for the correct folding and function of the DBD.

4.1.2.1 Tetramerization of p53 in the absence of DNA

The ability of p53 to undergo tetramerization was checked by native MS. Signals in the *m/z* range 3000-8000 were detected and assigned (Figure 4.6). No signals were observed in the *m/z* range above 8000, thus excluding the formation of p53 multimers with higher degrees of oligomerization. The protein was detected in four main charge distributions (monomer, dimer, trimer, tetramer). After prior calibration of the instrument using CsI, the determined molecular weights of the different p53

oligomeric species showed a small deviation (~3 amu) from the calculated values. As shown in Figure 4.7, p53 was observed as monomeric species (43.8 kDa) with charge states from 12+ to 15+, dimeric species (87.6 kDa) with charge states from 18+ to 23+, trimeric species (131.4 kDa) with charge states from 23+ to 26+, and tetrameric species (175.2 kDa) with charge states from 26+ to 31+. Additionally, signals from the chaperone DnaK protein (69.0 kDa) were observed that had already been identified as an impurity in the p53 preparation [120]. Also, signals were detected for an unknown species with a molecular weight of ca. 47.5 kDa.

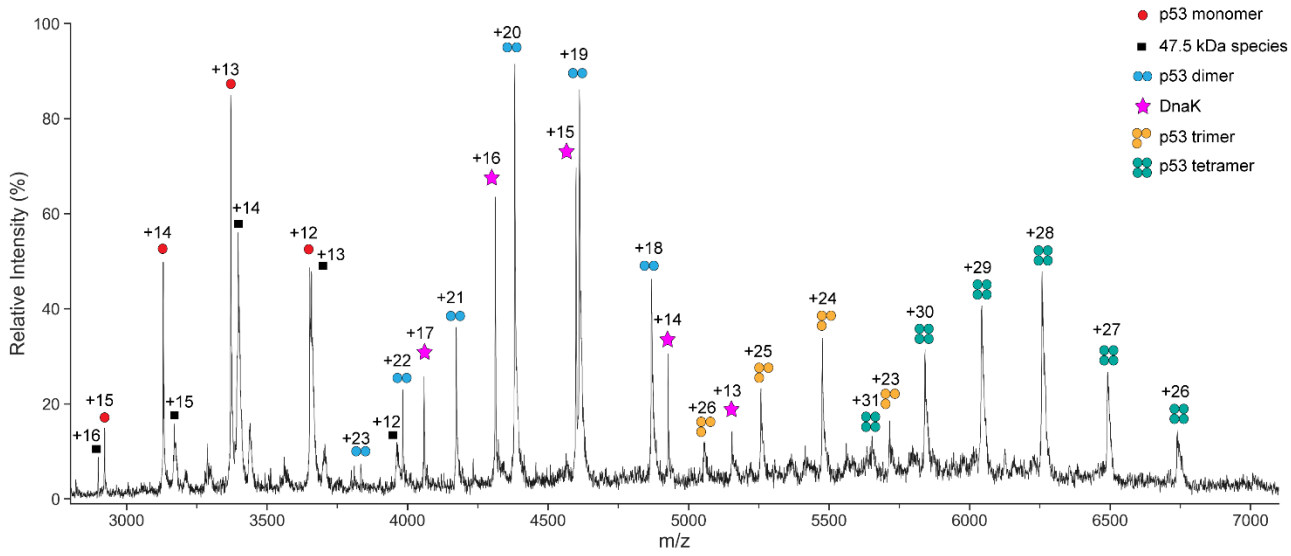


Figure 4.6: Native mass spectrum of p53 in absence of DNA.

Signals originating from the p53 monomer (red) in its different oligomeric states: dimer (blue), trimer (orange) and tetramer (green), were detected and assigned. In addition, signals for the chaperone DnaK (magenta) and an unknown species (black) were observed.

4.1.2.2 Specific DNA binding of p53

The second criterion for assessing the functionality of purified p53 is its ability to specifically bind DNA. This is an indicator for a functional DBD [74]. For this purpose, purified p53 was mixed with specific RE-DNA and incubated for 16 hours. Subsequently, the sample preparation and MS analysis were carried out as previously described (chapter 3.2.5.1). As DNA, an artificial response element (REP, 3.1.4) was used, which has an even lower K_D value for p53 binding compared to the naturally occurring response element with highest affinity. This RE-DNA consists of two palindromic repeats (half sites) flanked by a 'CGC' triplet to increase hybridization efficiency and reduce self-hybridization. The successful hybridization of the top and bottom DNA strands was checked via agarose gel (Figure A2, Appendix). Native MS measurement in the presence of RE-DNA was

conducted analogously as previously described in chapter 4.1.2.1. Signals for p53 dimer (light blue) and DnaK (magenta) were detected again. The charge state distribution of the p53 tetramer showed a shift towards higher *m/z* values between 6500 and 8000 (Figure 4.7), with charge states from 24+ to 29+ being visible. The experimental mass of this species was 191.6 kDa, which corresponds to a mass shift of ~16 kDa compared to the native MS analysis of p53 without DNA. This mass shift matches exactly to the molecular weight of the REP-DNA. Interestingly, p53 monomeric and trimeric species were not detected.

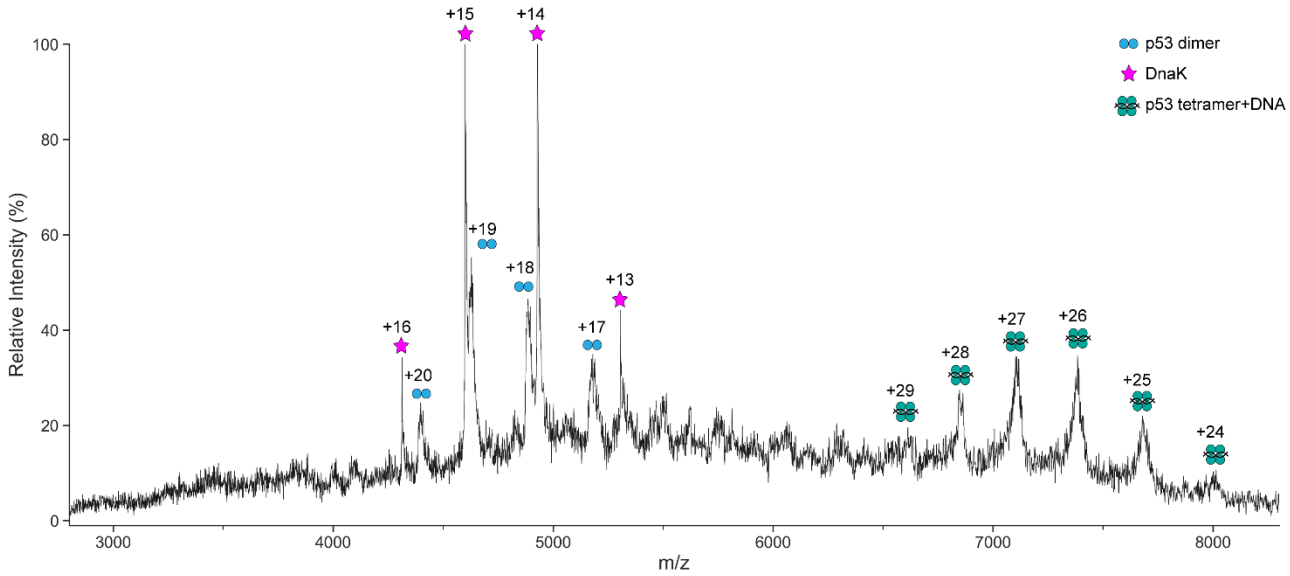


Figure 4.7: Native mass spectrum of p53 in the presence of RE-P DNA.

Signals from the p53 dimer (blue) as well as tetramer bound to DNA (green) were identified. In addition, signals for the chaperone DnaK protein (magenta) were observed.

Thus, the use of native MS allowed assessing tetramerization and DNA-binding abilities of purified full-length, wild-type human p53, ensuring the protein to be employed in structural studies using XL-MS.

4.1.3 Quantitative XL-MS using isotope-labeled BS^{2}G

The rationale behind these quantitative XL-MS experiments was to study plausible conformational changes in p53 upon DNA binding. The workflow of quantitative XL-MS and the results obtained from these experiments are shown in Figure 4.8.

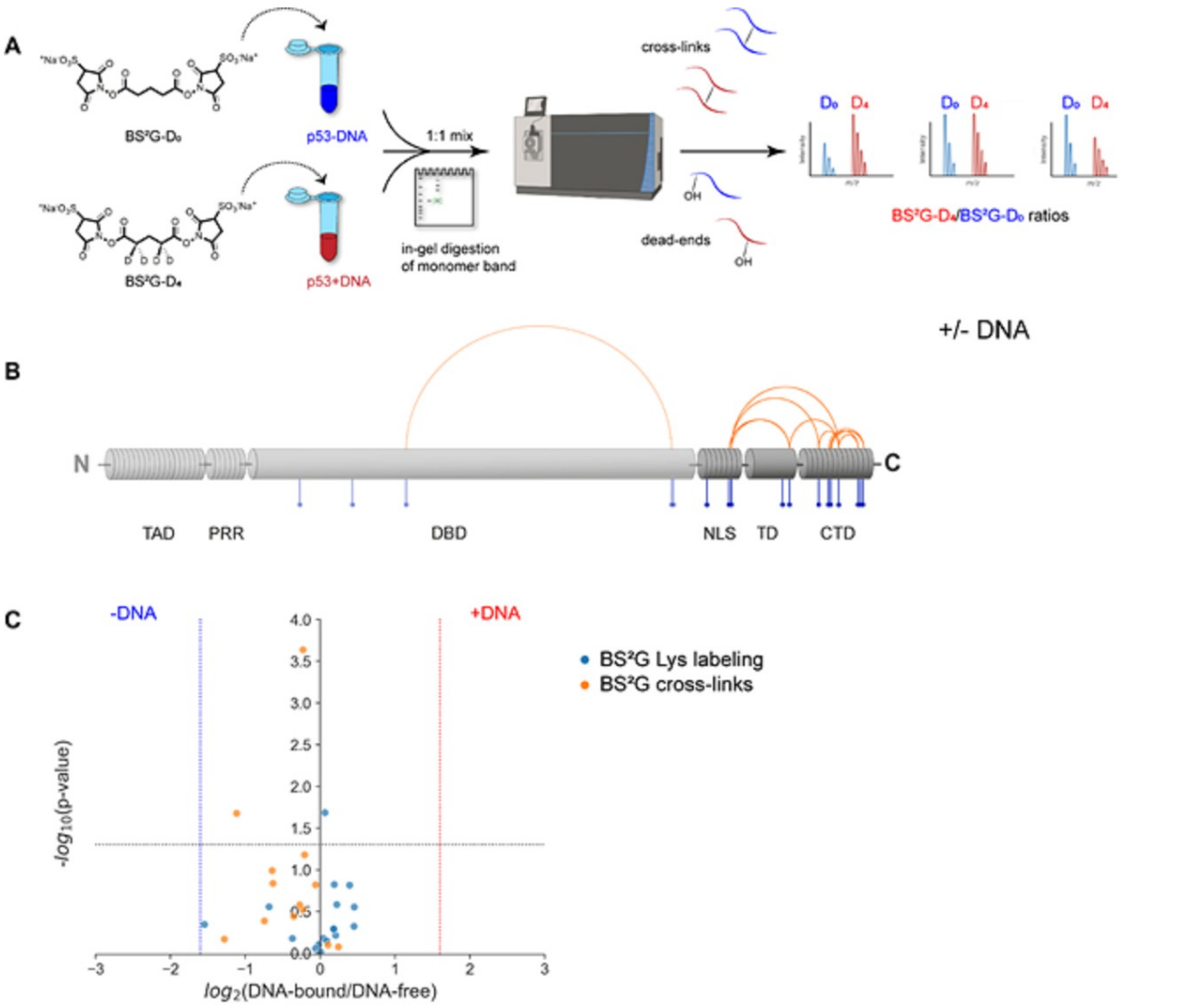


Figure 4.8: Quantitative XL-MS analysis of p53.

(A) Quantitative XL-MS workflow to study structural changes in p53 upon DNA binding [233]. (B) Schematic overview of BS^2G cross-linking (orange) and footprinting data (blue). (C) Comparison of XL-MS and footprinting data for DNA-free (left hand site) and DNA-bound (right hand site) p53.

The cross-linking reactions were performed in triplicates for each condition and repeated with reversed deuterium labeling. The cross-linking reactions were checked via SDS-PAGE (Figure A3, Appendix). Cross-linking of p53 resulted in different bands, corresponding to covalently fixed p53 dimers, trimers, and tetramers. From this gel, only the monomeric p53 band was excised and analyzed after in-gel enzymatic digestion (chapter 3.2.3.5). This was done for two main reasons:

1) Analyzing only intra-molecular cross-links within each p53 monomer originating from the tetrameric species that is present in solution;
2) Simplifying intra-/inter-molecular cross-link interpretation [120].

Twelve unique intra-molecular cross-links in p53 were identified, mainly covering the disordered NLS and CTD (Fig. 4.8 B, Table 4.1). BS^2G cross-links were quantified in the mass spectra to deduce potential structural changes in the DNA-free versus DNA-bound states of p53. Cross-links identified in at least two out of three replicates were used for quantification. Quantification was performed by comparing the intensities of 'light' (D_0) and 'heavy' (D_4) cross-links derived from the +REP-DNA and –REP-DNA conditions (chapter 3.2.7). Strikingly, the cross-link patterns did not show major differences between DNA-bound and DNA-free states, indicating similar conformational ensembles of p53 (Fig. 4.8 C). The same p53 samples were used to extract protein footprinting MS data by quantifying 'dead-end' cross-links (Fig. 4.8 B). 'Dead-end' cross-links are peptides that are modified by a partially hydrolyzed cross-linker (Fig. 1.10). These species do not provide direct distance information, but yield important insights into the overall topology of proteins based on solvent-accessible surface area (SASA) and local pK_a values of specific amino acid side chains. The combined information from inter-peptide and 'dead-end' cross-links is a highly valuable input for 3D-protein modeling. The mean labeling rates ranged between ~ 2 and 30%, and residues were classified as buried (< 5% labeling rate), partially accessible (< 10% labeling rate), and accessible (> 10% labeling rate). Conclusively, the footprinting results indicate that the accessibility of p53 CTD remains unaltered upon DNA binding (Fig. 4.8 B and C, Table 4.2). This finding is in agreement with XL-MS data and suggests that the C-terminus of p53 retains its overall topology upon DNA binding.

Table 4.1: BS^2G cross-links of p53 (+/- RE-DNA).

Identified BS^2G cross-links from the quantitative XL-MS experiment. Start/end of each peptide is indicated as superscript. Cross-linked residues for each cross-link are underlined.

Cross-linked sequences	Cross-linked residues
^{320}<u>K</u>KPLDGEYFTLQIR^{333} ^{352}DAQAG<u>K</u>EPGGSR^{363}	K320–K357
^{159}AMAIY<u>K</u>QSQHMTEVVR^{174} ^{292}<u>K</u>GEPHHELPPGSTK^{305}	Y163–K292
^{320}<u>K</u>KPLDGEYFTLQIR^{333} ^{373}<u>K</u>GQSTSR^{379}	K320–K373
^{373}<u>K</u>GQSTSR^{379} ^{382}<u>K</u>LMFK^{386}	K373–K382
^{383}LMF<u>K</u>TEGPDSD^{393} ^{364}AHSSHL<u>K</u>SK^{372}	K370–K386

^{383}LMF<u>K</u>TEGPDSD^{393} ^{373}<u>K</u>GQSTSR^{379}	K373–K386
^{307}ALPNNTSSSPQP<u>K</u><u>K</u>^{320} ^{352}DAQAG<u>K</u>EPGGSR^{363}	K319–K357
^{364}AHS<u>S</u>HLKSK^{372} ^{373}<u>K</u>GQSTSR^{379}	S367–K373
^{364}AHSSHL<u>K</u>^{370}(SK) ^{382}<u>K</u>LMFK^{386}	K370–K382
^{382}<u>K</u>LMFK^{386} ^{373}GQ<u>S</u>TSR^{379}	S376–K382
^{352}DAQAG<u>K</u>EPGGSR^{363} ^{373}<u>K</u>GQSTSR^{379}	K357–K373
^{364}AHSSHL<u>K</u>SK^{372} ^{373}GQ<u>S</u>TSR^{379}	K370–S376

Table 4.2: Footprinting results of lysine residues in p53.

Average labeling is compared between DNA-bound and DNA-free p53 and labeling ratios are calculated; p-value threshold for statistical significance was set to < 0.05.

Labeled residue number	Average DNA^- % labeling	Average DNA^+ % labeling	$\log_2$ (DNA^+/DNA^-)	$\log_{10}$(p-value)
K120	20.38	20.14	-0.02	0.10
K139	0.81	0.51	-0.68	0.55
K165	26.22	29.91	0.19	0.82
K291	3.29	3.15	-0.06	0.05
K292	8.92	12.22	0.45	0.32
K305	21.77	25.15	0.21	0.21
K319	19.31	25.39	0.39	0.81
K320	2.21	2.27	0.04	0.18
K321	9.04	9.45	0.06	1.68
K351	3.03	2.34	-0.37	0.17
K357	14.44	4.95	-1.54	0.34
K370	30.81	34.97	0.18	0.28

K372	7.56	8.83	0.22	0.58
K373	16.24	18.41	0.18	0.29
K381	17.88	17.98	0.01	0.01
K382	21.28	22.64	0.09	0.14
K386	22.54	30.95	0.46	0.55

4.1.3.1. Structural modeling of p53

Structural modeling of p53 was performed by Dr. Christian Tüting in the Kastritis lab (MLU Halle-Wittenberg). The cross-links identified in this study as well as in previous work [120] are predominantly located in the C-terminal region of p53, specifically in the TD with flanking IDRs. The data-driven modeling of p53 was therefore focused on the C-terminal region (residues 301 to 393, comprising NLS, TD and CTD). The initial tetrameric full-length p53 model was generated with AlphaFold2 and the C-terminal regions were isolated. The ordered TD showed high confidence and was in agreement with previously solved X-ray and NMR structures [71, 72, 84]. However, AlphaFold2 gave only low-confidence models for the disordered regions of p53 (NLS and CTD; Appendix Figure A4).

To optimize the structural models by satisfying the distance constraints, the initial AlphaFold2 models were subjected to MODELLER. Three different model refinement strategies were adopted:

1) ‘All restraints’, in which all distance constraints were applied to all four p53 chains symmetrically, generating 12,000 models;
2) ‘Reduced restraints’, in which 6 out of 12 experimentally determined cross-links were randomly selected for each p53 monomer, resulting in an asymmetric selection of tetrameric p53 structures, generating 2,000 models;
3) ‘Random restraints’, in which two residues within the amino acid sequence of p53 (residues 301–393) were randomly selected, 12 random constraints were defined and applied to all monomeric p53 chains symmetrically generating 2,000 models.

These strategies were used to define distance constraints and 16,000 tetrameric p53 models were generated. The refinement strategy is shown in Figure 4.9.

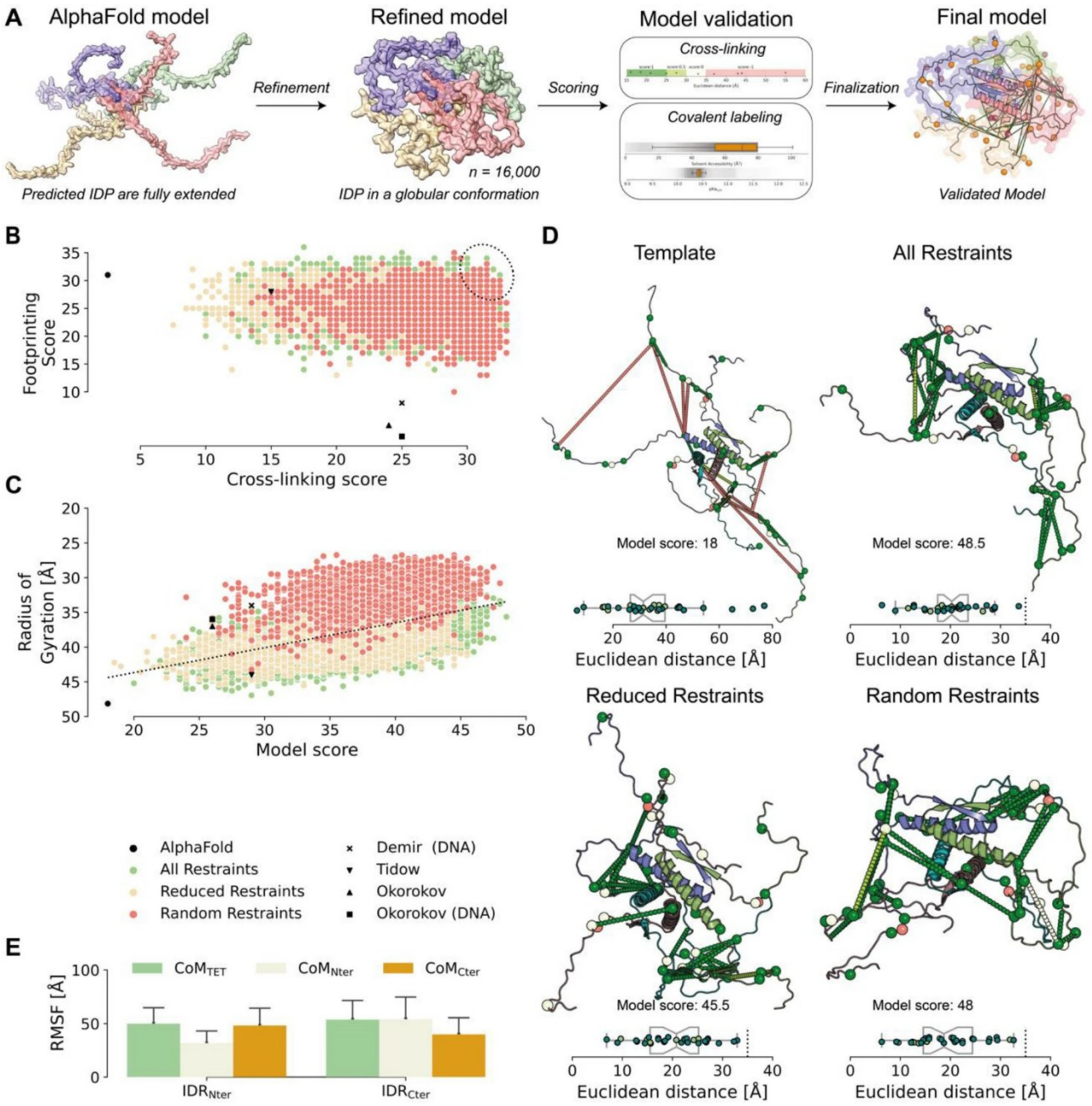


Figure 4.9: AI- and knowledge-based remodeling of the disordered C-terminal region of p53.

(A) Refinement workflow: the initial model was predicted using AlphaFold2 and was refined using experimentally and randomly derived constraints. 16,000 generated models were classified and validated using XL-MS and footprinting MS data. (B) Distribution of all models generated. The dotted ellipsoid displays the highest ranked models. (C) Correlation between radius of gyration (Rg) and model score. High compaction (low Rg) correlates with high scores of the models. (D) Presentation of input model (template) and top-ranked refined models (all restraints, reduced restraints, and random restraints). The dotted lines indicate cross-links, the spheres show labeled residues. The box-plots indicate Euclidian distances between the Cα-atoms of cross-linked residues; BS^2G cross-links (light green) and DSBU cross-links (teal). (E) Structural variation of disordered p53 termini. The center of mass (CoM) of the ordered TD, as well as the disordered NLS and CTD in each refined model were used to calculate the variations between the refined structures.

The models generated were validated with the experimentally determined cross-links (BS^2G, 12 cross-links) in addition to previously published cross-links (DSBU; 21 cross-links) [120], and BS^2G footprinting data. A stepwise scoring matrix was applied with a Cα-Cα distance threshold of 25 Å for BS^2G and 30 Å for DSBU. Distances of 10 Å above these values were included to account for the high flexibility in the IDRs, based on the overall distance distribution. In addition, footprinting data yielded information on the labeling efficiency of specific lysine residues, depending on their reactivity and SASA. The reactivity is encoded by the pK_a values of lysines and was calculated for each model with PropKa [234], while the accessibility can be described by the SASA and was calculated with the MSMS-tool [235]. To access the distribution of pK_a and SASA values of each lysine residue, all models were plotted against the experimentally determined accessibilities of lysines (Appendix Figure A5). Notably, the pK_a value did not allow differentiating the results, while the SASA showed a population of lysine residues with reduced accessibility, correlating with footprinting data (Table 4.2). For scoring, a stepwise rating matrix was applied that penalizes outliers. For model selection, XL-MS and footprinting data were plotted against each other, and the highest ranked solution of each refinement strategy was chosen (Figure 4.9 B). A clear improvement of scoring, both regarding cross-link distance and SASA, was assumed when compared to the AlphaFold2-generated input model and already published tetrameric p53 models (chapter 1.3.3, Figures 4.9 B, A6). This indicates an experimentally-driven optimization of computationally predicted AlphaFold2 models. One striking observation was that the refined p53 models appeared to be more compact than already existing models. To confirm this, the Rg was calculated for each model. There was a correlation between the scoring and the Rg of p53 (Figure 4.9 C). Selecting and aligning the top-ranked models revealed an ordered and structurally preserved TD, but the neighboring IDRs exhibited various conformational states (Figures 4.9 D, E). An identical or similar kind of compaction of the p53 tetramer has already been observed [98, 120, 236] but is not reflected in alternative models (Appendix Figure A6). Using the experimental data and the scoring, all previously proposed models of the CTD performed worse in the overall scoring than the refined compact models (Figure 4.9 C).

4.1.3.2. Structural dynamics of full-length p53 probed by HDX-MS

Additional validation of the generated p53 tetrameric models (chapter 4.1.3.1) was performed with HDX-MS data. HDX-MS displays protein dynamics and changes in a time-dependent manner and complements XL-MS and footprinting data [121, 122, 237]. In HDX, backbone amide protons of a polypeptide chain are limited in their H/D exchange if they are involved in hydrogen bonding

networks. The deuterium incorporation kinetics of the full-length p53 tetramer was mapped for the first time, yielding fast exchange rates for the IDRs, TAD, NLS, and CTD, and slow exchange rates for regions with known secondary structures, DBD and TD (Figure. 4.10 A) [233]. Plotting the accessibility of the backbone amide protons in the initial AlphaFold2 model as well as in the previously proposed models (Appendix Figure A6, Okorokov, Demir and Tidow models) and top-ranked refined models (Figure 4.9 D), the expected correlation with deuterium incorporation in HDX was confirmed (Figures 4.10 B, A6).

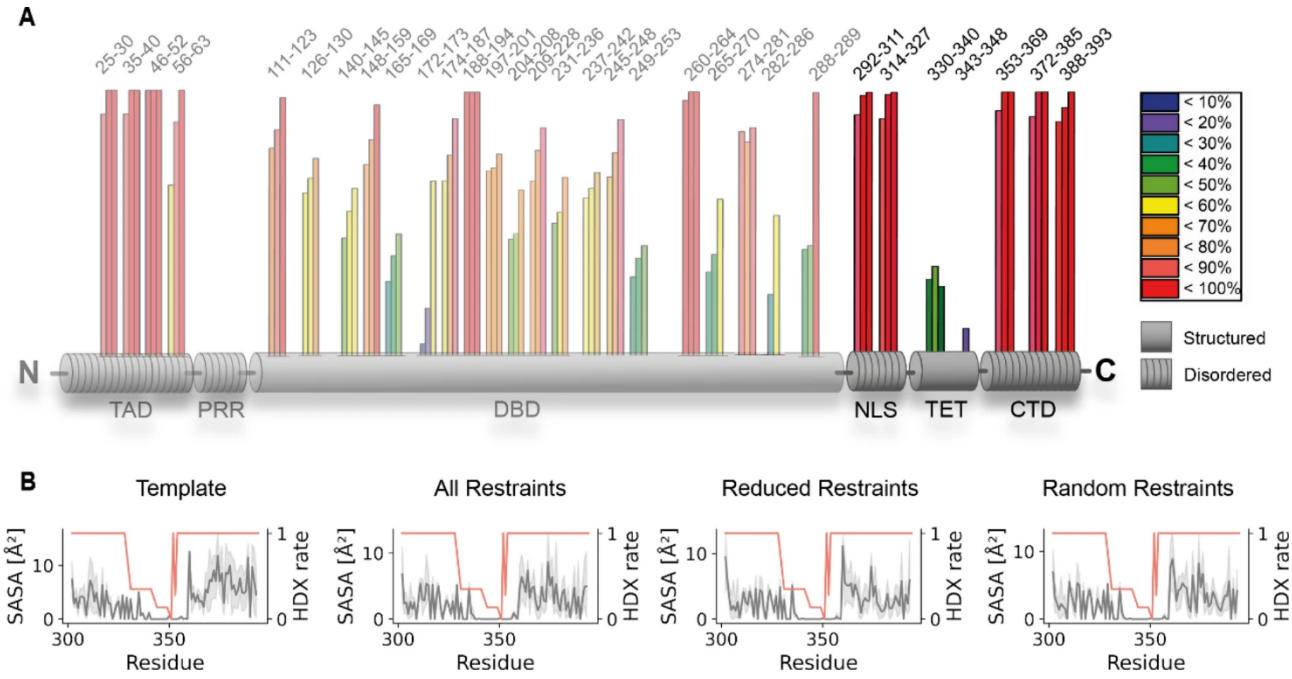


Figure 4.10: HDX-MS of backbone amide protons in DNA-free p53.

(A) HDX-MS in DNA-free full-length p53. Three bars (HDX after 30 s, 1 min, and 2 min) are presented for each peptide analyzed; amino acid numbers are shown above. Data from NLS, TD, and CTD confirm the top-ranked refined models. (B) Models (Fig. 4.9 D) show a high correlation between SASA (grey) and HDX (orange). The disordered and accessible nature of p53's IDRs is reflected by high SASA and HDX rates, while the structured TD (residues 326–356) exhibits low SASA and HDX rates.

4.1.4. Investigation of the C-terminal IDR of p53 in DNA binding

Due to the high degree of disorder in p53 (~40%) [236], structural studies of full-length p53 are scarce. Usually, protein constructs lacking the IDRs or IDR-derived peptides have been used instead. For this reason, all the models published so far [95-100] have limited structural information derived from empirical assays regarding the IDRs. These are usually a result of *ab initio*-modeling approaches. Recent advances of AI-based prediction tools, such as AlphaFold2, have enabled the

generation of structures for well-folded protein domains, while predicting the structures of IDRs/IDPs is still a daunting task. The aim of the work was to shed light on the structural roles of p53's C-terminal region in full-length, wild-type human p53 tetramer and its importance for DNA binding. For this, complementary techniques of structural MS, namely XL-MS, protein footprinting and HDX-MS, were used in an integrated approach with computational modeling. The purification of functional full-length human wild-type p53 based on an existing protocol [7, 120] facilitated obtaining structural information in terms of spatial distance constraints via XL-MS experiments. Specifically, a comparison of the topology of p53 with a focus on its C-terminal IDR in the absence and presence of RE-DNA were obtained. Due to the high content of lysines in the NLS, TD and CTD of p53, the homobifunctional cross-linker BS^2G targeting primary amine groups in proteins, i.e. lysine residues and N-termini, was employed. Furthermore, two BS^2G isotopologues, $BS^2G\text{-}D_0$ and $BS^2G\text{-}D_4$, are commercially available [238]. These reagents generate distinct MS patterns, simplifying cross-link identification and quantification. Cross-links obtained from BS^2G were used as input for subsequent 3D structure modeling using MODELLER [239], starting from the tetrameric p53 model generated with AlphaFold2 [29, 30]. The newly refined p53 models were validated via complementary XL-MS data of p53 using the MS-cleavable cross-linker DSBU, protein footprinting data as well as HDX-MS data [233]. By this, novel insights in the C-terminal IDR of p53 upon DNA binding could be obtained.

4.2. Development of a MS-cleavable, CuAAC-enrichable cross-linker: PAC4

4.2.1. Description of the cross-linker features

The first trifunctional CID-MS/MS-cleavable cross-linker characterized in the present work is the PAC4 (piperidinium-alkyne-C4) cross-linker developed in the group of Andrews [138]. As the name suggests, it possesses a central, positively charged MS-cleavable piperidinium function, on which an alkyne handle is installed for copper-catalyzed azide/alkyne cycloaddition (CuAAC) reaction for the enrichment of cross-linked products (Figure 4.11). The cross-linker is formulated as chloride salt.

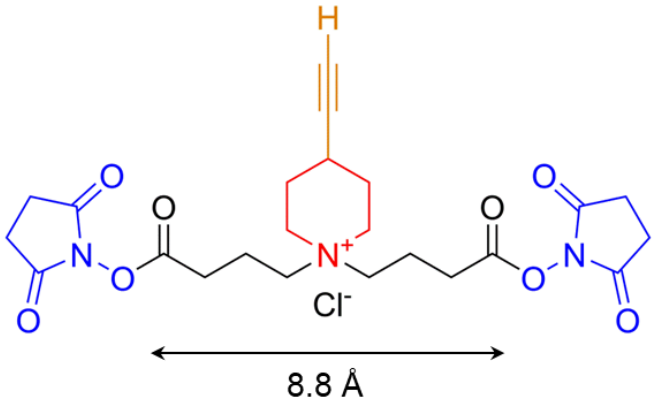


Figure 4.11: Design of the trifunctional CID-MS/MS cleavable cross-linker PAC4.

The cross-linker is characterized by a permanently positively charged CID-MS/MS-cleavable piperidinium function in the center (shown in red), two NHS esters as reactive groups (in blue) and an alkyne handle (in orange) to enrich cross-linked peptides via CuAAC click reaction.

To thoroughly understand the fragmentation behavior of the reagent, XL-MS experiments were performed using model peptides, such as Test Peptide 1 (TP1) and Angiotensin II (ATII) (chapter 3.1.5).

4.2.2. Cross-linking experiments with model peptides

4.2.2.1. TP1 cross-linking

As first model peptide, TP1 was used. This synthetic peptide, was developed specifically for testing novel cross-linkers [133, 240]. After the cross-linking reaction, tryptic digestion generated the cross-

linked products. Cross-linked TP1 was analyzed via direct-infusion ESI-MS (chapter 3.2.5.3) and mass spectra showed the presence of several cross-linked products of TP1 (Figure 4.12):

1) Inter-peptide type II TP1 cross-link (EADYLINKER-TESTDIKR), which was present in charge states 2+ (*m/z* 1222.635), 3+ (*m/z* 815.430), 4+ (*m/z* 611.823), 5+ (*m/z* 489.661). Most type II species were also present as adducts with Na^+, K^+ and combinations of the two cations;
2) 'Dead-end' type 0 cross-links, which were present both for the α-peptide EADYLINKER at charge states 2+, 3+ (*m/z* 757.403, 505.271 with neutral losses of water indicating possible type I links as well) and the β-peptide TESTDIKR with charge state 3+ (*m/z* 404.889); also NH_3-quenched species were observed.

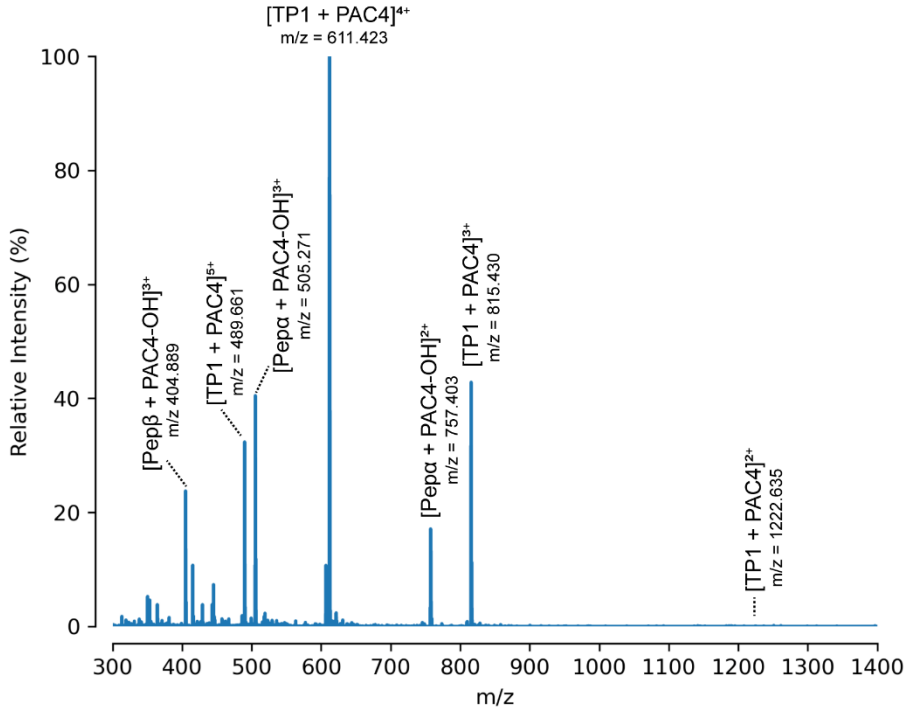


Figure 4.12: Mass spectrum of TP1 cross-linked with PAC4.

Different types of TP1 cross-linked products (type II, type 0) are present at different charge states. Precursor ions corresponding to these species were isolated and fragmented either using CID- or HCD-MS/MS in the orbitrap fusion tribrid mass spectrometer.

Precursor ions corresponding to these species were isolated in the quadrupole and fragmented either using CID or HCD in the orbitrap fusion tribrid mass spectrometer. MS/MS data were acquired in the orbitrap mass analyzer. These two fragmentation methods are the most commonly adopted for MS-cleavable cross-linkers on orbitrap instruments as they are usually very fast and do not require long duty cycles [241]. Hybrid dissociation techniques, such as a combination of CID with ETD (ETciD)

and HCD with ETD (EthcD) [214], were tested for the PAC4 cross-linker. Both required at least 150 ms of activation via the fluoranthene electron-transfer reagent for the dissociation to occur (data not shown) and were therefore not considered further. For this reason, the fragmentation behavior of the PAC4 linker was studied via CID and HCD in order to find the optimal energy to obtain both fragmentation of the MS-cleavable piperidinium function as well as the generation of highly informative peptide backbone fragments (a-, b-, and y-type ions). Specifically, dissociation experiments were carried out, in which the cross-linked species were subjected to CID and HCD energy ramps from 0 to 40% NCE using steps of 5%. From these experiments, the mechanism of MS-cleavability of PAC4 could be understood in detail (Figure 4.13).

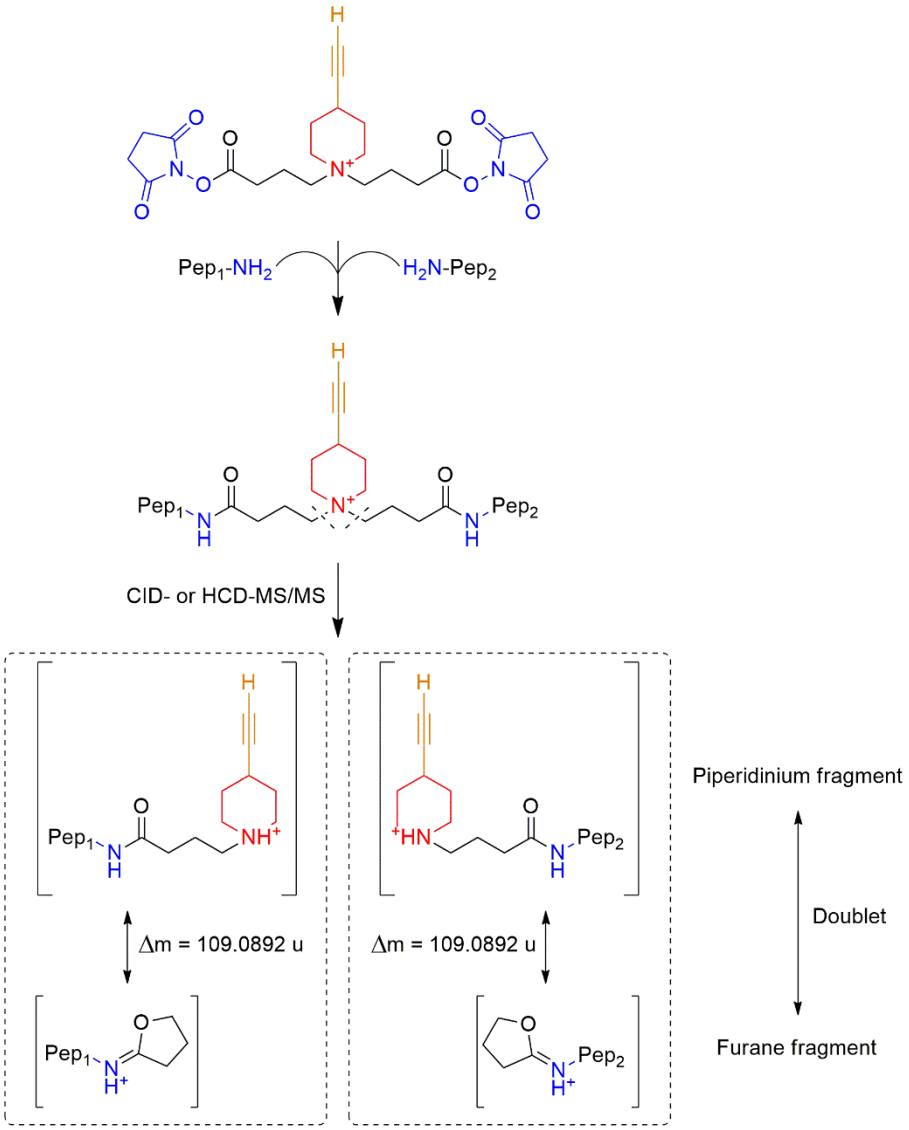


Figure 4.13: Fragmentation behavior (CID and HCD) of the PAC4 cross-linker [138].

Upon reaction of the NHS ester groups (in blue) with two amino acids (mainly lysine residues) of a protein in spatial proximity (25-30 Å), the cross-linked product is formed. CID or HCD delivers diagnostic 'doublet'

ions composed of a piperidinium and a furane fragment ion, one for each peptide involved in the cross-linking site, with a defined mass shift of ~109 u. The MS-cleavable function is depicted in red and the bond being cleaved is indicated by a dashed line.

The piperidinium group in the spacer of PAC4 is labile and undergoes gas-phase fragmentation via delocalization of the electron pair of the nitrogen of the newly formed amide bond. This results in a piperidinium fragment if the charge remains at the piperidine, or in a furane fragment if the charge is localized at the amide nitrogen. This leads to the generation of diagnostic 'doublet' ions with a mass shift of 109.0892 u, which are important for two main reasons:

1) Unambiguous assignment of cross-linked peptides as the mass of each peptide can be easily calculated, knowing the cross-link precursor mass and the mass of the cross-linker fragments.
2) Reduction of the quadratic search space problem (n^2), which increases the computational time of the analysis, to a linear search space (2n) (chapter 1.6.6.2.3) [212].

An exemplary fragment ion mass spectrum for the 4+ charged type II TP1 cross-linked peptide (*m/z* 611.823) is shown in Figure 4.14.

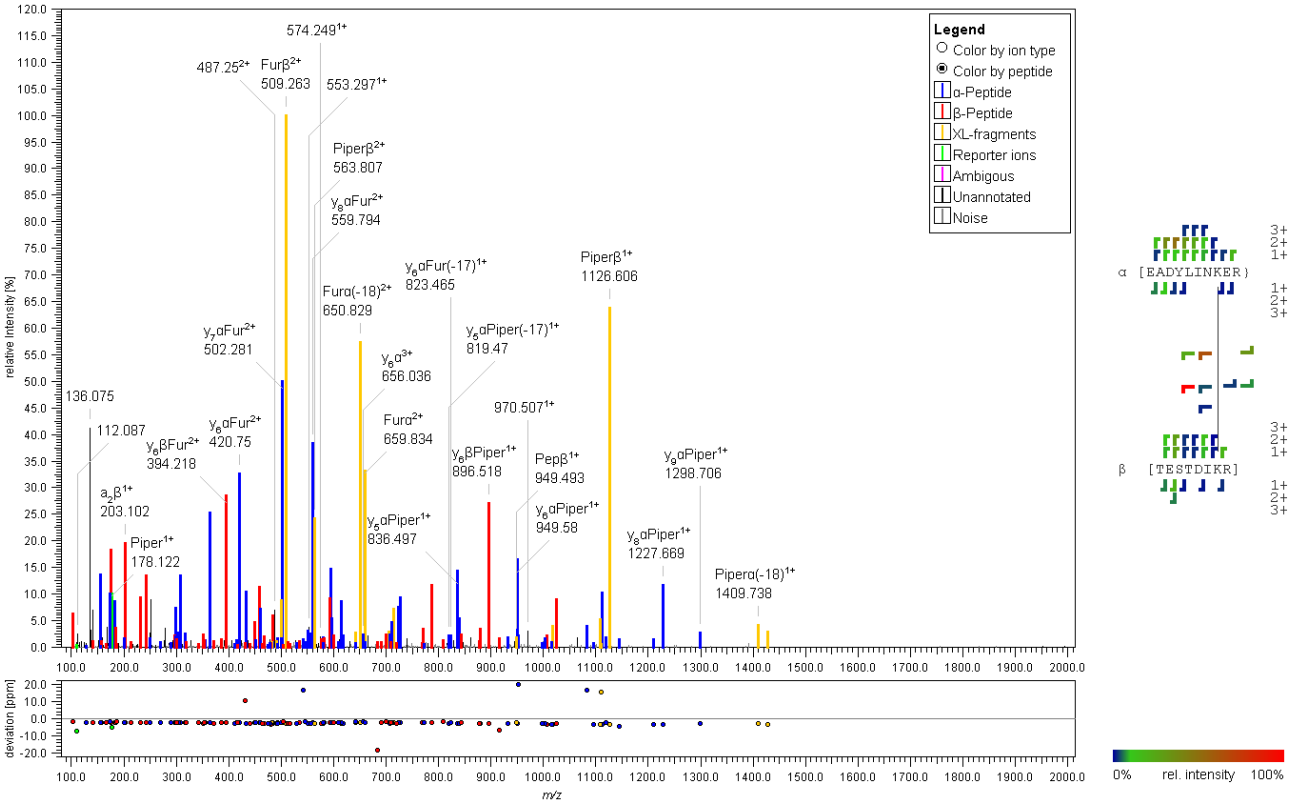


Figure 4.14: Fragment ion mass spectrum (HCD-MS/MS) of the 4+ charged type II TP1 cross-linked peptide (*m/z* 611.823), NCE 35 %, analysis was performed via MeroX [8, 213].

Exemplary deisotoped fragment ion mass spectrum for the 4+ charged TP1 cross-link. PAC4 diagnostic 'doublet' ions (Piper and Fur) are shown in yellow, b- and y-type ions series for the α (EADYLINKER) and β (TESTDIKR) peptides are presented in blue and red. PAC4 characteristic reporter ions (*m/z* 178.123 and 110.096) are indicated in green. Mass deviation (in ppm) for the fragment ions is shown in the panel below.

Some peculiar features of the PAC4 cross-linker's gas phase behavior were observed in these dissociation experiments:

1) The optimal collision energies in HCD experiments to observe both cross-linker doublets (piperidinium and furane) as well as the peptide backbone fragments were NCE = 30–35% for all charge states (2+ to 5+). CID resulted mainly in the formation of cross-linker doublets. Peptide backbone fragments increased for 4+ charge states at NCE = 40%.
2) In HCD experiments, NCE = 25% is required to induce fragmentation of the 4+ charged type II cross-linked species. At this energy, the most prominent ions were resulting from the doubly charged furane fragments of the α and β peptides, while the respective doubly charged piperidinium fragments had a lower intensity. At higher collision energy (NCE = 35–40%), the abundance of the doubly charged furane and piperidine fragments remained similar, while an abundant piperidinium fragment appeared for the β peptide and, with lower intensity for the α peptide. Upon collisional activation, the cleavage of the cross-linker apparently produces doubly charged furane and piperidine ions from the 4+ charged precursor ion. The piperidinium fragment ion can undergo further dissociations in HCD, such as neutral losses, creating the doubly charged furane fragment. At higher NCE, charges might not be distributed equally on each side of the cross-link, and as a result singly charged fragment ions (whose charge is located on the most basic site, the piperidine) are generated (Appendix Figure A7). The fragmentation of the triply charged type II precursor ion generates higher intensity piperidinium fragments already at lower collision energies (NCE = 25–30%), due to the odd number of charges on the cross-link (Appendix Figure A8);
3) The most interesting feature is the presence of PAC4 reporter ions from type I/II and type 0 ('dead end') cross-links in HCD. The ethinyl-piperidine fragment (*m/z* 110.096) and the piperidine acylium ion (*m/z* 178.123) are present in type I/II and type 0 fragment ion mass spectra, while the piperidine carboxylic acid fragment (*m/z* 196.133) is exclusively present in spectra of 'dead end' cross-links. This species can only arise from a partially hydrolyzed cross-linker and not from a K-K linkage (Figure 4.15, see also Appendix Figure A9). This feature could be quite appealing for developing targeted MS methods that distinguish between inter-peptide and 'dead end' cross-links.

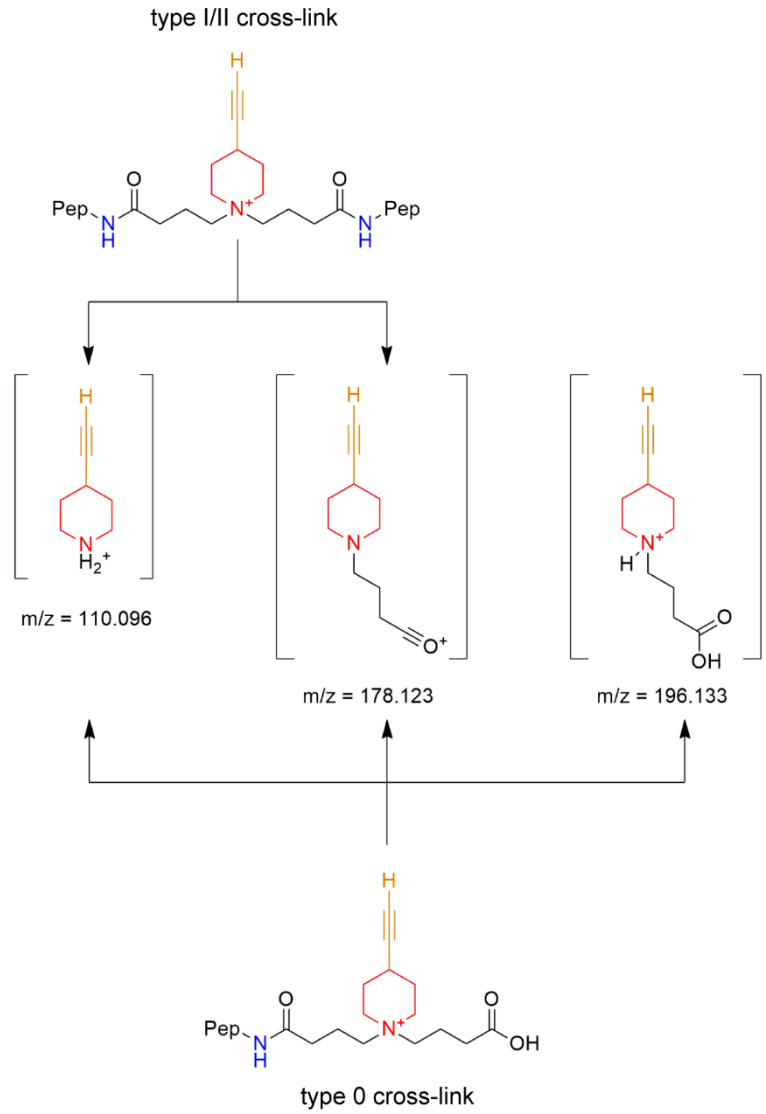


Figure 4.15: PAC4 reporter ions generated from type I/II and type 0 cross-linked products.

PAC4 generates characteristic reporter ions when subjected to HCD. Some of these reporter ions are shared between type I/II and type 0 cross-links (*m/z* 110.096 and 178.123) while a reporter ion at *m/z* 196.133 is unique to 'dead end' cross-links. This feature enables the discrimination of inter-peptide cross-links from 'dead end' cross-links.

An example of this approach has been recently published for the tBu-PhoX cross-linker, in which a fast survey MS/MS scan in the linear ion trap or in the orbitrap was used to generate reporter ions in the orbitrap lumos mass spectrometer. The relative intensities differ between type II (as well as the more similar type I cross-links) and type 0 cross-links and were used to trigger a high-resolution scan in the orbitrap only for the desired type II species. This approach has been termed Real-Time Library Search (RTLS) [242] and has been shown to improve cross-link identification from non-enriched

samples using short gradients, since 'dead end' cross-links are usually the most abundant species and therefore a large proportion of measurement time is spent on their acquisition.

4.2.2.2. ATII cross-linking

A second peptide for testing the PAC4 cross-linker, the model peptide angiotensin II (ATII), was used. This peptide does not contain lysine residues. Hence, the main reactive sites are the N-terminus of the peptide and the tyrosine (Y) in position 4, which is close to a histidine (H) residue that could mediate the cross-linking reaction [197]. Cross-linked ATII was analyzed via direct-infusion ESI-MS. Several ATII cross-linked products were observed in the mass spectrum (Figure 4.16):

1) A homeotypic type II cross-link (DRVYIHPF-DRVYIHPF) that is present at charge states 2+ (*m/z* 1169.113), 3+ (*m/z* 779.744), 4+ (*m/z* 585.060), 5+ (*m/z* 468.249).
2) A type 0 cross-link, present at charge states 1+ (*m/z* 1308.710), 2+ (*m/z* 654.858), 3+ (*m/z* 436.908). Also neutral losses of water (-18 u) were observed.
3) A doubly modified type 0 cross-link at charge states 2+ (*m/z* 786.929), 3+ (*m/z* 524.955), 4+ (*m/z* 393.968), in which both the N-terminal aspartate (D) and the Y at position 4 are modified. Additionally, mixed type 0 cross-links were found at all charge states as well as a doubly NH_3 modified type 0 cross-link at charge state 4+.

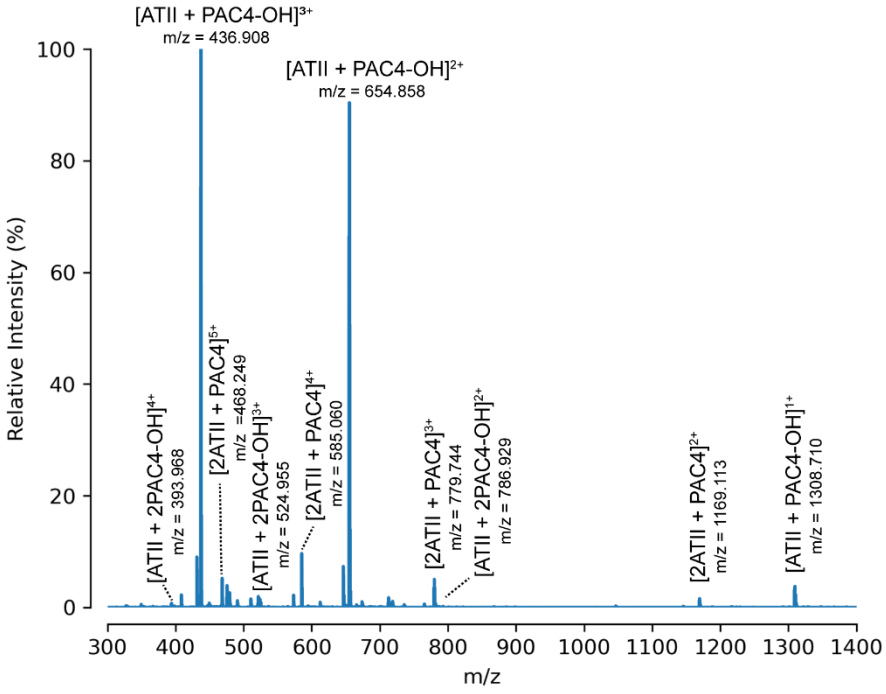


Figure 4.16: Mass spectrum for ATII cross-linked with PAC4.

Different types of ATII cross-linked products (type II, type 0) are present at different charge states. Precursor ions corresponding to these species were isolated and fragmented either using CID- or HCD-MS/MS in the orbitrap fusion tribrid mass spectrometer.

Collisional activation (CID, HCD) experiments were carried out as described for TP1 (chapter 3.2.5.3). One striking feature of ATII, previously observed in XL-MS experiments during (-)-ESI-MS of the homeotypic type II ATII cross-linked peptide, is an intramolecular rearrangement of the N-terminal D side chain and subsequent release of the second ATII peptide [240]. The formation of type 0 cross-link species (*m/z* 654.858–655.350 and *m/z* 436.908–437.243 at charge states 2+ and 3+) was also observed. These species might be a result of an intramolecular rearrangement of the arginine (R) at position 2 or the N-terminal D (Appendix Figure A10). An exemplary spectrum for the 3+ charged species of the homeotypic type II cross-link of ATII is shown in Figure 4.17.

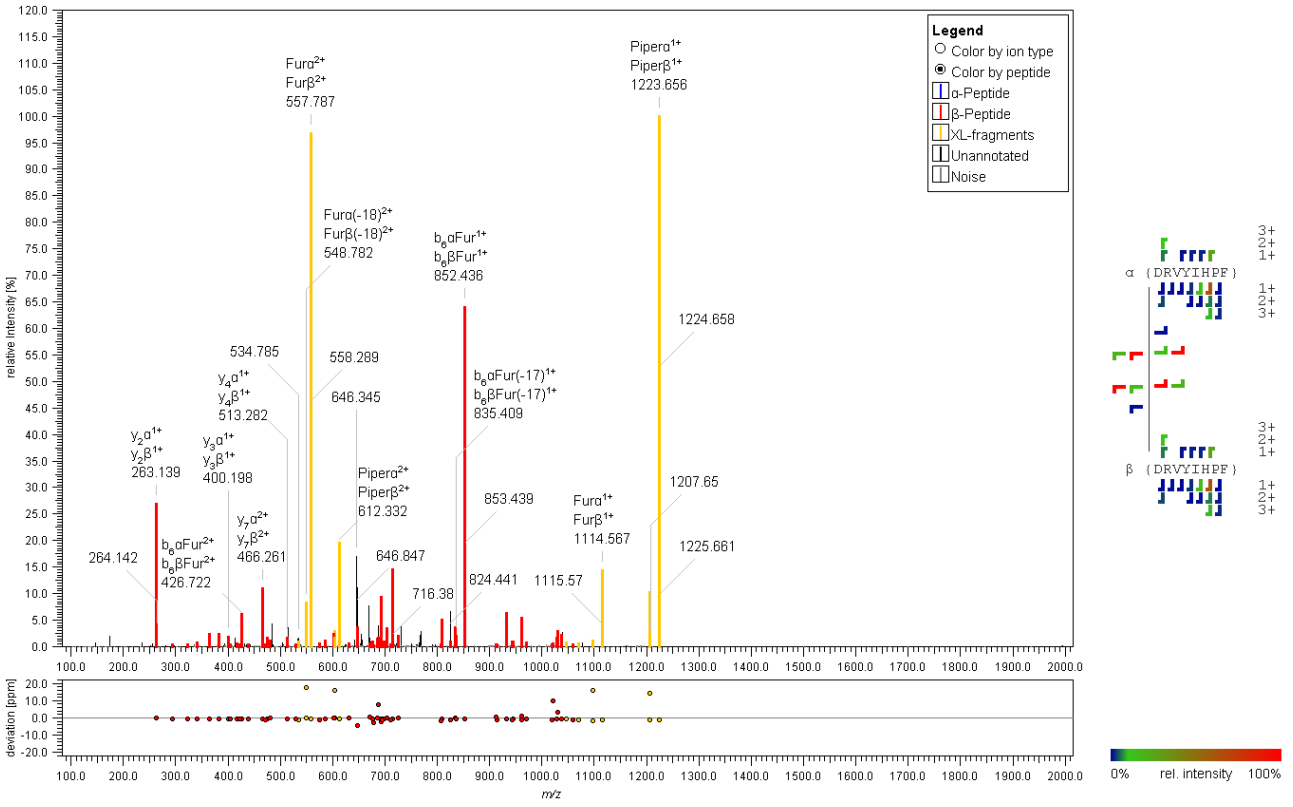


Figure 4.17: Fragment ion mass spectrum (CID-MS/MS) of the 3+ charged homeotypic type II cross-linked peptide of ATII (*m/z* 779.744), NCE = 35%, analyzed via MeroX [8, 213].

Exemplary fragment ion mass spectrum for the 3+ charged ATII cross-link. PAC4 diagnostic fragment ions are shown in yellow, b- and y-type ions series for the ATII peptide are presented in red.

4.2.2.3. Reporter ions analysis of PAC4 cross-linked products

As discussed above (chapter 4.2.2.1), an interesting feature of the PAC4 cross-linker is its capability of generating informative reporter ions that might be used to discriminate between type II and type 0 cross-linked products using HCD. Therefore, an in-depth analysis of PAC4's reporter ions was conducted using TP1 as a model peptide (Figure 4.18).

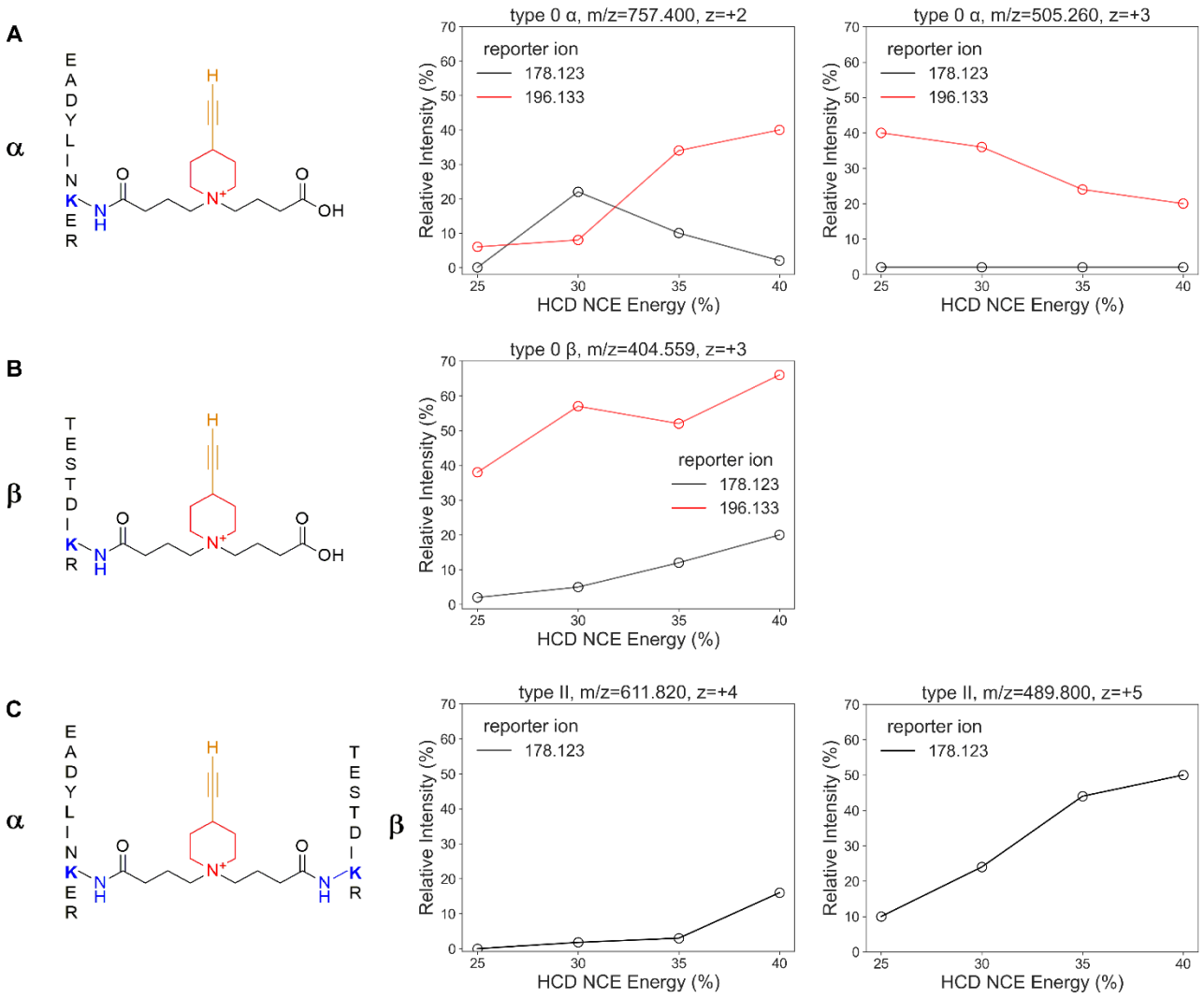


Figure 4.18: Reporter ion analysis for the PAC4 cross-links of TP1.

Reporter ion analysis of PAC4 TP1 cross-linked products: the piperidine-acylium at *m/z* 178.123 (black) and the piperidine-carboxylic acid fragment (red) at *m/z* 196.133. (A) Type 0 'dead-end' species of TP1 α-peptide EADYLINKER (2+ and 3+ charge states); (B) type 0 'dead-end' species of TP1 β-peptde TESTDIKR (3+ charge state); (C) type II cross-link of TP1 (4+ and 5+ charge states). The relative intensities of the reporter ions for all the different TP1 cross-linked products are plotted as a function of the NCE (%) applied in HCD-MS/MS experiments.

The intensity of the ethinyl-piperidine fragment (*m/z* 110.096) was low and it did not change dramatically between the two types of cross-links. For the 2+ species of the EADYLINKER type 0

cross-link, the piperidine-acylium ion at *m/z* 178.123 (black) reached its maximum relative intensity (22%) at HCD NCE=30%, and decreased at higher HCD energy (Figure 4.18 A, center). On the contrary, the piperidine-carboxylic acid ion at *m/z* 196.133 (red) that is generated from a type 0 ('dead-end') cross-link, showed an increase in the relative intensity with a maximum of 40% at HCD NCE=40%, marking a prominent difference between these two reporters for a type 0 cross-link. The 3+ species of the same peptide (Figure 4.18 A, right) showed a prominent 196.133 species with its intensity dropping from 40% (HCD NCE = 25%) to 20% (NCE=40%), while the 178.123 species remained low intense (~2%) for all collision energies applied. For the 3+ species of the TESTDIKR type 0 cross-link (Figure 4.18 B, center), the piperidine-acylium ion was ranging from a 2% relative intensity at HCD NCE=25% to a relative intensity of 20% at HCD NCE=40%. For the piperidine-carboxylic acid ion, the relative intensity ranged from 38% at HCD NCE=25% to 66% at HCD NCE=40%, again showing a dramatic difference in the relative intensities of these two diagnostic ions. The inter-peptide type II cross-link TESTDIKR-EADYLINKER 4+ species (Figure 4.18 C, center), showed formation of only the piperidine-acylium ion, with intensities ranging from 1% (HCD NCE = 25%) to almost 20% at HCD NCE=40%, while for the 5+ species (Figure 4.18 C, right) the intensity range was ranging between 10% to 50% (HCD NCE=40%). These observations indicate different behavior of type II and type 0 cross-links during collisional activation, which might serve as a feature to develop targeted MS methods, i.e., by triggering high resolution MS/MS scans based on the presence and relative intensities of these indicative PAC4 fragment ions.

4.2.3. Cross-linking experiments with proteins

4.2.3.1. Cross-linking experiments with bovine serum albumin and carbonic anhydrase 2

To test the applicability of PAC4 for protein cross-linking, bovine serum albumin (BSA) and bovine carbonic anhydrase 2 (CA2) were employed (see chapter 3.1.6). BSA is known to form dimers (and high-molecular weight aggregates) [243], while CA2 exists as monomer in solution [244]. Both proteins serve as proof-of-principle to understand if the PAC4 can efficiently cross-link monomeric as well as multimeric protein systems. Cross-linking experiments were conducted by titrating the proteins with increasing concentrations of the PAC4 cross-linker (see chapter 3.2.4.3 for details) and/or increasing reaction times. SDS-PAGE analysis (Appendix Figures A11 A, B) showed that both proteins were successfully cross-linked already at low cross-linker concentrations (10-25 fold excess for BSA, 25-50 fold excess for CA2). For BSA, a monomeric species with different electrophoretic

mobility appeared on the gel (which was not present in the control) as well as increasing amounts of BSA oligomers, indicating successful BSA modification. For CA2, no oligomeric bands were observed, but a clear shift in the electrophoretic mobility of the monomeric protein towards higher molecular weight species was detected.

4.2.3.2. Native MS with cross-linked bovine serum albumin and carbonic anhydrase 2

BSA and CA2 were also analyzed via native MS to exactly define the number of cross-linker modifications per protein molecule. Native MS can be used as complementary approach to select the correct concentration of cross-linker in order to not over cross-link protein structures and get artificial structural constraints. BSA and CA2 were analyzed using an in-house built online buffer exchange system [227] (see chapter 3.2.5.1). For CA2, Native MS of the control sample with no addition of cross-linker resulted in a clear, normal-shaped distribution spanning charge states from 6+ to 11+, centered around the 9+ charge state (Figure 4.19 A). Addition of the PAC4 cross-linker caused the appearance of several peaks at the right hand side of each CA2's charge state distribution. Each of these peaks could be assigned to PAC4-modified CA2 species. The number of PAC4 cross-links ranged from 2-3 at 10-fold to a maximum of 8-9 at 100-fold excess. Remarkably, 25-fold molar excess of the cross-linker seemed to be already a sufficient concentration for cross-linking of CA2, for which a substantial amount of unmodified protein could still be observed (Figure 4.19 B). At 50-fold molar excess of the cross-linker, the majority of CA2 molecules has reacted and at 100-fold molar excess CA2 was heavily modified by PAC4. Given these observations, a 25-fold molar excess of PAC4 was determined to be the optimal concentration for modifying this protein. It has to be noted that at 50-fold molar excess of the cross-linker and even more pronounced at 100-fold excess, charge state distribution of CA2 was affected. Higher charge states increased in intensity due to the charge imposed by the PAC4 cross-linker (Figure 4.19 A).

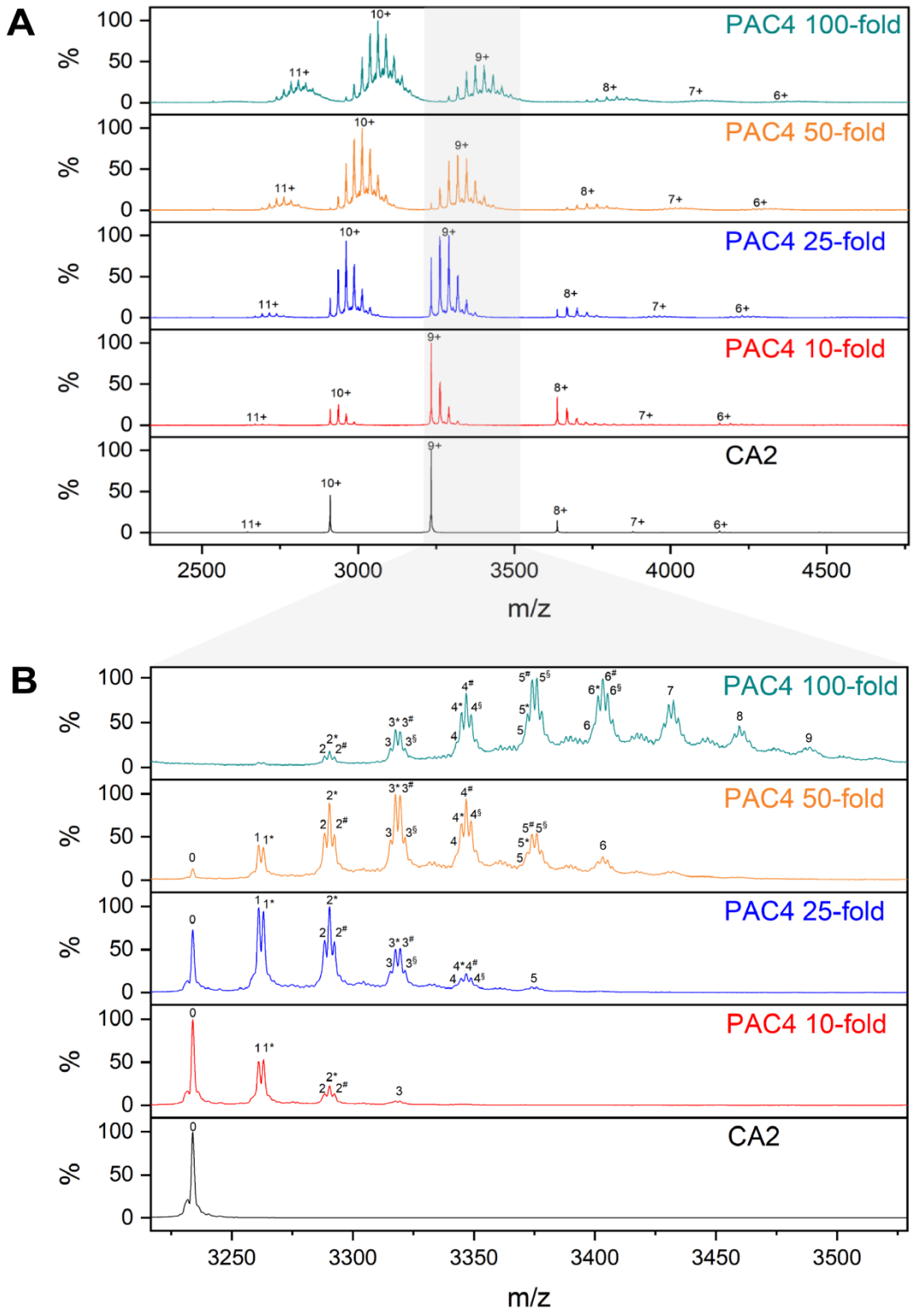


Figure 4.19: PAC4 modification of CA2 determined by native MS.

A) Native mass spectra of CA2 cross-linked with different molar excess (from 0 to 100-fold) of PAC4 cross-linker. Addition of PAC4 resulted in a number of cross-linked species. Charge states are indicated. B) Zoom-in on the most abundant charge state (9+) of CA2, after addition of PAC4 at different molar excesses. Each peak is labeled with the number of PAC4 modifications: X: number of type II/type I cross-links; X*: number

of type II/type I cross-links and one type 0 cross-link, $X^{\#}$: number of type II/type I modifications and two type 0 cross-links, $X^{\S}$: number of type II/type I modifications and three type 0 cross-links. Intensities are normalized to the most abundant species for each spectrum.

For BSA, a higher PAC4 molar excess was found to be optimal for protein modification (50- or 100-fold molar excess, Appendix Figure A12). The monomer:dimer ratio of BSA in any conditions tested was not found to be altered.

4.2.4. Optimization of a click chemistry protocol for enrichment of cross-links

4.2.4.1. CuAAC reaction with cross-linked TP1

To make an efficient use of the PAC4 cross-linker for the analysis of complex mixtures, the CuAAC enrichment of cross-links had to be enabled by the alkyne handle on the MS-cleavable piperidine function (Figure 4.11). As a simple system for testing the ability of PAC4 to undergo CuAAC click chemistry conversion, TP1 was used. The CuAAC reaction set up was characterized by the use of a Cu^{2+} source ($CuSO_4$) for catalysis, a reducing agent (sodium ascorbate), which was necessary to generate the active Cu^{+} center [245-247], and a coordinating ligand (Tris(3-hydroxypropyltriazolylmethyl)amine, THPTA) [248, 249] to accelerate the reaction (Appendix Figure A13). Two different types of beads, photo-cleavable azide agarose (PCAB) and disulfide azide agarose beads (DSAB), were evaluated (Appendix Figure A14, see also chapter 3.2.4.4.). The beads differ in their nature of releasing the cross-linked products. PCAB beads have a carbamate group which is cleaved upon UV irradiation, while DSAB beads possess a disulfide bond cleavable under reducing conditions. After the CuAAC reaction, the cross-linking reaction mixture was subjected to LC-MS analysis to profile all TP1 cross-linked species. If a cross-linked product was detected the precursor ion was subjected to sequential CID- and HCD-MS/MS (see chapter 3.2.5.4). All potential cross-linked products were found to be clicked successfully and were visible at different charge states. As expected, 1,2,3-triazole CuAAC-derivatized PAC4 reporter ions were found in the fragment ion spectra (Appendix Figure A15); two exemplary spectra are shown in Figure 4.21. More interestingly, the same reporter ions for clicked type 0 and type II TP1 cross-links appeared again, even with more pronounced intensity differences between the piperidine-acylium ion and the piperidine-carboxylic acid ion in MS/MS spectra of “dead-end” cross-links (Figures A16-A17). Type II TP1 cross-links only showed the piperidine-acylium ion, but no piperidine-carboxylic acid ion.

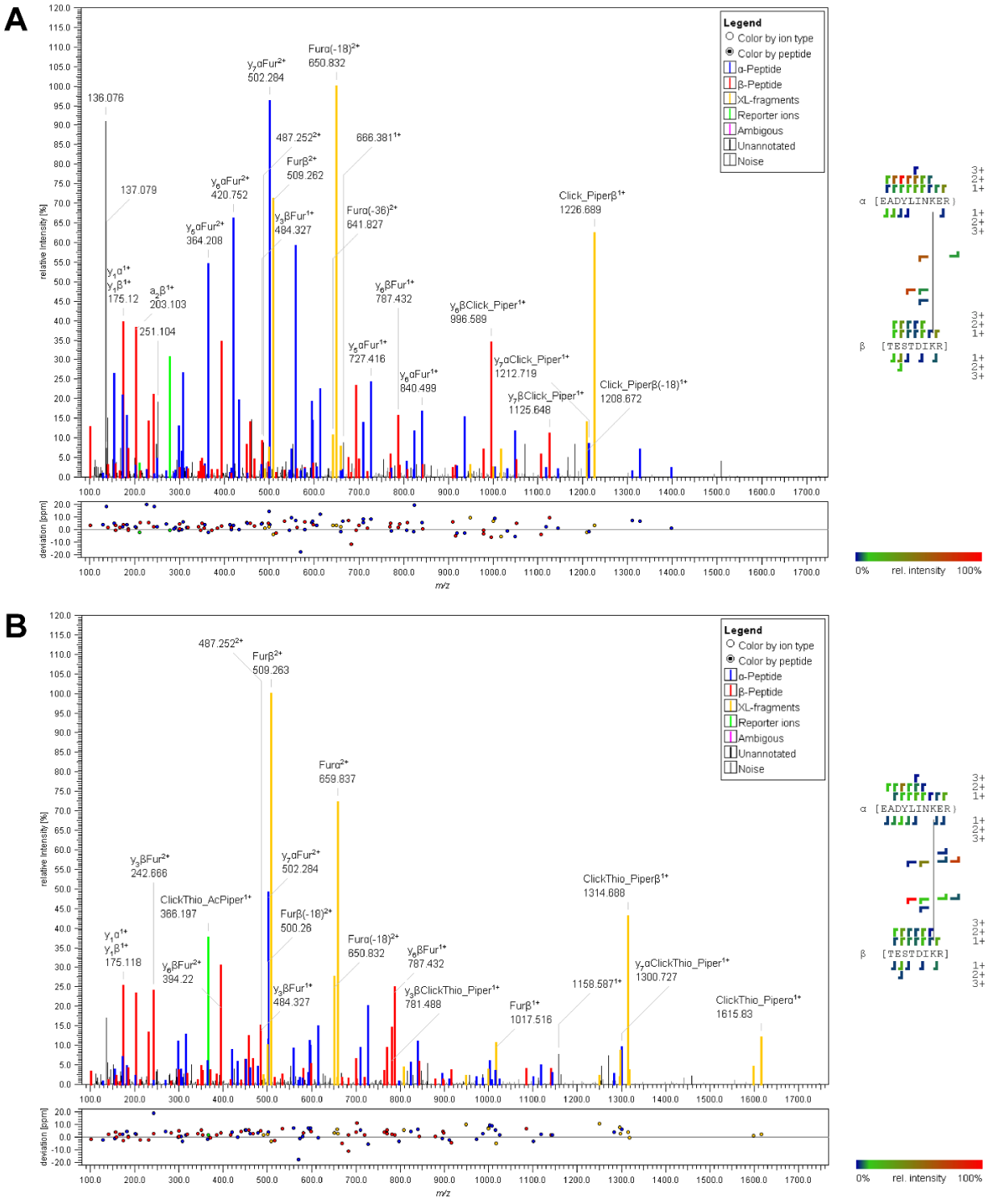


Figure 4.20: (A) Fragment ion spectra of the 4+ charged type II TP1 cross-linked peptide clicked with PCAB beads (signal at *m/z* 636.841) and (B) the 3+ charged type II TP1 cross-linked peptide clicked with DSAB beads (signal at *m/z* 878.119), HCD NCE=35%, analyzed with MeroX [8, 213].

Fragment ion spectra for the clicked TP1 type II cross-links. PAC4 diagnostic doublets ions are shown in yellow, b- and y-type ions series for the α (EADYLINKER) and β (TESTDIKR) peptides of TP1 are presented in blue and red. Reporter ions characteristic for PAC4 (at *m/z* 278.198 and 366.196) are indicated in green. The ion at *m/z* 136.076 is the iminium ion of tyrosine.

4.2.4.2. CuAAC enrichment optimization with cross-linked BSA

After successful CuAAC reaction of TP1, extensive optimization of the CuAAC protocol for cross-link peptides enrichment was performed. In detail, several empirical parameters were sequentially varied to evaluate both CuAAC reaction efficiency and enrichment of cross-links (Table 4.3).

Table 4.3: Summary of all tested conditions for CuAAC enrichment optimization experiments.

Condition	Rationale
Copper salts [$CuSO_4$, $Cu(CH_3CN)_4BF_4$]	Evaluate if a Cu^{2+} source or a direct active Cu^{+} salt is more beneficial for the CuAAC reaction
Reaction time	Time-course evaluation of CuAAC reaction
Temperature	Increased temperature reduces Cu^{+} sequestration from coordinating species, e.g. thiols, imidazoles in amino acids [250]
Pseudo-ligandless CuAAC	Use of 0.6% (v/v) acetonitrile without ligand-stabilized Cu^{+} active complexes [251]
Degassing	Higher CuAAC yield is obtained when the reaction is performed under inert gas conditions [251]
Beads	Evaluate both PCAB/DSAB beads with different cleavage principles for cross-links enrichment
Organic solvent	Use three-component mixtures to evaluate CuAAC performance [249]: 10%H_2O/10%ACN/70%DMSO, 10%H_2O/10%ACN/70%Acetone, 10%H_2O/10%ACN/70%MeOH
Sonication	Alternative beads resuspension method

Screening these conditions resulted in several observations:

- A direct Cu^{+} proved to be slightly more efficient than using a Cu^{2+} source;
- Two hours reaction time seemed sufficient for an efficient conversion;
- Increased temperature proved better for reproducibility;

- Using acetonitrile alone as stabilizing ligand is not beneficial for CuAAC cross-linked peptide conversion as this condition has been optimized for nucleic acids;
- PCAB beads showed poor robustness as clicked TP1 products were found in the supernatant. This means the photo-cleavable bond on the beads is broken prior to UV irradiation. On the contrary, DSAB beads showed high stability and high release of cross-linked products under reducing conditions;
- Degassing did not enhance the conversion efficiency of cross-links in the CuAAC reaction;
- Three-component mixtures using methanol as the primary co-solvent appeared to be effective in enriching a higher quantity of cross-links, while DMSO and acetone proved less efficient;

In terms of unique cross-links, the enrichment protocol using HEPES proved to be most efficient, enriching a higher amount of unique cross-links (70 vs 38 without enrichment, Figure 4.21 A). Using MeOH as co-solvent proved beneficial, while sonication seemed to negatively affect the recovery of cross-linked products. Regarding type 0 cross-links, the enrichment seemed to capture a higher amount of 'dead end' cross-links, with the highest number obtained for the HEPES/sonication condition. Concerning the number of cross-linked spectral matches (CSMs), using HEPES yielded the highest number of CSMs. This was 3.4-fold higher than without enrichment (581 vs 173) (Figure 4.21 B left). In respect to the type 0 PSMs, again the HEPES condition gave the highest number (2-fold higher than without enrichment). Type 0 cross-linked products as the most abundant cross-link species are generally more frequently identified than type I and type II cross-link products (Figure 4.21 B right). This was confirmed for the enriched samples, although to a lower extent. This proved a successful enrichment of cross-links as well as depletion of unmodified peptides, as reflected by a higher average number of CSMs compared to the condition without enrichment. The HEPES condition retrieved the highest number of unique cross-links with almost 26% of cross-links shared between different conditions (Figure 4.21 C). One homeotypic cross-link was common in the MeOH condition and without enrichment, while for the five unique cross-links found without enrichment the cross-link site was ambiguous. All cross-links observed were mapped onto the BSA crystal structure (pdb entry: 4f5s) (Figure 4.21 D). As no $^{14}N/^{15}N$ labeling was performed and no in-gel digestion was carried out, cross-links may arise from monomeric or dimeric BSA, justifying the over-length distances. The average as well as median Cα-Cα distances for all different conditions is below 30 Å, which is a reasonable Euclidean distance that can be bridged by the cross-linker. Conclusively, DSAB beads, THPTA, $Cu(CH_3CN)_4BF4$ and sodium ascorbate at 40 °C reaction temperature with HEPES, and methanol as major organic co-solvent (70% (v/v)) were established as final conditions for the CuAAC reaction.

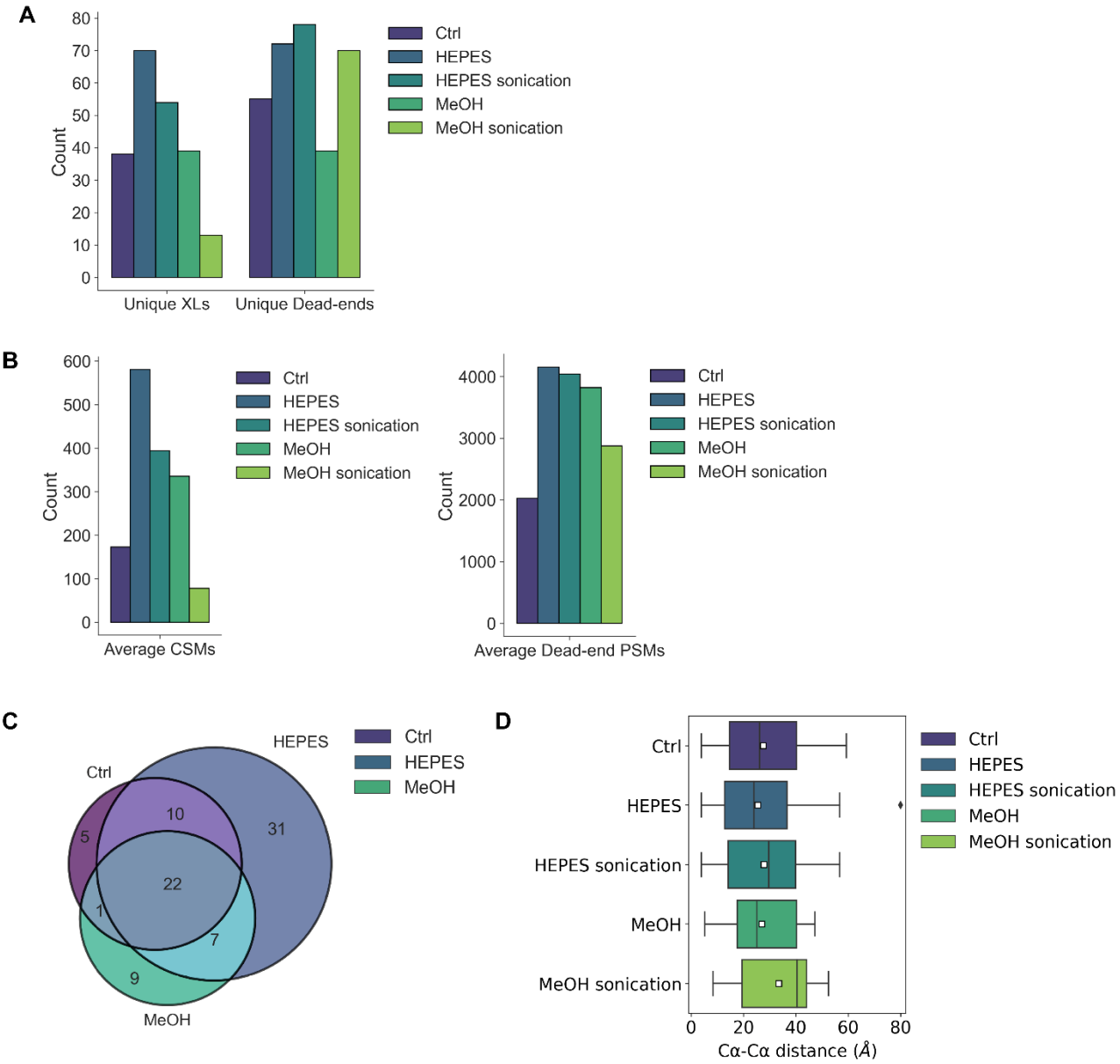


Figure 4.21: CuAAC enrichment optimization of PAC4 cross-links with BSA.

A) Comparison of unique cross-links/dead-ends before (Ctrl) and after enrichment using DSAB beads, THPTA, $Cu(CH_3CN)_4BF_4$ and sodium ascorbate at 40 °C with HEPES or MeOH as major organic co-solvent (70% (v/v)) and with two different methods for beads resuspension. Samples were analyzed on a timsTOF Pro instrument. B) Average count of CSMs and dead-end PSMs. C) Venn diagram of unique cross-links obtained in the non-enriched sample (Ctrl) and in the enriched samples using HEPES or MeOH as major organic co-solvent. D) Box plot analysis for the mapping of identified cross-linked peptides on the crystal structure of BSA (pdb entry: 4f5s). Median is shown as a vertical line, mean is depicted as white square.

4.3. Development of a trifunctional CID-MS/MS cleavable, IMAC-enrichable cross-linker: DSSI

4.3.1. Description of the cross-linker features

Disuccinimidyl dibutyric urea (DSBU) is an MS-cleavable cross-linker, finding multiple applications in structural biology, ranging from isolated protein complexes to comprehensive system-wide interactomics [8, 128, 146, 148]. DSBU facilitates a rapid and reliable identification of cross-links through the dissociation of its urea moiety in the gas phase. However, the urea moiety in urea-based cross-linkers is unsuitable for accommodating an affinity tag without disrupting the molecule's symmetry. To overcome this limitation, the central urea group was replaced with an imide. (Figure 4.22).

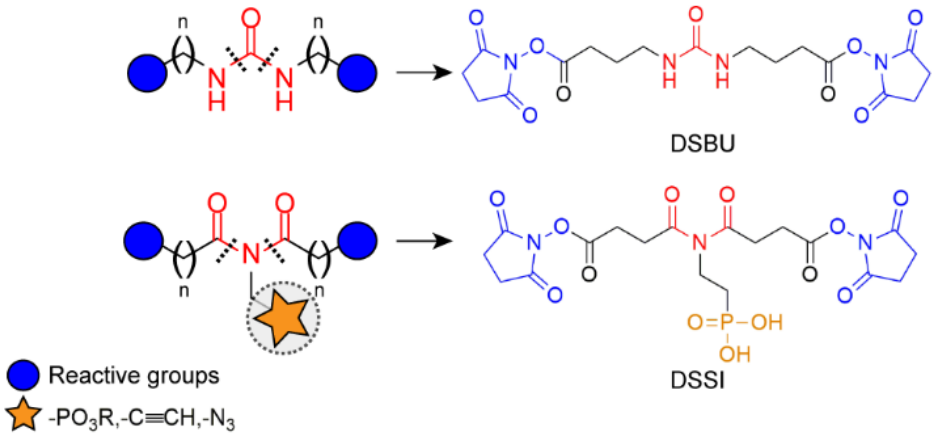


Figure 4.22: MS-cleavable urea- and imide-based cross-linkers.

Chemical structures of the DSBU and DSSI cross-linkers. The cross-linkers are characterized by a central urea/imide CID-MS/MS cleavable function (shown in red), two NHS esters as reactive groups (in blue). Additionally, DSSI is featured with an ethyl phosphonate tag (in orange) to enrich cross-linked peptides via IMAC chromatography.

This enables designing cross-linkers with a central trivalent nitrogen atom for attaching affinity handles. At the same time imide-based cross-linkers maintain the MS-cleavability and the spacer length modularity of their urea-based ancestors. The second trifunctional CID-MS/MS cleavable cross-linker characterized in the current work is termed disuccinimidyl disuccinicimide (DSSI). DSSI, the first imide-based cross-linker prototype, possesses the two disuccinimidyl C4 arms of DSBU alongside a central ethylene phosphonic acid as affinity tag and two NHS esters as reactive groups. As for the PAC4 cross-linker, details regarding the gas-phase behavior of the DSSI cross-linker were obtained after cross-linking with the TP1 model peptide.

4.3.2. Cross-linking experiments with Test Peptide 1

DSBU cross-linked products are cleaved upon collisional activation during MS/MS experiments, resulting in the cleavage of the HN-CO bonds of the urea moiety. The dissociation generates an amine and an isocyanate fragment per peptide visible in fragment ion mass spectra as two doublets with a characteristic mass shift of ~26 u [8, 128]. Gas phase dissociation of linear imides has not been systematically explored before. Nelson and McCloskey reported that, upon collisional activation, uracil and its derivatives [252] undergo ring-opening via CO-N bond cleavage yielding an amide and a protonated acylium ion. DSSI should hypothetically follow a similar fragmentation pattern, resulting in an MS-cleavable cross-linker (Figure 4.23). TP1 was cross-linked with DSSI to examine its fragmentation behavior during collisional activation (Appendix Figure A18). Similarly to cyclic imides, DSSI dissociates into ethylphosphonate amides and acylium ions (Figures 4.23, A19 A). Furthermore, one hydroxy group of the phosphonic acid undergoes an intramolecular rearrangement yielding the dehydrated form of the ethylphosphonate amide fragment and a carboxylic acid fragment for each peptide (Figures 4.23, A19 B). Thus, DSSI cross-links generate two diagnostic doublets of signals in fragment ion mass spectra, spaced by ~89 and ~125 u. Neutral losses of water molecules from the precursor ion were observed at low collision energy (Figure A18 A) and from the ethylphosphonate amide and the carboxylic acid fragments at higher collision energy (Figures A18 B-D). MS/MS experiments of DSSI cross-links revealed that the optimal collision energy range for linker and backbone fragmentation was between 25 and 35% HCD NCE, similar to that of DSBU (Appendix Figure A18). The comparable stabilities of the amide bonds in the peptide backbone and of the imide group of DSSI enable XL-MS analysis already at the MS/MS level. This has been shown to increase scan rate and sensitivity compared to multistage MS (MS^n) methods [241]. Additionally, DSSI generates characteristic reporter ions in the low *m/z* range, further enhancing confidence in cross-link identification (Appendix Figure A20).

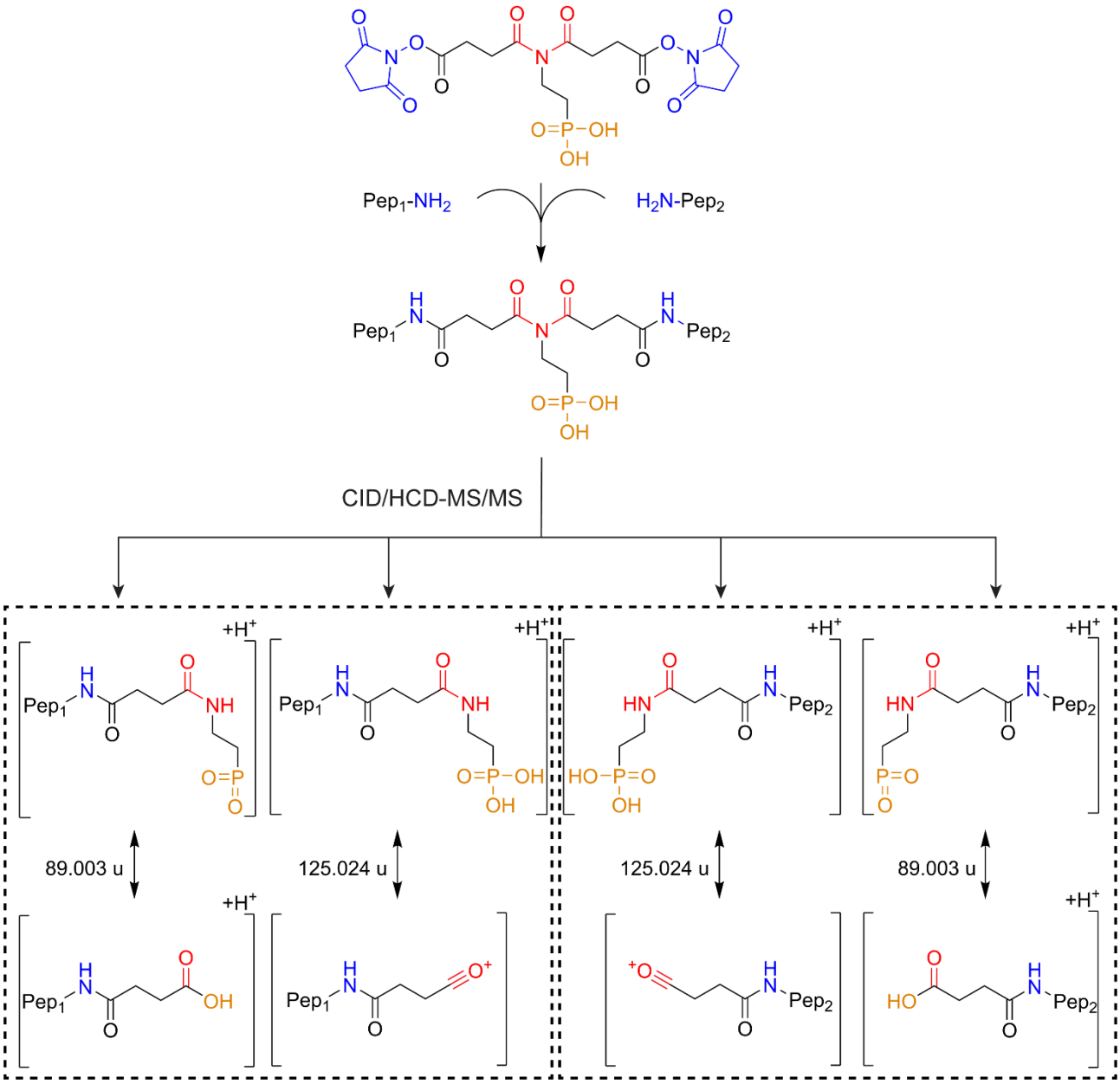


Figure 4.23: CID/HCD-MS/MS dissociation behavior of the DSSI cross-linker.

Upon collisional activation, the central imide group of DSSI dissociates, generating two sets of fragment ions for each cross-linked peptide. The doublets of signals in the fragment ion mass spectrum possess mass differences of 89 and 125 u respectively.

4.3.3. Cross-linking experiments with α-synuclein

The detailed characterization of DSSI gas phase behavior enabled an automated cross-link identification in complex samples by using the MeroX software [213]. α-synuclein (α-syn) was selected for assessing the reactivity of DSSI at physiological pH and to establish the downstream sample preparation workflow. α-syn is a small IDP relevant in Parkinson's disease [253]. It consists

of 140 amino acids and three domains, namely the N-terminal lipid-binding domain, the central non-amyloid component (NAC) domain and the C-terminal acidic domain [253, 254]. α-syn was cross-linked with DSSI for 1 hour at room temperature, digested for two hours by two consecutive additions of trypsin (fast in-solution digestion, fast ISD) [255], and the resulting peptides were analyzed by nano-HPLC-MS/MS. A MeroX search identified 77 unique residue pairs with ~84% overlap across three replicates (Figure 4.24 A). These were located within the N-terminal domain as well as between the N- and C-terminal regions of α-syn as it has been observed with DSBU in a previous study (Figure 4.24 B) [254]. This suggests a comparable reactivity of the two cross-linkers. Interestingly, a high number of DSSI cross-links was obtained using the fast ISD protocol for digestion. Thus, the fast ISD method was adopted for further DSSI XL-MS experiments.

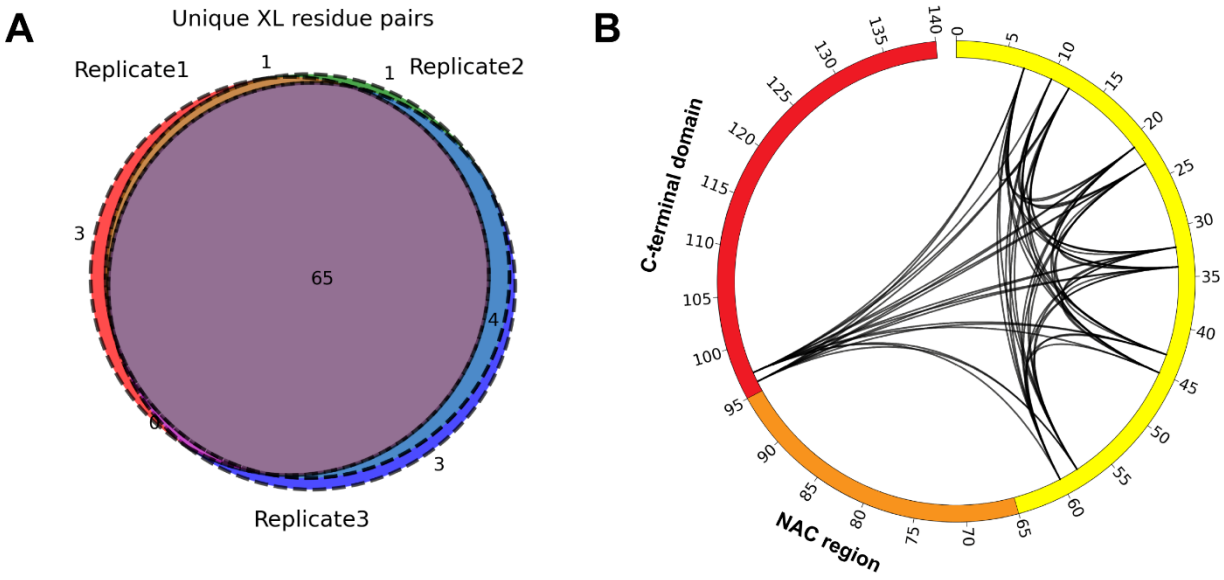


Figure 4.24: Application of DSSI for the study of the IDP α-synuclein.

A) Reproducibility of cross-linked residue pairs. B) Circos plot showing DSSI cross-links identified for α-synuclein.

4.3.4. Application of DSSI to HEK293T cell lysates

Before applying the protocol to more complex samples, the fast ISD approach was further evaluated for HEK293T cell lysates and compared to an overnight enzymatic digestion. Both digested samples were analyzed by reverse phase nano-HPLC followed by MS analysis on a timsTOF Pro instrument (Appendix Figure A21). The fast ISD workflow allowed the identification of 5,363 proteins with 81% reproducibility over three replicates. That was only 6% less than those observed at overnight ISD

while maintaining ~80% reproducibility across three replicates (Appendix Figure A21 A). Also, the number of identified peptides was comparable between the two digestion workflows (Figures A21 B, C). Therefore, the fast ISD was applied to DSSI cross-linked samples to maximize the analysis throughput. Cross-linking of HEK293T cell lysates was performed at a total protein concentration of ~1 g/L and 2 mM DSSI. The resulting cross-links were compared to those observed when using DSBU and the overnight ISD approach. The two protocols were compared at the same cross-link enrichment depth. Digested samples were fractionated by SEC to enrich for cross-linked peptides (Figure 4.25 A). SEC fractions A12 to B9 of each replicate were collected and individually subjected to LC-MS/MS analysis. The MeroX analysis allowed the identification of 2,478 unique DSSI cross-links related to 1,241 proteins at 1% FDR determined for CSMs. Cross-link peptides were non-uniformly distributed in the SEC fractions (Figure 4.25 B), ranging from 147 cross-links in fraction B9 to 1,440 in fraction B2. DSSI cross-links provided structural information on 582 protein−protein interactions (PPI) in HEK293T cells. DSSI cross-linked proteins were compared to those targeted by DSBU. For this, proteins cross-linked by DSSI and DSBU were analyzed with the clusteProfiler R package [256] and classified based on Gene Ontology (GO) [257]. Cross-linked proteins were enriched in the same cellular compartments within HEK293T cells, which is in agreement with the similar physico-chemical properties of the two cross-linkers (Figures A22 A, B). In the case of molecular functions, proteins involved in chromatin, tubulin and mRNA/rRNA binding were preferentially cross-linked by DSSI, while structural constituents of the cytoskeleton, proteins involved in ribonucleoprotein complex binding, NAD binding and RNA helicases were preferentially cross-linked by DSBU (Appendix Figures A22 C, D). By combining DSSI with DSBU, the cross-linked proteome and the cross-linking sites increase by 18% and 19% compared to DSBU alone (Appendix Figure A23), enhancing the depth of structural data and the coverage of the PPI network in cells.

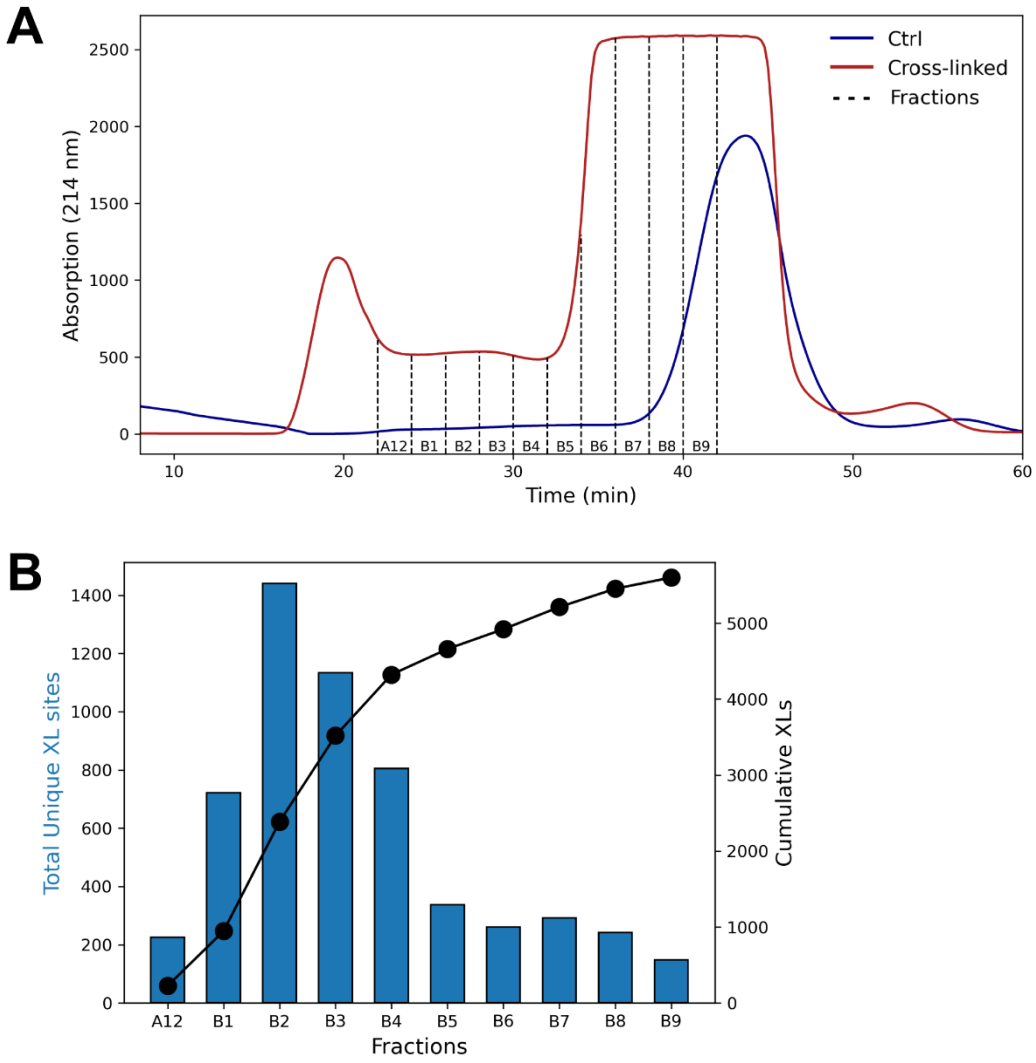


Figure 4.25: Proteome-wide XL-MS application of DSSI on HEK293T cells.

A) Exemplary peptide SEC fractionation of a control (ctrl, blue trace) and cross-linked (red trace) HEK293T lysate. Absorption at 214 nm was monitored over the time course of the run. B) DSSI total unique cross-linking sites identified from three biological replicates contained in the selected SEC fractions.

4.3.5. Structure-based validation of DSSI cross-links

DSSI has a similar spacer length like DSBU and therefore is expected to covalently bridge residues with a maximum Cα-Cα distance of 35 Å [233]. A structural validation of DSSI cross-links was performed by mapping them into the 3D structures of three large protein complexes, namely the chaperonin containing TCP-1 (CCT) complex, the 26S proteasome and the 80S ribosome (Figure 4.26 A) and comparing them with DSBU cross-links. The median cross-link distances ranged between 12 and 15 Å for both cross-linkers. 99% of DSSI cross-links were compatible with known high-resolution structures of the three protein assemblies (Figure 4.26 B).

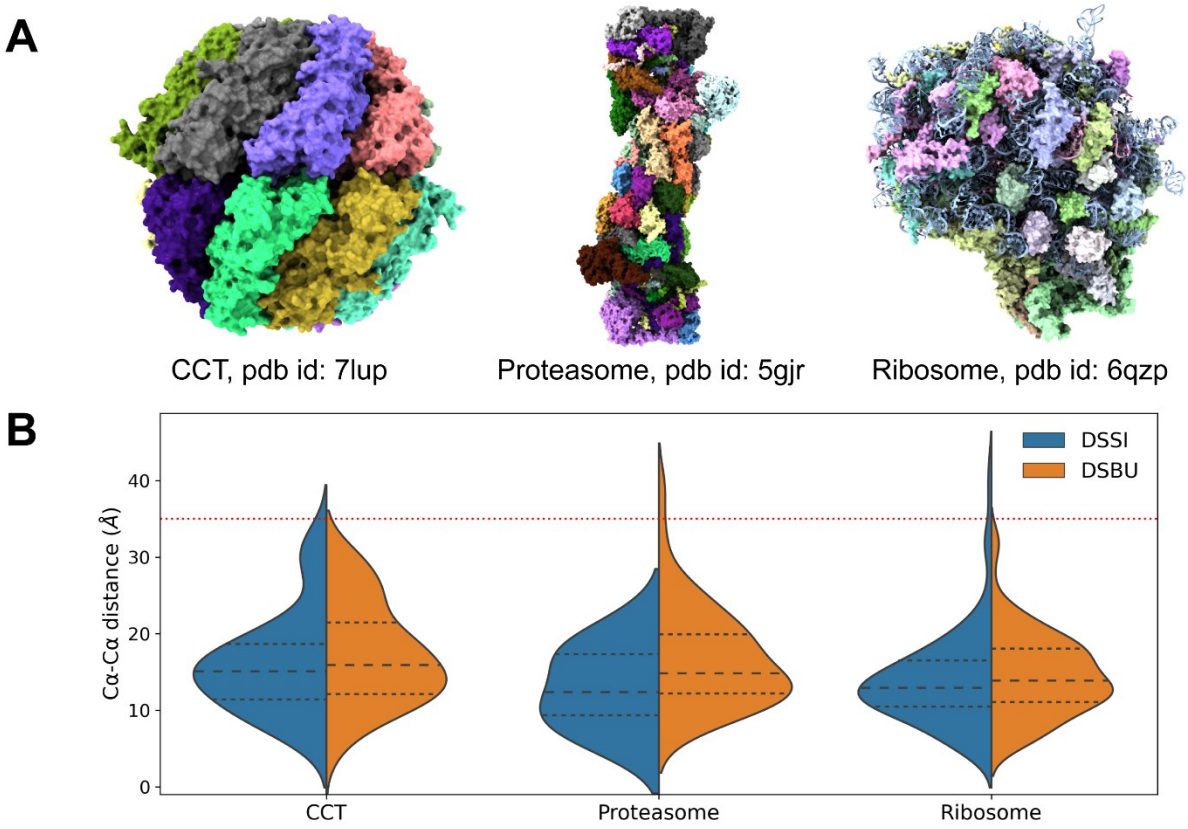


Figure 4.26: Structural XL-MS validation of DSSI cross-links and comparison with DSBU.

A) Selected protein complexes: chaperonin containing TCP-1 (CCT), ~950 kDa; 26S proteasome, ~2.5 MDa; 80S Ribosome, ~4.5 MDa. B) Violin plots comparing the Cα-Cα distance distributions of DSSI (blue) and DSBU (orange) cross-links. Dashed and dotted lines represent the median and the interquartile range.

4.3.6. Proteome-wide cross-linking of intrinsically disordered proteins

Several biologically relevant PPIs have been identified with DSSI, covering nuclear, mitochondrial, and cytoplasmic proteins. Interestingly, cross-links were also detected in proteins containing IDRs. Up to half of all human proteins contain structurally disordered regions, which present challenges for visualization using high-resolution methods [258]. The characterization of their highly dynamic conformational ensembles requires integrative approaches, combining different methods, including XL-MS experiments. The key role of XL-MS in capturing IDPs conformational ensembles is confirmed by the experiments on HEK293T cell lysates, where intra- and inter-protein cross-links involving 295 proteins listed in the DisProt database were identified [21-23]. Specifically, 102 of them display disordered content ranging from 20% to 100% (Appendix Table A1) including:

i) transcription factors/repressors and their modulators (NF-κB, YY1, TP53BP1, CREBBP);

ii) histones and their associated proteins (H4C1, H3-3A, H2BC11, HDAC1, EP300);

iii) proteins involved in RNA binding/processing and splicing factors (SNRPA-B, SRRM1-2, SRSF1, U2AF1);

iv) ribosomal proteins/factors (SERBP1, RPLP2, RPL4, RPL24, ElF4B, ElF4EBP1).

Among those, the interactome of the IDP serpine 1 mRNA-binding protein 1 (SERBP1, also known as plasminogen activator inhibitor 1 RNA-binding protein) was investigated. SERBP1 is a ~45 kDa RNA-binding protein involved in mRNA maturation, translational regulation, and other biological functions [259, 260]. It was initially defined as a hyaluronic acid binding protein, but later associated with crucial roles in brain function/development and also in cancer. Despite this, very little is known about its structural and dynamic features, as the protein is for a large part intrinsically disordered. SERBP1 recognizes G-rich RNA sequences by means of the two distinct RG/RGG repeats and neighboring residues and it does not possess any canonical RNA binding motifs, such as RNA-recognition motifs, zinc-finger motifs or K-homology (KH) domains. SERBP1-ribosome binding promotes ribosome hibernation. SERBP1 is an IDP (see Figure 4.27 A according to IUPRED and DISPRED disorder predictions [261, 262]), with a short α-helix comprising residues 290 to 300, which was also confirmed by NMR [259]. The AlphaFold [29] model for SERBP1 (Figure 4.27 B) displays a disordered, fully-extended structure with most parts having a low or very low pLDDT score. Two α-helix motifs (e.g., residues 27 to 42 and 290 to 300) are predicted with high confidence, while for other regions the prediction is ambiguous. As IDP flexibility does not hinder their chemical reactivity, 10 unique SERBP1 cross-links were identified, recapitulating the interaction of all its regions with 7 neighboring proteins (Figure 4.27 C). Initially, the cross-link between K299 of SERBP1 and K116 of RPS12 was mapped. This cross-link aligns well with the cryo-EM structure of the 80S ribosome [263], with a Cα-Cα distance of 10 Å, confirming the localization of SERBP1's sole visualized α-helix (pdb entry: 6z6m, Figure 4.28 D). On top of that, additional interactions were identified:

i) N-terminal region of SERBP1 interacting with RPS3A, RPS9, RPS14 and RPS28;

ii) the C-terminal domain interacting with RPS15 and RACK1, a scaffold protein involved in SERBP1 recruitment [264].

These PPIs possess a high STRING score and some of them have been identified in previous studies [194, 265]. SERBP1 can therefore occupy a broad surface of the 40S subunit, sampling a large conformational ensemble at the mRNA entry channel, as well as the A and P sites of the ribosome, thereby inhibiting translation. (Figure 4.28 E). The same was observed for SERBP1 homologues [263, 266, 267], suggesting that disorder plays a pivotal role for the regulatory function of these hibernation factors.

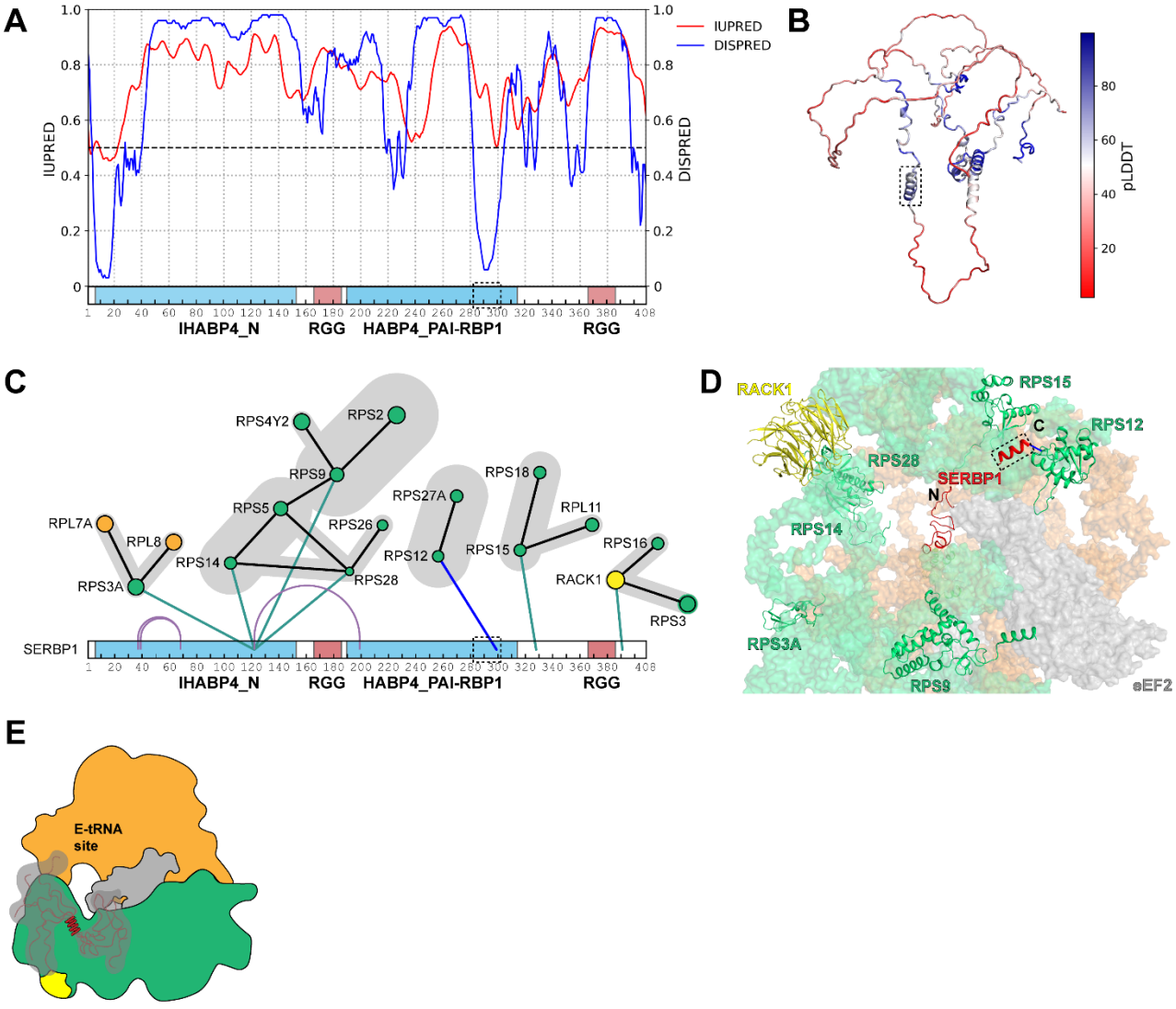


Figure 4.27: Evaluation of PPIs involving the IDP SERBP1.

A) IUPRED and DISOPRED disorder predictions of SERBP1, confirming its high disordered nature. The horizontal bar is displaying the SERBP1 sequence number and domain annotation. B) AlphaFold2 model of SERBP1, with the experimentally confirmed α-helix (dashed rectangle) spanning residues 290-300. C) SERBP1 inter-protein cross-links to several ribosomal proteins of the 40S subunit. The cross-link involving SERBP1 solved helix (dashed rectangle) with RPS12 is colored in blue. D) Bottom view of the cryo-EM structure of the 80S ribosome (pdb entry: 6z6m), showing the only mappable cross-link between SERBP1

helix and RPS12. All other cross-links involved unsolved regions in either SERBP1 or ribosomal proteins. Proteins of the 40S subunit are depicted in green, proteins of the 60S subunit in orange, RACK1 in yellow, eEF2 and other associated factors in gray. RNA is not shown for clarity. E) Schematic cartoon representation of the dynamic nature of the interaction between SERBP1 and ribosome. The high disorder content allows SERBP1 to sample a vast ensemble of conformations; SERBP1 binds to the mRNA entry channel and the A and P sites of the small 40S subunit (where it interacts with eEF2, in gray), inhibiting translation.

5 Discussion

5.1 Expression and characterization of p53 by structural MS

The tumor suppressor protein p53 is an IDP. Its disordered character regulates the rapid transition between p53's active and inactive states *in vivo* along with its short half-life in cells [51]. This poses problems for the recombinant production of the protein in organisms, such as *E. coli*. The stability of the full-length p53 is defined by the stability of the central DBD, which is 25 kJ/mol at 25 °C and 41 kJ/mol at 10 °C [76, 77]. Additionally, full-length p53 possesses a strong tendency to aggregate. Therefore, the expression of p53 was performed as HLT fusion protein [7, 120, 232]. The HLT tag is composed of an octa-histidine tag, followed by the first lipoyl domain of the dihydrolipoyl transacetylase of the pyruvate dehydrogenase complex from *B. stearothermophilus* and a TEV cleavage site (see chapter 4.1). This tag has already been successfully applied for the purification of p53 [7, 120, 232] and has proven suitable for the expression of proteins that tend to aggregate. A major finding of this work is the compaction of the whole NLS-TD-CTD arrangement in tetrameric full-length, wild-type human p53, which was initially indicated by the experimental data [7, 120, 232] and is also mirrored in the newly generated models [233]. It has to be noted that cross-links will capture conformational states depending on the cross-linker's spacer length and its reactivity. It cannot be ruled out that only a fraction of conformations is captured, while others will remain elusive. To extend the coverage of the entire conformational space, additional cross-linking principles with other chemistries and spacer lengths should be employed. The first published structural model by Okorokov et al. [95] was generated by cryo-EM utilizing murine p53 (Figure 1.3 B), while all other studies were performed with human p53. The first human-based, tetrameric p53 structural model derived by the Fersht group [97] differed dramatically in the DNA-free (extended cross-shape, Figure 1.3 A) and DNA-bound form (compact basket shape, Figure 1.3 C). The DNA-bound tetrameric p53 model was further refined by a follow-up study utilizing negative stain EM [99]. Here, different conformational ensembles (classes) were observed, which already gave hints of inter-domain proximities of CTD and NLS, as well as DBD and the DNA itself. Single-molecule FRET (smFRET) experiments from the same study also showed a spatial proximity between the very C-terminus of p53 and the L1-loop (residues 112–124) within the DBD in presence of RE-DNA. However, in the absence of DNA, an equilibrium might exist between the formerly published, extended cross-shaped p53 tetramer and a more compact arrangement as observed for the DNA-bound p53 tetramer. The existence of an equilibrium state is controversially discussed due to the limited distance information of smFRET data as only one labeled construct was used. Also, a spatial rearrangement of the flexible

CTD was not discussed, but only a complete rearrangement of the whole p53 tetramer. In contrast, the present data show that in both cases, DNA-free and DNA-bound states, a more compact p53 structure is present. Another publication agreeing with the findings of a more compact p53 tetramer, both in the DNA-bound and DNA-free states, is based on ion mobility coupled to MS [236]. Here, experimentally derived collisional cross sections for DNA-free, full-length p53 and DNA-bound p53 are smaller than the values calculated for the extended model structures [97, 236]. Also, collisional cross sections were found to be more similar between both states (with and without DNA) than expected from the models. At that time, the observed compaction was described as a gas phase collapse, which can be excluded in this case as XL-MS, footprinting, and HDX-MS data were collected in solution at near physiological pH conditions. In the most recent MD study of a reconstructed p53 tetramer in complex with RE-DNA, a compaction and even a contact of the CTD to the bound DNA molecule is described, supporting the findings of the present work (Figure 1.3 E) [98]. Together, all above mentioned p53 structural studies show similar results as the MS data and agree with the new p53 models.

Initially, the CTD was characterized as negative regulator impairing specific DNA binding [268-270]. Later, it was recognized to be a positive regulator being responsible for various functions of the p53 tetramer related to DNA binding [271, 272]. The CTD has been described to facilitate p53 binding to long non-specific DNA molecules as well as to chromatin [273], enabling lateral movement along the DNA via electrostatic interactions. This however does not seem to be entirely true as the CTD might be involved in specific DNA binding as well [274]. A two-mode binding model switching between a sliding- and recognition-mode was introduced, termed S- and R-modes [92, 275]. This two-mode DNA-binding model suggests a hemi-specific binding, with one dimer of tetrameric p53 binding to DNA in a site-specific manner, while the other p53 dimer binds non-specifically to non-cognate DNA transitioning into a fully specific, DNA-bound state upon response element recognition. Reflecting this work's data in the light of that model (S- and R-modes), a snapshot for specific DNA binding could be obtained. The conformational ensemble for one specific 26-bp RE-DNA bound to p53 could be captured, which might reflect the recognition mode (R-mode). A comparison of that two-mode DNA-binding model proposed by Tafvizi et al. [92] with the EM-based model of Melero et al. [99] revealed the following: in the latter model, EM maps were grouped into classes that depend not only on the sequence of the RE-DNA, but also on the length of the individual DNA fragment 18-mer vs. 44-mer vs. 60-mer. Also, the number of p53 conformational states observed differed depending on the nature of the bound DNA-fragment. In this work, an intermediate length (26-mer) RE-DNA was used with high specificity and affinity towards the DBD [236, 276]. This is a strong

hint that one of the specific DNA-binding modes was captured, in which the CTD of p53 might be involved to stabilize the R-mode. The findings are also in line with the reaction kinetics of the cross-linking and labeling reagents used herein that capture dynamics on the minutes to seconds time scale. There are however still discrepancies whether CTD and DBD are both binding to DNA, interdependently of each other, or if the CTD controls site-specific binding of the DBD to the DNA by inducing conformational changes within the DBD itself [274, 277]. If there were conformational changes upon DNA-binding on the sub-second time scale, like a channeling of the DNA from the CTD to the DBD, faster reaction kinetics available with photo cross-linkers (diazirines, incorporated photo-amino acids) would be needed [137, 205]. To capture the conformational dynamics of the highly complex binding events between full-length p53 and DNA, integrative approaches are inevitable. One technique alone will not be able to catch the full picture, but structural MS combined with computational modeling allows getting insights into the high complexity of p53's conformational ensembles. Finally, follow-up studies with recombinantly expressed PTM-modified full-length p53 variants (acetylated, phosphorylated, ubiquitinilated) and with different RE-DNA lengths as well as cellular studies will possibly shed light on the role that PTMs play in regulating the structural/functional properties of the protein.

5.2 Application of a CID-MS/MS cleavable, CuAAC-enrichable cross-linker: PAC4

PAC4, an MS-cleavable and CuAAC-enrichable cross-linker [138], was characterized in detail regarding its potential for proteome-wide XL-MS studies. Its CID/HCD gas phase chemistry was elucidated in detail using TP1 and ATII as model peptides (see chapters 4.2.2.1, 4.2.2.2). Besides the fragmentation of the labile central piperidinium group, PAC4 was found to form highly specific reporter ions. Interestingly, the intensities of these reporter ions were different depending on the nature of the cross-linked products. Type II cross-links exhibited higher intensity for the piperidine-acylium ion, while type 0 cross-links showed a prominence of the piperidine-carboxylic acid ion (Figure 4.18). This paves the way to develop future MS methods that target only the low intense and more interesting type II cross-links, while excluding type 0 cross-links, which the mass spectrometer usually spends most of its cycle time fragmenting. The information on MS-cleavability of the cross-linker as well as the generation of its reporter ions was introduced into the MeroX software, allowing an automated analysis of cross-links. With the aim of expanding the application of PAC4 to protein XL-MS studies, the PAC4 cross-linker was applied to BSA and CA2 as model proteins. Native MS

analysis revealed the cross-linker was able to modify both proteins at usual molar excesses applied in XL-MS experiments (Figures 4.19, Appendix Figure A12).

After probing PAC4's reactivity in solution, optimized conditions for a future proteome-wide XL-MS workflow had to be evaluated, which relies on CuAAC for cross-link enrichment. For this purpose, TP1 was first cross-linked and then reacted with two types of beads with different cleavable groups for cross-links release. PCAB beads possess a photocleavable carbamate moiety that can be cleaved using UV light, while DSAB beads contain a reducible disulfide bond for reductive release of the cross-linked peptides/proteins from the beads. In both cases, successful conversion of TP1 was achieved and a dedicated MS method was used to define the best range of collision energies for the follow-up optimization experiments. CuAAC reaction and enrichment optimization were carried out using BSA as a model protein. Again, CuAAC-clicked cross-links showed formation of similar types of reporter ions with even higher intensity differences compared to non-clicked PAC4 cross-links (Figures A16-17). To note, the piperidine-carboxylic acid ion cannot be formed in a K-K cross-link (the most common type for NHS cross-linkers) as two amide bonds result from the covalent modification of K residues. In the case of a K-STY cross-link, an ester is formed after modification of a hydroxy-containing amino acid. The ester is cleaved by collisional activation resulting in the generation of a piperidine-carboxylic acid fragment ion. Some of these cases were found when analyzing cross-links in BSA. This further enables the identification of the correct cross-linked residue pair when multiple sites are available on peptides, which most of the current software still struggle to perform.

DSAB beads in combination with a direct Cu^{+} source, THPTA as ligand, sodium ascorbate as reducing agent in a buffered aqueous solution proved the best performing enrichment method for PAC4 clicked cross-links (see chapter 4.2.4.2). In comparison to the control sample where no enrichment was performed, a higher amount of unique cross-links as well as CSMs were detected. The same was observed for type 0 cross-links, which resulted to be highly enriched as they are usually the most abundant type of cross-linked product. This proof-of-principle study can be extended to more complex scenarios, such as cell lysates to further evaluate the CuAAC enrichment performance, alone or combined with other enrichment methods. Finally, a comparison between magnetic beads with the ones used in this work (PCAB and DSAB beads) could be performed to further increase the robustness of the workflow.

5.3 Development of a CID-MS/MS cleavable, IMAC-enrichable cross-linker: DSSI

DSSI, the prototype of the class of imide-based cross-linkers, was designed and applied for *in vitro* protein and proteome-wide structural studies [278]. This new class of cross-linkers can harbor a tag for affinity purification of cross-links. In particular, DSSI maintains the NHS-ester groups and the spacer length of the widely used DSBU, but possesses a phosphonate handle for IMAC or TiO_2-based cross-link enrichment. The gas-phase chemistry of DSSI was deciphered and its MS-cleavability parametrized for automated cross-link identification. Its diagnostic fragment ion doublets can be recognized by customized software tools, such as MeroX, for fast and reliable identification of cross-links in complex samples. The reactivity of DSSI was assessed by cross-linking α-syn and HEK293T cell lysates. A two-hour trypsin digestion protocol was employed to increase the analysis throughput. In total, 2,478 unique cross-links were identified, involving structural information on 1,241 proteins. DSSI targets proteins in a wide range of cellular compartments and expands the scope of XL-MS. DSSI cross-links were validated based on the high-resolution structures of three large protein complexes, the CCT complex, the 26S proteasome, and the 80S ribosome (see chapter 4.3.5). The Cα-Cα distance distribution of DSSI cross-links matches that of DSBU, according to their similar spacer lengths. Beyond characterizing PPIs in HEK293T cells, structural insights into IDPs were obtained, including SERBP1 and its complexes.

6 Outlook

Future work will focus on the development of proteome-wide XL-MS workflows. It can be envisioned to employ an IMAC enrichment of DSSI-cross-linked species, followed by SEC and/or strong cation-exchange (SCX) [279], hydrophilic strong anion exchange (hSAX) [280], and high-pH reverse phase chromatography. Proteome-wide XL-MS experiments with trifunctional cross-linkers aim at sampling whole proteomes of different cells in a shorter amount of time. Additionally, comparisons of different biological conditions can be performed to infer changes in cellular interactomes upon different stimuli, e.g. non- vs. drug-treated cells, normal vs. cancer cell lines [281]. The current bottleneck of DSSI - as well as the PhoX linker [10] - is their low permeability across cell membranes due to the presence of a dianionic species at physiological pH. This problem could be solved by converting the phosphonic acid into a tertiary di-tertbutyl ester to create a membrane-permeable reagent for performing cellular XL-MS studies. This has been successfully shown for the tBu-PhoX cross-linker [125]. The trifunctional MS-cleavable and enrichable DSSI and other imide-

based cross-linkers could potentially contribute in the future to unveil structural details of cell interactomes.

Literature references

1. Emil, F., *Einfluss der configuration auf die wirkung der enzyme.* Ber. Dtsch. Chem. Ges., 1894.
2. Uversky V.N., J.R. Gillespie, and A.L. Fink, *Why are "natively unfolded" proteins unstructured under physiologic conditions?* Proteins, 2000. 41(3): 415-27.
3. Uversky N.V., et al., *Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins and human diseases.* Chem. Rev., 2014 114(13):6844-79.
4. Joerger A. C. and A.R. Fersht, *Structural Biology of the Tumor Suppressor p53.* Annu. Rev. Biochem., 2008. 77:557-582.
5. Vousden K.H. and X. Lu, *Live or let die: the cell's response to p53.* Nat. Rev. Cancer, 2002. (8):594-604. 594-604. .
6. Piersimoni L., et al., *Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein–Protein Interactions─A Method for All Seasons.* Chem. Rev., 2021. 122(8), 7500–7531.
7. Arlt C., Ihling C. H., and A. Sinz, *Structure of full-length p53 tumor suppressor probed by chemical cross-linking and mass spectrometry.* Proteomics, 2015 15(16):2746-55.
8. Götze M., et al., *A Simple Cross-Linking/Mass Spectrometry Workflow for Studying System-wide Protein Interactions.* Anal. Chem., 2019 91(15):10236-10244.
9. Chavez J.D., et al., *Chemical cross-linking mass spectrometry analysis of protein conformations and supercomplexes in heart tissue.* Cell Syst. , 2018. 6(1): 136–141.
10. Steigenberger B., et al., *PhoX: An IMAC-Enrichable Cross-Linking Reagent.* ACS Cent. Sci. , 2019. 5(9), 1514–1522.
11. Stadlmeier M., et al., *A Click-Chemistry-Based Enrichable Crosslinker for Structural and Protein Interaction Analysis by Mass Spectrometry.* Chembiochem., 2020. 21(1-2):103-107.
12. Taniuchi, H. and C.B. Anfinsen, *An experimental approach to the study of the folding of staphylococcal nuclease.* J. Biol. Chem., 1969. 244(14): p. 3864-75.
13. Anfinsen, C.B., *Principles that Govern the Folding of Protein Chains.* Science, 1973. 181(4096):223-30.
14. Levinthal, C., *How to fold graciously.* In Mossbauer Spectroscopy in Biological Systems, Proceedings of a Meeting held at Allerton House, Monticello, Illinois.(Debrunner, P., Tsibris, J.C.M. & Münck, E., eds.), 1969. 22, University of Illinois Press, Urbana.
15. Karplus, M., *The Levinthal paradox: yesterday and today.* Folding & Design, 1997. 2:S69–S75.
16. Bryngelson J. and P. Wolynes, *Intermediates and Barrier Crossing in a Random Energy Model (with Applications to Protein Folding).* J. Phys. Chem., 1989. 93, 6902-6915.
17. Dill, K.A. and H.S. Chan, *From Levinthal to pathways to funnel.* Nat. Struct. Biol., 1997. 4(1):10-9.
18. P.E. Wright and H.J. Dyson, *Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm.* J. Mol. Biol., 1999. 293(2):3.

19. Dunker A.K., et al., *Intrinsically disordered protein.* J. Mol. Graph. Model., 2001. 19(1), 26-59.
20. Chouard, T., *Structural biology: Breaking the protein rules.* Nature, 2011. 471(7337):151-3.
21. Piovesan D., et al., *DisProt 7.0: a major update of the database of disordered proteins.* Nucleic Acids Research, 2017. 45(D1):D219-D227.
22. Hatos A., et al., *DisProt: intrinsic protein disorder annotation in 2020.* Nucleic Acids Research, 2020. 48(D1):D269-D276.
23. Quaglia F., et al., *DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation.* Nucleic Acids Research, 2022. 50(D1):D480-D487.
24. Piovesan D. and S. Tosatto, *Mobi 2.0: an improved method to define intrinsic disorder, mobility and linear binding regions in protein structures.* Bioinformatics, 2018. 34(1), 122–123.
25. Oates M.E., et al., *D^2P^2: database of disordered protein predictions.* Nucleic Acids Research, 2012. 41:D508-16.
26. Zhao B., et al., *DescribePROT: database of amino acid-level protein structure and function predictions.* Nucleic Acids Research 2020. 49(D1):D298-D308.
27. Garner E., et al., *Predicting Disordered Regions from Amino Acid Sequence: Common Themes Despite Differing Structural Characterization.* Genome Inform. Ser. Workshop, 1998. 9:201-213.
28. He Bo, et al., *Predicting intrinsic disorder in proteins: an overview.* Cell Research 2009. 19:929-949.
29. Jumper J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. 596(7873):583-589.
30. Evans R., et al., *Protein complex prediction with AlphaFold-Multimer.* BioRxiv, 2022.
31. Piersimoni L., et al., *Lighting up Nobel Prize-winning studies with protein intrinsic disorder.* Cellular and Molecular Life Sciences 2022. 79(8):449.
32. Zhenling P., et al., *Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life.* Cell. Mol. Life Sci., 2015. 72:137–151.
33. Bondos S.E., Dunker A.K., and V.N. Uversky, *On the roles of intrinsically disordered proteins and regions in cell communication and signalin.* Cell Commun. Signal., 2021. 19(1):88. .
34. Zeno W.F., et al., *Molecular Mechanisms of Membrane Curvature Sensing by a Disordered Protein.* J. Am. Chem. Soc. , 2019. 141, 10361−10371.
35. Dunker A.K., et al., *Intrinsic Disorder and Protein Function.* Biochemistry, 2002. 41(21):6573-82.
36. Iakoucheva L.M., et al., *The importance of intrinsic disorder for protein phosphorylation.* Nucleic Acids Research, 2004. 32(3):1037-1049.
37. van der Lee R., et al., *Classification of Intrinsically Disordered Regions and Proteins.* Chem. Rev., 2014. 114, 6589−6631.
38. Lane D.P. and L.V. Crawford, *T antigen is bound to a host protein in SV40-transformed cells.* Nature, 1979 278(5701):261-3.

39. Linzer D.I. and A.J. Levine, *Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells.* Cell, 1979. 17(1):43-52.
40. Smith A.E., Smith R., and E. Paucha, *Characterization of different tumor antigens present in cells transformed by simian virus 40.* Cell, 1979. 18(2):335-46.
41. Melero J.A., et al., *Identification of new polypeptide species (48-55K) immunoprecipitable by antiserum to purified large T antigen and present in SV40-infected and -transformed cells.* Virology, 1979. 93(2):466-80.
42. Werness B.A., Levine A.J., and P.M. Howley, *Association of human papillomavirus types 16 and 18 E6 proteins with p53.* Science, 1990. 248(4951):76-9.
43. Scheffner M., et al., *The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53.* Cell, 1990. 63(6):1129-36.
44. Sarnow P., et al., *Adenovirus E1b-58kd tumor antigen and SV40 large tumor antigen are physically associated with the same 54 kd cellular protein in transformed cells.* Cell, 1982 28(2):387-94.
45. Hupp T.R., et al., *Regulation of the specific DNA binding function of p53.* Cell, 1992. 71(5):875-86.
46. Farmer G., et al., *Wild-type p53 activates transcription in vitro.* Nature, 1992. 358(6381):83-6.
47. Flatt P.M., et al., *p53-dependent expression of PIG3 during proliferation, genotoxic stress, and reversible growth arrest.* Cancer Lett., 2000. 156(1):63-72.
48. Levine A.J., Hu W., and Z. Feng, *The P53 pathway: what questions remain to be explored?* Cell Death Differ., 2006. 13(6):1027-36.
49. Pitolli C., et al., *p53-Mediated Tumor Suppression: DNA-Damage Response and Alternative Mechanism.* Cancers (Basel), 2019 11(12):1983.
50. Williams A.B. and B. Schumacher, *p53 in the DNA-Damage-Repair Process.* Cold Spring Harb. Perspect. Med., 2016. 6(5).
51. Gaglia G., et al., *Activation and control of p53 tetramerization in individual living cells.* Proc. Natl. Acad. Sci. U S A, 2013. 110(38):15497-501.
52. Brooks C.L. and W. Gu, *p53 ubiquitination: Mdm2 and beyond.* Mol. Cell., 2006. 21(3):307-15.
53. Toledo F. and G.M. Wahl, *Regulating the p53 pathway: in vitro hypotheses, in vivo veritas.* Nat. Rev. Cancer, 2006. (12):909-23.
54. Meek D.W. and C.W. Anderson, *Posttranslational modification of p53: cooperative integrators of function.* Cold Spring Harb. Perspect. Biol., 2009. 1(6):a000950.
55. DeHart C.J., et al., *Extensive Post-translational Modification of Active and Inactivated Forms of Endogenous p53.* Mol. Cell. Proteomics, 2014. (1):1-17.
56. Kussie P.H., et al., *Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain.* Science, 1996. 274(5289):948-53.
57. W. Gu, X.L. Shi, and R.G. Roeder, *Synergistic activation of transcription by CBP and p53.* Nature, 1997. 387(6635):819-23.

58. Dunker A.K., et al., *Flexible nets. The roles of intrinsic disorder in protein interaction networks.* FEBS J., 2005. 272(20):5129-48.

59. Li D. W-C, et al., *Protein serine/threonine phosphatase-1 dephosphorylates p53 at Ser-15 and Ser-37 to modulate its transcriptional and apoptotic activities.* Oncogene, 2006. 25, 3006–3022.

60. Mohan A., et al., *Analysis of Molecular Recognition Features (MoRFs).* J. Mol. Biol., 2006. 362(5):1043-59.

61. Vacic V., et al., *Characterization of molecular recognition features, MoRFs, and their binding partners.* J. Proteome Res., 2007. 6(6): 2351–2366.

62. Popowicz G.M., et al., *Molecular basis for the inhibition of p53 by Mdmx.* Cell Cycle, 2007. 6(19):2386-92.

63. Di Lello P., et al., *Structure of the Tfb1/p53 complex: Insights into the interaction between the p62/Tfb1 subunit of TFIIH and the activation domain of p53.* Mol. Cell., 2006. 22(6):731-740.

64. Teufel D.P., et al., *Four domains of p300 each bind tightly to a sequence spanning both transactivation subdomains of p53.* Proc. Natl. Acad. Sci. U S A, 2007. 104(17):7009-14.

65. Polley S., et al., *Differential recognition of phosphorylated transactivation domains of p53 by different p300 domains.* J. Mol. Biol., 2008. 376(1):8-12.

66. Lambert P.F., et al., *Phosphorylation of p53 Serine 15 Increases Interaction with CBP.* J. Biol. Chem., 1998. 273(49):33048-53.

67. Walker K.K. and A.J. Levine, *Identification of a novel p53 functional domain that is necessary for efficient growth suppression.* Proc. Natl. Acad. Sci. U S A, 1996. 93(26):15335-40.

68. Kay B.K., Williamson M.P., and M. Sudol, *The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains.* FASEB J., 2000. (2):231-41.

69. Toledo F., et al., *Mouse mutants reveal that putative protein interaction sites in the p53 proline-rich domain are dispensable for tumor suppression.* Mol. Cell. Biol., 2007. (4):1425-32.

70. Adzhubei A.A., Sternberg M.J.E., and A.A. Makarov, *Polyproline-II helix in proteins: structure and function.* J. Mol. Biol., 2013 425(12):2100-32.

71. Cañadillas J.M.P., et al., *Solution structure of p53 core domain: Structural basis for its instability.* Proc. Natl. Acad. Sci. U S A, 2006. 103 (7) 2109-2114.

72. Y. Cho, et al., *Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations.* Science, 1994. 265(5170):346-55.

73. Wang Y., Rosengarth A., and H. Luecke, *Structure of the human p53 core domain in the absence of DNA.* Acta Crystallogr. D. Biol. Crystallogr., 2007. 63:276-81.

74. Butler J.S. and S.N. Loh, *Structure, Function, and Aggregation of the Zinc-Free Form of the p53 DNA Binding Domain.* Biochemistry 2003. 42, 2396-2403.

75. Duan J. and L. Nilsson, *Effect of Zn2+ on DNA recognition and stability of the p53 DNA-binding domain.* Biochemistry, 2006. 45(24):7483-92.

76. Bullock A.N., et al., *Thermodynamic stability of wild-type and mutant p53 core domain.* Proc. Natl. Acad. Sci. U S A 1997. 94(26):14338-42.

77. Natan E., et al., *Ultraslow oligomerization equilibria of p53 and its implications.* Proc. Natl. Acad. Sci. U S A, 2009. 106(34):14327-32.

78. Kaghad M., et al., *Monoallelically expressed gene related to p53 at 1p36, a region frequently deleted in neuroblastoma and other human cancers.* Cell Commun. Signal., 1997. 90(4):809-19.

79. Yang A., et al., *p63, a p53 homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities.* Mol. Cell., 1998. 2(3):305-16.

80. el-Deiry W.S., et al., *Definition of a consensus binding site for p53.* Nat. Genet., 1992. 1(1):45-9.

81. Riley T., et al., *Transcriptional control of human p53-regulated genes.* Nat. Rev. Mol. Cell. Biol., 2008. 9(5):402-12.

82. Kitayner M., et al., *Structural basis of DNA recognition by p53 tetramers.* Mol. Cell., 2006. 22(6):741-753.

83. Liang S.-H. and N.F. Clarke, *A Bipartite Nuclear Localization Signal Is Required for p53 Nuclear Import Regulated by a Carboxyl-terminal Domain.* J. Biol. Chem., 1999. 274(46), 32699–32703.

84. Jeffrey P.D., Gorina S., and N.P. Pavletich, *Crystal Structure of the Tetramerization Domain of the p53 Tumor Suppressor at 1.7 Angstroms.* Science, 1995. 267(5203).

85. Clore G.M., et al., *Refined solution structure of the oligomerization domain of the tumour suppressor p53.* Nat. Struct. Biol., 1995. 2(4):321-33.

86. de Oliveira G.A.P., et al., *The Status of p53 Oligomeric and Aggregation States in Cancer.* Biomolecules 2020. 10(4),548.

87. Gencel-Augusto J. and G. Lozano, *p53 tetramerization: at the center of the dominant negative effect of mutant p53.* Genes Dev., 2020. 34(17-18): 1128-1146.

88. Kamada R., et al., *Cancer-associated p53 Tetramerization Domain Mutants.* J. Biol. Chem., 2010. 286(1): 252-258.

89. Mujtaba S., et al., *Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation.* Mol. Cell., 2004. 13(2):251-63.

90. Friedler A., et al., *Modulation of binding of DNA to the C-terminal domain of p53 by acetylation.* Structure, 2005. 13(4):629-36.

91. Weinberg R.L., et al., *Regulation of DNA binding of p53 by its C-terminal domain.* J. Mol. Biol., 2004. 342(3):801-11.

92. Tafvizi A., et al., *A single-molecule characterization of p53 search on DNA.* Proc. Natl. Acad. Sci. U S A, 2010. 108(2):563-568.

93. Murata A., et al., *One-Dimensional Sliding of p53 Along DNA Is Accelerated in the Presence of Ca(2+) or Mg(2+) at Millimolar Concentrations.* J. Mol. Biol., 2015. 427(16):2663-78.

94. Krois A.S., et al., *Mapping Interactions of the Intrinsically Disordered C-Terminal Regions of Tetrameric p53 by Segmental Isotope Labeling and NMR.* Biochemistry, 2022. 61, 23, 2709–2719.

95. Okorokov A.L., et al., *The structure of p53 tumour suppressor protein reveals the basis for its functional plasticity.* EMBO J., 2006. 25(21):5191-200.

96. Aramayo R., et al., *Quaternary structure of the specific p53-DNA complex reveals the mechanism of p53 mutant dominance.* Nucleic Acids Res., 2011. 39(20):8960-71.

97. Tidow H., et al., *Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex.* Proc. Natl. Acad. Sci. U S A, 2007. 104(30):12324-9.

98. Demir Ö., Ieong P.U., and R.E. Amaro, *Full-length p53 tetramer bound to DNA and its quaternary dynamics.* Oncogene, 2017. 36(10):1451-1460.

99. Melero R., et al., *Electron microscopy studies on the quaternary structure of p53 reveal different binding modes for p53 tetramers in complex with DNA.* Proc. Natl. Acad. Sci. U S A, 2011. 108(2):557-62.

100. Nishimura M., et al., *Structural basis for p53 binding to its nucleosomal target DNA sequence.* PNAS Nexus, 2022. 1(4):pgac177.

101. Chemes L.B., et al., *Circular dichroism techniques for the analysis of intrinsically disordered proteins and domains.* Methods Mol. Biol., 2012. 895:387-404.

102. Micsonai A., et al., *Disordered-Ordered Protein Binary Classification by Circular Dichroism Spectroscopy.* Front. Mol. Biosci., 2022. 9:863141.

103. Gibbs E.B., Cook E.C., and S.A. Showalter, *Application of NMR to studies of intrinsically disordered proteins.* Arch. Biochem. Biophys., 2017. 628:57-70.

104. Kosol S., et al., *Structural Characterization of Intrinsically Disordered Proteins by NMR Spectroscopy.* Molecules 2013. 18, 10802-10828.

105. Gil S., et al., *NMR Spectroscopic Studies of Intrinsically Disordered Proteins atNear-Physiological Conditions.* Angew. Chem. Int. Ed., 2013. 52, 11808–11812.

106. Borgia A., et al., *Extreme disorder in an ultra-high-affinity protein complex.* Nature, 2018 555(7694): 61–66.

107. Brucale M., Schuler B., and B. Samorì, *Single-Molecule Studies of Intrinsically Disordered Proteins.* Chem. Rev., 2014. 114, 3281−3317.

108. Gomes G.-N.W., et al., *Conformational Ensembles of an Intrinsically Disordered Protein Consistent with NMR, SAXS, and Single-Molecule FRET.* J. Am. Chem. Soc., 2020. 142, 15697-15710.

109. Wrapp D., et al., *Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation.* Science, 2020. 367(6483):1260-1263.

110. Klusch N., et al., *Cryo-EM structure of the respiratory I + III2 supercomplex from Arabidopsis thaliana at 2 Å resolution.* Nature Plants, 2023. 9, 142–156.

111. Stark, H., *GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM.* Methods Enzymol., 2010. 481:109-26.

112. Sigworth, F.J., *Principles of cryo-EM single-particle image processing.* Microscopy, 2016. 57–67.

113. Scheres, S.H.W., *RELION: Implementation of a Bayesian approach to cryo-EM structure determination.* J. Struct. Biol., 2012. 180(3): 519-530.

114. Zivanov J., et al., *New tools for automated high-resolution cryo-EM structure determination in RELION-3.* Elife, 2018. 7:e42166.

115. Nwanochie E. and V.N. Uversky, *Structure Determination by Single-Particle Cryo-Electron Microscopy: Only the Sky (and Intrinsic Disorder) is the Limit.* Int. J. Mol. Sci., 2019. 20(17):4186.

116. Beveridge R. and A.N. Calabrese, *Structural Proteomics Methods to Interrogate the Conformations and Dynamics of Intrinsically Disordered Proteins.* Front. Chem., 2021. 9:603639.

117. Sinz, A., *Cross-Linking/Mass Spectrometry for Studying Protein Structures and Protein-Protein Interactions: Where Are We Now and Where Should We Go from Here?* Angew. Chem. Int. Ed. Engl., 2018. 57(22):6390-6396.

118. Beveridge R., et al., *Ion Mobility Mass Spectrometry Measures the Conformational Landscape of p27 and its Domains and how this is Modulated upon Interaction with Cdk2/cyclin A.* Angew. Chem. Int. Ed. Engl., 2019. 58(10):3114-3118.

119. Beveridge R., et al., *Ion Mobility Mass Spectrometry Uncovers the Impact of the Patterning of Oppositely Charged Residues on the Conformational Distributions of Intrinsically Disordered Proteins.* J. Am. Chem. Soc., 2019. 141(12):4908-4918.

120. Arlt C., et al., *An Integrated Mass Spectrometry Based Approach to Probe the Structure of the Full-Length Wild-Type Tetrameric p53 Tumor Suppressor.* Angew. Chem. Int. Ed. Engl., 2017. 56(1):275-279.

121. Balasubramaniam D. and E.A. Komives, *Hydrogen-exchange mass spectrometry for the study of intrinsic disorder in proteins.* Biochim. Biophys. Acta, 2013. 1834(6):1202-9.

122. Ozohanics O. and A. Ambrus, *Hydrogen-Deuterium Exchange Mass Spectrometry: A Novel Structural Biology Approach to Structure, Dynamics and Interactions of Proteins and Their Complexes.* Life (Basel), 2020 10(11):286.

123. Trabjerg E., et al., *Conformational Analysis of Large and Highly Disulfide-Stabilized Proteins by Integrating Online Electrochemical Reduction into an Optimized H/D Exchange Mass Spectrometry Workflow.* Anal. Chem., 2015. 87(17), 8880–8888.

124. Seetaloo N. and J.J. Phillips, *Millisecond Hydrogen/Deuterium-Exchange Mass Spectrometry Approach to Correlate Local Structure and Aggregation in α-Synuclein.* Anal. Chem., 2022. 94(48):16711-16719.

125. Graziadei A. and J. Rappsilber, *Leveraging crosslinking mass spectrometry in structural and cell biology.* Structure, 2022 30(1):37-54.

126. Tang X., et al., *Mass Spectrometry Identifiable Cross-Linking Strategy for Studying Protein-Protein Interactions.* Anal. Chem., 2005. 77, 311-318.

127. Soderblom E.J. and M.B. Goshe, *Collision-Induced Dissociative Chemical Cross-Linking Reagents and Methodology: Applications to Protein Structural Characterization Using Tandem Mass Spectrometry Analysis.* Anal. Chem., 2006. 78, 8059-8068.

128. Müller M.Q., et al., *Cleavable cross-linker for protein structure analysis: reliable identification of cross-linking products by tandem MS.* Anal. Chem., 2010. 82(16):6958-68.

129. Kao A., et al., *Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes.* Mol. Cell. Proteomics, 2011. 10(1):M110.002212.

130. Petrotchenko E.V., J.J. Serpa, and C.H. Borchers, *An isotopically coded CID-cleavable biotinylated cross-linker for structural proteomics.* Mol. Cell., 2011. 10(2):M110.001420.

131. Clifford-Nunn B., Showalter H.D.H., and P.C. Andrews, *Quaternary Diamines as Mass Spectrometry Cleavable Crosslinkers for Protein Interactions.* J. Am. Soc. Mass Spectrom., 2012. 23(2): 201–212.

132. Burke A.M., et al., *Synthesis of two new enrichable and MS-cleavable cross-linkers to define protein-protein interactions by mass spectrometry.* Org. Biomol. Chem., 2015. 13(17):5030-7.

133. Ihling C. H., et al., *Dissociation behavior of a bifunctional tempo-active ester reagent for peptide structure analysis by free radical initiated peptide sequencing (FRIPS) mass spectrometry.* J. Mass. Spectrom., 2015. 50(2):396-406.

134. Gutierrez C.B., et al., *Developing an Acidic Residue Reactive and Sulfoxide-Containing MS-Cleavable Homobifunctional Cross-Linker for Probing Protein−Protein Interactions.* Anal. Chem., 2016. 88, 8315−8322.

135. Hage C., et al., *The First Zero-Length Mass Spectrometry-Cleavable Cross-Linker for Protein Structure Analysis.* Angew Chem. Int. Ed. Engl., 2017. 56(46):14551-14555.

136. Gutierrez C.B., et al., *Development of a Novel Sulfoxide-Containing MS-Cleavable Homobifunctional Cysteine-Reactive Cross-Linker for Studying Protein-Protein Interactions.* Anal. Chem., 2018. 90(12):7600-7607.

137. Iacobucci C., et al., *Carboxyl-Photo-Reactive MS-Cleavable Cross-Linkers: Unveiling a Hidden Aspect of Diazirine-Based Reagents.* Anal. Chem., 2018. 90(4):2805-2809.

138. Hagen S.E., et al., *Synthesis of CID-cleavable protein crosslinking agents containing quaternary amines for structural mass spectrometry.* Org. Biomol. Chem. , 2018 16(37): 6867–6870.

139. Gutierrez C.B., et al., *Enabling Photoactivated Cross-Linking Mass Spectrometric Analysis of Protein Complexes by Novel MS-Cleavable Cross-Linkers.* Mol. Cell. Proteomics, 2021. 20, 100084.

140. Jiang P-L., et al., *A Membrane-Permeable and Immobilized Metal Affinity Chromatography (IMAC) Enrichable Cross-Linking Reagent to Advance In Vivo Cross-Linking Mass Spectrometry.* Angew. Chem. Int. Ed. Engl., 2022. 61(12):e202113937.

141. Jiao F., et al., *Exploring an Alternative Cysteine-Reactive Chemistry to Enable Proteome-Wide PPI Analysis by Cross-Linking Mass Spectrometry.* Anal. Chem., 2023. 95(4):2532-2539.

142. Kolbowski L., et al., *Light-Induced Orthogonal Fragmentation of Crosslinked Peptides.* JACS Au, 2023. 3(8):2123-2130.

143. Chen J., et al., *A Glycosidic-Bond-Based Mass-Spectrometry-Cleavable Cross-linker Enables In Vivo Cross-linking for Protein Complex Analysis.* Angew. Chem. Int. Ed. Engl., 2023. 62(24):e202212860.

144. Belsom A., et al., *Complementary Benzophenone Cross-Linking/Mass Spectrometry Photochemistry.* Anal. Chem., 2017. 89, 5319−5324.

145. Mohr J.P., et al., *Multidimensional Cross-Linking and Real-Time Informatics for Multiprotein Interaction Studies.* J. Proteome Res., 2024. 23(1):107-116.

146. Iacobucci C., et al., *A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein–protein interactions.* Nat. Protocols, 2018. 13, 2864–2889.

147. Steigenberger B., et al., *Benefits of Collisional Cross Section Assisted Precursor Selection (caps-PASEF) for Cross-linking Mass Spectrometry.* Mol. Cell. Proteomics, 2020. 19(10), 1677–1687.

148. Ihling C.H., et al., *Cross-Linking/Mass Spectrometry Combined with Ion Mobility on a timsTOF Pro Instrument for Structural Proteomics.* Anal. Chem., 2021. 93(33), 11442–11450.

149. Jiao F., et al., *Two-Dimensional Fractionation Method for Proteome-Wide Cross-Linking Mass Spectrometry Analysis.* Anal. Chem., 2022. 94(10):4236-4242.

150. Hao Y., et al., *4D-diaXLMS: Proteome-wide Four-Dimensional Data-Independent Acquisition Workflow for Cross-Linking Mass Spectrometry.* Anal. Chem., 2023. 95(37):14077-14085.

151. Sinz, A., *Crosslinking Mass Spectrometry Goes In-Tissue.* Cell Syst., 2018. 6(1):10-12.

152. Nagy B., et al., *Structure of the dihydrolipoamide succinyltransferase (E2) component of the human alpha-ketoglutarate dehydrogenase complex (hKGDHc) revealed by cryo-EM and cross-linking mass spectrometry: Implications for the overall hKGDHc structure.* Biochim. Biophys. Acta Gen. Subj., 2021. 1865(5), 129889.

153. Schmidt C. and U. Henning, *Combining cryo-electron microscopy (cryo-EM) and cross-linking mass spectrometry (CX-MS) for structural elucidation of large protein assemblies.* Curr. Opin. Struct. Biol., 2017. 46:157-168.

154. Manthei K.A., et al., *Structural analysis of lecithin:cholesterol acyltransferase bound to high density lipoprotein particles.* Commun. Biol., 2020. 3(1):28.

155. Liu, H., *AlphaFold and Structural Mass Spectrometry Enable Interrogations on the Intrinsically Disordered Regions in Cyanobacterial Light-harvesting Complex Phycobilisome.* J. Mol. Biol., 2022. 434(21), 167831.

156. Stahl K., et al., *Protein structure prediction with in-cell photo-crosslinking mass spectrometry and deep learning.* Nat. Biotechnol., 2023. 41(12):1810-1819.

157. Hillenkamp F., et al., *Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers.* Anal. Chem., 1991. 63(24):1193A-1203A.

158. Karas M. and F. Hillenkamp, *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons.* Anal. Chem., 1988 60(20):2299-301.

159. Fenn J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989 246(4926):64-71.

160. Gross, J.H., *Mass Spectrometry: A Textbook*. 2017: Springer Berlin Heidelberg.

161. Banerjee S. and S. Mazumdar, *Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte.* Int. J. Anal. Chem., 2012. 282574.

162. Dole M., et al., *Molecular Beams of Macroions.* J. Chem. Phys., 1968. 49, 2240–2249.

163. Konermann L., et al., *Unraveling the mechanism of electrospray ionization.* Anal. Chem., 2013. 85(1):2-9.

164. Iribarne J.V. and B.A. Thomson, *On the evaporation of small ions from charged droplets* J. Chem. Phys., 1976. 64, 2287–2294

165. Nguyen S. and J.B. Fenn, *Gas-phase ions of solute species from charged droplets of solutions.* Proc. Natl. Acad. Sci. U S A, 2007. 104 (4) 1111-1117.

166. de la Mora, F., *Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism.* Analytica Chimica Acta, 2000. 406, (1), 93-104.

167. Paul W. and H. Steinwedel, *A New Mass Spectrometer Without Magnetic Field.* . Z. Naturforsch 1953. 8A:448–450.

168. Wolff M.M. and W.E. Stephens, *A Pulsed Mass Spectrometer with Time Dispersion.* Review of Scientific Instruments 1953. 24, 616

169. W.C. Wiley and I.H. McLaren, *Time-of-Flight Mass Spectrometer with Improved Resolution.* Rev. Sci. Instrum., 1955. 26, 1150–1157

170. R.S. Brown and J.J. Lennon, *Mass Resolution Improvement by Incorporation of Pulsed Ion Extraction in a Matrix-Assisted Laser Desorption/Ionization Linear Time-of-Flight Mass Spectrometer.* Anal. Chem., 1995. 67, 13, 1998–2003.

171. Mamyrin, B.A., et al., *The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution.* Zh. Eksp. Teor. Fiz, 1973. 64, 82-89.

172. Berkenkamp S., et al., *Measurements of mean initial velocities of analyte and matrix ions in infrared matrix-assisted laser desorption ionization mass spectrometry.* J. Am. Soc. Mass Spectrom., 2002. 13(3):209-20.

173. Kingdon, K.H., *A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures.* Phys. Rev., 1923. 21, 408.

174. Knight, R.D., *Storage of ions from laser-produced plasmas.* Appl. Phys. Lett., 1981. 38, 221–223

175. Hu Q., et al., *The Orbitrap: a new mass spectrometer.* J. Mass Spectrom., 2005. 40(4):430-43.

176. Kim M.S. and F.W. McLafferty, *Collisional activation and metastable ion characteristics. 59. Efficiency of collisional activation of gaseous organic ions.* J. Am. Chem. Soc., 1978. 100, 11, 3279–3282.

177. Csonka I.P., et al., *Proton mobility and main fragmentation pathways of protonated lysylglycine.* Rapid Commun. Mass Spectrom., 2001. 15(16):1457-72.

178. Bythell B.J., et al., *Proton-driven amide bond-cleavage pathways of gas-phase peptide ions lacking mobile protons.* J. Am. Chem. Soc., 2009. 131(39):14057-65.

179. Harrison A.G. and T. Yalcin, *Proton mobility in protonated amino acids and peptides.* Int. J. Mass Spectrom. Ion Processes, 1997. 165–166, 339-347.

180. Ashok R. Dongre´, et al., *Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model.* J. Am. Chem. Soc., 1996. 118, 35, 8365–8374.

181. P. Roepstorff and J. Fohlman, *Proposal for a common nomenclature for sequence ions in mass spectra of peptides.* Biomed. Mass Spectrom., 1984 11(11):601.

182. Olsen J.V., et al., *Higher-energy C-trap dissociation for peptide modification analysis.* Nat. Methods, 2007. 4(9):709-12.

183. Senko M.W., et al., *Novel parallelized quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates.* Anal. Chem., 2013. 85(24):11710-4.

184. Zubarev R.A., Kelleher N.L., and F.W. McLafferty, *Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process.* J. Am. Chem. Soc., 1998. 120, 13, 3265–3266.

185. Mehmood S., Allison T.M., and C.V. Robinson, *Mass spectrometry of protein complexes: from origins to applications.* Annu. Rev. Phys. Chem., 2015 66:453-74.

186. Heck, A.J.R., *Native mass spectrometry: a bridge between interactomics and structural biology.* Nat. Methods, 2008 5(11):927-33.

187. Tamara S., den Boer M.A., and A.J.R. Heck, *High-Resolution Native Mass Spectrometry.* Chem. Rev., 2022. 122(8):7269-7326.

188. Sinz A., et al., *Chemical cross-linking and native mass spectrometry: A fruitful combination for structural biology.* Protein Sci., 2015 24(8):1193-209.

189. Chen Y., et al., *Protein Folding: Then and Now.* Arch. Biochem. Biophys., 2008. 469(1):4-19.

190. Oliveira Junior A.B., et al., *Exploring Energy Landscapes of Intrinsically Disordered Proteins: Insights into Functional Mechanisms.* J. Chem. Theory Comput., 2021. 17(5):3178-3187.

191. Sabatini D., Bensch K., and R.J. Barrnett, *Cytochemistry and electron microscopy. The preservation of cellular ultrastructure and enzymatic activity by aldehyde fixation.* J. Cell. Biol., 1963 17(1):19-58.

192. Rappsilber J. and M. Mann, *Analysis of the topology of protein complexes using cross-linking and mass spectrometry.* CSH Protoc., 2007. pdb.prot4594.

193. Sinz A. and K. Wang, *Mapping protein interfaces with a fluorogenic cross-linker and mass spectrometry: application to nebulin-calmodulin complexes.* Biochemistry, 2001. 40(26):7903-13.

194. Liu F., et al., *Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry.* Nat. Methods, 2015. 12(12):1179-8.

195. Chen F., Nielsen S., and R. Zenobi, *Understanding chemical reactivity for homo- and heterobifunctional protein cross-linking agents.* J. Mass. Spectrom., 2013. 48(7):807-12.

196. Müller F., Graziadei A., and J. Rappsilber, *Quantitative Photo-crosslinking Mass Spectrometry Revealing Protein Structure Response to Environmental Changes.* Anal. Chem., 2019. 91(14): 9041–9048.

197. Mädler S., et al., *Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities.* J. Mass Spectrom., 2009 44(5):694-706.

198. Koolen H.H.F., et al., *Imidate-based cross-linkers for structural proteomics: increased charge of protein and peptide ions and CID and ECD fragmentation studies.* J. Am. Soc. Mass. Spectrom., 2014 25(7):1181-91.

199. Nury C., et al., *A novel bio-orthogonal cross-linker for improved protein/protein interaction analysis.* Anal. Chem. , 2015. 87(3):1853-60.

200. Sheikh M.C., et al., *First synthesis of non-symmetrical "phthalimidoyl active ester" bi-dentate cross-linking reagents having an acid chloride, 2-benzothiazole, or 1-benzotriazol group.* Org. Biomol. Chem., 2008. 6(24):4505-8.

201. Mintseris J. and S.P. Gygi, *High-density chemical cross-linking for modeling protein interactions.* Proc. Natl. Acad. Sci. U S A, 2020 117(1):93-102.

202. Leitner A., et al., *Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes.* Proc. Natl. Acad. Sci. U S A., 2014. 111(26): 9455–9460.

203. West A.V., et al., *Labeling Preferences of Diazirines with Protein Biomolecules.* J. Am. Chem. Soc., 2021 143(17):6691-6700.

204. Sinz, A., *Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions.* Mass Spectrom. Rev., 2006 25(4):663-82.

205. Suchanek M., Radzikowska A., and C. Thiele, *Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells.* Nat. Methods, 2005 2(4):261-7.

206. Young T.S. and P.G. Schultz, *Beyond the canonical 20 amino acids: expanding the genetic lexicon.* J. Biol. Chem., 2010. 285(15):11039-44.

207. Jones A.X., et al., *Improving mass spectrometry analysis of protein structures with arginine-selective chemical cross-linkers.* Nat. Commun., 2019. 10(1):3911.

208. Cui L., et al., *Tyrosine-Reactive Cross-Linker for Probing Protein Three-Dimensional Structures.* Anal. Chem., 2021 93(10):4434-4440.

209. Lu M., et al., *The Magic of Linking Rings: Discovery of a Unique Photoinduced Fluorescent Protein Crosslink.* J. Am. Chem. Soc., 2022 144(24):10809-10816.

210. Schilling B., et al., *MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides.* J. Am. Soc. Mass Spectrom., 2003. 14(8):834-50.

211. Steigenberger B., et al., *To Cleave or Not To Cleave in XL-MS?* J. Am. Soc. Mass. Spectrom., 2020 31(2):196-206.

212. Piersimoni L. and A. Sinz, *Cross-linking/mass spectrometry at the crossroads.* Anal. Bioanal. Chem., 2020. 412(24):5981-5987.

213. Götze M., et al., *Automated assignment of MS/MS cleavable cross-links in protein 3D-structure analysis.* J. Am. Soc. Mass Spectrom., 2015. 26(1):83-97.

214. Liu F., et al., *Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification.* Nat. Commun., 2017. 8:15473.

215. Hoopmann M.R., et al., *Improved Analysis of Cross-Linking Mass Spectrometry Data with Kojak 2.0, Advanced by Integration into the Trans-Proteomic Pipeline.* J. Proteome Res., 2023 22(2):647-655.

216. Mendes M.L., et al., *An integrated workflow for crosslinking mass spectrometry.* Mol. Syst. Biol., 2019. 15: e8994.

217. Yılmaz S., et al., *Accurate and Automated High-Coverage Identification of Chemically Cross-Linked Peptides with MaxLynx.* Anal. Chem., 2022. 94, 3, 1608–1617.

218. Pirklbauer G.J., et al., *MS Annika: A New Cross-Linking Search Engine.* J. Proteome Res., 2021. 20(5):2560-2569.

219. Chen Z.-L., et al., *A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides.* Nat. Commun., 2019 10(1):3404.

220. Sinz, A., *Divide and conquer: cleavable cross-linkers to study protein conformation and protein-protein interactions.* Anal. Bioanal. Chem., 2017 409(1):33-44.

221. Kolb H.C., Finn M.G., and K.B. Sharpless, *Click Chemistry: Diverse Chemical Function from a Few Good Reactions.* Angew. Chem. Int. Ed. Engl., 2001 40(11):2004-2021.

222. Jewett J.C., Sletten E.M., and C.R. Bertozzi, *Rapid Cu-Free Click Chemistry with Readily Synthesized Biarylazacyclooctynones.* J. Am. Chem. Soc., 2010 132(11):3688-90.

223. Rey M., et al., *Advanced In Vivo Cross-Linking Mass Spectrometry Platform to Characterize Proteome-Wide Protein Interactions.* Anal. Chem., 2021. 93(9):4166-4174.

224. Sambrook, J. and D.W. Russell, *Molecular Cloning: A Laboratory Manual.* . Vol. 3rd Edition, Vol. 1 2001: Cold Spring Harbor Laboratory Press, New York.

225. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Anal. Biochem., 1976. 72:248-54.

226. Neuhoff V., et al., *Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250.* Electrophoresis 1988. 9(6):255-62.

227. VanAernum Z.L., et al., *Rapid online buffer exchange for screening of proteins, protein complexes and cell lysates by native mass spectrometry.* Nat. Protoc., 2020. 15(3):1132-1157.

228. Meier F., et al., *Parallel Accumulation–Serial Fragmentation (PASEF): Multiplying Sequencing Speed and Sensitivity by Synchronized Scans in a Trapped Ion Mobility Device.* J. Proteome Res. , 2015. 14, 12, 5378–5387.

229. Meier F., et al., *Online Parallel Accumulation–Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer.* Mol. Cell. Proteomics, 2018. 17(12): 2534–2545.

230. Meier F., Park M.A., and M. Mann, *Trapped Ion Mobility Spectrometry and Parallel Accumulation-Serial Fragmentation in Proteomics.* Mol. Cell. Proteomics, 2021. 20:100138.

231. Ziemianowicz D.S., Sarpe V., and D.C. Schriemer, *Quantitative Analysis of Protein Covalent Labeling Mass Spectrometry Data in the Mass Spec Studio.* Anal. Chem., 2019. 91(13):8492-849.

232. Arlt, C., *Massenspektrometrische Untersuchungen am Tumorsuppressorprotein p53* in *Marthin Luther University Halle-Wittenberg*. 2017.

233. Di Ianni A., et al., *Structural assessment of the full-length wild-type tumor suppressor protein p53 by mass spectrometry-guided computational modeling.* Sci. Rep., 2023 13(1):8497.

234. Olsson M.H.M., et al., *PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions.* J. Chem. Theory Comput., 2011. 7, 2, 525–537.

235. Sanner M.F., Olson A.J., and J.C. Spehner, *Reduced surface: an efficient way to compute molecular surfaces.* Biopolymers, 1996 38(3):305-20.

236. Pagel K., et al., *Intrinsically disordered p53 and its complexes populate compact conformations in the gas phase.* Angew. Chem. Int. Ed. Engl., 2013 52(1):361-5.

237. Trabjerg E., et al., *Conformational analysis of large and highly disulfide-stabilized proteins by integrating online electrochemical reduction into an optimized H/D exchange mass spectrometry workflow.* Anal. Chem., 2015 87(17):8880-8.

238. Ihling C.H., et al., *Isotope-labeled cross-linkers and Fourier transform ion cyclotron resonance mass spectrometry for structural analysis of a protein/peptide complex.* J. Am. Soc. Mass Spectrom., 2006. 17(8):1100-1113.

239. Webb B. and A. Sali, *Protein Structure Modeling with MODELLER.* Methods Mol. Biol., 2021. 2199:239-255.

240. Hage C., et al., *Dissociation Behavior of a TEMPO-Active Ester Cross-Linker for Peptide Structure Analysis by Free Radical Initiated Peptide Sequencing (FRIPS) in Negative ESI-MS.* J. Am. Soc. Mass Spectrom., 2016. 28(1):56-68.

241. Stieger C.E., Doppler P., and K. Mechtler, *Optimized Fragmentation Improves the Identification of Peptides Cross-Linked by MS-Cleavable Reagents.* J. Proteome Res., 2019. 18(3):1363-1370.

242. Ruwolt M., et al., *Real-Time Library Search Increases Cross-Link Identification Depth across All Levels of Sample Complexity.* Anal. Chem., 2023 95(12):5248-5255.

243. Pindrus M.A., et al., *Effect of Aggregation on the Hydrodynamic Properties of Bovine Serum Albumin.* Pharm. Res., 2017. 34(11):2250-2259.

244. Saito R., et al., *Structure of bovine carbonic anhydrase II at 1.95 A resolution.* Acta Crystallogr. D. Biol. Crystallogr., 2004 60(Pt 4):792-5.

245. Rostovtsev V.V., et al., *A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes.* Angew. Chem. Int. Ed. Engl., 2002. 41(14):2596-9.

246. Worrell B.T., Malik J.A., and V.V. Fokin, *Direct evidence of a dinuclear copper intermediate in Cu(I)-catalyzed azide-alkyne cycloadditions.* Science, 2013. 340(6131):457-60.

247. Jin L., et al., *Isolation of bis(copper) key intermediates in Cu-catalyzed azide-alkyne "click reaction".* Sci. Adv., 2015. 1(5):e1500304.

248. Kislukhin A.A., et al., *Relative performance of alkynes in copper-catalyzed azide-alkyne cycloaddition.* Bioconjug. Chem., 2013 24(4):684-9.

249. Presolski S.I., et al., *Tailored ligand acceleration of the Cu-catalyzed azide-alkyne cycloaddition reaction: practical and mechanistic implications.* J. Am. Chem. Soc., 2010. 132(41):14570-6.

250. Presolski S.I., Hong V.P., and M.G. Finn, *Copper-Catalyzed Azide-Alkyne Click Chemistry for Bioconjugation.* Curr. Protoc. Chem. Biol., 2011. 3(4):153-162.

251. Mack S., et al., *Pseudo-Ligandless Click Chemistry for Oligonucleotide Conjugation.* Curr. Protoc. Chem. Biol., 2016 8(2):83-95.

252. Nelson C.C. and J.A. McCloskey, *Collision-induced dissociation of uracil and its derivatives.* J. Am. Soc. Mass. Spectrom., 1994. 5, 5, 339–349.

253. Stefanis, L., *α-Synuclein in Parkinson's Disease.* Cold Spring Harb. Perspect. Med. , 2012. 2(2).

254. Ubbiali D., et al., *Direct Observation of "Elongated" Conformational States in α-Synuclein upon Liquid-Liquid Phase Separation.* Angew. Chem. Int. Ed. Engl., 2022 61(46).

255. Matzinger M., et al., *Robust and Easy-to-Use One-Pot Workflow for Label-Free Single-Cell Proteomics.* Anal. Chem., 2023 95(9):4435-4445.

256. Yu G., et al., *clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters.* OMICS, 2012. 16(5): 284–287.

257. The Gene Ontology Consortium, et al., *Gene Ontology: tool for the unification of biology.* Nat Genet., 2000. 25(1): 25–29.

258. Xue B., A.K. Dunker, and V.N. Uversky, *Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life.* Journal of Biomolecular Structure and Dynamics, 2012. 30:2, 137–149.

259. Baudin A., et al., *Structural Characterization of the RNA-Binding Protein SERBP1 Reveals Intrinsic Disorder and Atypical RNA Binding Modes.* Front. Mol. Biosci., 2021. 8.

260. Muto A., et al., *The mRNA-binding protein Serbp1 as an auxiliary protein associated with mammalian cytoplasmic ribosomes.* Cell Biochem Funct., 2018. 1–11.

261. Dosztányi, Z., *Prediction of protein disorder based on IUPred.* Protein Sci., 2018 27(1):331-340.

262. Ward J.J., et al., *The DISOPRED server for the prediction of protein disorder.* Bioinformatics, 2004 20(13):2138-9.

263. Wells J.N., et al., *Structure and function of yeast Lso2 and human CCDC124 bound to hibernating ribosomes.* . PLoS Biol., 2020. 18(7):e3000780.

264. Bolger, G.B., *The RNA-binding protein SERBP1 interacts selectively with the signaling protein RACK1.* Cell Signal., 2017. 35: 256–263.

265. Kipper K., Mansour A., and A.Pulk, *Neuronal RNA granules are ribosome complexes stalled at the pre-translocation state.* J. Mol. Biol., 2022 434(20):167801.

266. Anger A.M., et al., *Structures of the human and Drosophila 80S ribosome.* Nature, 2013. 497(7447):80-5.

267. Brown A., et al., *Structures of translationally inactive mammalian ribosomes.* Elife, 2018. 7:e40486.

268. Anderson M.E., et al., *Reciprocal interference between the sequence-specific core and nonspecific C-terminal DNA binding domains of p53: implications for regulation.* Mol. Cell. Biol., 1997 17(11):6255-64.

269. Hupp T.R. and D.P. Lane, *Allosteric activation of latent p53 tetramers.* Curr. Biol., 1994 4(10):865-75.

270. Jayaraman J. and C. Prives, *Activation of p53 sequence-specific DNA binding by short single strands of DNA requires the p53 C-terminus.* Cell., 1995 81(7):1021-9.

271. Nie L., Sasaki M., and C.G. Maki, *Regulation of p53 nuclear export through sequential changes in conformation and ubiquitination.* J. Biol. Chem., 2007 282(19):14616-25.

272. An W., Kim J., and R.G. Roeder, *Ordered cooperative functions of PRMT1, p300, and CARM1 in transcriptional activation by p53.* Cell., 2004 117(6):735-48.

273. Espinosa J.M. and B.M. Emerson, *Transcriptional regulation by p53 through intrinsic DNA/chromatin binding and site-directed cofactor recruitment.* Mol. Cell., 2001. 8(1):57-69.

274. Laptenko O., et al., *The p53 C terminus controls site-specific DNA binding and promotes structural changes within the central DNA binding domain.* Mol. Cell., 2015. 57(6):1034-1046.

275. Tafvizi A., et al., *Tumor suppressor p53 slides on DNA with low friction and high stability.* Biophys. J., 2008 95(1):L01-3.

276. Veprintsev D.B. and A.R. Fersht, *Algorithm for prediction of tumour suppressor p53 affinity for binding sites in DNA.* Nucleic Acids Research, 2008 36(5):1589-98.

277. D'Abramo M., et al., *The p53 tetramer shows an induced-fit interaction of the C-terminal domain with the DNA-binding domain.* Oncogene, 2016. 35(25):3272-81.

278. Di Ianni A., et al., *Evaluating Imide-Based Mass Spectrometry-Cleavable Cross-Linkers for Structural Proteomics Studies.* JACS Au, 2024. 4(8):2936-2943.

279. Lenz S., et al., *Reliable identification of protein-protein interactions by crosslinking mass spectrometry.* Nat. Commun., 2021. 12(1):3564.

280. Giese S.H., et al., *Retention time prediction using neural networks increases identifications in crosslinking mass spectrometry.* Nat. Commun., 2021. 12: 3237.

281. Jiao F., et al., *DSBSO-Based XL-MS Analysis of Breast Cancer PDX Tissues to Delineate Protein Interaction Network in Clinical Samples.* J. Proteome Res., 2024. Online ahead of print.

Appendix Figures

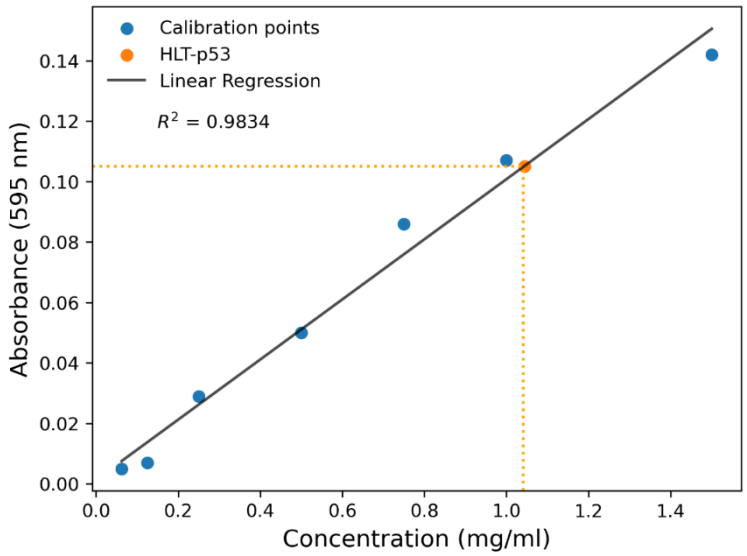


Figure A1: Estimation of HLT-p53 concentration via Bradford assay.

Calibration curve for the determination of the overall HLT-p53 concentration was built using bovine serum albumin (BSA) as protein standard covering a range between 0.125 mg/mL and 1.5 mg/mL. A dye solution in Milli-Q H_2O was prepared (1:5 dilution) and mixed to the samples with the suggested ratios. Absorbance at 595 nm was used here as a readout.

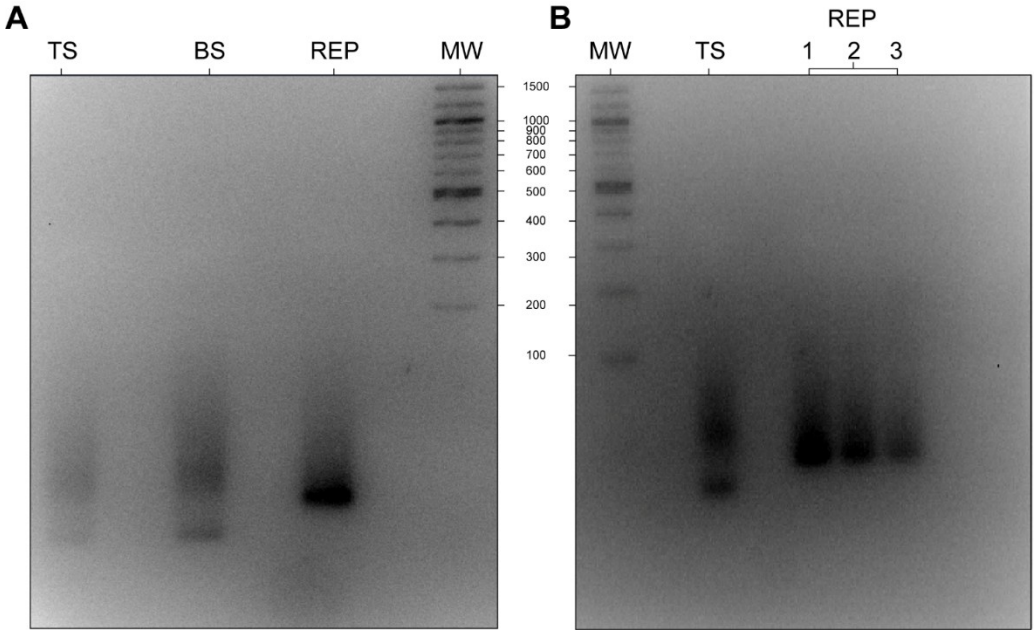


Figure A2: Hybridization of the REP used in this work.

A) Hybridization of REP, TS - Top Strand, BS - Bottom Strand, REP - Response Element P. B) Evaluation of the reproducibility of the annealing procedure.

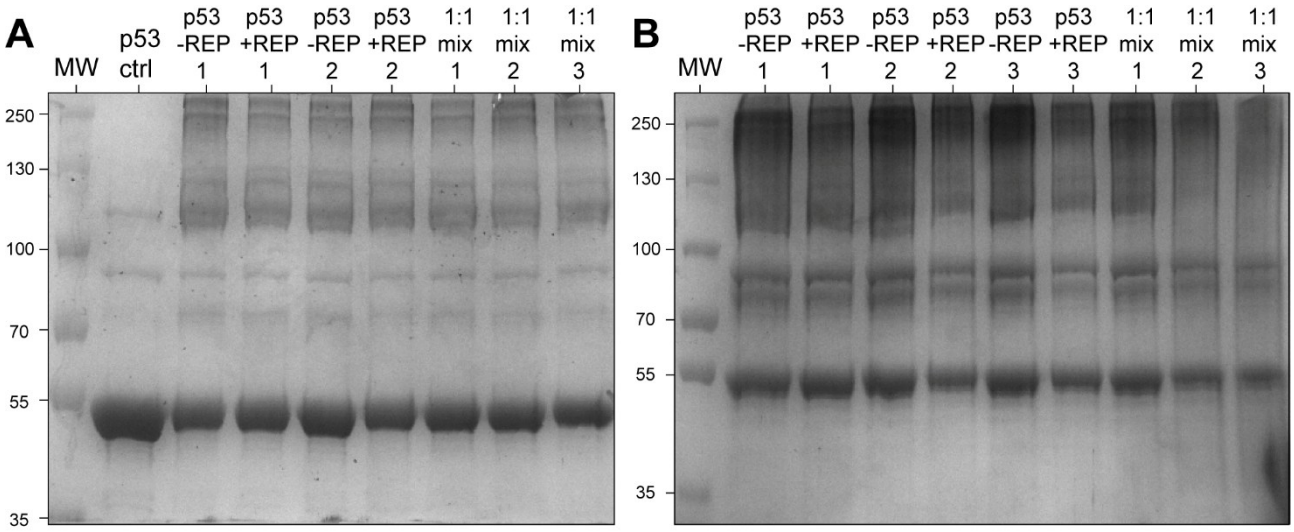


Figure A3: SDS-PAGE analysis of quantitative XL-MS experiments of p53 with BS^{2}G-D$_0$/D$_4$.

A) p53 without REP: BS^{2}G-D$_0$, with REP: BS^{2}G-D$_4$. B) p53 without REP: BS^{2}G-D$_4$, with REP: BS^{2}G-D$_0$.

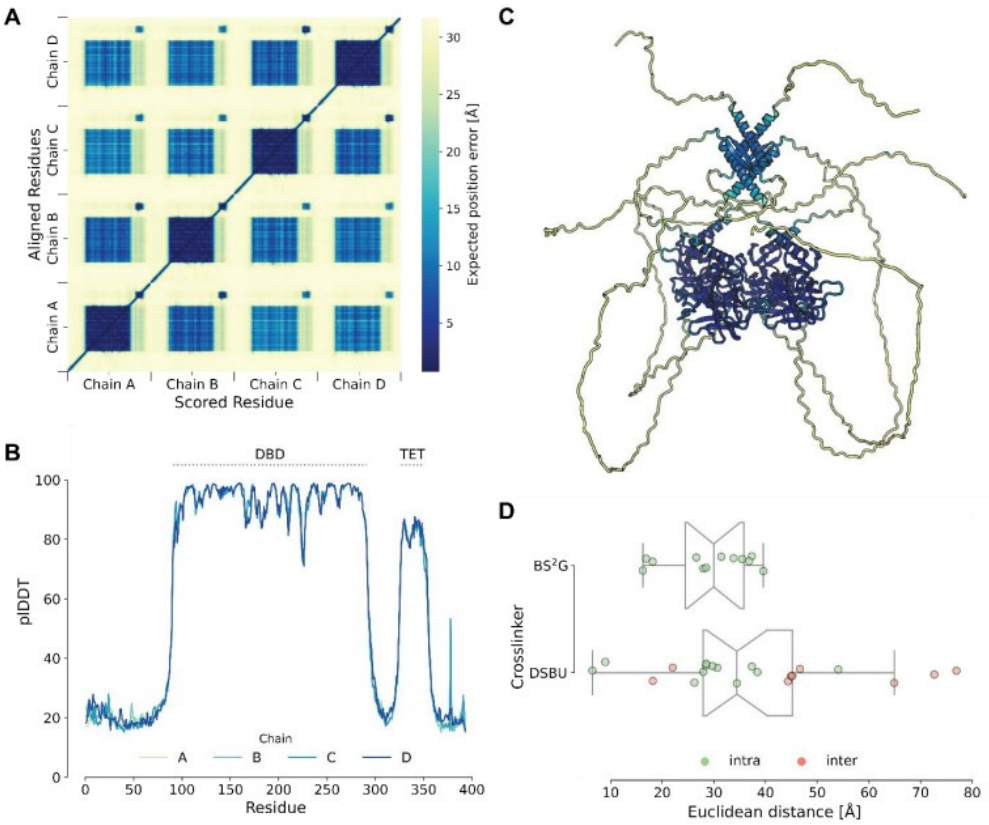


Figure A4: AlphaFold2 prediction of tetrameric p53.

A) Heatmap of the predicted aligned error of tetrameric p53. Only the structure of the (DBD; blue) has a high score. B) AlphaFold2 confidence scores (pLDDT) of each residue. Only the DBD and the TD are predicted with high confidence, indicating an ordered structure. C) AlphaFold2 predicted tetrameric p53 atomic model in cartoon representation; coloring is based on pLDDT scoring. D) Distance distribution of XL-MS data; the shortest distance of each cross-link is plotted.

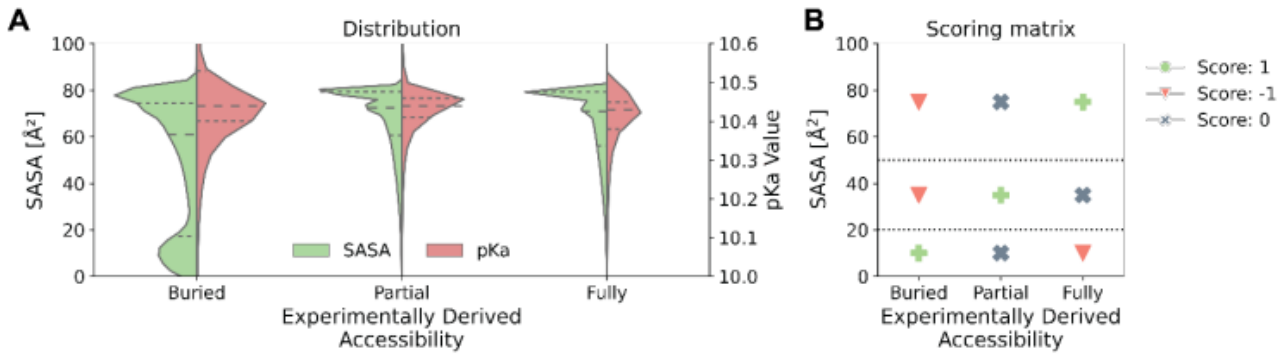


Figure A5: Scoring of labeling efficiency.

A) SASA and pKa values of labeled lysine residues in p53, plotted as function of footprinting data (experimentally derived accessibility). Although pKa values are unaltered there seems to be a population of lysine residues that are not accessible to labeling, as is visible from changing SASA values. B) Scoring matrix of footprinting data.

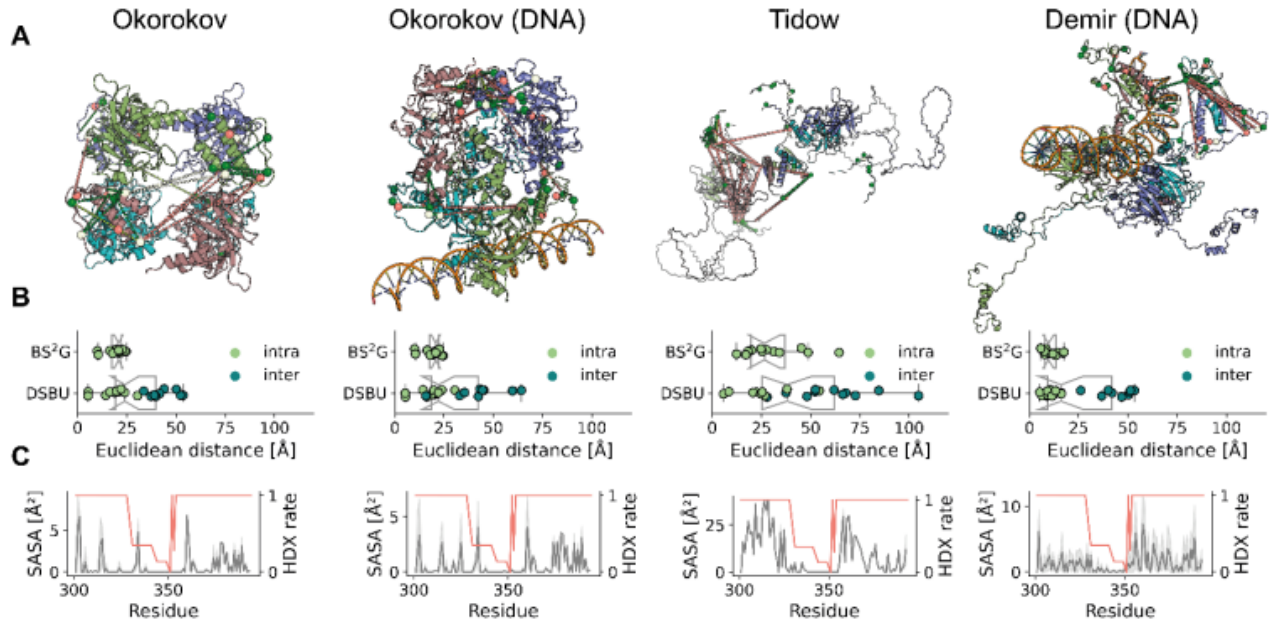


Figure A6: Existing models of p53 higher-order structures poorly explain XL-MS and footprinting data.

A) Atomic tetrameric p53 models according to Okorokov et al, 2006 (+/- DNA) [95, 96]; Tidow et al, 2007 (-DNA) [97]; Demir et al, 2017 (+DNA) [98]. Cross-links are indicated as dashed lines between Cα atoms of residues involved; residues labeled in footprinting experiments are indicated by colored spheres. B) Distance distributions of BS^2G and DBSU cross-links show distance violations between p53 dimers in the existing models of p53 tetramer; intra: cross-links within one p53 monomer, inter: cross-links between p53 dimers. C) Existing models show a high correlation between SASA (grey) and HDX (orange). The disordered and accessible nature of p53's IDRs is reflected by high SASA and HDX rates, while the structured TD (residues 326- 356) exhibits low SASA and HDX rates.

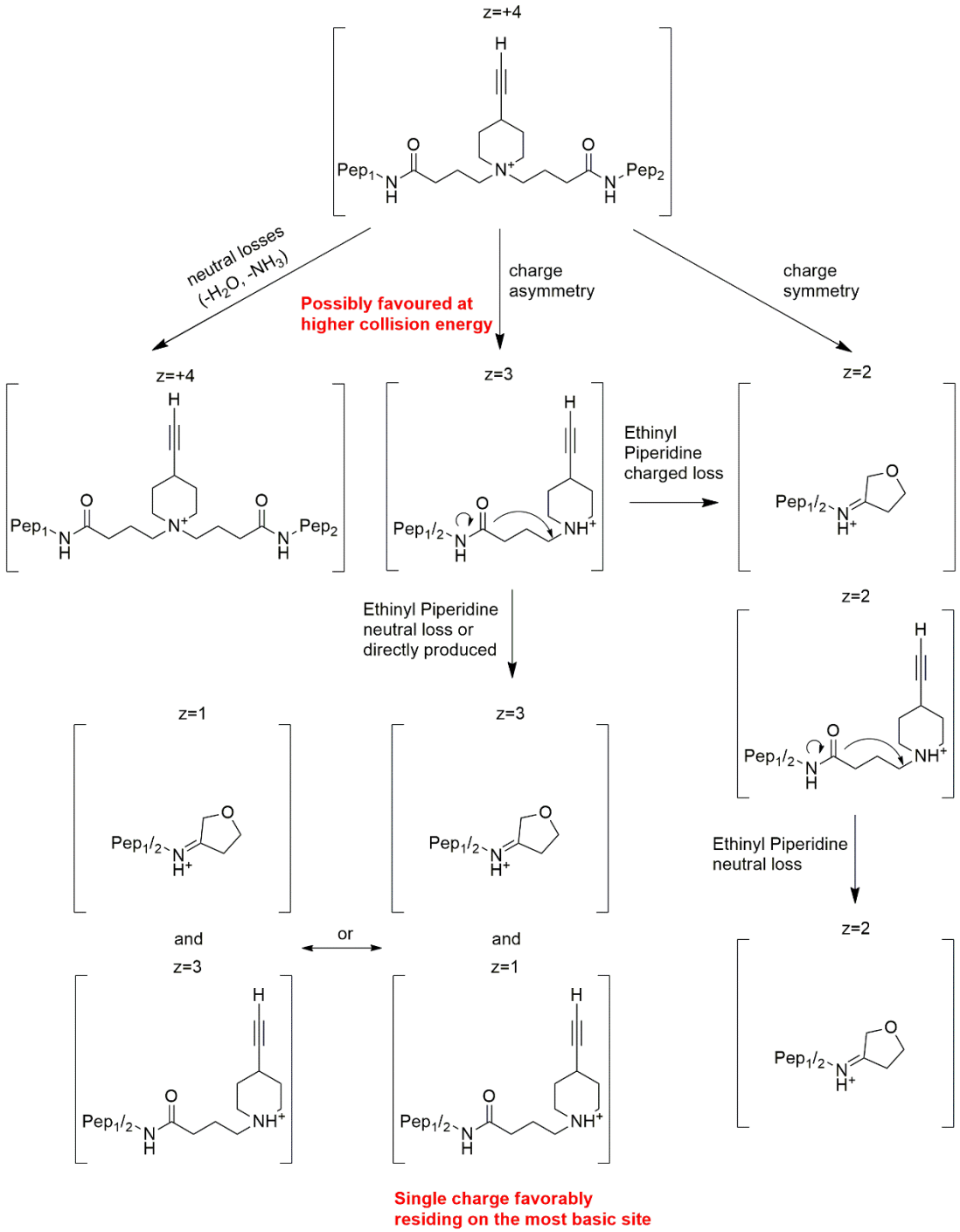


Figure A7: Collisional activated, even-electron dissociation pathways for a quadruply charged PAC4 cross-linked peptide.

A quadruply charged PAC4 cross-linked peptide can generate different product ions. Besides neutral losses on the 4+ precursor, if the charge is distributed symmetrically on the fragment ions, a 2+ furane-Pep and a 2+ piperidinium-Pep are generated. The second can promptly undergo further rearrangement (possible in HCD) through a neutral loss, generating again a 2+ furane-Pep ion. Increasing collision energy, singly charged piperidinium-Pep ions become more intense. This might be due to favored charge asymmetry, with formation of a 3+ furane-Pep (which can dissociate again) and a +1 piperidinium-Pep ion.

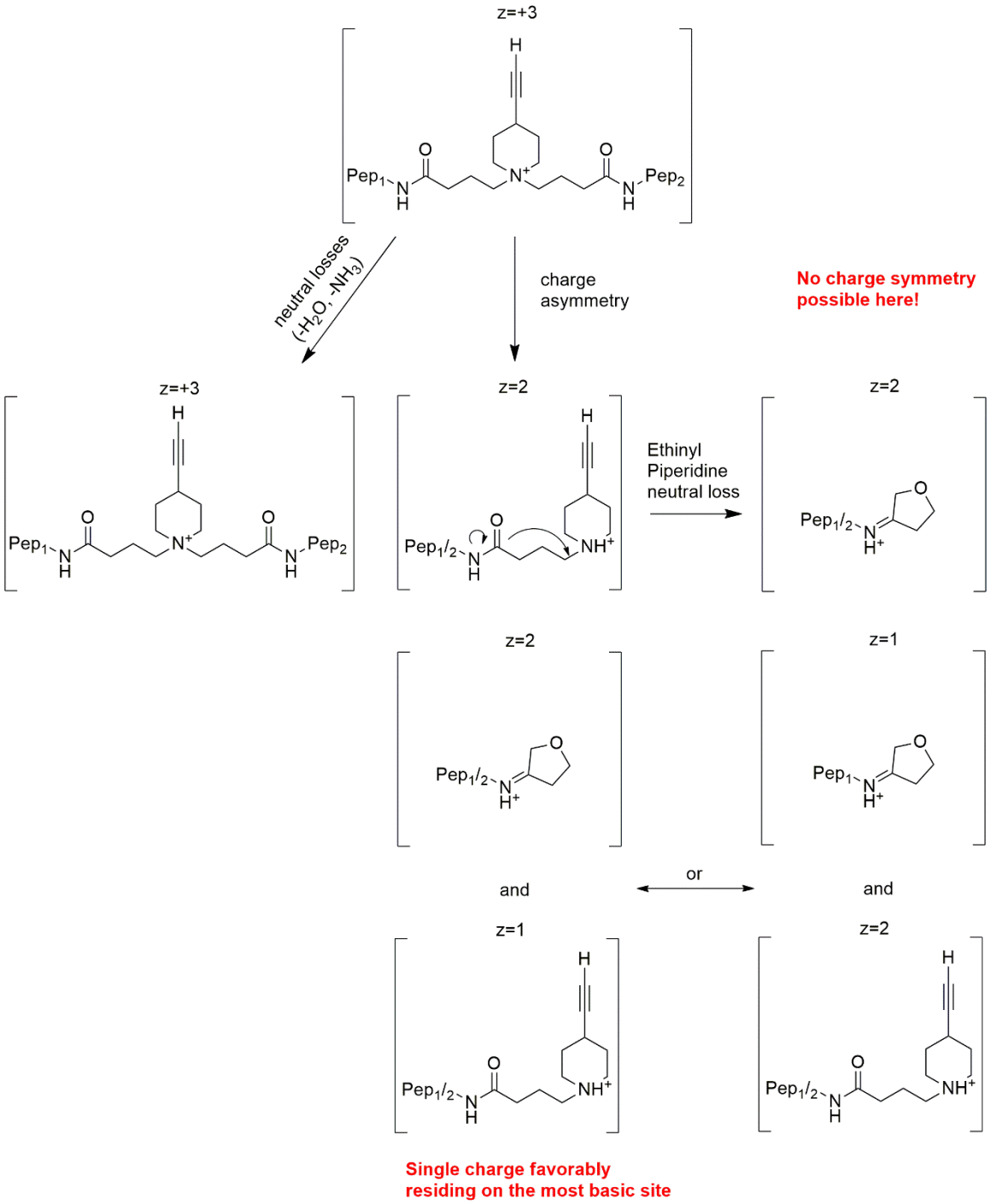


Figure A8: Collisional activated, even-electron dissociation pathways for a triply charged PAC4 cross-linked peptide.

A triply charged PAC4 cross-linked peptide can generate different product ions. Besides neutral losses on the 3+ precursor, the charge cannot be distributed symmetrically on the product ions. 2+ furane-Pep can be generated directly or through neutral loss from the 2+ piperidinium-Pep. The remaining +1 charge should then reside on the most basic site (the piperidine ring), leading to a higher intensity +1 piperidinium-Pep ion with increasing collision energy.

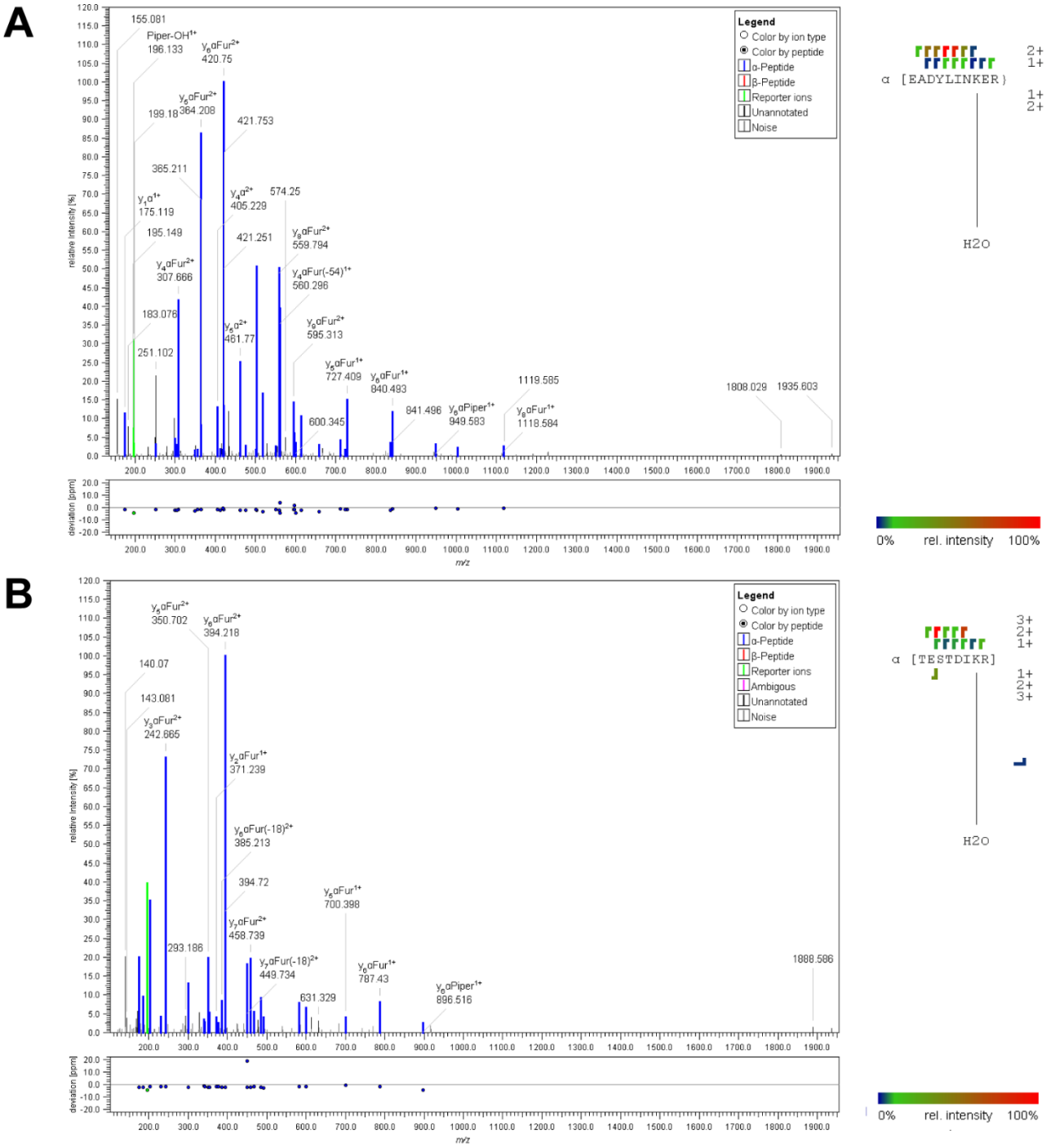


Figure A9: PAC4 cross-linked peptides show a prominent piperidine-carboxylic acid reporter ion in fragment ion mass spectra.

A) HCD fragment ion mass spectrum for the 'dead end' cross-link of peptide α (EADYLINKER). B) HCD-fragment ion mass spectrum for the 'dead end' cross-link of peptide β (TESTDIKR). Both spectra display a prominent piperidine-carboxylic acid reporter ion at *m/z* 196.133 (in green) that is absent in type II fragment ion mass spectra for TP1.

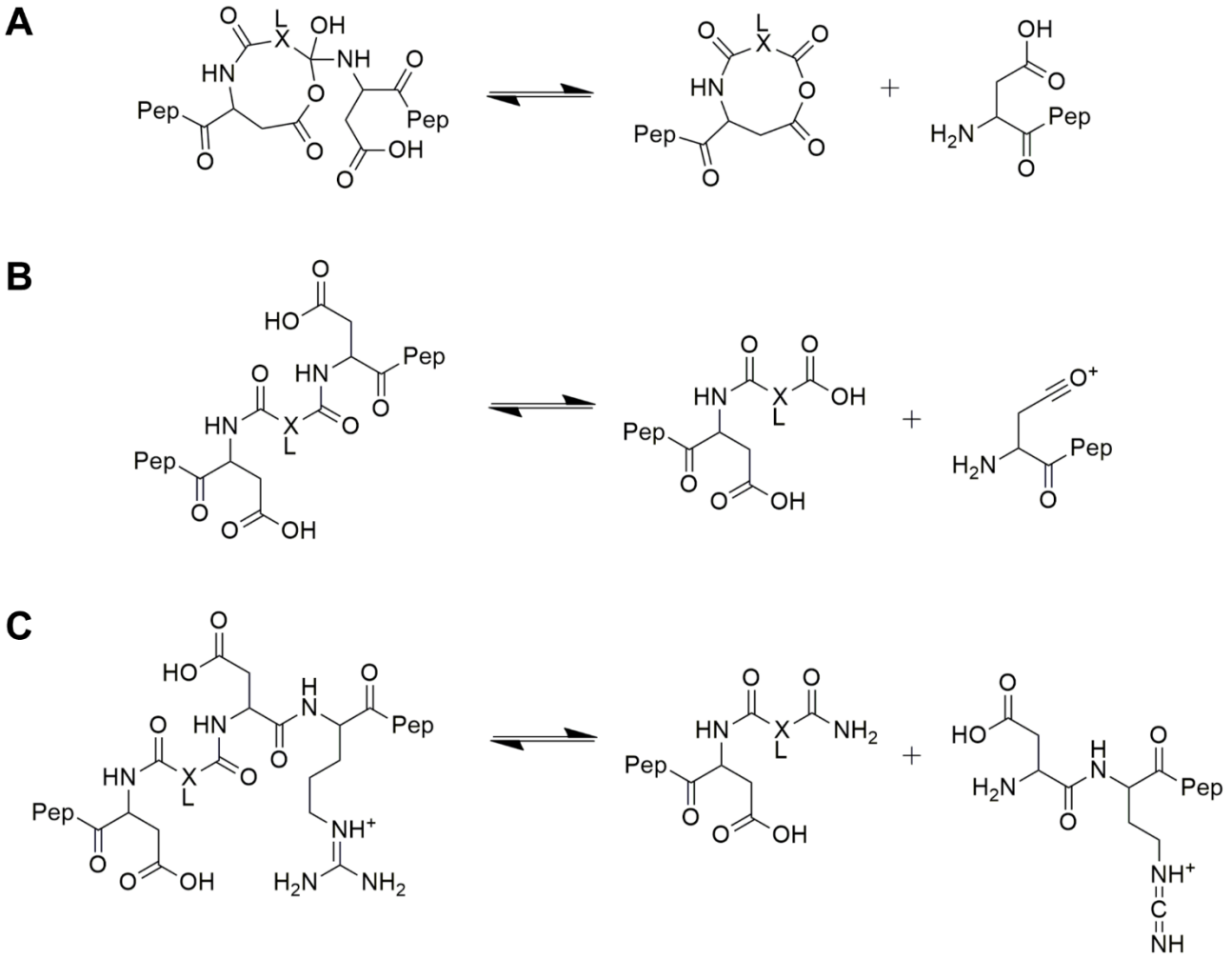


Figure A10: Cross-linked ATII exhibits unusual dissociation behavior upon collisional activation due to the presence of the N-terminal aspartate and the second arginine residues.

A) Formation of an isobaric species to a type I cross-linked peptide from a type II cross-linked peptide. B) Formation of a 'dead end' cross-link species from a type II cross-linked peptide. C) Formation of a 'dead end' amidated species from a type II cross-linked peptide.

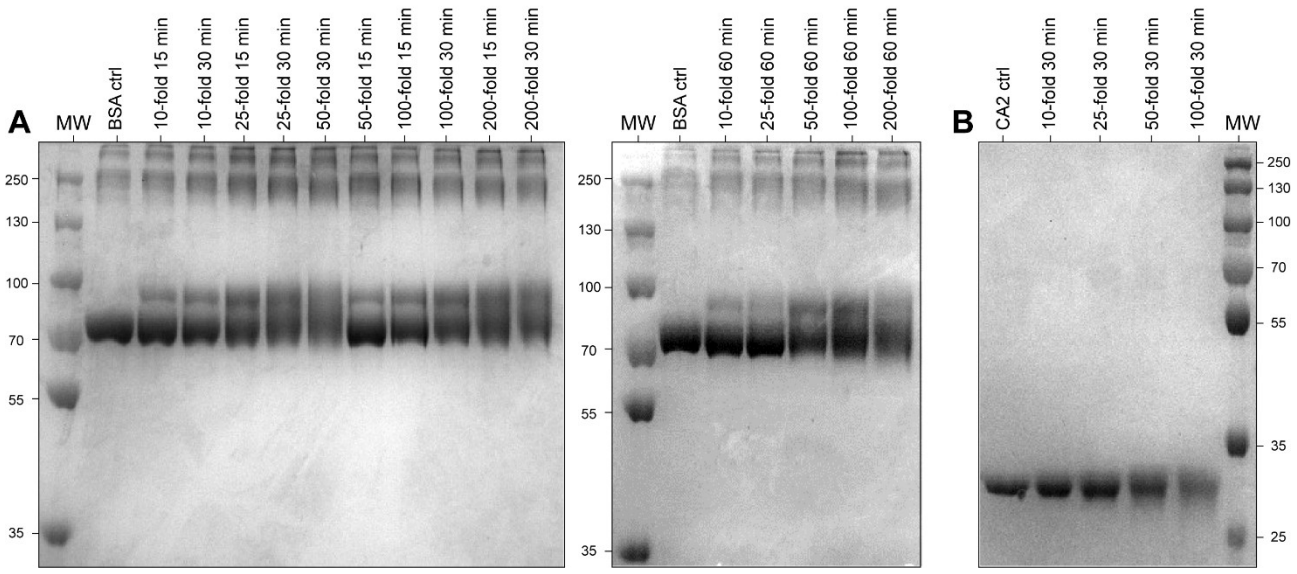


Figure A11: PAC4 testing on model proteins BSA and CA2.

A) PAC4 was tested in different molar excess (10-, 25-, 50-, 100-, 200-fold) and times (15, 30 and 60 minutes) on BSA. B) The cross-linker was then applied to CA2 with 10-, 25-, 50-, 100-fold excess for 30 minutes.

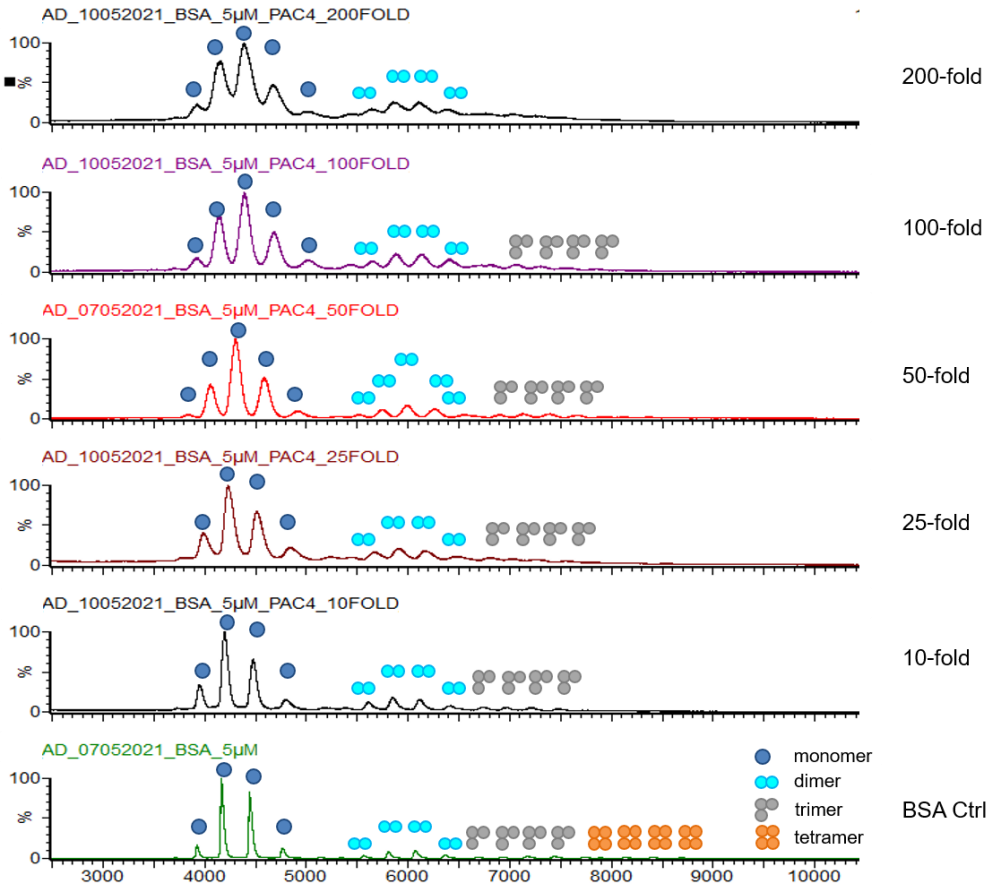


Figure A12: Native MS to monitor the number of PAC4 modifications on BSA.

Increasing PAC4 concentration, the BSA monomer, dimer and trimer distributions become broader and shift to higher *m/z* values, indicating successful modification of the protein.

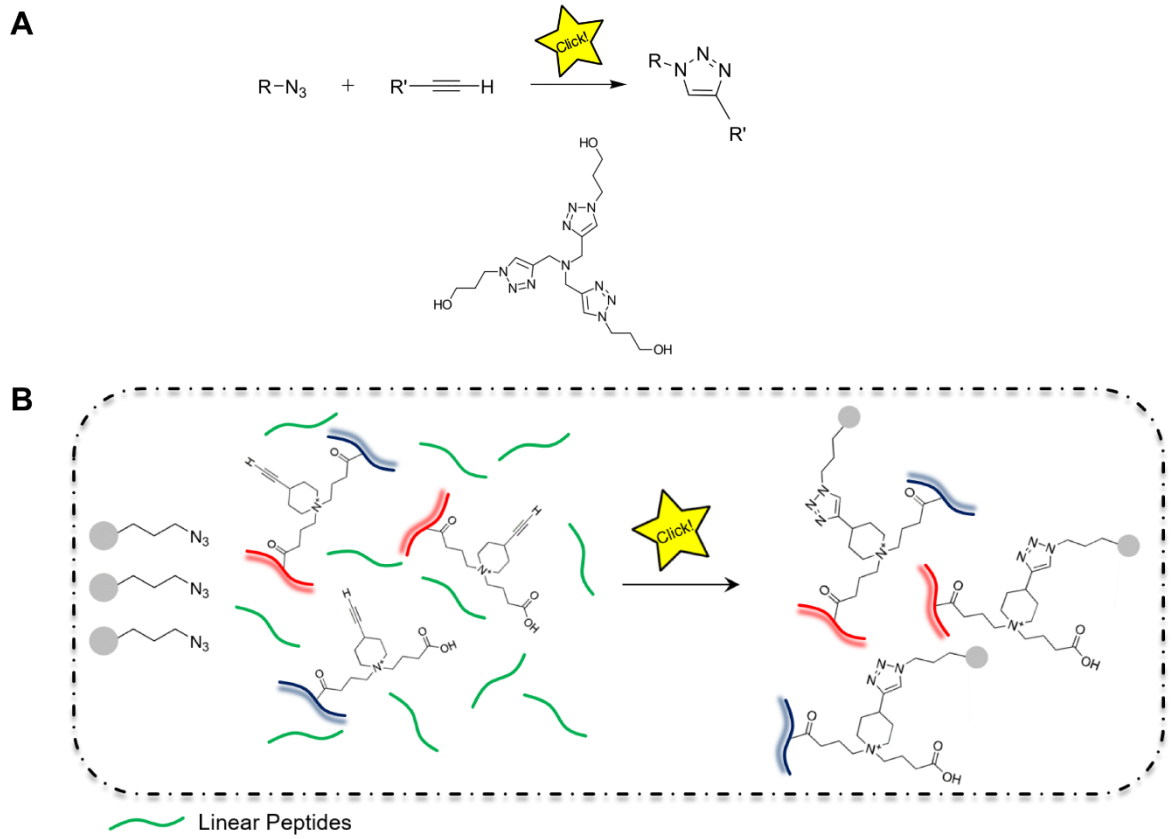


Figure A13: CuAAC reaction and enrichment of PAC4 cross-linked peptides.

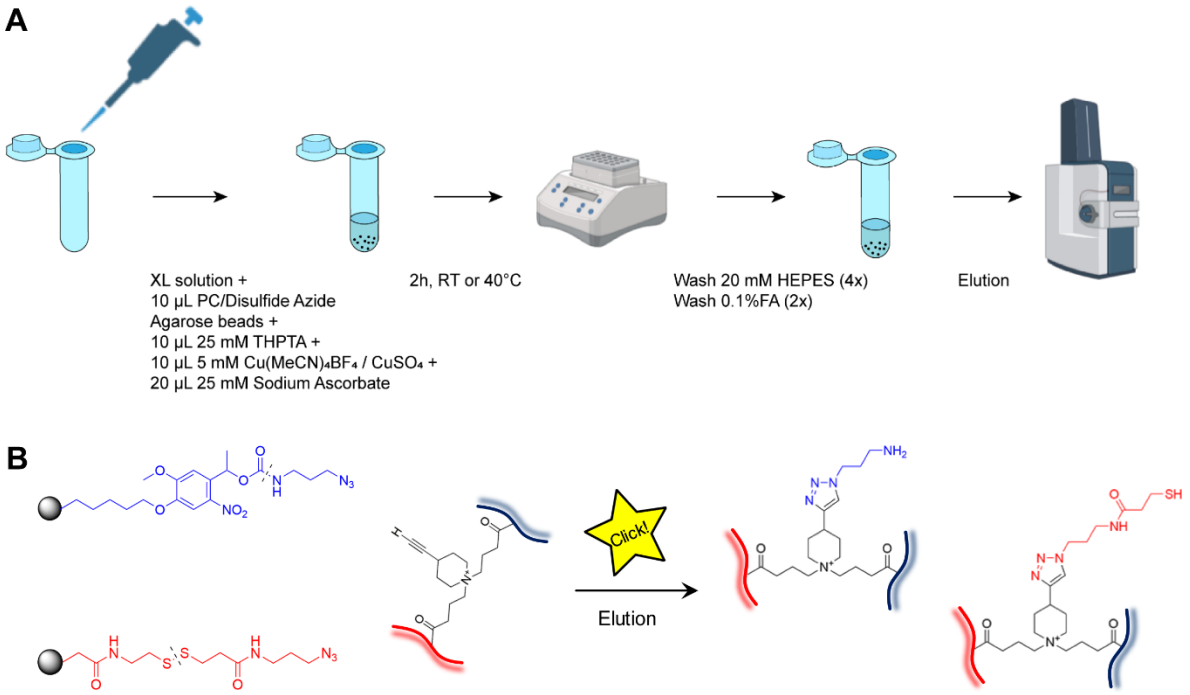


Figure A14: CuAAC enrichment protocol for the two different types of beads used in this work.

In this work, beads bearing a photocleavable (PCAB, blue) or a reducible group (DSAB, red) were evaluated for their enrichment performance.

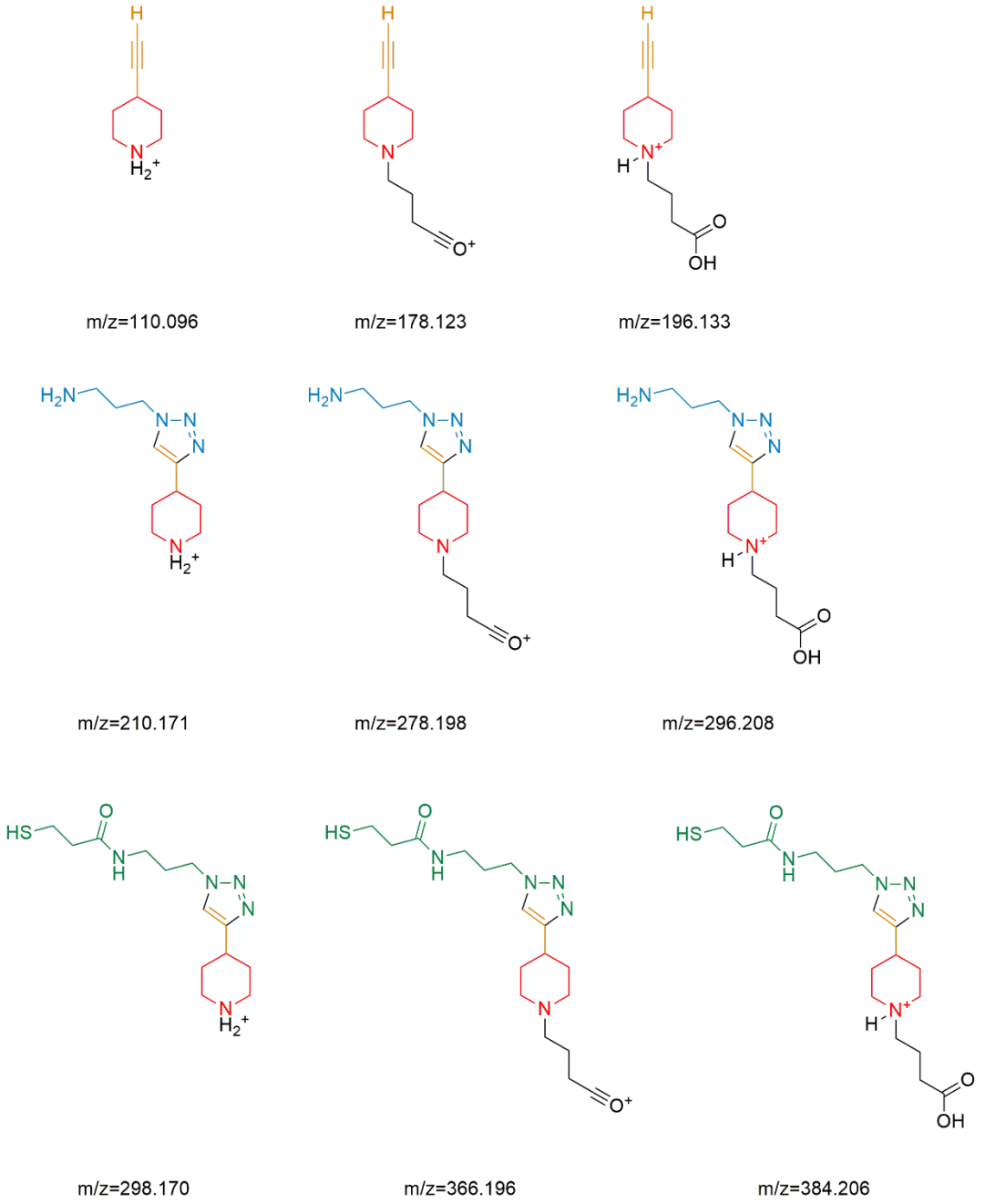


Figure A15: PAC4 reporter ions observed for cross-linked only peptides and CuAAC-enriched cross-linked peptides.

Similarly to PAC4 cross-links (top), clicked PAC4 cross-links enriched with both PCAB (center) and DSAB beads (bottom) generate reporter ions in the low *m/z* region to further increase confidence in the cross-link assignment.

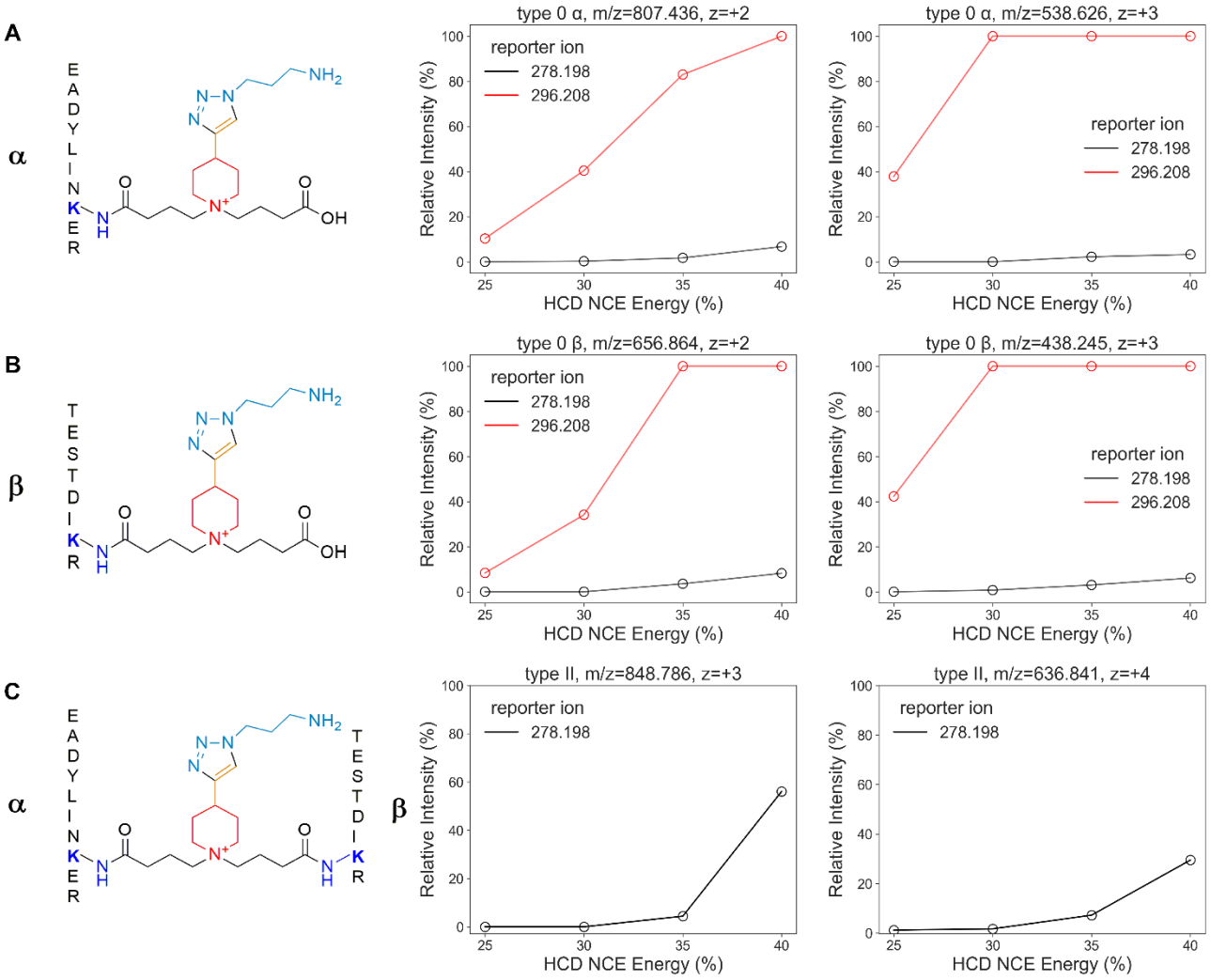


Figure A16: Reporter ion analysis of PCAB-enriched PAC4 TP1 cross-links.

Reporter ion analysis of PCAB-enriched PAC4 TP1 cross-linked products: The piperidine-acylium at *m/z* 278.198 (black) and the piperidine-carboxylic acid (red) at *m/z* 296.208 for (A) the type 0 'dead-end' species of TP1 peptide α EADYLINKER (2+, 3+ charge states); (B) the type 0 'dead-end' species of TP1 peptide β TESTDIKR (2+, 3+ charge states); (C) the type II cross-link of TP1 (3+, 4+ charge states). The relative intensity of each reporter ion for TP1 cross-linked products is plotted as a function of the HCD NCE (%) energy applied.

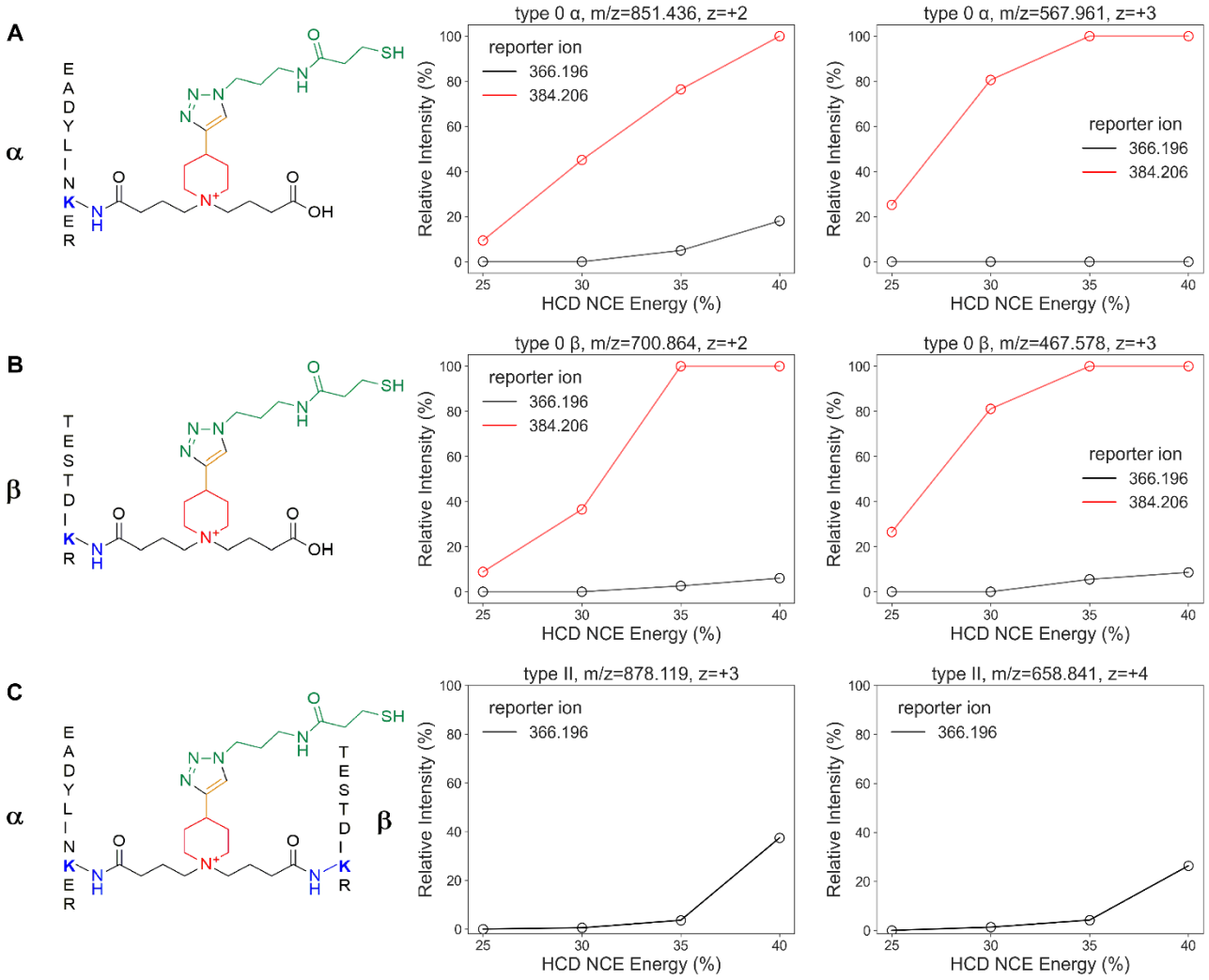


Figure A17: Reporter ion analysis of DSAB-enriched PAC4 TP1 cross-links.

Reporter ion analysis of DSAB-enriched PAC4 TP1 cross-linked products: The piperidine-acylium ion at *m/z* 366.196 (black) and the piperidine-carboxylic acid ion (red) at *m/z* 384.206 for (A) the type 0 'dead-end' species of TP1 peptide α EADYLINKER (2+, 3+ charge states); (B) the type 0 'dead-end' species of TP1 peptide β TESTDIKR (2+, 3+ charge states); (C) the type II cross-link of TP1 (3+, 4+ charge states). The relative intensity of each reporter ion for TP1 cross-linked products is plotted as a function of the HCD NCE (%) energy applied.

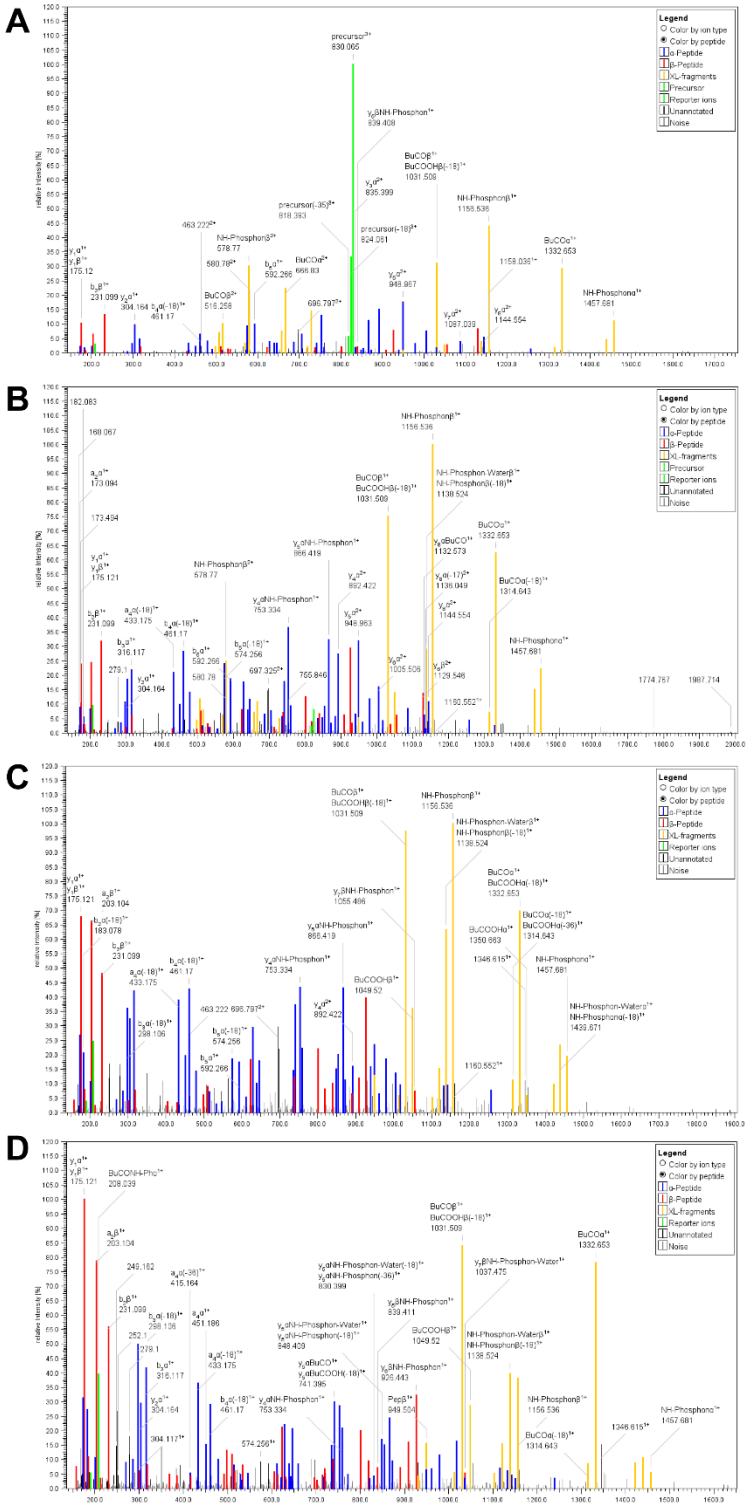


Figure A18: Fragment ion mass spectra of Test Peptide 1 (Creative Molecules Inc.) cross-linked with DSSI (Precursor ion at *m/z* 830.066, charge 3+, theoretical mass [M+H$^+$= 2488.160] acquired on an orbitrap fusion mass spectrometer, Thermo Fisher Scientific, San Jose, CA) at different HCD collision energies.

A) NCE=20%, B) NCE=25%, C) NCE=30%, D) NCE=35%. Diagnostic doublets of the cross-linker are shown in yellow, b- and y-type ions for the α and β peptides are presented in blue and red, respectively. Cross-linker reporter ions are shown in green.

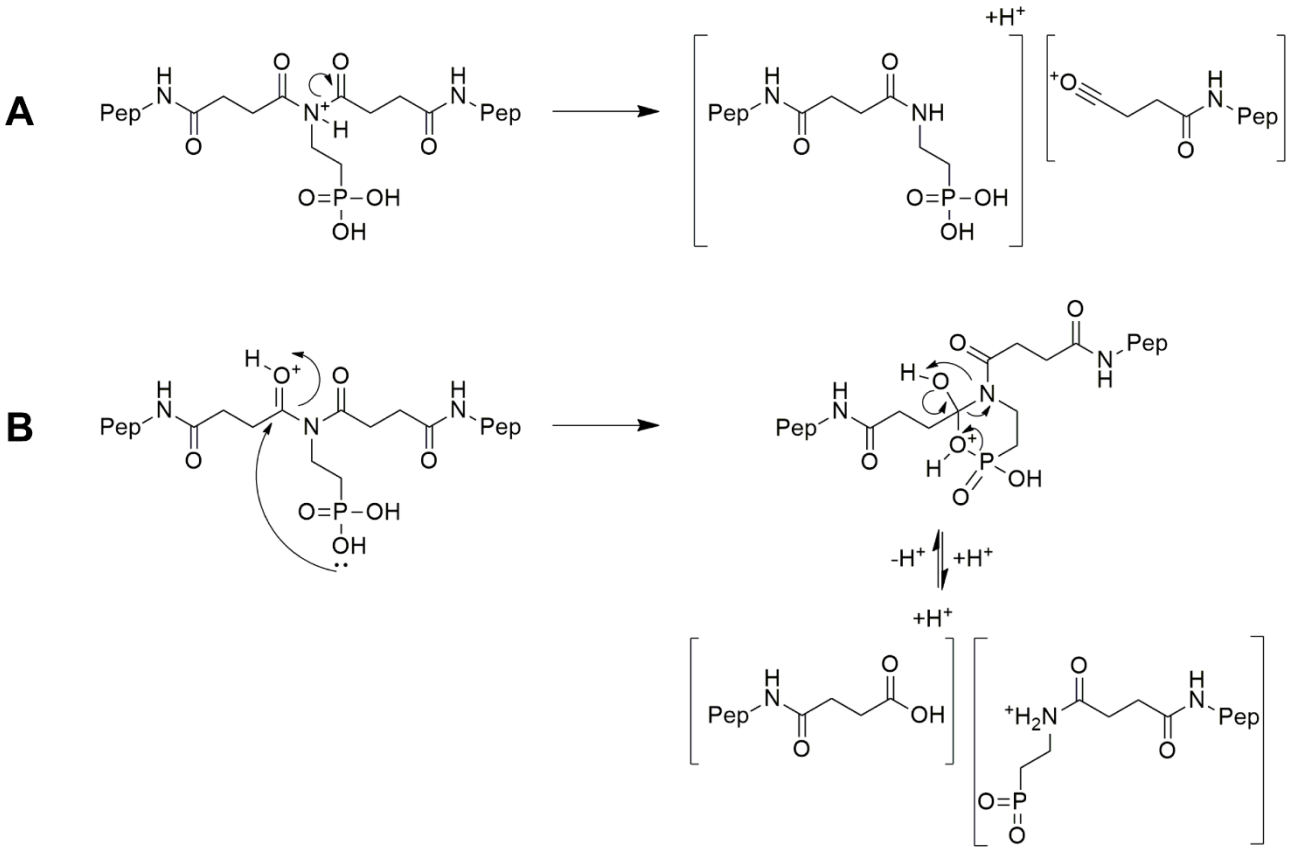


Figure A19: Proposed gas phase CID/HCD-induced dissociation pathways of the DSSI cross-linker.

A) DSSI's central imide moiety dissociates according to CID pathways reported for cyclic imides.

B) Hydroxy-mediated dissociation pathway due to the presence of the phosphonic acid.

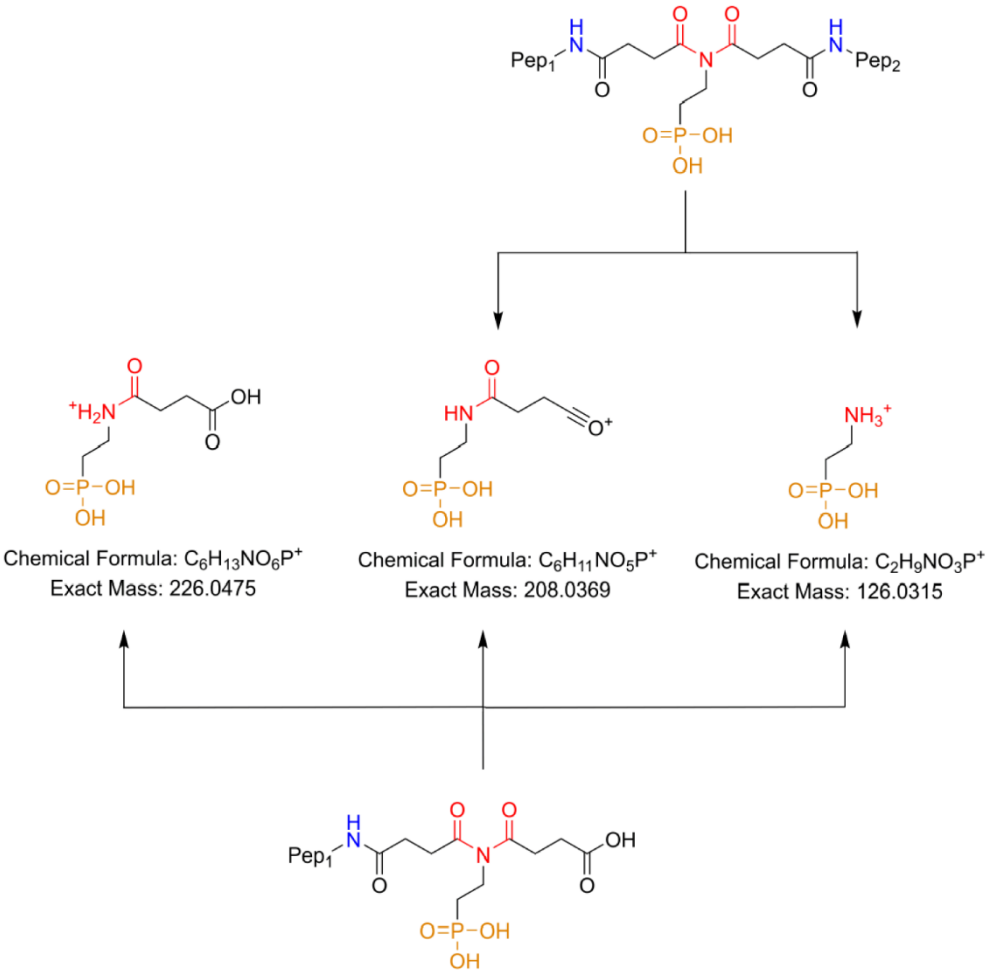


Figure A20: DSSI generates specific reporter ions for type II/I and type 0 cross-links.

Under high energy CID conditions, DSSI generates three different reporter ions at *m/z* 226.048, 208.037, and 126.032 depending on the type of cross-linked product.

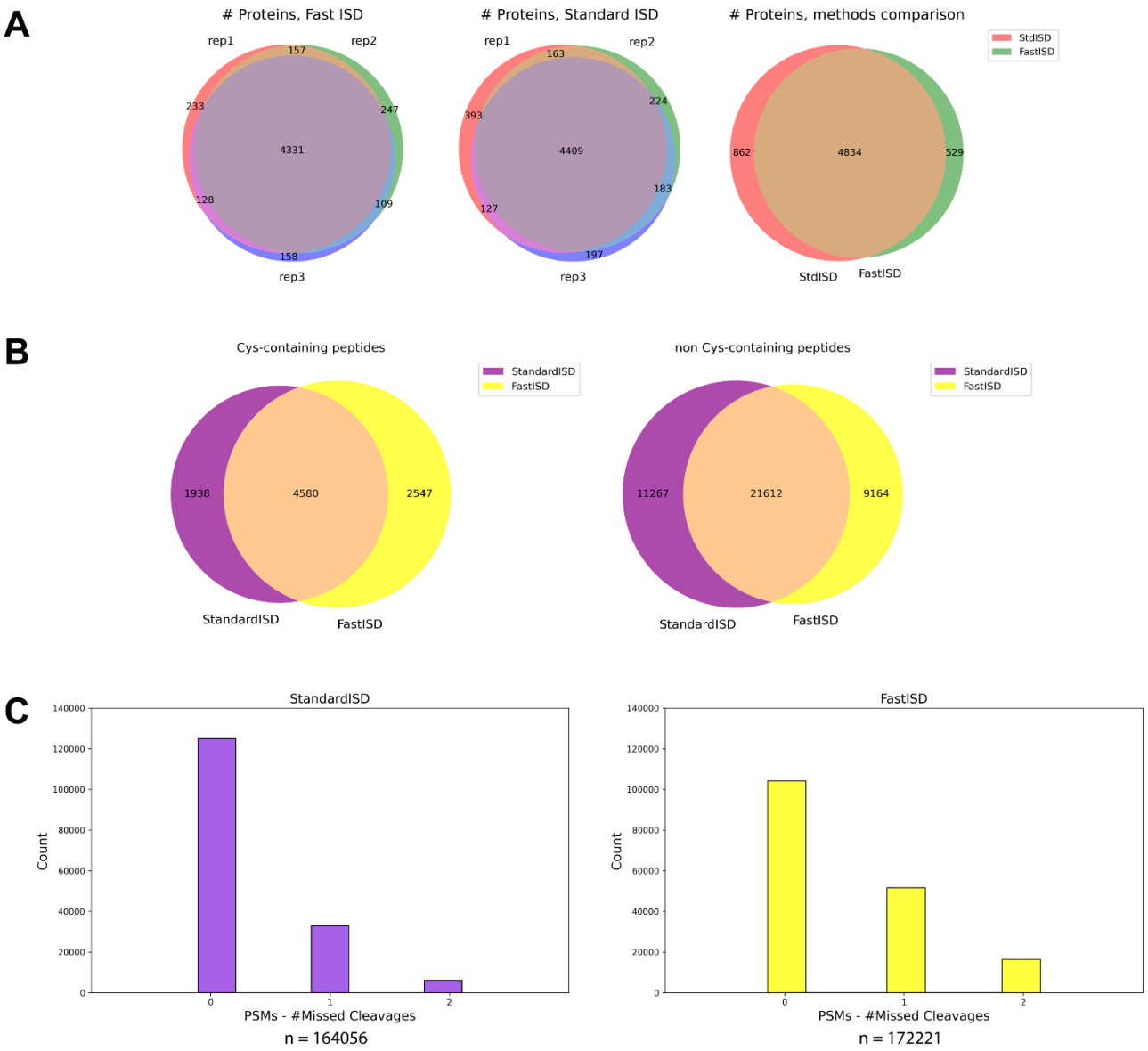


Figure A21: Evaluation of the fast in-solution digestion workflow on HEK293T cells.

A) Venn diagrams present the reproducibility of the digestion workflows. The overlap of identified proteins across three fast in-solution digestion replicates is presented on the left. The overlap of identified proteins across three standard in-solution digestion replicates is presented in the center. The overlap of identified proteins across the two digestion workflows is presented on the right. B) Comparison of identified Cys- and non Cys-containing peptides for the standard and fast ISD workflows. C) Histograms presenting the number of peptide spectral matches (PSMs), grouped by the number of missed cleavages, in the standard and fast in-solution digestion.

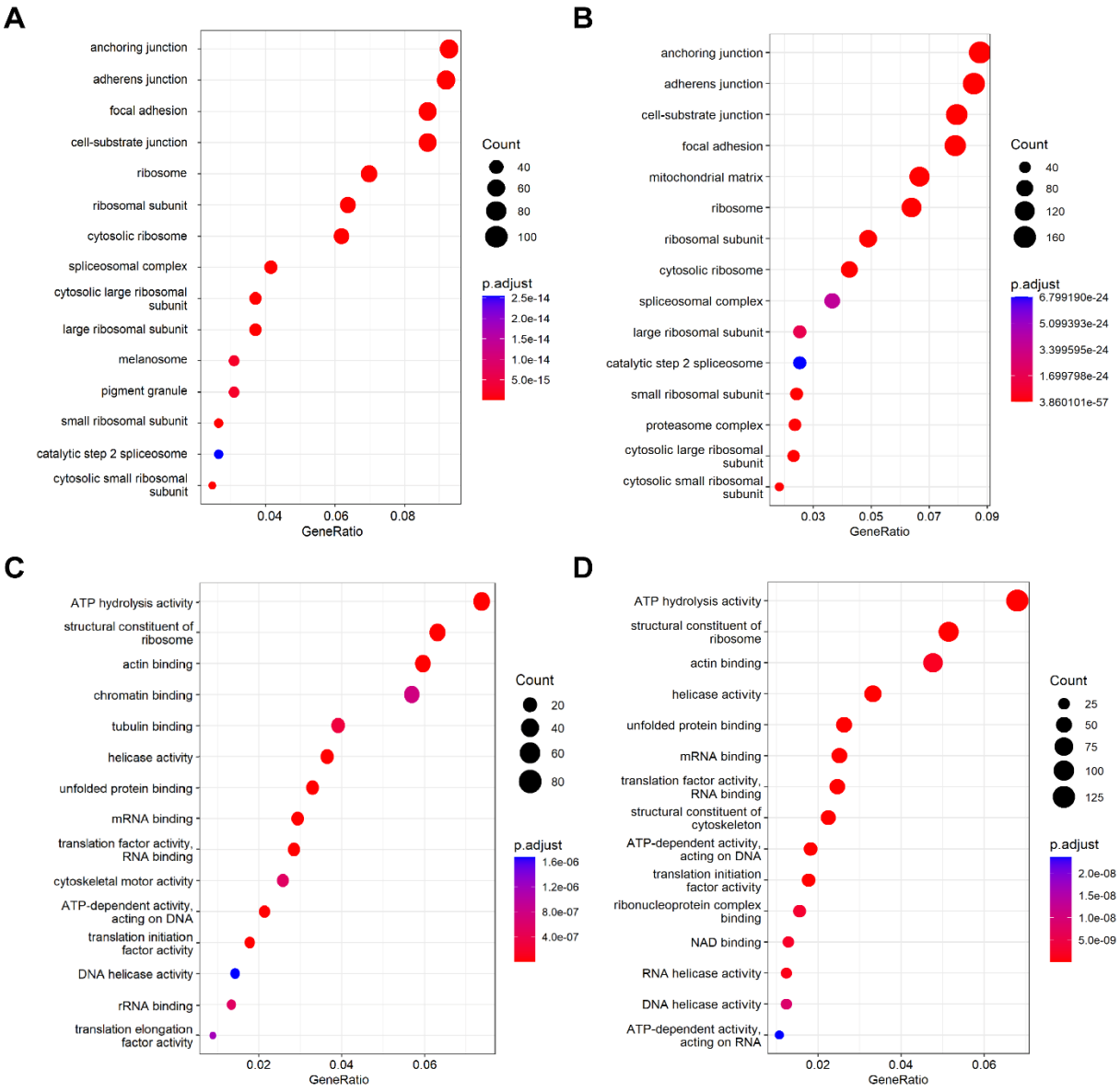


Figure A22: GO analysis of DSSI and DSBU cross-linked proteins.

A) GO analysis of cellular components for DSSI cross-linked proteins. B) GO analysis of cellular components for DSBU cross-linked proteins. C) GO analysis of molecular functions for DSSI cross-linked proteins. D) GO analysis of molecular functions for DSBU cross-linked proteins.

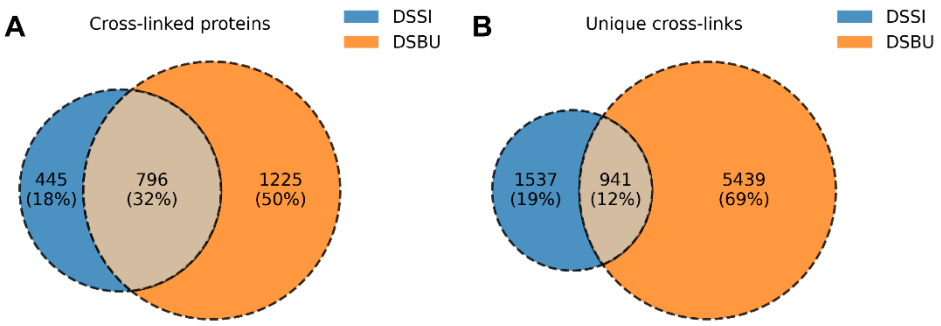


Figure A23: Venn diagram analysis of DSSI and DSBU cross-linked proteins and unique cross-links.

UniProt ID	DisProt ID	Protein name	Disorder content
Q13541	DP00028	Eukaryotic translation initiation factor 4E-binding protein 1	100
Q9NX55	DP00546	Huntingtin-interacting protein K	100
Q14978	DP01178	Nucleolar and coiled-body phosphoprotein 1	100
O60927	DP00219	E3 ubiquitin-protein ligase PPP1R11	100
P06454	DP01677	Prothymosin alpha	100
P62328	DP00357	Thymosin beta-4	100
Q9UQ35	DP03245	Serine/arginine repetitive matrix protein 2	99.06
P27797	DP00333	Calreticulin	95.92
P23588	DP03242	Eukaryotic translation initiation factor 4B	91.49
Q8IYB3	DP02359	Serine/arginine repetitive matrix protein 1	89.49
P38398	DP00238	Breast cancer type 1 susceptibility protein	83.2
Q96T21	DP00420	Selenocysteine insertion sequence-binding protein 2	79.51
Q09472	DP00633	Histone acetyltransferase p300	76.93
Q12888	DP02954	TP53-binding protein 1	76.62
P14678	DP03851	Small nuclear ribonucleoprotein-associated proteins B and B'	71.67
P25490	DP02595	Transcriptional repressor protein YY1	71.26
Q04637	DP02398	Eukaryotic translation initiation factor 4 gamma 1	70.61
Q9NYB9	DP02386	Abl interactor 2	69.79
P08621	DP02171	U1 small nuclear ribonucleoprotein 70 kDa	65.68
P50148	DP03235	Guanine nucleotide-binding protein G(q) subunit alpha	60.17
Q07157	DP02818	Tight junction protein ZO-1	59.78
Q8NC51	DP03637	Plasminogen activator inhibitor 1 RNA-binding protein	59.07
Q8WYP5	DP01549	Protein ELYS	55.12
P80303	DP03424	Nucleobindin-2	55
P52948	DP02123	Nuclear pore complex protein Nup98-Nup96	54.71
Q9H8Y8	DP02592	Golgi reassembly-stacking protein 2	53.98
P15311	DP00775	Ezrin	52.9
P13647	DP02606	Keratin, type II cytoskeletal 5	52.88
Q06787	DP00134	Synaptic functional regulator FMR1	52.37
P15927	DP01361	Replication protein A 32 kDa subunit	51.11
Q13153	DP03433	Serine/threonine-protein kinase PAK 1	48.44
Q16186	DP00839	Proteasomal ubiquitin receptor ADRM1	48.4
Q14011	DP03917	Cold-inducible RNA-binding protein	48.26
P06748	DP01474	Nucleophosmin	47.96
Q01081	DP01456	Splicing factor U2AF 35 kDa subunit	47.92
P53999	DP01804	Activated RNA polymerase II transcriptional coactivator p15	47.24
P05387	DP00793	60S acidic ribosomal protein P2	46.09
Q13177	DP03447	Serine/threonine-protein kinase PAK 2	45.8
Q00169	DP02327	Phosphatidylinositol transfer protein alpha isoform	45.19
P22626	DP01109	Heterogeneous nuclear ribonucleoproteins A2/B1	43.06
Q676U5	DP02148	Autophagy-related protein 16-1	42.67
P49790	DP01799	Nuclear pore complex protein Nup153	40.75
P11387	DP00075	DNA topoisomerase 1	40
O00268	DP01170	Transcription initiation factor TFIID subunit 4	39.72

P55957	DP01662	BH3-interacting domain death agonist	38.97
P09012	DP01857	U1 small nuclear ribonucleoprotein A	38.65
P56962	DP03670	Syntaxin-17	38.08
P47813	DP00903	Eukaryotic translation initiation factor 1A, X-chromosomal	36.81
Q9BUQ8	DP02332	Probable ATP-dependent RNA helicase DDX23	36.71
Q13148	DP01108	TAR DNA-binding protein 43	36.71
Q9H832	DP00953	Ubiquitin-conjugating enzyme E2 Z	35.88
P62318	DP01704	Small nuclear ribonucleoprotein Sm D3	35.71
O96028	DP03884	Histone-lysine N-methyltransferase NSD2	35.6
O75533	DP01863	Splicing factor 3B subunit 1	35.43
O14745	DP02000	Na(+)/H(+) exchange regulatory cofactor NHE-RF1	32.68
P41227	DP03148	N-alpha-acetyltransferase 10	31.91
P36578	DP01654	60S ribosomal protein L4	30.91
O96017	DP01797	Serine/threonine-protein kinase Chk2	30.2
P06899	DP01158	Histone H2B type 1-J	30.16
Q6FI81	DP01137	Anamorsin	30.13
P06400	DP01426	Retinoblastoma-associated protein	29.74
P04183	DP02006	Thymidine kinase, cytosolic	29.49
P62826	DP01364	GTP-binding nuclear protein Ran	29.17
P46108	DP00748	Adapter molecule crk	28.95
P84243	DP03908	Histone H3.3	27.94
Q8TEA8	DP02452	D-aminoacyl-tRNA deacylase 1	27.75
Q9UBE0	DP00485	SUMO-activating enzyme subunit 1	27.75
Q13162	DP02764	Peroxiredoxin-4	27.68
Q08209	DP00092	Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform	27.64
O43663	DP02316	Protein regulator of cytokinesis 1	27.42
O60493	DP03394	Sorting nexin-3	27.16
Q07955	DP01641	Serine/arginine-rich splicing factor 1	27.02
P19525	DP03944	Interferon-induced, double-stranded RNA-activated protein kinase	26.68
Q02413	DP01476	Desmoglein-1	26.31
Q14181	DP02759	DNA polymerase alpha subunit B	26.09
Q9Y6G9	DP02547	Cytoplasmic dynein 1 light intermediate chain 1	26
Q86UE8	DP02475	Serine/threonine-protein kinase tousled-like 2	25.91
Q9Y6K9	DP02269	NF-kappa-B essential modulator	25.54
O00571	DP02192	ATP-dependent RNA helicase DDX3X	25.23
Q92793	DP02004	CREB-binding protein	25.1
P41567	DP00910	Eukaryotic translation initiation factor 1	24.78
P83916	DP02856	Chromobox protein homolog 1	24.32
P46937	DP00702	Transcriptional coactivator YAP1	24.21
Q9HAU5	DP00949	Regulator of nonsense transcripts 2	23.98
Q15185	DP00358	Prostaglandin E synthase 3	23.75
P62805	DP02034	Histone H4	23.3
Q13188	DP01461	Serine/threonine-protein kinase 3	23.22
Q9UBT2	DP00486	SUMO-activating enzyme subunit 2	23.13
P09429	DP01493	High mobility group protein B1	22.79
O60885	DP01898	Bromodomain-containing protein 4	22.76

Q99598	DP03150	Translin-associated protein X	22.76
Q04206	DP00085	Transcription factor p65	22.5
P83731	DP01935	60S ribosomal protein L24	22.29
P58004	DP03004	Sestrin-2	22.29
P55036	DP01192	26S proteasome non-ATPase regulatory subunit 4	22.28
Q15910	DP01817	Histone-lysine N-methyltransferase EZH2	22.25
Q13547	DP02233	Histone deacetylase 1	21.99
P05455	DP00229	Lupus La protein	20.34
O00411	DP02703	DNA-directed RNA polymerase, mitochondrial	20.33
Q13042	DP01452	Cell division cycle protein 16 homolog	20.32
O96019	DP03958	Actin-like protein 6A	20.28
Q16543	DP01420	Hsp90 co-chaperone Cdc37	20.11

Table A1: Intrinsically disordered proteins cross-linked in HEK293T cells.

Acknowledgments

At the end of this PhD journey, I would like to extend my deepest gratitude to my supervisor, Prof. Dr. Andrea Sinz. Thanks for giving me the opportunity to join the lab, taking up challenging projects and introducing me to the world of protein mass spectrometry. Her passion for scientific research, even after so many achievements, is what always impressed me about her. Finally, her ability to push and motivate me to never stop increasing my knowledge and expertise in mass spectrometry is what I am most grateful for. Additionally, thanks for allowing me to join the RTG 2467 research program on intrinsically disordered proteins. This gave me the opportunity to get to know PhD students from different labs and backgrounds, discuss science and exchange ideas about my projects. I would also like to thank all the lab members of the Sinz group that have been part of my journey. A special thanks goes to Dr. Claudio Iacobucci, for the great mentorship and support on how to further improve results and manage difficult situations. A warm thank you as well to Dr. Christian Ihling, for sharing his technical expertise in mass spectrometry, for his calm and assistance when instruments were not working, and to Dr. Marc Kipping, for his knowledge in mass spectrometry and ion mobility; Dr. Christian Arlt and Dirk, for their expertise in protein biochemistry. A deep gratitude goes also to Mrs. Martina Mannd, Mrs. Herbrich-Peters, Xiaohan and Claudia for their contribution to the organization of the lab. Next, I would like to sincerely thank all my lab mates, with whom I spent enjoyable moments in and outside the lab. Starting from the “old” ones, Daniele and Lolita: thanks for being there for me, your presence was fundamental when I just arrived in Germany and remained very important over the years. Alan (fra’), during the PhD our friendship grew stronger and stronger, until the point I now consider you a true friend who I can always speak to. Thanks for reminding me, once again, how beautiful it is to encounter and embrace different cultures and points of view. Wiebke (also known as Wiebke eins or Pharmaziebke), I will never be able to thank you enough for everything you did for me, and I hope I gave you back at least a small part. Your unconditional and constant presence in these years, your support, your friendship, have been essential during my stay in Germany. Lydia, thanks for your kindness, your calm and lovely humour – you are the best office company someone can wish for and a great friend outside the lab. Juan, you are the perfect example that proves true friendship can be built in a short period of time. We share common moral principles and life pillars (such as family) and most importantly, nerd (that’s how we call each other), we also share the same passion for research and mass spectrometry – reason why we spent several late evenings at uni instead of going back home. Vishnu, thanks for your wisdom and contagious curiosity, you were always able to make me smile. Janett, Wiebke Schültze (referred to as Biochemiebke or Wiebke zwei), Florian, Hendrik, Christoph, Diego and Erik – thank you all for the great time I

pleasantly enjoyed spending with you. A heartfelt thanks goes also to all my good friends outside the Sinz lab: Luca, Jaime, Eleonora, Marina, Peter and Tala. Finally, to all the aforementioned people so far, I will keep and cherish all the beautiful moments and experiences we had together.

Likewise, I would like to express my sincere gratitude to all my old friends and family. Andrea, I cannot consider myself luckier than having you side by side since we were children. This continued until university, up to the point we are now pursuing a PhD (of course, both in mass spectrometry). Thanks for cheering me up and being an active source of inspiration. Without you, this work would not have been possible. Pingu, you taught me a lot of things and I cannot be more thankful to life for all the love you keep giving me. We are proof that if you really want something to work out, you fight for it with all your soul and spirit. Without your invaluable presence, this achievement would not have been possible. Finally, I want to thank my parents, Felice, and all my relatives for their continuous support.

Publications

1) Piersimoni L., et al., *Lighting up Nobel Prize-winning studies with protein intrinsic disorder.* Cellular and Molecular Life Sciences 2022. 79(8):449.

2) Di Ianni A., et al., *Structural assessment of the full-length wild-type tumor suppressor protein p53 by mass spectrometry-guided computational modeling.* Sci. Rep., 2023 13(1):8497.

3) Di Ianni A., Di Ianni A., et al., *Leveraging Cross-Linking Mass Spectrometry for Modeling Antibody-Antigen Complexes*. J. Proteome Res. 2024 23(3):1049-1061

4) Di Ianni A., et al., *Evaluating Imide-Based Mass Spectrometry-Cleavable Cross-Linkers for Structural Proteomics Studies.* JACS Au, 2024. 4(8):2936-2943.

Curriculum Vitae

Personal Data

Nationality	Italian

Education

10/2019-Present	PhD Martin Luther University Halle-Wittenberg, Halle (Saale), Germany
12/2017-Present	Pharmacist Professional Qualification
10/2012-09/2017	Unique Cycle Degree in Chemistry and Pharmaceutical Technology
2006-2012	Scientific High School Diploma

Work experience

07/2024-Present	Structural Proteomics Specialist, Human Technopole Foundation - Milan, Italy
02/2024-04/2024	Research Associate, University of L'Aquila
10/2019-Present	PhD Candidate – Protein Mass Spectrometry, Department of Pharmaceutical Chemistry and Bioanalytics, Center for Structural Mass Spectrometry, Martin Luther University Halle-Wittenberg
04/2019-09/2019	Research Trainee – Analytical Chemistry, University of Parma

04/2019-09/2019	Pharmacist, Farmacie Comunali Riunite, Reggio Emilia
06/2017-12/2017	Computational Chemist, Chiesi Farmaceutici S.p.A., Parma Chemistry Research and Drug Design Department

Language skills

- Italian (native speaker)
- English (professional working proficiency)
- German (elementary working proficiency)
- Spanish (elementary working proficiency)

Declaration

I confirm that I have written this dissertation independently, without any unapproved external assistance or involvement from others, and I have only utilized sources properly cited within the work. All text sections, whether quoted directly or paraphrased from published or unpublished materials, as well as any information gathered from spoken sources, are appropriately referenced and listed following the correct bibliographic guidelines. In carrying out this research, I adhered to the principles of proper scientific conduct.

Halle (Saale) _____

Alessio Di Ianni