

---

Mass spectrometry-based methods for the analysis of  
proteins and protein complexes of neuronal model systems

---

Dissertation  
zur Erlangung des  
Doktorgrades der Naturwissenschaften (Dr. rer. nat.)  
der  
Naturwissenschaftlichen Fakultät I – Biowissenschaften –  
der Martin-Luther-Universität  
Halle-Wittenberg,

vorgelegt  
von Frau **Marie Barth**  
geboren am 17. Februar 1994 in Dresden

eingereicht am: 20.01.2022

verteidigt am: 31.05.2022

**Gutacher\*innen**

1. Jun.-Prof. Dr. Carla Schmidt, Martin-Luther-Universität Halle-Wittenberg
2. Prof. Dr. Sven-Erik Behrens, Martin-Luther-Universität Halle-Wittenberg
3. Prof. Dr. Katrin Marcus, Ruhr-Universität Bochum



## Table of contents

List of abbreviations .....	i
One letter code of amino acids .....	iii
List of figures .....	iv
List of tables .....	vi
<b>Summary</b> .....	vii
<b>Zusammenfassung</b> .....	ix
<b>1. Introduction</b> .....	<b>1</b>
1.1 Structure and function of neurons .....	1
1.2 The synapse and the synaptic vesicle cycle .....	3
1.3 SH-SY5Y cells as neuronal model system .....	5
1.4 Identification of proteins by liquid chromatography-coupled mass spectrometry .....	7
1.5 Proteome analysis and protein quantification strategies .....	10
1.6 Structural proteomics .....	15
1.6.1 Native mass spectrometry .....	16
1.6.2 Cross-linking mass spectrometry .....	18
1.6.1 Labeling mass spectrometry.....	22
1.7 Aim of this study.....	24
<b>2. Materials and methods</b> .....	<b>27</b>
2.1 Materials .....	27
2.1.1 Chemicals.....	27
2.1.2 Cell lines, proteins and biological material .....	27
2.1.3 Chromatographic material and consumables .....	28
2.1.4 Buffers, media, ready-to-use solutions and kits .....	28
2.1.5 Instruments.....	29
2.1.6 Software .....	29
2.2 Methods.....	30
2.2.1 Cell culture .....	30
2.2.2 Differentiation of SH-SY5Y cells .....	30
2.2.3 Cross-linking using formaldehyde .....	31
2.2.4 Labeling of model proteins with NHS-acetate .....	32
2.2.5 Labeling of model proteins with DEPC .....	32
2.2.6 Labeling of synaptic vesicle proteins with sulfo-NHS-acetate or NHS-acetate .	33
2.2.7 Labeling of synaptic vesicle proteins with DEPC.....	33
2.2.8 Sample preparation for LC-MS/MS based proteomics .....	33
2.2.9 Sample preparation for LC-MS/MS analysis of cross-linked samples .....	34

2.2.10 In-gel digestion .....	34
2.2.11 Ethanol precipitation.....	35
2.2.12 In-solution digestion in the presence of urea .....	36
2.2.13 In-solution digestion in the presence of RapiGest.....	36
2.2.14 Desalting of peptides using ZipTips .....	36
2.2.15 Desalting of peptides using Pierce <sup>TM</sup> peptide desalting columns .....	37
2.2.16 Enrichment of cross-linked peptide pairs .....	37
2.2.17 LC-MS/MS analysis .....	37
2.2.18 Proteomic data analysis using MaxQuant.....	39
2.2.19 Relative quantification of protein expression of proteins with a specific subcellular localization .....	39
2.2.20 Quantification of global protein abundances .....	40
2.2.21 Identification of cross-linked peptide pairs .....	40
2.2.22 Identification of chemically labeled amino acids .....	41
2.2.23 Quantification of modified amino acids of model proteins .....	42
2.2.24 Calculation of the solvent accessible surface area (SASA) and visualization of labeled residues on model proteins .....	43
2.2.25 Quantification of labeled residues of synaptic vesicle proteins.....	43
2.2.26 Data availability .....	44
<b>3. Results .....</b>	<b>45</b>
3.1 Proteomic characterization of SH-SY5Y cells .....	45
3.1.1 Workflow for proteome analysis of undifferentiated and differentiated SH-SY5Y cells .....	45
3.1.2 Relative quantification of proteins with a specific subcellular localization .....	46
3.1.3 Proteome analysis of RA- and RA/PMA-differentiated cells uncovers markers for early neuronal differentiation .....	48
3.1.4 KEGG pathway analysis revealed structural rearrangements during neuronal differentiation .....	51
3.1.5 Proliferation is a main characteristic of undifferentiated SH-SY5Y cells.....	53
3.2 Protein cross-linking in SH-SY5Y cells .....	56
3.2.1 Workflow for in-cell cross-linking .....	56
3.2.2 Identification of cross-linked peptide pairs .....	56
3.2.3 Cross-links identified in ribosomal proteins.....	59
3.2.4 Protein interaction networks of SH-SY5Y cells.....	62
3.3 Quantitative labeling of model proteins .....	67
3.3.1 Chemical labeling workflow of model protein .....	67
3.3.2 Characterization of fragment spectra of labeled peptides .....	68

3.3.3 Labeling efficiency of NHS-acetate and DEPC.....	73
3.3.4 Quantification of labeled residues for determination of solvent accessible amino acids on the surface of protein complexes .....	75
3.3.5 Ion mobility MS of modified protein complexes.....	78
3.4 Solvent accessibility of synaptic vesicle proteins probed by chemical labeling .....	81
3.4.1 Labeling of synaptic vesicle proteins using three different chemical labeling reagents .....	81
3.4.2 Labeling of the synaptic V-ATPase complex .....	85
3.4.3 Labeling percentage of individual residues of synaptic vesicle proteins.....	89
3.4.4 Solvent accessibility of synaptic vesicle proteins .....	92
<b>4. Discussion and outlook.....</b>	<b>99</b>
4.1 Proteome analysis of undifferentiated and differentiated SH-SY5Y cells .....	99
4.2 In-cell cross-linking of SH-SY5Y cells .....	104
4.3 Chemical labeling for identification of solvent accessible amino acid residues .....	106
4.4 Labeling of synaptic vesicle proteins.....	110
<b>5. References.....</b>	<b>119</b>
<b>6. Appendix .....</b>	<b>137</b>
6.1 Supplementary Tables.....	137
6.2 Supplementary Figures .....	146
Acknowledgment.....	148
Curriculum vitae.....	149
Publication list .....	150
Affidavit .....	154
Eidesstattliche Erklärung.....	155



## List of abbreviations

di-CEt-histidine	1,3-dicarboethoxyhistidine
EDC	1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride
MES	2-(N morpholino)ethansulfonic acid
CAA	2-chloroacetamide
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
ACN	acetonitrile
ARP 2/3	actin related protein 2/3
AP-2 complex	adaptor-related protein complex 2
ADP	adenosine diphosphate
ATP	adenosine triphosphate
ATP	adenosine triphosphate
ADH	alcoholdehydrogenase
AGC	automatic gain control
BS3	bis(sulfosuccinimidyl)suberat
BDNF	brain-derived neurotrophic factor
CEt	carboethoxy
CID	collision-induced dissociation
CSP	cysteine string protein
DIA	data independent acquisition
DDA	data-dependent acquisition
DNA	deoxyribonucleic acid
DEPC	diethylpyrocarbonate
DSS	disuccinimidyl suberate
DTT	dithiothreitol
ECD	electron capture dissociation
ETD	electron transfer dissociation
ESI	electrospray-ionization
ER	endoplasmic reticulum
emPAI	exponentially modified protein abundance index
ECM-receptor	extracellular matrix receptor
FDR	false discovery rate
FPOP	fast photochemical oxidation of proteins
FCS	fetal calf serum
FASP	filter-aided sample preparation
FA	formic acid
formyl-CEt histidine	formyl-carboethoxyhistidine
GPCR	G protein-coupled receptors
GDP	guanosine diphosphate
GTP	guanosine triphosphate
HCD	high-energy collisional dissociation

FAIMS	high-field asymmetric waveform ion mobility spectrometry
HDX	Hydrogen-Deuterium exchange
HRF	hydroxyl radical foot printing
IAA	iodoacetamide
iTRAQ	isobaric tags for absolute and relative quantification
ISOC1	isochorismatase domain containing protein 1
KEGG	Kyoto Encyclopedia of Genes and Genomes
LFQ	label free quantification
LC	liquid chromatography
LC-MS/MS	liquid chromatography-coupled mass spectrometry
MS	mass spectrometry
mRNA	messenger ribonucleic acid
MCM	minichromosome maintenance complex component
EMEM-medium	minimum essential medium eagle
MRM	multiple reaction monitoring
NHS-acetate	N-hydroxysuccinimidyl acetate
NSAF	normalized spectral abundance factor
PRM	parallel reaction monitoring
PSM	peptide spectrum match
PMA	phorbol-12-myristat-13-acetat
PBS	phosphate buffered saline
PTM	post-transcriptional modification
PAF	protein abundance factor
PAI	protein abundance index
PK	pyruvate kinase
RIM	Rab3-interacting molecule
Rab3a	Ras-related protein Rab-3A
RA	retinoic acid
SPEED	Sample Preparation by Easy Extraction and Digestion
SCAM1	secretory carrier-associated membrane protein 1
SRM	selected reaction monitoring
SWATH	sequential window acquisition of all theoretical fragment ion spectra
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SP3	single-pot, solid-phase-enhanced sample-preparation
AT1A3	sodium/potassium-transporting ATPase subunit alpha-3
SDS-PAGE	sodiumdodecyl sulphate-polyacrylamide gel electrophoresis
SNARE	soluble N-ethylmaleimide-sensitive-factor attachment receptor
SASA	solvent accessible surface area
SILAC	stable isotope labeling by amino acids in cell culture
S-NHS-acetate	sulfo-N-hydroxysuccinimidyl acetate
SV2A	synaptic vesicle glycoprotein 2A
TMT	tandem mass tag



iBAQ	the intensity based absolute quantification
TOF	time-of-flight
TFA	trifluoroacetic acid
TCEP	tris(2-carboxyethyl)phosphine
urethane-CET histidine	urethane-carboethoxyhistidine
V-ATPase	vacuolar H <sup>+</sup> -ATPase complex
VAcHT	vesicular acetylcholine transporter
VGLU1	vesicular glutamate transporter 1
VGLUT	vesicular glutamate transporters
VMAT	vesicular monoamine transporters
VGAT	vesicular $\gamma$ -aminobutyric acid transporter
GABA	$\gamma$ -aminobutyric acid

### One letter code of amino acids

amino acid	one letter code
alanine	A
arginine	R
asparagine	N
aspartic acid	D
asparagine or aspartic acid	B
cysteine	C
glutamic acid	E
glutamine	Q
glutamine or glutamic acid	Z
glycine	G
histidine	H
isoleucine	I
leucine	L
lysine	K
methionine	M
phenylalanine	F
proline	P
serine	S
threonine	T
tryptophan	W
tyrosine	Y
valine	V

## List of figures

Figure 1: Structural cartoon of a neuron .....	2
Figure 2: Synaptic vesicle cycle within the chemical synapse.....	4
Figure 3: Protein identification by mass spectrometry.....	8
Figure 4: Workflow for proteomic sample preparation and analysis.....	11
Figure 5: MS-based protein quantification methods .....	12
Figure 6: Representative native MS spectrum .....	17
Figure 7: Cross-linking workflow .....	19
Figure 8: Formaldehyde cross-linking reaction mechanism .....	20
Figure 9: Labeling of proteins.....	23
Figure 10: Differentiation of SH-SY5Y cells .....	45
Figure 11: Workflow – Proteomic analysis of SH-SY5Y cells .....	46
Figure 12: Relative quantification of proteins with a specific subcellular localization .....	47
Figure 13: Relative quantification of undifferentiated versus RA- and RA/PMA- differentiated SH-SY5Y cells .....	49
Figure 14: Enriched KEGG pathways in differentiated cells.....	52
Figure 15: Enriched KEGG pathways in undifferentiated cells.....	55
Figure 16: Workflow for in-cell cross-linking of proteins using formaldehyde .....	56
Figure 17: Example spectrum for the identification of an inter-molecular protein cross-link .....	58
Figure 18: Comparison of identified cross-linked proteins, inter-protein cross-links and intra-protein cross-links .....	59
Figure 19: Inter- and intra-molecular cross-links of ribosomal proteins .....	61
Figure 20: Protein interaction network of actin gamma 1.....	63
Figure 21: Detailed protein interaction networks of actin gamma 1 .....	64
Figure 22: Workflow for chemical labeling of model proteins using NHS-acetate or DEPC .....	68
Figure 23: Example spectrum of NHS-acetate modified peptide .....	69
Figure 24: Example spectra of DEPC labeled peptides .....	70
Figure 25: Possible mechanisms of neutral loss from DEPC labeled residues .....	72
Figure 26: Reproducibility of NHS-acetate or DEPC labeling of model proteins .....	73
Figure 27: Data analysis strategy .....	76
Figure 28: Identification of solvent accessible amino acid residues of ADH and PK .....	77
Figure 29: Ion mobility MS of NHS-acetate and DEPC labeled ADH .....	80
Figure 30: Labeling workflow for synaptic vesicle proteins .....	82

Figure 31: Overview of labeled proteins and residues .....	83
Figure 32: Comparison of labeled lysine, serine, threonine and tyrosine residues using S-NHS/NHS-acetate or DEPC .....	84
Figure 33: Labeling of solvent accessible residues of the V-ATPase .....	86
Figure 34: Detailed assessment of modified residues in the 'V0' domain of the V-ATPase complex .....	88
Figure 35: Cartoon models of labeled synaptic vesicle proteins .....	90
Figure 36: Comparison of identified modified residues identified in synaptic vesicles proteins using S-NHS-acetate or NHS-acetate .....	92
Figure 37: Labeled residues of synapsin-1, Rab3a and CSP .....	94
Figure 38: Modified residues of synaptophysin, synaptoporin, synaptogyrin-1 and SV2A by DEPC and S-NHS/NHS-acetate .....	98
Figure 39: High-resolution structures of ADH and PK.....	108
Figure 40: Solvent accessibility and protein interactions in the V-ATPase complex .....	113
Figure 41: Protein interactions identified in flexible luminal loops.....	117
Supplementary Figure 1: Protein interaction network observed by formaldehyde cross-linking.....	146
Supplementary Figure 2: Example spectra of DEPC labeled peptides .....	147

## List of tables

Table 2.1 Differentiation media .....	31
Table 2.2 Digestion buffer .....	35
Table 2.3 Liquid-chromatography gradient .....	38
Table 2.4 Cross-link scores .....	41
Table 3.1 Changes in protein expression upon differentiation (differentiated cells).....	50
Table 3.2 Changes in protein expression upon differentiation (undifferentiated cells).....	54
Table 3.3. Overview on the number of modified amino acids of ADH and PK when using NHS-acetate and DEPC for surface labeling .....	74
Supplementary Table 1: Enriched KEGG pathways in undifferentiated cells (vs. RA- differentiated cells).....	137
Supplementary Table 2: Enriched KEGG pathways in undifferentiated cells (vs. RA/PMA- differentiated cells) .....	138
Supplementary Table 3: Enriched KEGG pathways in RA-differentiated cells (vs. undifferentiated cells).....	139
Supplementary Table 4: Enriched KEGG pathways in RA/PMA-differentiated cells (vs. undifferentiated cells).....	140
Supplementary Table 5: Inter- and intra-molecular cross-links of ribosomal proteins observed in undifferentiated SH-SY5Y cells .....	141
Supplementary Table 6: Inter- and intra-molecular cross-links of ribosomal proteins observed in RA-differentiated SH-SY5Y cells .....	142
Supplementary Table 7: Inter- and intra-molecular cross-links of ribosomal proteins observed in RA/PMA-differentiated SH-SY5Y cells .....	144

## Summary

In this thesis, different neuronal systems were studied using mass spectrometry-based (MS-based) methods. More precisely, the neuroblastoma cell-line SH-SY5Y and synaptic vesicles purified from rat brain were selected. SH-SY5Y cells are a commonly employed model system to study neuronal function and disease. This includes neuronal disorders occurring, for instance, in Parkinson disease or Alzheimer. Even though SH-SY5Y cells are widely explored, a complete description of the resulting proteomes and cellular reorganization during differentiation is still missing.

In this thesis, the proteomes of cells obtained under standard growth conditions and by two differentiation protocols employing retinoic acid (RA) or a combination of RA and phorbol-12-myristat-13-acetate (PMA) were compared. Relative protein abundance was obtained by an MS-based label-free quantification approach. First, the abundance of proteins with specific subcellular localization was compared revealing an increase in protein abundance in differentiated cells for proteins localized in the endoplasmic reticulum, plasma membrane and the lysosome. Similar changes in protein expression of RA- and RA/PMA-differentiated cells during neuronal differentiation indicate that similar cells are obtained by the two protocols. The proteins were relatively quantified between the three conditions. An increase of the expression of proteins related to RA administration or neuronal development, synaptic proteins as well as proteins of the antioxidant defense mechanism was observed in differentiated cells compared to undifferentiated cells. In contrast, upregulated proteins in undifferentiated cells are associated with cell proliferation. In conclusion, undifferentiated SH-SY5Y cells represent neuroblastoma cells which have not yet developed a neuronal character while both differentiated cells with neuronal character consequently serving as better neuronal model system were obtained from both differentiation protocols.

In-cell cross-linking using formaldehyde further allowed capturing protein interactions in various cellular organelles of SH-SY5Y cells. The employed workflow and identified cross-links were validated by plotting cross-links obtained in ribosomal proteins on a previously published high-resolution structure of the ribosome. The calculated cross-linking distances were in good agreement with the high-resolution structure and, therefore validated the employed workflow. Furthermore, protein-protein interactions involving histone, mitochondrial and cytoskeletal proteins were identified. Specifically, structural reorganization involving regulating protein factors of the actin cytoskeleton upon differentiation were observed.

Covalent labeling is widely used to identify solvent accessible amino acid residues of proteins or protein complexes. The obtained information adds valuable information for generating structural models of proteins. In this thesis, two chemical labeling strategies using: i) N-hydroxysuccinimidyl acetate (NHS-acetate) and ii) diethylpyrocarbonate (DEPC) as labeling reagents were employed. Characterizing the mass spectra of modified peptides uncovered neutral losses from DEPC-modified amino acids improving the assignments of the peptide fragment spectra. A quantitative labeling workflow was then established to determine labeling percentage and unambiguously distinguish solvent accessible amino acid residues from stochastically labeled residues.

The established labeling workflow was then applied to study proteins of synaptic vesicles. Residues of solvent accessible domains of synaptic vesicle proteins were found to be modified using the labeling reagents: i) sulfo-N-hydroxysuccinimidyl acetate (S-NHS-acetate), ii) NHS-acetate and iii) DEPC. This included typical synaptic vesicle proteins such as synaptobrevin-2, synaptotagmin and proteins with unknown structure including synapsin-1, Ras-related protein Rab-3A and cysteine string protein. For the latter proteins, their structure was predicted using AlphaFold. The observed absence of labeling in specific domains of these proteins indicated lipid- and protein-binding sites. Surprisingly, residues located in luminal loops of tetra-spanning proteins and synaptic vesicle glycoprotein 2A were found to be labeled. The explicit functions of these proteins are unknown. The labeling data suggests structural rearrangements of these proteins under specific conditions resulting in solvent accessibility. Specifically, their dynamic structure might contribute to the previously reported increase in size of synaptic vesicles. In conclusion, the labeling workflow is applicable to unravel structural properties of proteins with unknown function.

## **Zusammenfassung**

In dieser Arbeit wurden verschiedene neuronale Systeme mit Massenspektrometrie-basierten (MS-basierten) Methoden untersucht. Hierfür wurden die Neuroblastom-Zelllinie SH-SY5Y und aus Rattenhirn präparierte synaptische Vesikel ausgewählt. SH-SY5Y-Zellen sind ein häufig verwendetes Modellsystem zur Untersuchung neuronaler Funktionen und Erkrankungen. Dazu gehören neurodegenerative Prozesse, die beispielsweise bei der Parkinson-Krankheit oder Alzheimer auftreten. Obwohl SH-SY5Y-Zellen weitgehend erforscht sind, fehlt eine vollständige Beschreibung der resultierenden Proteome und der zellulären Reorganisation während der Differenzierung.

In dieser Arbeit wurden die Proteome von Zellen verglichen, die unter Standardwachstumsbedingungen und durch zwei Differenzierungsprotokolle unter Verwendung von Retinsäure (RA) oder einer Kombination aus RA und Phorbol-12-Myristat-13-Acetat (PMA) erhalten wurden. Die relative Proteinhäufigkeit wurde durch eine MS-basierte markierungsfreie Quantifizierungsstrategie bestimmt. Zunächst wurde die Häufigkeit von Proteinen mit spezifischer subzellulärer Lokalisation verglichen, welches eine Zunahme der Proteinhäufigkeit in differenzierten Zellen der Proteine zeigte, die im endoplasmatischen Retikulum, der Plasmamembran und dem Lysosom lokalisiert sind. Die beobachtete Proteinexpression von RA- und RA/PMA-differenzierten Zellen, weist auf eine ähnliche neuronale Differenzierung hin. Die Proteine wurden zwischen den drei Bedingungen relativ quantifiziert. In differenzierten Zellen wurde im Vergleich zu undifferenzierten Zellen eine erhöhte Expression von Proteinen im Zusammenhang mit der RA-Gabe oder der neuronalen Entwicklung, synaptischen Proteinen sowie Proteinen des antioxidativen Abwehrmechanismus beobachtet. Im Gegensatz dazu sind hochregulierte Proteine in undifferenzierten Zellen mit Zellproliferation assoziiert. Zusammenfassend stellen undifferenzierte SH-SY5Y-Zellen Neuroblastomzellen dar, die noch keinen neuronalen Charakter entwickelt haben, während beide Differenzierungsprotokolle ähnlich differenzierte Zellen ergaben, die besser als neuronale Modellsysteme geeignet sind.

Die chemische Quervernetzung mit Formaldehyd ermöglichte die Identifizierung von Proteininteraktionen in verschiedenen Zellorganellen von SH-SY5Y-Zellen. Der verwendete Arbeitsablauf und die identifizierten Quervernetzungen wurden validiert, indem Quervernetzungen von ribosomalen Proteinen auf eine zuvor veröffentlichte hochauflösende Struktur des Ribosoms aufgetragen wurden. Die berechneten Quervernetzungsabstände

stimmten gut mit der hochauflösenden Struktur überein und validierten daher den angewendeten Arbeitsablauf. Darüber hinaus wurden Protein-Protein-Interaktionen von Histon-, Mitochondrien- und Zytoskelettproteinen identifiziert. Insbesondere wurde eine strukturelle Reorganisation des Aktinzytoskeletts nach der Differenzierung beobachtet.

Kovalente Markierung wird weit verbreitet verwendet, um lösungsmittelzugängliche Aminosäurereste von Proteinen oder Proteinkomplexen zu identifizieren. Die erhaltenen Informationen können zur Erstellung von Strukturmodellen von Proteinen genutzt werden. In dieser Arbeit wurden zwei chemische Markierungsstrategien unter Verwendung von i) N-Hydroxysuccinimidylacetat (NHS-Acetat) und ii) Diethylpyrocarbonat (DEPC) als Reagenzien verwendet. Die Charakterisierung der Massenspektren von modifizierten Peptiden deckte Neutralverluste von DEPC-modifizierten Aminosäuren auf, welche die Zuordnung der Peptidfragmentspektren verbesserte. Anschließend wurde eine quantitative Markierungsstrategie etabliert, um den prozentualen Anteil der Markierung zu bestimmen und lösungsmittelzugängliche Aminosäurereste eindeutig von stochastisch markierten Resten zu unterscheiden.

Die etablierte Markierungsstrategie wurde anschließend angewendet, um Proteine von synaptischen Vesikeln zu untersuchen. Reste von lösungsmittelzugänglichen Domänen synaptischer Vesikelproteine wurden unter Verwendung der Reagenzien: i) Sulfo-N-hydroxysuccinimidylacetat (S-NHS-Acetat), ii) NHS-Acetat und iii) DEPC modifiziert. Dazu gehörten typische Vesikelproteine wie Synaptobrevin-2, Synaptotagmin und Proteine mit unbekannter Struktur, darunter Synapsin-1, Ras-verwandtes Protein Rab-3A und Cystein-String-Protein. Für letztere Proteine wurde ihre Struktur mit AlphaFold vorhergesagt. Das beobachtete Fehlen einer Markierung in spezifischen Domänen dieser Proteine deutete auf Lipid- und Proteinbindungsstellen hin. Überraschenderweise wurde festgestellt, dass Reste, die sich in luminalen Schleifen von Tetra-überspannenden Proteinen und synaptischem Vesikel-Glykoprotein 2A befinden, markiert sind. Die expliziten Funktionen dieser Proteine sind unbekannt. Die Markierungsdaten legen strukturelle Dynamiken dieser Proteine unter spezifischen Bedingungen nahe, welche zu einer Lösungsmittelzugänglichkeit führt. Insbesondere ihre dynamische Struktur könnte zu der zuvor berichteten Größenzunahme von synaptischen Vesikeln beitragen. Zusammenfassend lässt sich sagen, dass die Markierungsstrategie anwendbar ist, um strukturelle Eigenschaften von Proteinen mit unbekannter Funktion zu identifizieren.



# 1. Introduction

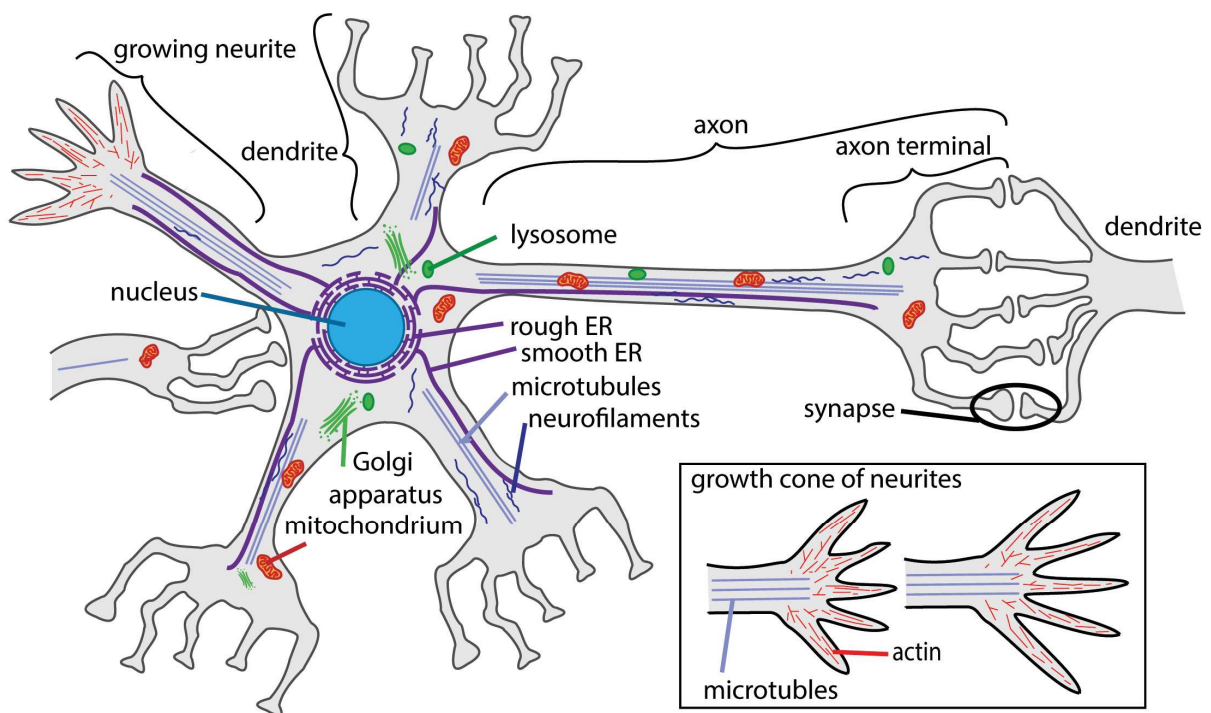
Elucidating the explicit functions of neurons is the basis for many research areas. In neurons, proteins are required for many tasks and their expression and expression level is regulated by multiple factors. In addition, a protein's structure is closely related to its function. Analyzing protein expression as well as the structure of proteins and the protein complexes they form reveals important pathways within cells and uncovers functional protein assemblies. One powerful method to obtain this information is mass spectrometry (MS). It allows quantification of the proteome as well as investigating the structure of proteins.

## 1.1 Structure and function of neurons

Neurons are required for processing and transmission of intercellular signals. Different kinds of neurons exist; they are structurally (multipolar, bipolar and pseudounipolar) or functionally (sensory, motor and interneuron) classified and fulfil specific tasks. In general, a neuron can be divided into four units: dendrites, cell body (soma), axon and axon terminal which is also called synaptic terminal (**Figure 1**) [1]. In the following paragraphs these structural units as well as organelles will be introduced.

The dendrites as well as the soma constitute the postsynaptic neuron and receive signals [2, 3]. When a neurotransmitter binds to the corresponding neurotransmitter receptor, the opening of its ligand gated ion channel is initiated and an ion influx occurs. In addition, neurotransmitters can bind to G protein-coupled receptors (GPCRs), triggering a cascade of processes mediated by second messengers, leading to opening of specific ion channels. The opening either results in the influx of positive charge, leading to depolarization and an excitatory postsynaptic potential, or negative charge, leading to hyperpolarization and inhibitory postsynaptic potential [4].

As mentioned above, not only the dendrites, but also the whole cell body can function as postsynaptic side. In addition, protein synthesis occurs within the cell body. Gene expression starts with transcription and is then followed by translation within the rough endoplasmic reticulum (ER). Proteins are further processed within the Golgi apparatus. Within the Golgi, the proteins as well as lipids are prepared for further transport by packaging into vesicles. The Golgi is not only located within the soma, but also in dendrites as Golgi outposts and satellites [5].



**Figure 1: Structural cartoon of a neuron**

The schematic of a neuron is shown. Neurite growth is mediated by a growth cone (box). The growing neurites specialize in becoming either the axon including the axon terminal or dendrites. The following organelles are highlighted: nucleus (blue), rough and smooth endoplasmic reticulum (ER) (purple), lysosome and Golgi apparatus (green) and mitochondria (orange). In addition, cytoskeletal components are shown: neurofilaments (blue), microtubules (purple) and actin (red).

The Axon forwards the action potential via a depolarization wave (positive charge) or repolarization wave (negative charge) [6]. Along microtubules within the axon, transport of organelles such as mitochondria, lysosomes and vesicles containing neurotransmitters or proteins takes place. Specifically, kinesin is responsible for anterograde transport along microtubules towards the axon terminal and dynein for retrograde axonal transport for example of vesicles containing degraded and misfolded proteins or growth factors, that may again stimulate the expression of specific genes [7, 8]. The axonal membrane contains voltage gated sodium channels. These are stimulated by the flow of positive charge and even increase this flow by opening their pores leading to a sodium influx [9]. After stimulation and the resulting high positive ion population within the axon, voltage gated potassium channels are activated, leading to efflux of potassium and therefore repolarization [9].

When the flow of positive charge reaches the axon terminal, specific voltage-gated channels located at the presynaptic side lead to influx of calcium ions [10]. In the chemical synapse, this triggers the fusion of neurotransmitter loaded synaptic vesicles with the membrane and the release of neurotransmitters into the synaptic cleft [10]. This process will be described

in more detail in the following section (**Section 1.2**). Neurotransmitters then bind to receptors located in the postsynaptic membrane as described above. After this process, neurotransmitters are degraded or reuptake is initiated.

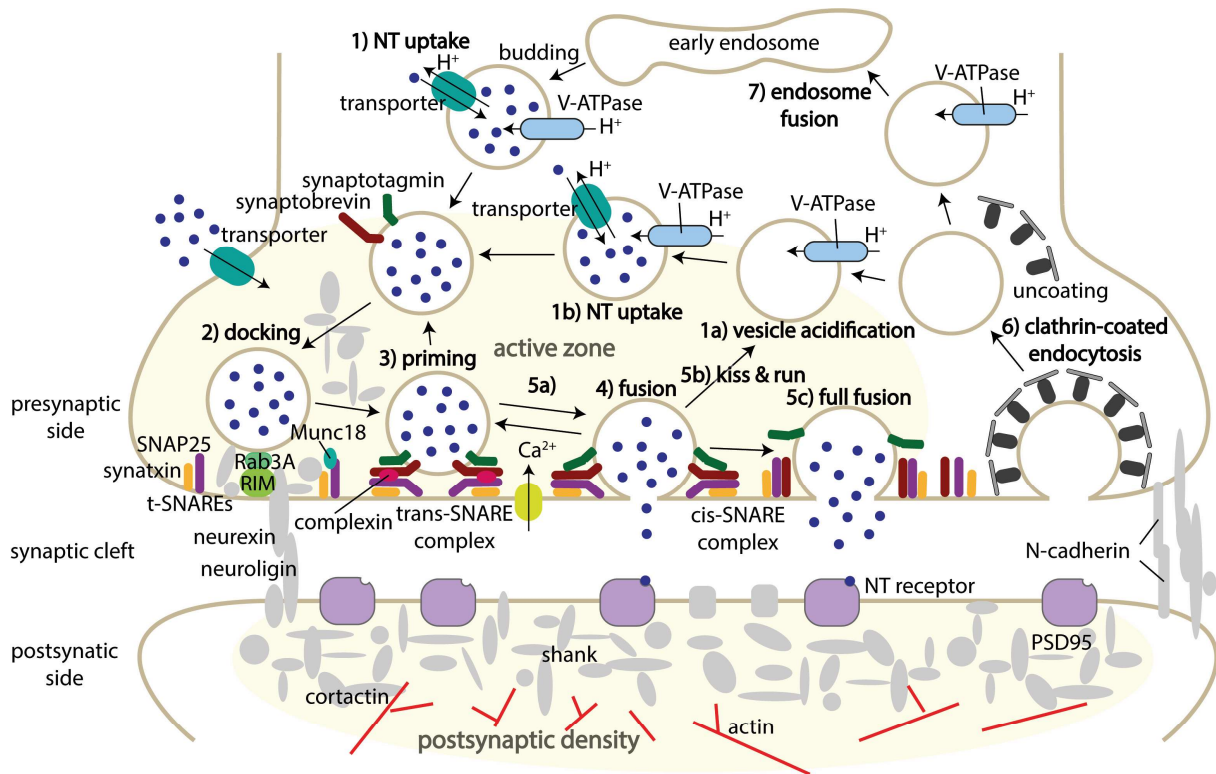
Mitochondria play an important role within neurons. They are located in the cell body as well as the dendrites and axon terminals. Their main function is energy supply in the form of adenosine triphosphate (ATP). The neurons high energy demand is associated with synaptic vesicle cycle as well as neurotransmitter synthesis [11]. Furthermore, mitochondria contain enzymes for neurotransmitter degradation [12].

In neurons, lysosomes are located in the cell body as well as in lower abundance in the distal parts of axons and dendrites [13-15]. They are required for macromolecular degradation of proteins, lipid membranes, DNA, RNA and carbohydrates and enable recycling of the obtained components. Important roles of axonal and dendritic lysosomes are the clearance of dysfunctional synaptic proteins and regulating the structural plasticity of dendritic spines [16-18].

The highly compatibilized structure as well as dynamic growth of neurites, which can be either dendrites or axons, is facilitated by a dynamic cytoskeleton. Three classes of filaments form the cytoskeleton: intermediate filaments, such as neurofilaments; microfilaments built by actin, and microtubules based on tubulin. During early neuronal development nestin and vimentin are the major subunits of intermediate filaments [19]. Upon neuronal maturation, the expression changes to neurofilament types (light, medium and heavy polypeptides). In contrast to intermediate filaments, microtubules and actin are more dynamic. Both microtubules and actin play an import role during neurite elongation within the lamellipodium and filopodium of the growth cone (**Figure 1**, box). In addition, microtubules stabilize the actin shaft and, as mentioned above, are required as transport route to the growth cone or axon terminal.

## **1.2 The synapse and the synaptic vesicle cycle**

The synapse is a specialized structure enabling neurotransmission between neurons. In the case of chemical synapses, the release and recycling of synaptic vesicles is essential for their function. Synaptic vesicles are abundant organelles within the presynapse and all vesicles in one synapse contain roughly the same amount of neurotransmitter. The trafficking cycle of synaptic vesicles includes several steps (see **Figure 2** for details).



## Figure 2: Synaptic vesicle cycle within the chemical synapse

The schematic of a chemical synapse including the synaptic vesicle cycle is shown. The main steps of the synaptic vesicle cycle are numbered. Abbreviation: NT, neurotransmitter; Rab3a, Ras-related protein Rab-3A; RIM, Rab3 interacting molecule; SNAP25, synaptosomal-associated protein, 25kDa; V-ATPase, proton-pumping vacuolar H<sup>+</sup>-ATPase and PSD95, postsynaptic density protein 95; SNARE, soluble N-ethylmaleimide-sensitive-factor attachment receptor.

First, synaptic vesicles are loaded with neurotransmitters via secondary active transporters that utilize an electrochemical gradient of ions as energy source (**Figure 2**, step 1 and 1a). There are specific neurotransmitter transporters: the vesicular  $\gamma$ -aminobutyric acid transporter (VGAT) transports  $\gamma$ -aminobutyric acid (GABA) and glycine [20, 21], the vesicular monoamine transporters (VMAT) transport monoamines such as 5-hydroxytryptamine, dopamine, adrenaline, noradrenaline and histamine [22], the vesicular acetylcholine transporter (VACHT) [22] transports acetylcholine and the vesicular glutamate transporters (VGLUT) transports glutamate [23]. Filled vesicles are then transported to the so-called ‘active zone’ of the presynapse. The distribution of vesicles near the active zone is controlled by several proteins including, for instance, bassoon and piccolo [24]. In the active zone, the vesicles dock to the presynaptic membrane (**Figure 2**, step 2). Docking of synaptic vesicles is mediated by Ras-related protein Rab-3A (Rab3a) [25] and Rab3-interacting molecule (RIM) further promoting Munc13 function and vesicle priming [26]. In the next step, docked vesicles are activated resulting in primed vesicles (**Figure 2**, step 3). During priming, formation of the trans-SNARE complex is formed. This complex contains SNAP25

and syntaxin located in the presynaptic membrane and synaptobrevin-2 located in the synaptic vesicle membrane [27]. In addition, synaptotagmin and complexin fulfill important regulatory roles. Complexin functions as a priming factor for SNARE complexes and as an activator of the SNARE complexes for subsequent fusion [28]. During fusion, the proteins of the SNARE complex are fully assembled into the so called 'cis-complex'. Calcium influx through voltage-gated calcium channels leads to calcium binding to synaptotagmin and triggers synchronous fusion of synaptic vesicles with the presynaptic membrane [29] (**Figure 2**, step 4). Subsequently, there are three possibilities of vesicle recycling: i) synaptic vesicles are reacidified and refilled with neurotransmitters without undocking [30], ii) vesicles undock and recycle locally ('kiss-and-run') to reacidify and refill with neurotransmitters [31] or iii) vesicles fully fuse with the presynaptic membrane (**Figure 2**, step 5a-c). In the latter case, alternative recycling mechanisms of synaptic vesicles such as clathrin-mediated endocytosis [32] (see **Figure 2**, step 6), activity-dependent bulk endocytosis [33] and ultrafast endocytosis [34] are possible. Note that the synaptic vesicle cycle is not fully understood yet and many additional vesicular synaptic vesicle or presynaptic proteins play or may play a functional role.

Upon synaptic vesicle fusion with the presynaptic membrane, neurotransmitters are released into the synaptic cleft and bind to receptors located at the postsynaptic side. These receptors are ionotropic or metabotropic receptors. Ionotropic receptors, also known as ligand-gated ion channels, are membrane proteins that form a pore upon neurotransmitter binding allowing ions to pass into the cell. As an example the inhibitory neurotransmitter GABA binds to GABA A type receptors leading to an influx of chloride ions and hyperpolarization of the cell [35]. Metabotropic receptors are GPCRs that are activated by neurotransmitter binding followed by a series of reactions leading to ion channel opening [4]. Coupling of the receptor activation and further signaling events are mediated by other proteins located in postsynaptic density. In addition to the mentioned signaling enzymes, membrane and scaffold proteins as well as the cytoskeletal component actin form an interaction network within the postsynaptic density (**Figure 2**, postsynaptic density).

### **1.3 SH-SY5Y cells as neuronal model system**

To date, detailed functional analysis of neurons relies on the application of neuronal model systems. These systems are commonly animal- or cell line-based. Depending on the research

question, animal-based research enables a more complex functional analysis of different kind of cells. In particular, the effect of drugs is studied. However, moral and ethical aspects have to be considered when animals are used. Cell lines are an alternative to animal-based research. When cell lines are used, a specific type of cell is studied in a defined environment. With the reduced complexity of the system, for instance, the specific effect of chemical reagents on this type of cell can be explored in detail. In many research areas, legal, ethical and moral aspects as well as economic benefits lead to the selection of cell line-based systems. In this thesis, synaptic vesicles from rat brain were studied and a cell line-based neuronal model system was used. The used cell line SH-SY5Y will be introduced in more detail in the following paragraphs.

The SH-SY5Y cell line is a neuroblast-like clonal subtype of the SK-N-SH cell line [36]. The original cell line was established from a neuroblastoma of a female four year old girl [37]. Neuroblastoma is a cancer occurring primarily in children and especially at very young age. It develops from precursor cells of the sympathetic nervous system. However, the exact development is unknown. Several genetic mutations such as anaplastic lymphoma kinase and paired-like homeobox 2B gene have been identified in neuroblastoma cells [38]. The heterogeneous distribution of mutations in cancer cells from different patients yet, indicates no specific mutation that leads to cancer development. Depending on the stage, the 10-year overall survival ranges from 38 % for high-risk patients to 91 % for low-risk patients [39].

In addition to the analysis of neuroblastoma, SH-SY5Y cells are used as model system to study neuronal differentiation and neuronal diseases [40-42]. Several differentiation protocols exist. Among available reagents, the most prominent differentiation reagent is retinoic acid (RA). RA treatment leads to regulation of transcription of neurotrophin receptor genes, Wnt signaling pathways and protein kinase A-dependent pathways [43-45]. After differentiation, adrenergic-, dopaminergic- and predominantly cholinergic-like neuronal subtypes were identified [40, 41]. The potential of RA for differentiation of neuroblastoma cells, therefore, led to the implementation of RA-treatment during therapy of high-risk patients [46]. Other differentiation protocols use, for instance, brain-derived neurotrophic factor (BDNF), dibutyryl cyclic adenosine monophosphate or staurosporine [47-49]. The treatment of SH-SY5Y cells with RA and phorbol-12-myristat-13-acetat (PMA) predominantly results in dopaminergic-like neurons [40, 50]. The obtained differentiated cells express specific neurotransmitter transporters and receptors and typical mature marker proteins such as synaptophysin and microtubule-associated protein 2 [42, 51].

Proteomic studies compared undifferentiated with RA/BDNF-differentiated cells [42] or analyzed changes in protein expression across a seven-day treatment with RA [52]. However, the proteomes of cells obtained from various differentiation strategies have not been analyzed.

## **1.4 Identification of proteins by liquid chromatography-coupled mass spectrometry**

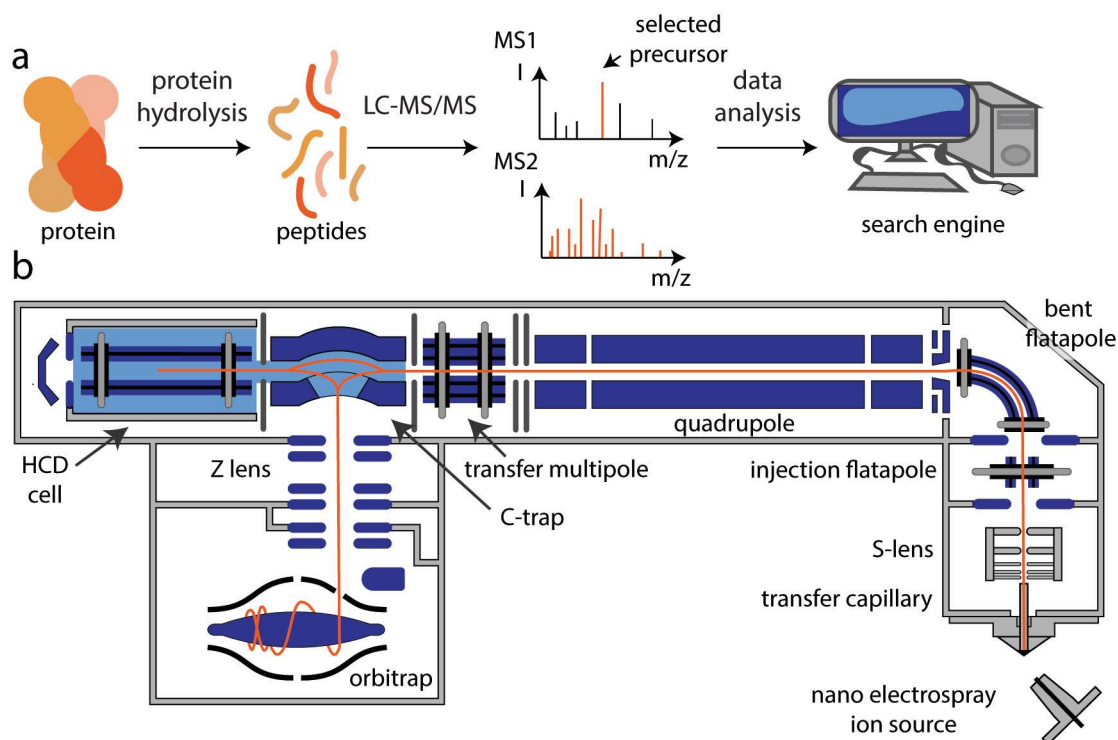
The identification of a protein using MS is either based on a top down or bottom up approach. When top down is used, the protein is transferred into the gas-phase using different ionization techniques and its molecular mass is determined using a mass spectrometer. In addition, fragmentation of the protein into peptides and amino acids enables the identification of the protein. The more popular bottom up workflow will be described in more detail in this chapter.

In a typical bottom up workflow, the protein of interest is first hydrolyzed using a specific protease such as trypsin (**Figure 3a**). The identification of proteins is then predominantly performed using liquid chromatography-coupled mass spectrometry (LC-MS/MS). For this, the obtained peptide mixture is separated by reverse phase nano LC-systems and directly eluted into the mass spectrometer.

### **Electrospray ionization of peptides**

Electrospray-ionization (ESI) is a soft ionization technique [53]. During ESI, the analyte solution is passed through a spray emitter. Applying an electrical potential of up to 3 kilovolts to the emitter pulls the ions out of the analyte solution towards the counter electrode [54]. At the tip of the emitter, the solution is distorted into a Taylor cone that emits a spray of small charged droplets [55]. The evaporation of solvent leads to shrinking of the droplets. When the size and charge of the droplet reaches its stability limit (the so-called ‘Rayleigh limit’) the droplet ‘explodes’ into several smaller droplets (Coulomb explosions) [56]. The final generation of ions is described by different models. The ionization of small molecules such as peptides is described by the ion evaporation model. This model is based on shrinking droplets through evaporation of the solvent until the surface charge density is close to the Rayleigh limit and small solvated ions are ejected from the

droplet surface [57]. Remaining solvent molecules evaporate when the ion travels through the sampling interface of the mass spectrometer.



**Figure 3: Protein identification by mass spectrometry**

**a)** The typical bottom up workflow for protein identification is shown. Proteins are hydrolyzed using a specific protease. Subsequently, the obtained peptides are analyzed by LC-MS/MS. Precursor  $m/z$ 's are recorded in the MS1 spectrum. For MS2 experiments, a specific precursor ion (orange) is selected for fragmentation and corresponding fragment ions are recorded. The acquired data is then analyzed using different search engine. **b)** Schematics of a Q Exactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer used in this thesis. Ions are generated using a nano electrospray ion source and are passed through lenses into the mass spectrometer. For MS1 experiments, the quadrupole operates in scanning mode and ions are passed through the C-trap and are analyzed and detected in the orbitrap. During MS2 experiments, an ion with a specific  $m/z$  is selected in the quadrupole and fragmented in the high-energy collisional dissociation cell (HCD cell), followed by analysis and detection of fragment ions in the orbitrap. (Figure 3b was adapted from Thermo Fisher.)

### Ion separation, fragmentation and detection

Subsequent to ion generation, the ions are submitted to a mass analyzer. The quadrupole [58], the time-of-flight (TOF) [59] and the orbitrap are commonly used mass analyzer [60]. The ions are then detected in the ion detector such as the orbitrap [60], secondary electron multiplier [61] or microchannel plates [62]. The recorded mass-to-charge ratio, that is the  $m/z$  or in more detail the  $(m+z)/z$ , where  $m$  is the mass of the uncharged analyte and  $z$  is the charge, is displayed in the mass spectrum (MS1). Triple-quadrupole, quadrupole-TOF or quadrupole-orbitrap mass spectrometers enable the selection of a specific ion and subsequent fragmentation and analysis of the generated fragment ions (MS2) [63]. Fragmentation of the precursor ion is usually achieved by collision-induced



dissociation (CID) [64], high-energy collisional dissociation (HCD) [65], electron transfer dissociation (ETD) [66] or electron capture dissociation (ECD) [67]. The  $m/z$ 's of the fragment ions are then recorded (MS2 spectrum).

### **The Q Exactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer**

In this thesis, a Q Exactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer was used (**Figure 3b**). Here, ions are generated at atmospheric pressure using nano ESI and are passed through the front-end of the instrument. For this, the lenses focus the ion beam and the bent flatapole ion guide reduces noise by preventing the forwarding of neutral molecules before transmission of the ions into the hyperbolic quadrupole [58]. The hyperbolic quadrupole operates in two modes. In the scanning mode, ions of a specific  $m/z$  range pass the quadrupole and are further passed through the C-trap into the orbitrap. In this mode, the so called MS1 spectrum is recorded. Then, one of the recorded ions can be selected in the quadrupole for the following MS2 experiment. In this case, the quadrupole functions as a mass filter. In detail, a radio frequency voltage with a direct current offset voltage is applied to stabilize a specific  $m/z$  [68]. This  $m/z$  is carried through the quadrupole on a stable oscillating trajectory [68]. Then the ions with a specific  $m/z$  are transferred to the C-trap [69] and are submitted to fragmentation in the HCD cell [65]. The obtained fragment ion population is transferred back to the C-trap and ejected into the Orbitrap [69]. In the orbitrap the ions oscillate between an outer electrode and an inner spindle electrode [60]. The orbital motion around the inner electrode is detected and converted by applying Fourier transformation into  $m/z$  values for each ion [60].

### **Protein identification by database search**

For protein identification, the acquired raw data is usually converted to a peak list that contains information on the  $m/z$  of the peptides precursors and their fragment ions as well as their relative intensity and charge. This peak list is used for a search against a target database containing theoretical masses of *in silico* generated peptides and their fragment masses. Different software has been developed to identify the best peptide match. In this thesis, Andromeda [70] within MaxQuant was used, which applies a probability scoring model for identification. Other algorithms are based on cross correlation scoring [71] or hyper-geometric scoring models [72]. The obtained peptide spectrum matches (PSMs) are further used for protein identification. In particular unique peptides of a protein are required for precise protein identification.

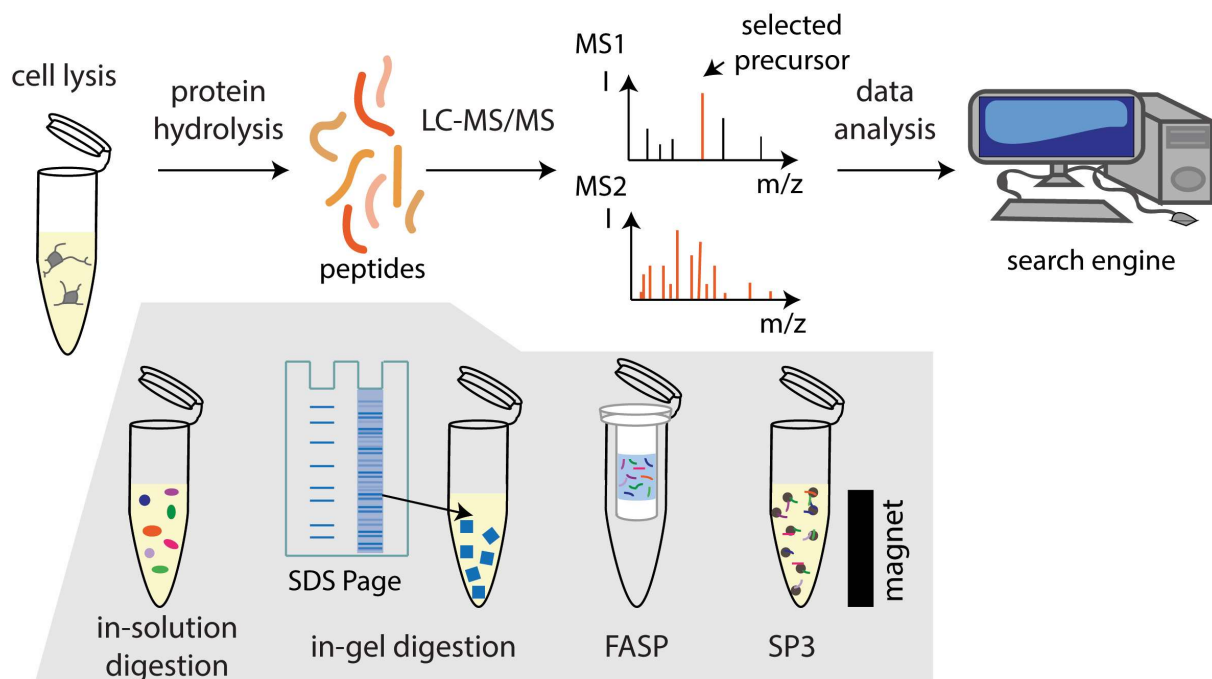
In addition, the quality of PSMs is assessed by a false discovery rate (FDR) that calculates the ratio between false PSMs and the total number of PSMs. Different strategies exist for calculation of the FDR: target-decoy search [73, 74] using a decoy database for the identification of false positives and mixture model-based methods [75]. However, false negative identification of peptides leads to a lower number of identified proteins. These false negative identifications occur due to peptides with post-transcriptional modifications (PTMs) or inaccurate databases of all proteins of the specific organism. To cope with these challenges, *de novo* sequencing is used in some cases to determine the composition of the peptide directly from the precursor mass and the corresponding MS2 spectrum [76-80].

## 1.5 Proteome analysis and protein quantification strategies

The proteome of an organelle, a cell or entire organism describes the set of expressed proteins, including their PTMs at a specific timepoint and under specific conditions. In contrast to the genome, it is very dynamic and responds to signals or environmental changes. The identification and quantification of proteins of a biological system or between different systems in defined conditions is of particular interest. Different MS-based methods exist for protein identification and quantification. However, only a fraction of the proteins of a complex proteome are usually identified and even a lower number of them is reliably quantified [81]. Therefore, several sample preparation strategies have been developed and were optimized.

### Sample preparation for proteomic analysis

Initially cells are lysed and, then, proteins are digested in-solution (**Figure 4**). This approach leads to a very complex mixture of peptides, that is difficult to analyze by MS without additional prefractionation. Therefore, an approach was developed, in which proteins are first separated by sodiumdodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), followed by in-gel digestion of gel-bands [82, 83]. Following this workflow, sample components such as buffer ingredients, in particular SDS, that interfere with the LC-MS/MS analysis are removed. However, very small sample amounts are challenging to analyze with this strategy.



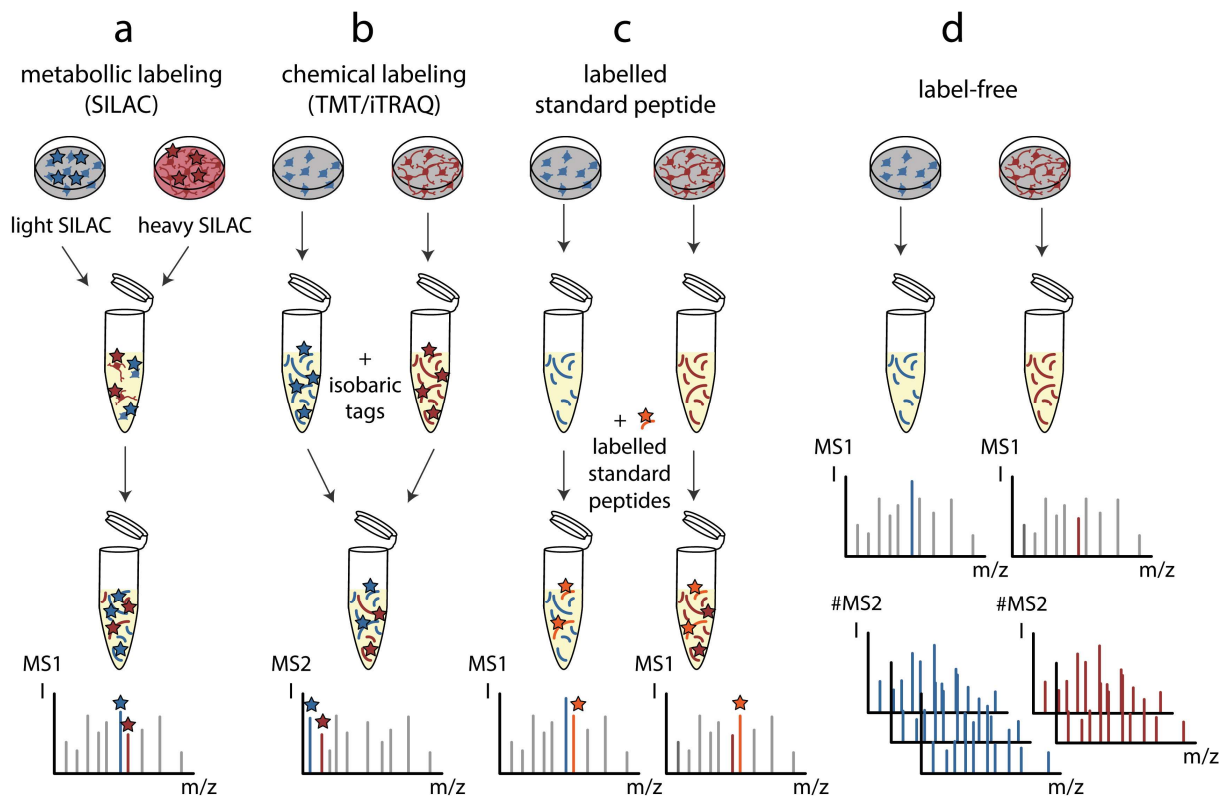
**Figure 4: Workflow for proteomic sample preparation and analysis**

For proteomic experiments, cells or tissue are lysed using different protocols. Then, several approaches for protein hydrolysis are applied. These include, for instance, in-solution digestion, in-gel digestion, filter-aided sample preparation (FASP) or single-pot, solid-phase-enhanced sample-preparation (SP3). The obtained peptides are then analyzed by LC-MS/MS. Subsequently, the acquired data is analyzed using different search engines.

The advances of LC separations, prior to MS analysis resulted in intensive studies on optimal in-solution digestion. In particular, detergent free sample preparation techniques are employed. Sample Preparation by Easy Extraction and Digestion (SPEED) [84] utilizes trifluoroacetic acid (TFA) for highly efficient cell lysis and enables subsequent protein digestion in the same reaction tube. In addition, filter-aided sample preparation (FASP) [85], combining the advantages of in-gel, and in-solution digestion and single-pot, solid-phase-enhanced sample-preparation (SP3) [86], utilizing paramagnetic beads thereby enabling removal of several interfering buffer components, have been established recently.

### Quantification of proteins

For quantification of the abundance of proteins, four main quantification strategies are employed: metabolic labeling, chemical labeling, samples spiked with labeled standard peptides or label-free quantification (**Figure 5**). These techniques will be introduced in the following paragraphs.



**Figure 5: MS-based protein quantification methods**

The workflows for different protein quantification strategies are shown. Quantification strategies where a label (indicated with a star) is introduced are shown (a-c). **a)** Metabolic labeling. Cells grown in ‘light’ SILAC medium and cells grown in ‘heavy’ SILAC medium are combined and prepared for MS analysis. For relative quantification, the intensities of the isotope cluster in the MS1 spectra are compared. **b)** Chemical labeling. Isobaric tags, such as ‘tandem mass tags’ and ‘isobaric tags for absolute and relative quantification’ are used for chemical labeling of peptides. In the experimental set-up a different label is used for each condition and samples are then combined. In the MS2 spectrum differentially isotope encoded reporter ions are observed for each tag. **c)** Labeled standard peptide. The protein digest is spiked with a known amount of stable isotope labeled peptide of the corresponding peptide of the protein of interest. The intensity in the MS1 spectrum of the endogenous peptide to the spiked peptide is used to calculate absolute protein abundance. **d)** Label-free quantification. Label-free quantification is either based on the peptide’s intensity in the MS1 spectrum or counting of MS2 spectra of a particular peptide. Abbreviations: SILAC, stable isotope labeling by amino acids in cell culture; TMT, tandem mass tag and iTRAQ, isobaric tags for absolute and relative quantification.

**Metabolic labeling for relative protein quantification** Metabolic labels are introduced by using  $^{15}\text{N}$ -enriched culture medium or stable isotope labeling by amino acids in cell culture (SILAC) and enable relative protein quantification [87, 88]. In both cases, the labeled components are incorporated into the newly synthesized proteins. The SILAC approach will be described in more detail. In the initial SILAC protocol, cells of two differentially treated cell culture conditions are compared. For this, cells grown in ‘light’ SILAC medium and cells grown in ‘heavy’ SILAC medium are combined and prepared for MS analysis (**Figure 5a**). When  $^{13}\text{C}_6$ -arginine and  $^{13}\text{C}_6$ -lysine are used, all tryptic peptides, except for the protein’s C-terminus, include at least one labeled amino acid. For identification of a peptide sequence, one of the coeluting ‘light’ and ‘heavy’ peptides is selected for fragmentation. The generated MS2 spectra are used during database search for the identification of PSMs. For relative

quantification, the intensities of the isotope cluster in the MS1 spectra are compared (**Figure 5a**). Synthesis and turnover of proteins can be analyzed by pulsed SILAC [89, 90]. For this, labeled amino acids are added to the growth medium for a specific period of time allowing monitoring of differences in *de novo* protein synthesis. In addition, super SILAC which uses SILAC-labeled cell lines is used for relative quantification of non-labeled tissue or even organisms [91].

One advantage of metabolic labeling is that quantification errors introduced during sample preparation and mass spectrometric analysis are minimized by combining the differentially treated samples before these steps. This enables the analysis of small changes in protein levels including PTMs. However, metabolic labeling is cost and time consuming and is limited to the number of available labels (i.e. labeled amino acids).

**Chemical labeling for relative protein quantification** Another main relative quantification strategy is based on chemical or enzymatic labeling of proteins or peptides. One labeling strategy will be introduced in more detail in this paragraph. For peptide labeling, the two isobaric tags ‘tandem mass tags’ (TMTs) [92] and ‘isobaric tags for absolute and relative quantification’ (iTRAQ) [93] are often used (see **Figure 5b** for details). Both labeling reagents target primary amines and, therefore, all peptide N-termini and lysine residues are modified. In the experimental set-up, a different label is used for each condition. Differentially isotope encoded reporter ions are then observed for each tag in the MS2 spectrum (**Figure 5b**). This allows the comparison of different conditions. In comparison to deuterated labels, which lead to small shifts in retention time of the eluting peptide during LC-MS/MS analysis, isobaric labeled peptides precisely elute at the same time [94]. Hence, spectra complexity is not increased as for approaches such as SILAC.

**Quantification of proteins using labeled peptide standards** In contrast to other quantification strategies, spiking the sample with a synthetic protein or isotope labeled synthetic standard peptide enables absolute quantification of selected proteins [95-97]. For this, the protein digest is spiked with a known amount of a stable isotope labeled peptide of the corresponding endogenous peptide of the protein of interest (**Figure 5c**). A quantitative comparison of the intensity in the MS1 spectrum of the endogenous peptide to the spiked peptide is then used to calculate absolute protein abundance [97]. However, one disadvantage of this approach is that only proteins of interest are quantified. The whole proteome cannot be quantified following this strategy.

**Label-free quantification techniques** In contrast to the above-mentioned quantification strategies, label-free quantification does not require the introduction of labels. Relative quantitative information on the protein abundance is either obtained by counting the number of PSMs for each identified protein or by utilizing the peptide's peak intensities recorded in the mass spectrometer (**Figure 5d**). Both methods are challenging. The obtained PSM count is influenced by the experimental settings, such as the scan speed of the mass spectrometer or the application of dynamic exclusion of precursor ions for a defined time range. In addition, the length of a protein and its potential proteolytic cleavage sites influence the number of observable PSMs for each individual protein. For the correction of these effects, methods aiming to improve the quantification accuracy have been developed. These methods utilize the protein abundance factor (PAF) [98, 99], the exponentially modified protein abundance index (emPAI) [100], the normalized spectral abundance factor (NSAF) [101] or the MS2-based Normalized Spectral Index to normalize the spectral count (StPeter) [102]. In contrast to spectral counting, the MS1-based quantification is based on integrating the ion's intensity over their chromatographic elution profile. For most peptides, the extracted and calculated chromatographic peak area is proportional to the corresponding peptide concentration [103]. For accurate quantification, especially of complex samples, narrow LC peak width and retention time stability is an important prerequisite [104]. This can be achieved by using nano or micro-flow LC instruments, robust columns and by avoiding column saturation [105]. Furthermore, different ionization efficiencies of peptides due to different properties and/or co-eluting peptides and contaminants have an effect on optimal quantification [106]. For accurate protein quantification between samples, several data processing steps are required. The data processing steps include i) peak detection by distinguishing peaks from neighboring peaks and background noise, ii) peak integration to obtain peak areas, iii) deconvolution to cope with charge detection and isotope patterns, iv) chromatographic alignment of elution profiles between samples for a comparative analysis and v) normalization for accurate quantification. Several commercial software packages and freely available tools such as MaxQuant [70, 107], which is used in this thesis, have been developed. The main advantage of label-free quantification is the in theory unlimited number of samples that can be compared. Furthermore, time-consuming steps for introducing a label and the costs for labeling reagents are avoided. Specific approaches are even used to enable absolute quantification. These approaches are either based on spectral counts or signal intensities. The protein abundance index (PAI) [108], for example, is calculated by dividing the number of observed peptides by the number of theoretically observable unmodified

peptides. Intensity-based methods use the intensity of the three most intense peptides of a protein (Top3), the peptide with the highest intensity or calculate the intensity based absolute quantification value (iBAQ value) [109] by dividing the sum of intensities of the observed peptides of a protein by the number of theoretically observable peptides. For an optimal quantification, these approaches depend on a reproducible identification of the same peptide in all samples and due to different peptide properties (see above) only an estimate absolute quantification is possible. Despite the challenges of label-free quantification, it is a prominently used and reliable alternative to labeling-based quantification approaches.

### **Acquisition techniques for protein identification and quantification**

In most proteomic studies data-dependent acquisition (DDA) is used for protein identification and quantification [110]. Usually, a fixed number of the most intense peptide ions is selected for fragmentation. An alternative to DDA is selected reaction monitoring (SRM) (also known as multiple reaction monitoring (MRM) [111], parallel reaction monitoring (PRM) [112] or data independent acquisition-based (DIA-based) sequential window acquisition of all theoretical fragment ion spectra (SWATH) [113]. In SRM/MRM, precursor and product ion pairs are monitored over time for precise quantification [114]. PRM is similar and known or predicted peptide ions are specifically targeted for fragmentation [112]. DIA-based SWATH in contrast employs another strategy by isolating and fragmenting ranges of  $m/z$  sequentially [113]. This method aims to identify a high number of proteins and to accurately quantify them.

### **Analysis of proteomic data sets**

The next step after obtaining a quantitative large-scale proteomic data set is its analysis with the goal to extract biologically relevant information. In most cases, changes in relative protein abundance between different conditions are compared. For identification of statistically significant changes in protein abundance, independent replicates of each condition are a prerequisite for subsequent statistical tests such as the t-test or ANOVA. In specific cases not only protein abundance is compared, but also changes in PTMs such as phosphorylation sites. The identified up- or downregulated proteins or changing PTM abundance are further used for downstream analysis. For this gene ontology terms, pathways or interaction networks are analyzed [115].

## **1.6 Structural proteomics**

Typically, structural techniques such as x-ray crystallography, nuclear magnetic resonance or electron microscopy are used to obtain high resolution structures of proteins and protein complexes. However, some proteins and protein complexes remain challenging to analyze with these classical techniques. Structural proteomics is an alternative approach to gain structural information. In most cases, techniques such as native MS, cross-linking MS and labeling approaches are applied to obtain structural information.

### **1.6.1 Native mass spectrometry**

Native MS is a technique that is used to study intact protein complexes in the gas-phase. The acquired mass spectra reveal information on the composition as well as on subunit interactions, topology and stability of the complexes.

#### **Sample preparation and electrospray ionization of protein complexes**

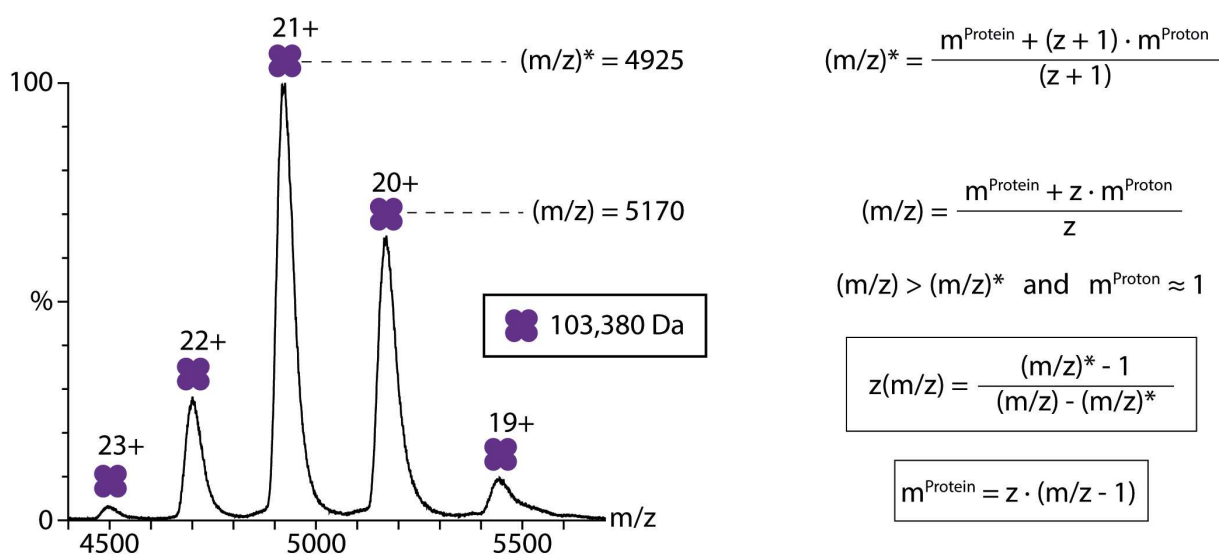
Typically, ions are generated by nano-ESI and are then transferred into a mass spectrometer which preserves noncovalent interactions of protein assemblies in the gas-phase [116, 117]. Two prerequisites for ESI of large protein assemblies and their intact analysis in the gas-phase are a volatile sample solution and a modified instrument for the transmission of high mass complexes. For sample preparation the buffer of the sample is exchanged to a volatile ESI-compatible solution that preserves non-covalent protein interactions, such as ammonium acetate solutions. For this, miniaturized gel filtration columns, molecular weight cut-off filters or dialysis devices are employed. In particular aqueous ammonium acetate is widely used in native MS, because it is a volatile electrolyte ( $\text{NH}_4^+ \text{CH}_3\text{-COO}^-$ ) and simultaneously stabilizes folded proteins and their complexes as background electrolyte [118]. During nano-ESI, the ‘charged residue model’ describes the ionization process of large molecules, such as proteins. Very small droplets that usually contain only one “native” protein assembly are generated and upon complete solvent evaporation the charge is transferred to the molecule [119, 120].



## Analysis of large ions

The ions generated by ESI are then transferred and detected in the mass spectrometer. For this, a modified quadrupole-TOF [121] or more recently a high-resolution quadrupole-Orbitrap mass spectrometer [122] is employed. The observed MS1 spectra typically show a series of charge states with a Gaussian distribution (**Figure 6**).

The acquired mass spectrum (**Figure 6**) is then used to determine the mass of the protein complex. In detail, the recorded  $m/z$  of two neighboring peaks differs by one proton. With this information the calculation of the charge of a complex and its mass is possible (see **Figure 6** for details). The annotation and calculation of the molecular weight of the protein assembly is facilitated by different deconvolution software tools [123-127]. Note that the observed charge state distribution correlates with the folding state of the protein. Compared to folded proteins denatured proteins show a wider charge state distribution. MS2 experiments are further used to select a specific precursor (in native MS experiments a specific charge state of the protein complex or subcomplex) for CID or HCD experiments. Increasing collisional voltages then leads to the dissociation of peripheral subunits. The peripheral subunits are usually highly charged, while the remaining complex (so called ‘stripped-complex’) carries less charges, due to the asymmetric charge subdivision of the former complex.



**Figure 6: Representative native MS spectrum**

The native MS spectrum of tetrameric concanavalin A is shown (lhs). A Gaussian distribution of charge states (19+ to 23+) is observed and a molecular weight of 103,380 Da was calculated for the concanavalin A tetramer. The given equations are used to calculate the molecular weight of the protein complex (rhs). (Figure taken from Barth and Schmidt 2020 [128])

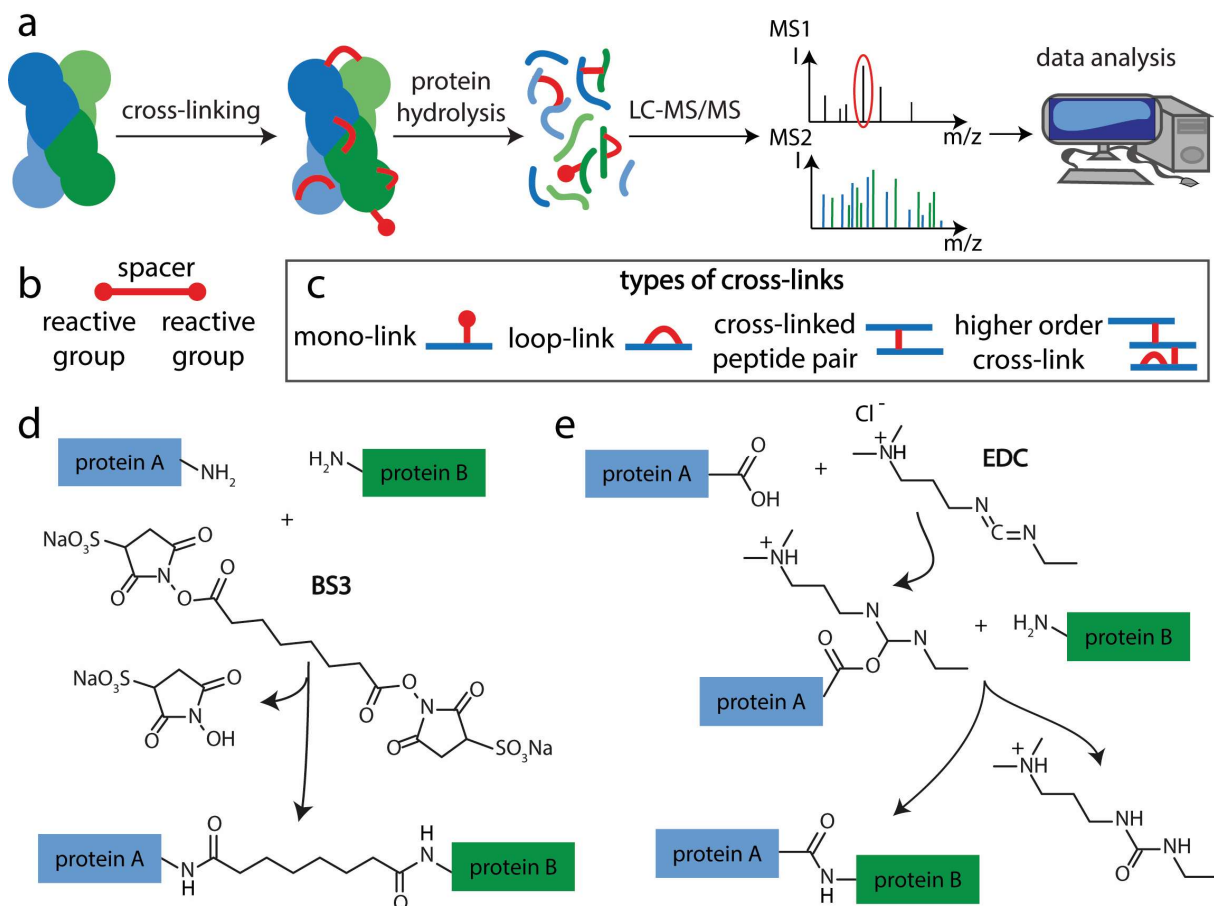
Native MS has been successfully applied to study stoichiometries, interactions and stability protein complexes as well as ligand binding to proteins. The applicability of native MS ranges from small proteins such as ubiquitin up to large complexes such as the 9 MDa Flock House virus [129]. In conclusion, native MS is a versatile and valuable structural biology method.

### **1.6.2 Cross-linking mass spectrometry**

Cross-linking MS is a tool to study protein-protein interactions. This approach has a wide range of applicability. Protein cross-linking of proteins, large protein complexes, proteins within organelles and even proteins of whole cells have shown the diverse application of cross-linking MS [130-132].

#### **Protein cross-linking workflow**

The first step of the cross-linking workflow is the cross-linking reaction. For this, two approaches are commonly followed, namely photo-induced and chemical cross-linking. The goal of both is to covalently link two amino acid side chains in close proximity. The cross-linked proteins are then enzymatically hydrolyzed using proteases such as trypsin (see workflow in **Figure 7a** for details). After hydrolysis linear peptides, mono-linked linear peptides, loop-linked linear peptides, cross-linked peptide pairs and higher order cross-linked peptides are obtained (**Figure 7c**). This mixture is then analyzed by LC-MS/MS and subsequent data analysis is used for identification of cross-links.



### Figure 7: Cross-linking workflow

**a)** The typical cross-linking workflow is shown. After cross-linking, the proteins are hydrolyzed and obtained peptides are analyzed by LC-MS/MS. The acquired data is then further processed using different software. **b)** Most cross-linkers consist of two reactive groups that are connected by a linker. **c)** After cross-linking and protein hydrolysis linear peptides as well as different kind of cross-linked peptides are obtained. **d)** Reaction mechanism of the cross-linker bis(sulfosuccinimidyl)suberat (BS3). **e)** Reaction mechanism of the zero-length cross-linker 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC).

### Cross-linking reagents

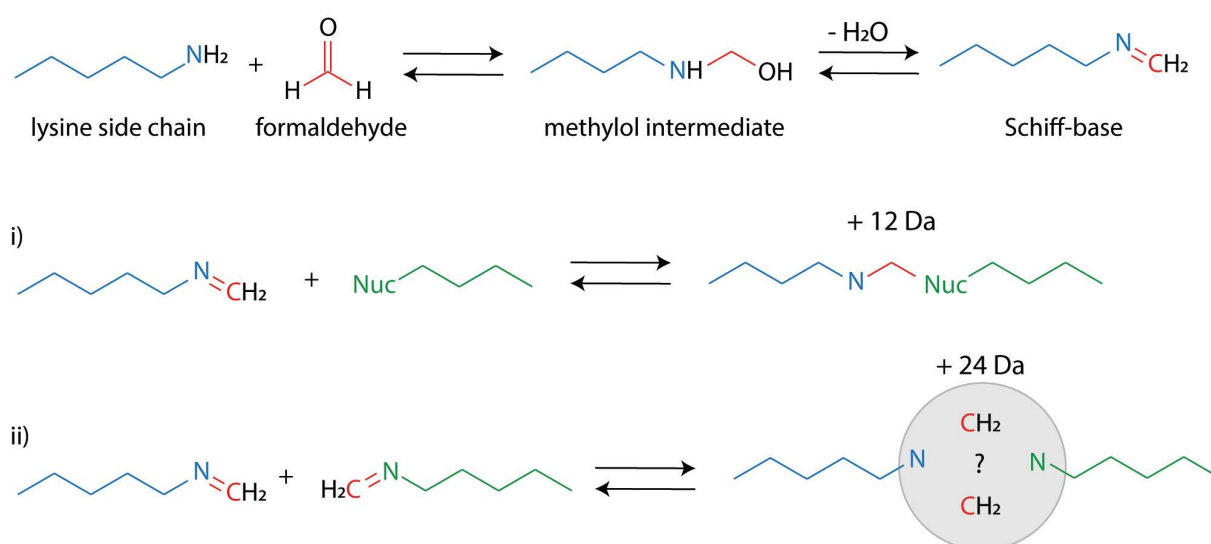
Most cross-linking reagents contain two reactive groups and a spacer (**Figure 7b**). The reactive groups introduce the covalent linkage while the length of the spacer determines the distance of the two cross-linked side chains.

Prominent N-hydroxysuccinimide ester-based cross-linking reagents are bis(sulfosuccinimidyl)suberat (BS3) (**Figure 7c**) [133] and disuccinimidyl suberate (DSS) [134]. The functional groups of these reagents react mainly with primary amines such as the amino group of the amino acid lysine and the protein's N-terminus. Zero-length cross-linkers act as coupling reagents without incorporating a spacer. 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) is a commonly used zero-length cross-linker and cross-links carboxyl groups of aspartic acid, glutamic acid, or the C-terminal

of the protein to primary amines (**Figure 7d**) [135]. The main advantage of zero-length cross-linkers is the defined distance restraint in the range of a salt-bridge. In addition, tri- and multifunctional cross-linking reagents are available. These contain, for instance, cleavable affinity tags for enrichment.

### Cross-linking with formaldehyde

In this thesis, formaldehyde was used for cross-linking and will be introduced in more detail in the following part. Recently, a cross-linking MS workflow has been established for formaldehyde [136]. Formaldehyde, similar to DSS (see above), is membrane-permeable, making it suitable for in-cell cross-linking. In the first reaction step, a nucleophilic group in the peptide, for instance the  $\epsilon$ -amino group of a lysine side chain, attacks the carbonyl carbon of formaldehyde (**Figure 8**).



### Figure 8: Formaldehyde cross-linking reaction mechanism

The reaction mechanism for protein cross-linking using formaldehyde (red) is shown. The  $\epsilon$ -amino group of a lysine side chain (blue), attacks the carbonyl carbon of formaldehyde resulting in the formation of a methylol intermediate. Dehydration of the methylol intermediate then leads to the formation of an imine (also called ‘Schiff-base’) intermediate. i) Another nucleophile reacts with the Schiff-base by a nucleophilic attack on its methylene carbon resulting in a stable methylene bridge to the other residue and a reaction product with an additional mass of 12 Da. ii) The reaction of two Schiff-base intermediates leads to the formation of a cross-linked product including a 24 Da mass shift.

Dehydration of the methylol intermediate then leads to the formation of an imine (also called ‘Schiff-base’) intermediate. Subsequently, two possible reactions occur: i) another nucleophile reacts with the Schiff-base by a nucleophilic attack on its methylene carbon resulting in a stable methylene bridge to the other residue (12 Da mass shift) or ii) the

reaction of two Schiff-base intermediates leads to the formation of a product containing probably two methylene bridges (24 Da mass shift) [136-138].

### **Improving the detection of low abundant cross-linked peptide pairs**

As mentioned above, a complex mixture of linear peptides, mono-linked linear peptides, loop-linked linear peptides, cross-linked peptide pairs and higher order cross-linked peptides is obtained after protein hydrolysis. Of these, cross-linked peptide pairs are of particular interest. Due to their low abundance, enrichment strategies such as size exclusion chromatography [139] or strong cation exchange chromatography [140] are applied. When trifunctional cross-linkers containing an additional affinity tag are used, the tag can be utilized for selective enrichment. The fractions containing cross-linked peptide pairs are then analyzed by LC-coupled MS. Due to the higher charge of cross-linked peptide pairs when compared with their linear or mono- and loop-linked peptides, ions with a lower charge state are often excluded during MS analysis. Recently the coupling of a high-field asymmetric waveform ion mobility spectrometry (FAIMS) device at the front end of the mass spectrometer was used as a gas-phase separation technique to filter for ions with higher charge states and thereby improving cross-linked peptide pair identification [141].

### **Identification of cross-linked peptide pairs**

The identification of cross-linked peptide pairs is in principle similar to the identification of linear peptides (see **Section 1.4**). First, all possible peptides are predicted *in silico* and, in contrast to linear peptides, all possible peptide pair combinations are then predicted. Depending on the number of proteins, the number of possible combinations increases exponentially with each protein. Therefore, specific software (e.g. Kojak [142], pLink2 [143], xQuest [140], Xlink-Identifier [144], XiSearch [145], Formaldehyde Cross-link Analyser [136] and MeroX [146]) has been developed to identify cross-linked peptide pairs and to cope with the high risk of random matches. Within the group of cleavable cross-linking reagents, specific reporter ions are generated upon fragmentation and are utilized for validation [147]. In general, for reduction of false identifications, manual validation of spectra or the application of a false discovery rate (FDR) is required. The target-decoy method with a false mass for cross-linker [148] or decoy database [140] are commonly used for FDR estimation and controlling.

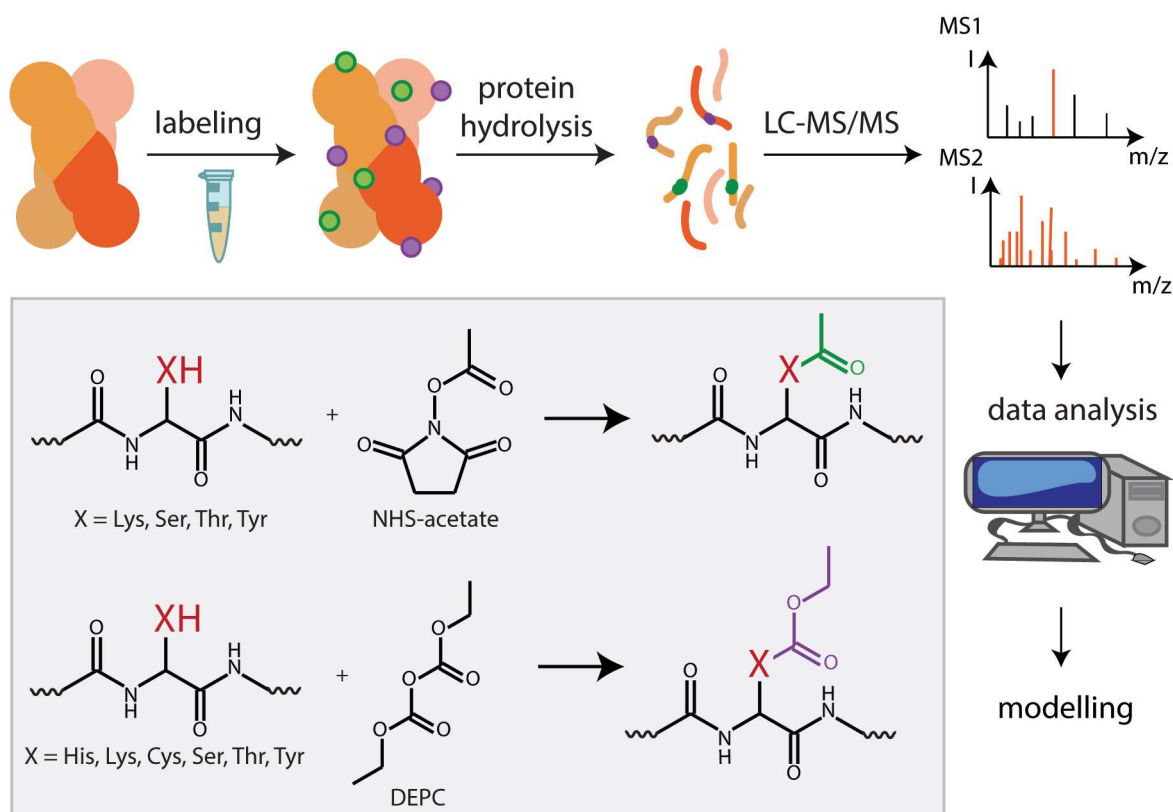
### **Analysis of identified cross-linked sites**

The obtained set of cross-links contains intra- as well as inter-protein cross-links. Both can be visualized using different diagrams, for instance bar-, circle- or network-plots [149]. When high-resolution structures of the protein or proteins are available, cross-links are plotted on these structures and distances between cross-linked sites can be compared with the expected length depending on the introduced linker length. Importantly, the identified cross-links and obtained distance restraints facilitate docking of proteins to high-resolution structures as well as modelling of protein structures [150]. In summary, cross-linking MS reveals protein-protein interactions with biological relevance.

### **1.6.3 Labeling mass spectrometry**

Labeling MS is a technique enabling the analysis of the solvent accessibility of amino acid residues of a folded protein or protein complex. The typical labeling workflow starts with a covalent chemical modification of residues (**Figure 9**). The labeled protein is then hydrolyzed, and the obtained peptides are analyzed by LC-MS/MS. Subsequently, modified sites are identified using different software or search engines.

Three main labeling techniques are frequently applied. Of these, the most classical approach is Hydrogen-Deuterium exchange MS (HDX-MS). In HDX, the exchange of the protein's hydrogen atoms to deuterium is measured. Information on the solvent accessibility of the protein's backbone is mainly obtained when using HDX. Complementary techniques to HDX are, for instance, hydroxyl radical foot printing (HRF) or fast photochemical oxidation of proteins (FPOP). These techniques are based on radicals reacting with solvent exposed amino acid side chains of a protein and are able to modify 19 of the 20 common amino acids [151]. This is on one hand an advantage because a good coverage of the protein's amino acid sequence is achieved, on the other hand the modifications of several residues lead to complex products including peptides containing different modified amino acid residues [152]. Therefore, data analysis is challenging. The third group of labeling techniques is based on covalent labeling using chemical reagents. This technique was used in this thesis and will be introduced in the following paragraphs.



### Figure 9: Labeling of proteins

The typical workflow for labeling of proteins and MS-based analysis is shown. Labeled proteins are enzymatically hydrolyzed and peptides are analyzed by LC-MS/MS. Subsequent data analysis enables modeling of protein or protein complex structures. Labeling reactions of the labeling reagents NHS-acetate and DEPC are described (grey box). (Figure adapted from Barth et al., 2020 [153])

### Chemical labeling

The modification of a residue depends on the reactivity of this residue towards the labeling reagent and on its localization within the protein structure. In general, residues located at the solvent accessible surface area (SASA) of a protein are more likely to be modified than buried residues or residues involved in protein-protein or protein-ligand interactions. A wide range of specific chemical labeling reagents exists. Among these, specific labeling reagents are for instance: dicarbonyls or kethoxal for modification of arginine residues [154-160], carbodiimides for modification of carboxylic acids [161-163], different reagents such as sulfhydryl reagents for cysteine modifications [164-169] and tetranitromethane, iodine and N-acetylimidazole for tyrosine modifications [155, 170]. Organic acid anhydrides or N-hydroxysuccinimide derivatives are employed for acetylation of lysine residues [171-174]. The modified sites are then identified using specific software (e.g. MaxQuant [175, 176], Mascot [177], MSS-CLEAN [178], PEAKS [179], ProteinPilot [180], pFind [181, 182], SEQUEST [183], COMPASS [184, 185], TPP [186]) or commonly employed software.

In this thesis, N-hydroxysuccinimidyl acetate (NHS-acetate), sulfo-NHS-acetate (S-NHS-acetate) and diethylpyrocarbonate (DEPC) are employed as labeling reagents. NHS-acetate and S-NHS-acetate are reactive towards primary amines and therefore acetylate lysine residues and the protein's N-terminus (**Figure 9**, box) [187, 188]. In addition, side reactions with serine, threonine and tyrosine residues occur. Solvent accessible amino acid residues of proteins have been successfully studied using both NHS-acetate [189, 190] as well as S-NHS-acetate [191, 192]. Of these, S-NHS-acetate is water soluble and not membrane permeable, while NHS-acetate is not water soluble and membrane permeable.

DEPC adds a carboethoxy (CEt) group to histidine, lysine, tyrosine, serine, threonine, cysteine and arginine residues [193-196] (**Figure 9**, box). In addition to carboethoxylation of histidine, other histidine modifications also occur; these are 1,3-dicarboethoxyhistidine (di-CEt-histidine), formyl-carboethoxyhistidine (formyl-CEt-histidine) and urethane-carboethoxyhistidine (urethane-CEt-histidine) [197]. Labeling with DEPC therefore results in modification of up to 25% of the amino acids sequence of an average protein [196, 198, 199]. However, the labeling reactivity and correlation with SASA is different between different amino acids. Accordingly, histidine and lysine modifications correlate with the SASA of a protein, while serine and threonine modifications show a poor correlation between SASA and reactivity [200]. For serine and threonine residues, the protein's local environment might influence the reactivity [201]. This might be the case when hydrophobic residues are in close proximity and higher local DEPC concentrations are available for labeling of reactive residues [201]. DEPC labeling was applied in several studies to study solvent accessibility of residues and, therefore, to elucidate the topology of proteins or protein complexes, protein conformations and binding events [202-205].

## **1.7 Aim of this study**

The aim of this thesis is to develop MS-based methods to study neuronal model systems. The focus is, first, on the quantitative characterization of the proteome of differentiating SH-SY5Y cells and, second, on the development of structural proteomic techniques to identify protein interactions and unravel the structure of proteins and protein complexes.

In the first part of this thesis, a neuronal model system should be established. For this, the human neuroblastoma cell line was selected. These cells have been used as neuronal model system in previous studies under standard growth conditions as well as following treatment with differentiation reagents. Therefore, the objective is to compare both undifferentiated



and differentiated cells. For this, differentiated cells with RA or a combination of RA and PMA (RA/PMA) will be investigated. Relative protein quantification will be performed to uncover changes in protein expression between the different cell culture conditions. In addition, relative quantification of proteins with a specific subcellular localization will be performed to unravel changes upon differentiation. Identification of enriched KEGG pathways should unravel altered pathways. These findings will be used to generate a detailed proteomic characterization of the obtained cells.

The characterized neuronal model cells will then be used to study protein interactions. For this, the applicability of the recently established formaldehyde cross-linking workflow for cross-linking within undifferentiated and differentiated cells will be explored. When in-cell cross-linking is successfully performed, protein-protein interactions within the cell will be identified. This should unravel changes in protein interactions upon differentiation.

Following proteomic and cross-linking studies, the development of a quantitative labeling method of solvent accessible amino acids of proteins is desirable. Previous experiments performed during my master thesis were the basis for further experiments and specific analysis of solvent accessible amino acids. Establishing a quantitative data analysis strategy for chemical labeling with NHS-acetate and DEPC is the goal of this subproject. This quantitative strategy will enable distinction of buried and solvent accessible residues of proteins embedded in phospholipid bilayers.

Synaptic vesicles play an important role within the synapse. The proteins embedded in or associated with the vesicle membrane mediate, for instance, neurotransmitter uptake, vesicle docking to the presynaptic membrane, fusion of the vesicles with the presynaptic membrane for neurotransmitter release and recycling of the vesicles. However, the explicit function of several synaptic vesicle proteins is still unknown. Labeling of proteins in intact synaptic vesicles allows the identification of solvent accessibility of proteins and their domain orientation within the synaptic vesicle membrane further providing indications of functional properties of the vesicle proteins.



## 2. Material and methods

### 2.1 Material

#### 2.1.1 Chemicals

acetonitrile (ACN), liquid chromatography-mass spectrometry (LC-MS) grade	Thermo Fisher Scientific, Waltham, USA
all trans retinoic acid (RA)	Sigma Aldrich, St. Louis, USA
ammonium bicarbonate	Sigma Aldrich, St. Louis, USA
diethylpyrocarbonate (DEPC)	Sigma Aldrich, St. Louis, USA
dithiothreitol (DTT)	Sigma Aldrich, St. Louis, USA
ethanol, 99.8% (v/v), LC-MS grade	Carl Roth, Karlsruhe
formaldehyde	Sigma Aldrich, St. Louis, USA
formic acid (FA)	Thermo Fisher Scientific, Waltham, USA
glutamine	Danaher Corporation, Washington D.C., USA
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Sigma Aldrich, St. Louis, USA
imidazole	Sigma Aldrich, St. Louis, USA
iodoacetamide (IAA)	Sigma Aldrich, St. Louis, USA
N-hydroxysuccinimidyl acetate (NHS-acetate)	Tokyo Chemical Industry, Präfektur Tokio, Japan
penicillin/streptomycin	Thermo Fisher Scientific, Waltham, USA
phorbol-12-myristat-13-acetate (PMA)	Sigma Aldrich, St. Louis, USA
RapiGest	Waters Corporation, Milford, USA
sodium acetate	Sigma Aldrich, St. Louis, USA
sulfo-N-hydroxysuccinimidyl acetate (S-NHS-acetate)	Thermo Fisher Scientific, Waltham, USA
trifluoroacetic acid (TFA)	Thermo Fisher Scientific, Waltham, USA
TrisBase (Tris(hydroxymethyl)aminomethan)	Sigma Aldrich, St. Louis, USA
tris(2-carboxyethyl)phosphine (TCEP)	Sigma Aldrich, St. Louis, USA
urea, 99.8%, analytical grade	GERBU Biotechnik, Heidelberg
water, LC-MS grade	Thermo Fisher Scientific, Waltham, USA
2-chloroacetamide (CAA)	Sigma Aldrich, St. Louis, USA

#### 2.1.2 Cell lines, proteins and biological material

alcoholdehydrogenase (ADH)	Sigma Aldrich, St. Louis, USA
fetal calf serum (FCS) from bovine	American Type Culture Collection, Manassas, USA
penicillin/streptomycin	Thermo Fisher Scientific, Waltham, USA
pyruvate kinase (PK) from rabbit muscle	Sigma Aldrich, St. Louis, USA
SH-SY5Y cell line	Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ)
synaptic vesicles from rat brain	Prof. Dr. Reinhard Jahn (Max Planck Institute for Biophysical Chemistry, Göttingen)

trypsin	Promega, Mannheim
trypsin	Roche, Mannheim
trypsin-EDTA (0.5%)	Thermo Fisher Scientific, Waltham, USA

### 2.1.3 Chromatographic material and consumables

6-well plates Nunc™ USA	Thermo Fisher Scientific, Waltham, USA
C18 capillary column (50 cm, HPLC column Acclaim™ PepMap™ 100, C18, 75 µm I.D., particle size 3 µm) USA	Thermo Fisher Scientific, Waltham, USA
Pierce™ peptide desalting spin columns USA	Thermo Fisher Scientific, Waltham, USA
Superdex™ peptide 3.2/300 column USA Sera-Mag SpeedBeads (cat. no. 45152105050250) USA	General Electric, Boston, General Electric, Boston,
Sera-Mag SpeedBeads (cat. no. 65152105050250) USA T25 flasks Nunc™ EasYFlask™ USA T75 flasks Nunc™ EasYFlask™ USA	General Electric, Boston, Thermo Fisher Scientific, Waltham, Thermo Fisher Scientific, Waltham,
µ-Precolumn C18 (Acclaim™ PepMap™ 100, C18, 300 µm I.D., particle size 5 µm) USA	Thermo Fisher Scientific, Waltham, USA
ZipTip C18	Merck Group, Darmstad

### 2.1.4 Buffers, media, ready-to-use solutions and kits

2-(N morpholino)ethansulfonic acid (MES) buffer USA InstantBlue® Coomassie Protein Stain UK minimum essential medium eagle (EMEM-medium) USA NuPAGE LDS sample buffer USA	Thermo Fisher Scientific, Waltham, Abcam, Cambridge, Sigma Aldrich, St. Louis, Thermo Fisher Scientific, Waltham,
NuPAGE LDS reducing agent USA	Thermo Fisher Scientific, Waltham, USA
NuPAGE Bis-Tris gels 4 –12% USA phosphate buffered saline (PBS) Louis, USA SeeBlue Plus 2 prestained standard protein marker Waltham,	Thermo Fisher Scientific, Waltham, Sigma Aldrich, St. Thermo Fisher Scientific, USA

### 2.1.5 Instruments

18 MP microscope digital camera	Swift Optical Instruments, Inc., Schertz, USA
ÄKTA pure chromatography system	General Electric, Boston, USA
DionexUltiMate 3000 RSLCnano System	Thermo Fisher Scientific, Waltham, USA
Primovert light microscope	Carl Zeiss IQS Deutschland GmbH, Oberkochen
Q Exactive Plus Hybrid	
Quadrupole-Orbitrap mass spectrometer	Thermo Fisher Scientific, Waltham, USA

### 2.1.6 Software

artMS (version 1.9.1)	<a href="http://artms.org">http://artms.org</a>
Cytoscape (version 3.8.2)	Shannon, P. et al. [206]
Enrichr	<a href="https://maayanlab.cloud/Enrichr">https://maayanlab.cloud/Enrichr</a> [207-209]
Formaldehyde Cross-link Analyser	Tayri-Wilk T, et al. [136]
GETAREA	Fraczkiewicz, R. & Braun, W. [210]
MaxQuant (version 1.6.3.3 and 1.6.17.0)	Cox, J. & Mann, M. [211]
MSstats (version 3.22.0)	Choi, M. et al. [212]
PDBePISA (version 1.52)	Krissinel, E. & Henrick, K. [213]
PRIDE	Perez-Riverol et al. [214]
QCloud2	Olivella, R. et al. [215]
Proteome Discoverer (version 2.4.1.15)	Thermo Fisher Scientific, Waltham, USA
PyMOL (version 1.3)	Schrödinger, L.L.C. [216]
R (version 4.0.2)	R core team [217]
RStudio (version 1.2.1335)	RStudio team [218]
UCSF Chimera (version 1.15)	Pettersen, E.F. et al. [219]
venn-tool	<a href="http://bioinformatics.psb.ugent.be/webtools/Venn">http://bioinformatics.psb.ugent.be/webtools/Venn</a>
Xcalibur (version 4.2.47)	Thermo Fisher Scientific, Waltham, USA
Xlink Analyzer (version 1.1.4)	Kosinski, J. et al. [220]
XlinkCyNET (version 1.2.5)	Lima, D.B., Zhu, Y. & Liu, F. [221]

## 2.2 Methods

### 2.2.1 Cell culture

The human cell line SH-SY5Y was cultured in 6-well plates or flasks in culture medium consisting of minimum essential medium eagle (EMEM medium) supplemented with 15% (v/v) fetal calf serum (FCS), 1% (v/v) penicillin/streptomycin and 2 mM glutamine. In detail, frozen cells obtained from the DSMZ were thawed at 37 °C and directly transferred into 10 ml prewarmed (37 °C) medium. The cells were centrifuged at  $800 \times g$  for 2 min and the supernatant was removed. The cell pellet was resuspended in 12 ml prewarmed medium (37 °C) and transferred into a T25 flask. Subsequently, cells were incubated in an atmosphere of 5% CO<sub>2</sub> at 37 °C. Cells were passaged every two to three days when approximately 80% confluence was reached. Cell growth was monitored using an inverted Primovert light microscope at  $20 \times$  magnification. For passaging, cells in a T25 flask were washed with 10 ml warm (37 °C) phosphate buffered saline (PBS). 1 ml of 0.05% trypsin in PBS was added and cells were incubated for two to three minutes in an atmosphere of 5% CO<sub>2</sub> at 37 °C. Detachment of cells was monitored using an inverted Primovert light microscope. 10 ml prewarmed cell culture medium was added to the detached cells, followed by centrifugation at  $800 \times g$  for 2 min. The culture medium was discarded and fresh culture medium was added for resuspension of the cells. The cell containing solution was divided 1:3 to 1:5 into new flask containing 12 ml culture medium (37 °C). Passaged cells were again incubated in an atmosphere of 5% CO<sub>2</sub> at 37 °C.

For cryo-conservation of cells, confluent cells were detached from the wall of the flask using trypsin, followed by the addition of culture medium and centrifugation as described above. The obtained cell pellet was dissolved in 950 µl culture medium and 50 µl dimethylsulfoxid were added. Cells were frozen in liquid nitrogen and stored at -120 °C.

### 2.2.2 Differentiation of SH-SY5Y cells

Cells were cultured under standard growth conditions (**Section 2.2.1**) and further used for differentiation. When the cells reached approximately 80% confluence, they were prepared as described above (**Section 2.2.1**). Before splitting the cells into new flasks, the cell number was estimated. For this, cells were resuspended in 5 ml culture medium and 5 µl were added to 5 µl of 0.5% trypanblue in PBS and transferred into a Neubauer hemocytometer. Cells in at least two quadrants of 1×1 mm were counted, and the mean was calculated. For the final

calculation of the cell number/ml, the mean is multiplied by the volume and dilution factor ( $2 \times 10^4$ ). A cell solution containing 100,000 cells was plated in 6-well plates and for cross-linking analysis, 800,000 cells were plated in T75 flasks. Freshly prepared culture medium (37°) was added and cells were incubated as described above. On the following day the culture medium was replaced by differentiation medium 1 (see **Table 2.1** for details).

**Table 2.1 Differentiation media.** The composition of differentiation medium 1 and 2 is given.

<b>component</b>	<b>differentiation medium 1</b>	<b>differentiation medium 2</b>
<b>EMEM-medium</b>	96.5%	96.5%
<b>FBS</b>	2.5%	2.5%
<b>penicillin/streptomycin</b>	1%	1%
<b>glutamine</b>	2 mM	2 mM
<b>RA</b>	10 $\mu$ M	10 $\mu$ M
<b>PMA</b>	-	80 nM

For RA-differentiation, freshly prepared differentiation medium 1 was added on day one, three and five. In detail, the culture medium was discarded and warm (37 °C) differentiation medium was added. For RA/PMA-differentiation, the medium was exchanged to freshly prepared differentiation medium 1 on days one and three and for differentiation medium 2 on days 4 and 6. Cell growth and differentiation was monitored using an inverted Primovert light microscope at 20  $\times$  magnification equipped with an 18 MP microscope digital camera. RA-differentiated cells were cross-linked and/or harvested on day 6 and RA/PMA-differentiated cells were cross-linked and/or harvested on day 7. Not cross-linked cells were detached from the plate using trypsin-EDTA (0.5%) as described above (**Section 2.2.1**). The obtained cell pellet was washed three times with PBS. The cell pellet was frozen in liquid nitrogen and stored at  $-80$  °C.

### **2.2.3 Cross-linking using formaldehyde**

Undifferentiated and differentiated cells (see **Section 2.2.1** and **2.2.2** for details) were used for cross-linking experiments. Cells were washed three times with warm (37 °C) PBS. For cross-linking, 3 ml of 4.5% (v/v) formaldehyde in PBS were added to the cells followed by incubation for 15 min at 37 °C. Subsequently, the cross-linked cells were washed three times

with ice-cold PBS and scraped off the T75 flask into PBS by pelleting centrifugation at  $800 \times g$  for 2 min at 4 °C. The obtained cell pellet was stored at  $-80$  °C.

#### **2.2.4 Labeling of model proteins with NHS-acetate<sup>1</sup>**

15 mM NHS-acetate stock solution was prepared in 33% (v/v) ACN. For chemical labeling, the stock solution was diluted 1:10 with water. For ADH and PK protein stock solutions, proteins were dissolved in PBS, pH 7.4, at a tetrameric protein concentration of 10  $\mu$ M. Then, a 0-, 50-, 100-, 250-, 500-, 1000- or 1500-fold molar excess of NHS-acetate was added to 50 pmol ADH or PK in a final volume of 20  $\mu$ l. The solution was incubated for 15 min at 23 °C. 1  $\times$  NuPAGE LDS sample buffer and 1  $\times$  reducing reagent were added to a final volume of 30  $\mu$ l. The samples were incubated for 10 min at 70 °C and gel electrophoresis using 4 – 12% NuPAGE Bis-Tris gels and 2-(N morpholino)ethansulfonic acid (MES) buffer was performed at 200 V for 35 min. The gel was stained overnight using Coomassie staining solution and destained in water until the background appeared clear. Subsequently, in-gel digestion (see **Section 2.2.10** for details) and LC-MS/MS analysis (**Section 2.2.17**) followed.

#### **2.2.5 Labeling of model proteins with DEPC<sup>2</sup>**

Protein stock solution of ADH and PK were prepared at a tetrameric protein concentration of 10  $\mu$ M in 25 mM HEPES, pH 7.5. The DEPC stock solution was diluted with water to 0.1, 0.5, 2.5 and 5 mM. For DEPC labeling, a 0-, 2-, 10-, 50- and 100-fold molar excess of DEPC was added to 50 pmol ADH or PK in a final volume of 15  $\mu$ l followed by incubation for 1 min at 37 °C and 500 rpm. The labeling reaction was quenched by addition of 1  $\mu$ l 10 mM imidazole. Ethanol precipitation (**Section 2.2.11**), in solution enzymatic hydrolysis in the presence of urea (**Section 2.2.12**) and LC-MS/MS analysis (**Section 2.2.17**) were performed as described.

---

<sup>1</sup> These experiments were previously conducted (Master thesis: Chemical modification of protein for structure elucidation, Marie Barth, 2018). The project was continued in this thesis.

<sup>2</sup> Two replicates were prepared and analyzed previously (Master thesis: Chemical modification of protein for structure elucidation, Marie Barth, Martin Luther University, 2018). Additional replicates were performed in this thesis.



### **2.2.6 Labeling of synaptic vesicle proteins with S-NHS-acetate or NHS-acetate**

Synaptic vesicles from rat brain were provided from by Prof. Dr. Reinhard Jahn (Max Planck Institute for Biophysical Chemistry, Göttingen). Vesicles corresponding to approximately 30 µg protein were incubated with 5 or 10 mM sulfo-NHS-acetate (S-NHS-acetate) or NHS-acetate for 15 min at 23 °C. Subsequently, ethanol precipitation (**Section 2.2.11**), in-solution enzymatic hydrolysis in the presence of RapiGest (**Section 2.2.13**) and LC-MS/MS analysis (**Section 2.2.17**) were performed as described.

### **2.2.7 Labeling of synaptic vesicle proteins with DEPC**

Vesicles corresponding to approximately 30 µg protein were incubated with 5 or 10 mM DEPC for 10 min at 37 °C. Subsequent, ethanol precipitation (**Section 2.2.11**), in-solution digestion in presence of urea (**Section 2.2.12**) and LC-MS/MS analysis (**Section 2.2.17**) were performed.

### **2.2.8 Sample preparation for LC-MS/MS-based proteomics**

Proteomic samples were prepared following SPEED [84] protocol, that is based on acid cell lysis. First, 4 vol. of trifluoroacetic acid (TFA) were added to the cell pellet followed by incubation for approximately 2 min at room temperature until cells were completely lysed. In addition, the solution was pipetted several times until the viscosity was similar to water. Samples were then neutralized by addition of 10 vol. (corresponding to the amount of TFA used for cell lysis) of 2 M tris(hydroxymethyl)aminomethane (TrisBase). For reduction and alkylation of disulfide bonds 1.1 vol. (according to the amount of TFA used for cell lysis) of 29 mM tris(2-carboxyethyl)phosphine (TCEP) and 37 mM 2-chloroacetamide (CAA) were added and the sample was incubated for 5 min at 95 °C. The cell lysate was diluted 1:5 with water and trypsin (Promega) was added at an enzyme:protein ratio of 1:50. Proteins were hydrolyzed overnight at 37 °C and 600 rpm. The obtained peptides were desalted using Pierce<sup>TM</sup> peptide desalting spin columns (**Section 2.2.15**) and analyzed by LC-MS/MS (**Section 2.2.17**).

### **2.2.9 Sample preparation for LC-MS/MS analysis of cross-linked samples**

For formaldehyde cross-linked samples, the sample preparation protocol SPEED was slightly modified to avoid high temperatures. In detail, 2 M TrisBase were added on ice and reduction and alkylation was performed at 37 °C for 15 min and 300 rpm (see **Section 2.2.8**). The SP3 technology, which is based on paramagnetic beads, was used for further sample preparation. For this, Sera-Mag SpeedBeads were washed three times with water. The cell lysate was then added to the beads. 1 vol. 100% (v/v) ethanol was added followed by incubation for 5 min at 24 °C and 1,000 rpm. Beads and bound protein were subsequently washed three times with 80% (v/v) ethanol. For protein hydrolysis, trypsin (Promega) in 25 mM ammonium bicarbonate, pH 8.5, was added at an enzyme:protein ratio of 1:50. The sample was sonicated for 30 s in a water bath to disaggregate the beads followed by incubation at 37 °C and 1,000 rpm overnight. The sample was centrifuged at 20,000 × g for 1 min at room temperature. The supernatant containing the generated peptides was collected by separating beads and peptides using a magnet. The peptides were dried in a vacuum centrifuge. Enrichment for cross-linked peptide pairs was performed as described (see **Section 2.2.16** for details).

### **2.2.10 In-gel digestion**

Protein bands were excised from the gel and cut into pieces of approximately 1 × 1 mm. 150 µl of water were added and the sample was incubated for 5 min at 26 °C and 1,050 rpm. The gel pieces were spun down in a centrifuge (for 1 min at 16,200 × g) and the supernatant was discarded. 150 µl ACN were added followed by incubation for 15 min at 26 °C and 1,050 rpm. The gel pieces were spun down (for 1 min at 16,200 × g) and the supernatant was discarded and gel pieces were dried in a vacuum centrifuge for 5 min. For reduction of disulfide bridges, 1 mM dithiothreitol (DTT) in 100 mM ammonium bicarbonate, pH 8.0, were added to the gel pieces followed by incubation at 56 °C for 50 min. The gel pieces were spun down (for 1 min at 16,200 × g) and the supernatant was removed. 150 µl ACN were added followed by incubation for 15 min at 26 °C and 1,050 rpm. Again, the gel pieces were spun down (for 1 min at 16,200 × g) and the supernatant was removed. For alkylation of reduced cysteines residues, 100 µl of 55 mM iodoacetamide (IAA) in 100 mM ammonium bicarbonate, pH 8.0, were added and incubated for 20 min at 26 °C and 1,050 rpm. Subsequently, 150 µl of 100 mM ammonium bicarbonate, pH 8.0, and incubated for 15 min at 26 °C and 1,050 rpm. 150 µl ACN were added and incubated as described in the previous

step. The gel pieces were spun down by centrifugation for 1 min at  $16,200 \times g$  and the supernatant was removed. Again, 150  $\mu\text{l}$  ACN were added and incubated as described. The gel pieces were spun down for 1 min at  $16,200 \times g$  and the gel pieces were dried in a vacuum centrifuge for 5 min after the removal of supernatant.

**Table 2.2 Digestion buffer.** The composition of digestion buffer 1 and 2 is shown.

component	digestion buffer 1	digestion buffer 2
trypsin (Roche)	0.0125 $\mu\text{g}/\mu\text{l}$	-
ammonium bicarbonate, pH 8.0	41.67 mM	47.62 mM
calcium chloride	4.17 mM	4.76 mM

Dry gel pieces were hydrated on ice for 45 min with approximately 20  $\mu\text{l}$  digestion buffer 1 containing trypsin (**Table 2.2**) for 45 min on ice. Subsequently, 50  $\mu\text{l}$  of digestion buffer 2 (**Table 2.2**) were added followed by incubation overnight at  $37^\circ\text{C}$  followed.

Gel pieces were spun down (for 1 min at  $16,200 \times g$ ) and 50  $\mu\text{l}$  of water were added and incubated for 15 min at  $37^\circ\text{C}$  and 1,050 rpm. Then, 50  $\mu\text{l}$  of ACN were added and incubated as described in the previous step. The gel pieces were spun down (for 1 min at  $16,200 \times g$ ) and the supernatant was collected in a new sample tube. 50  $\mu\text{l}$  of 5% FA were added to the gel pieces and incubated for 15 min at  $37^\circ\text{C}$  and 1,050 rpm. Again, 50  $\mu\text{l}$  of ACN were added and incubated as described in the previous step. The gel pieces were spun down by centrifugation at  $16,200 \times g$  for 1 min and the supernatant was pooled with the previously collected supernatant. The samples containing the extracted peptides were evaporated to dryness using a vacuum centrifugation. The peptides were stored at  $-20^\circ\text{C}$ .

### 2.2.11 Ethanol precipitation

Proteins were precipitated by addition of 3 vol. ice-cold ethanol and 1/10 vol. 3 M sodium acetate, pH 5.3, followed by incubation at  $-20^\circ\text{C}$  overnight. Then, proteins were pelleted by centrifugation at  $16,200 \times g$  for 30 min and washed with 80% (v/v) ethanol. The centrifugation step was repeated. The obtained protein pellet was dried in a vacuum centrifuge.

### **2.2.12 In-solution digestion in the presence of urea**

The protein pellet obtained from ethanol precipitation was dissolved in 20  $\mu$ l of 8 M urea in 25 mM Tris-HCl, pH 7.9, and incubated for 15 min at room temperature. For reduction of disulfide bridges, 20  $\mu$ l of 10 mM DTT in 8 M urea/ 25 mM Tris-HCl, pH 7.9, was added and incubated for 30 min at 23 °C. Alkylation by addition of 20  $\mu$ l 60 mM IAA in 8 M urea/ 25 mM Tris-HCl, pH 7.9, and incubation for 30 min at 23 °C followed. The sample solution was diluted with 100 mM ammonium bicarbonate, pH 8.5 to a final concentration of 2 M urea. The proteins were hydrolyzed with trypsin (Promega) at a 1:20 enzyme:protein ratio at 37 °C overnight. The generated peptides were desalted using ZipTips (**Section 2.2.14**).

### **2.2.13 In-solution digestion in the presence of RapiGest**

The protein pellet obtained from ethanol precipitation was dissolved in 10  $\mu$ l of 1% (m/v) RapiGest in 25 mM ammonium bicarbonate, pH 8.5 and incubated for 15 min at room temperature. For reduction of disulfide bridges, 10  $\mu$ l of 50 mM DTT in 25 mM ammonium bicarbonate, pH 8.5 were added and followed by incubation at 37 °C for 1 h. Alkylation by addition of 10  $\mu$ l 100 mM IAA in 25 mM ammonium bicarbonate, pH 8.5, and incubation for 1 h at 37 °C was then performed. RapiGest was diluted to 0.1% (m/v) with 25 mM ammonium bicarbonate, pH 8.5, and trypsin (Promega) was added at a 1:20 enzyme:protein ratio followed by incubation at 37 °C overnight. RapiGest was hydrolyzed by addition of 20  $\mu$ l 5% (v/v) TFA and incubation at 37 °C for 2 h. The samples were then centrifuged at 16,200  $\times$  g for 30 min. The supernatant was collected and peptides were dried in a vacuum centrifuge. The obtained peptides were stored at -20 °C.

### **2.2.14 Desalting of peptides using ZipTips**

Peptides obtained from digestion in the presence of urea were desalted using ZipTips (Merck). For this, the C18 material was equilibrated twice with 30  $\mu$ l 60% ACN/0.05% (v/v) FA. ZipTips were adjusted to loading conditions by washing three times with 30  $\mu$ l 0.05% (v/v) FA. The peptide solution (see **Section 2.2.12**) was loaded onto the ZipTip and washed three times with 30  $\mu$ l 0.025% (v/v) FA. Elution was performed using 20  $\mu$ l 60% (v/v) ACN/0.1% (v/v) FA. The elution step was repeated once and eluates containing peptides were combined and dried in a vacuum centrifuge. The peptides were stored at -20 °C.

### **2.2.15 Desalting of peptides using Pierce™ peptide desalting columns**

The peptides were desalted using Pierce™ peptide desalting spin columns according to the manufacturer's protocol. Briefly, the spin column was centrifuged at  $5,000 \times g$  for 1 min to remove the storage buffer. The spin column was then washed twice by adding 300  $\mu$ l ACN followed by centrifugation at  $5,000 \times g$  for 1 min. The column was washed twice with 0.1% TFA (v/v). The peptides were sequentially loaded onto the spin column in 300  $\mu$ l aliquots. After each loading step, centrifugation at  $3,000 \times g$  for 1 min was performed. The column was washed three times with 0.1% TFA (v/v) by centrifugation at  $3,000 \times g$  for 1 min. Peptides were eluted with 300  $\mu$ l of 50% (v/v) ACN/0.1% (v/v) TFA by centrifugation at  $3,000 \times g$  for 1 min. The elution step was repeated once and the desalted peptides were dried in a vacuum centrifuge. The peptides were stored at  $-20 \text{ }^{\circ}\text{C}$ .

### **2.2.16 Enrichment of cross-linked peptide pairs**

For enrichment of cross-linked peptide pairs, peptide size exclusion chromatography was employed. For this, peptides were dissolved in 30% (v/v) ACN/ 0.1% (v/v) TFA. Cross-linked peptide pairs were enriched using an ÄKTA pure chromatography system equipped with a Superdex™ peptide  $3.2 \times 300$  mm column and 30% (v/v) ACN, 0.1% (v/v) TFA as mobile phase. The peptides were separated isocratically with a flow rate of 50  $\mu$ l/min and elution of peptides was monitored at 280 nm. Fractions of 50  $\mu$ l were collected. Early fractions contained cross-linked peptide pairs and late fractions linear peptides. Peptides were dried in a vacuum centrifuge.

### **2.2.17 LC-MS/MS analysis**

Peptides were analyzed by nano-flow reversed-phase liquid chromatography on a DionexUltiMate 3000 RSLCnano System coupled with a Q Exactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer. For this, peptides were first dissolved in 2% (v/v) ACN/ 0.1% (v/v) FA. For liquid chromatography 0.1% (v/v) formic acid (FA) was used as mobile phase A and 80% (v/v) ACN/ 0.1% (v/v) FA was used as mobile phase B. Peptides were loaded onto a trap column ( $\mu$ -Precolumn C18 Acclaim™ PepMap™ 100, C18, 300  $\mu$ m I.D., particle size 5  $\mu$ m) with a flow rate of 10  $\mu$ l/min. The peptides were then separated on an analytical C18 capillary column (50 cm, HPLC column Acclaim™ PepMap™ 100, C18,

75  $\mu\text{m}$  I.D., particle size 3  $\mu\text{m}$ ) with a flow rate of 300 nl/min. A gradient of 4 – 90% (v/v) mobile phase B over different time lengths was applied (see **Table 2.3** for details).

**Table 2.3 Liquid-chromatography gradient.** The gradient consisting of 0.1% (v/v) formic acid (FA) as mobile phase A and 80% (v/v) ACN/0.1% (v/v) FA as mobile phase B is shown for different gradient time length. For cross-linking samples, the 120 min gradient was adjusted step-wise for early (early fractions), middle (see 120 min gradient) and late fractions (late fractions).

A [%]	B [%]	gradient length [min]					
		90 min	120 min	120 min early fractions	120 min late fractions	180 min	300 min
96	4	0	0	0	0	0	0
92	8	3	3	3	3	3	3
85	15	-	-	25	75	-	-
75	25	60	90	90	90	150	270
50	50	64	94	94	94	154	274
10	90	65	95	95	95	155	275
10	90	69	99	99	99	159	279
96	4	70	100	100	100	160	280
96	4	90	120	120	120	180	300

LC-gradients varying in length were applied as follows: 90 min for peptides obtained from labeled model proteins, 120 min for samples containing cross-linked peptide pairs, 180 min for peptides of proteome samples and 300 min for peptides obtained from labeled synaptic vesicle proteins. Fractions of peptide size exclusion chromatography were analyzed using different gradient length, i.e. the 120 min gradient was adjusted step-wise for early, middle and late fractions. Following chromatographic separation, the peptides were directly eluted into the mass spectrometer. Typical mass spectrometric conditions were: data depended mode; capillary voltage, 2.8 kV; capillary temperature, 275 °C and polarity, positive. Survey full scan MS spectra were acquired in a mass range of 350 – 1600 m/z, with a resolution of 70,000 and an automatic gain control (AGC) target at 3e6. The maximum injection time was set to 100 ms (all gradients) and 80 ms (180 min gradient). The 20 most intense peaks were selected for fragmentation in the HCD cell with an AGC target of 1e5 and a fixed first mass of 105 m/z. The resolution of MS2 spectra was 17,500 and for cross-linking samples 35,000. The maximum injection time (MS2 spectra) was 50 ms (for labeled samples), 150 ms (for

proteomic samples) and 200 ms (for cross-linking samples). Ions with a charge of 1 and > 8 were excluded from fragmentation; for cross-linking samples charge states 1 – 3 and > 8 were excluded. In addition, previously selected ions were dynamically excluded for 30 s. Internal calibration of the Orbitrap was performed using the lock mass  $m/z$  445.120025 [222]. QCloud2 [215] was used to control instrument longitudinal performance.

### **2.2.18 Proteomic data analysis using MaxQuant**

For proteomic data analysis, 6 biological replicates of each cell culture condition (undifferentiated, RA- and RA/PMA-differentiated cells) were used. The software MaxQuant (version 1.6.17.0) [107] was used for database search against the human uniprot database (Uniprot, Proteome ID: UP000005640, 73947 entries, version date: 1<sup>st</sup> April 2019). The following standard parameters were used: fixed modification, carbamidomethyl (cysteine); variable modifications, oxidation (methionine) and acetylation (protein N-terminus); max missed cleavage sites, 2; min peptide length, 7; max peptide mass 6,000 Da; peptide FDR, 0.01; protein FDR, 0.01 and enzyme, trypsin/P (cleavage C-terminal of lysine or arginine also when the C-terminal amino acid is proline). In addition, the iBAQ, MaxQuant LFQ and ‘match between runs’ options were enabled.

### **2.2.19 Relative quantification of protein expression of proteins with a specific subcellular localization**

Relative quantification of protein expression of proteins with a specific subcellular localization was performed in RStudio (version 1.2.1335) [218] using R (version 4.0.2) [217]. iBAQ values were calculated by dividing the sum of intensities of the observed peptides of a protein by the number of theoretically observable peptides for each protein [109]. These values were obtained from MaxQuant [107] and normalized for each experiment to obtain relative iBAQ values by dividing the iBAQ by the sum of iBAQ values of all proteins of each experiment. The subcellular localization of proteins with known main localization was assigned using the localization of these proteins from Beltran et al. as a reference [223]. The iBAQ values of all proteins with the same localization was summed for each experiment. The mean and standard error of the iBAQ sum for each localization and culture condition was calculated. Significant differences between the different culture

conditions for each localization were determined by performing a two-tailed t-test. The obtained p-values were adjusted using the Bonferroni correction method [224].

### **2.2.20 Quantification of global protein abundances**

Label-free quantification of proteins was used for calculation of the relative protein abundance. The 'evidence.txt' file obtained from MaxQuant contains information on identified peptides and their intensities and was used for data processing in RStudio (version 1.2.1335)[218] using R (version 4.0.2) [217] with artMS (version 1.9.1) (<http://artms.org>) and MSstats (version 3.22.0) [212]. The following parameters were used for artMS analysis: relative quantification method, global protein abundance; normalization method, equalize medians; cutoffCensored, minFeature; annotation species, human. Differences in protein expression between cell culture conditions was determined using a two-tailed t-test. For calculation of the adjusted p-value the Benjamini and Hochberg method [225] was used. Proteins with a  $\log_2(\text{fold change}) < -0.8$  or  $> 0.8$  and an adjusted p-value  $< 0.05$  were considered as significantly up- or down-regulated. Enriched KEGG pathways within the group of up- or down regulated proteins were determined using Enrichr [207-209].

### **2.2.21 Identification of cross-linked peptide pairs**

A database containing the 800 most abundant proteins was generated for each cell culture condition. For this, the mean of the relative iBAQ of 6 biological replicates of each condition was used. Then raw data files were converted to mascot generic file format (mgf) using Thermo Proteome Discoverer (version 2.4.1.15). Cross-linking analysis was performed using the 'Formaldehyde Cross-link Analyser' [136] employing the following parameters: MS1 tolerance, 6 ppm; MS2 tolerance, 8 ppm; number of missed cleavage sites, 4; the 'use decoy sequence' option was enabled. The results table contains hits of cross-lined peptide pairs with a characteristic mass-shift of 12 or 24 Da [136, 137]. Cross-links with a ratio of ( $\# \text{ observed fragments} / \# \text{ amino acids of the cross-linked peptide pair}$ )  $> 1.5$  as well as a total number of fragments for each peptide  $> 18$  were considered to be high-confident cross-links [136]. Cross-links with a total number of fragments for each peptide  $> 15$  were considered to be intermediate confident cross-links [136]. For identification of the overlap of high and intermediate confident cross-links between the three replicates for each condition the web-based venn-tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) was used. Depending



on the reproducibility of the identified cross-links between replicates, scores were assigned (for details see **Table 2.4**).

**Table 2.4 Cross-link scores.** The overlap of identified high-confident and intermediate confident cross-links within the three replicates for each condition was used to assign a cross-link score.

number and type of cross-link	score
one high confident cross-link each in three replicates	1
one high confident cross-link in two of three replicates	2
one high confident cross-link in one replicate and at least one intermediate confident cross-link in another replicate	3
one intermediate confident cross-link each in three replicates	4
one intermediate confident cross-link in two replicates	5
one high confident cross-link in one replicate	6
one intermediate confident cross-link in one replicate	7

For further analysis, only cross-links with a score  $\geq$  four were used. Network and bar plots of the identified cross-links were generated in Cytoscape (version 3.8.2) [206] using the Cytoscape app XlinkCyNET (version 1.2.5) [221]. In addition cross-links of proteins with a known high-resolution structure were analyzed using UCSF Chimera (version 1.15) [219] and the software tool Xlink Analyzer (version 1.1.4) [220].

### 2.2.22 Identification of chemically labeled amino acids<sup>3</sup>

Identification of labeled amino acids using NHS-acetate, S-NHS-acetate or DEPC as labeling reagents was performed with the software MaxQuant (version 1.6.3.3) [107]. For the labeled model proteins ADH (yeast *Saccharomyces cerevisiae*) and PK (rabbit), the corresponding protein amino acid sequence was used as database. The database for the identification of labeled amino acids of synaptic vesicles proteins was generated by Sabine Wittig using the 400 most abundant synaptic vesicle proteins [226]. Standard search parameters were used: mass accuracy for precursor ions in main search, 4.5 ppm; mass accuracy for fragment ions, 0.5 Da; max. missed cleavage sites, 2; min. peptide length, 7;

<sup>3</sup> The raw data of NHS-acetate labelled ADH and PK of three biological replicates, as well as the first two replicates of DEPC labelled ADH and PK were previously obtained (Master thesis: Chemical modification of protein for structure elucidation, Marie Barth, Martin Luther University, 2018). The project was continued during this thesis.

max. peptide mass, 6,000 Da; peptide FDR, 0.01; protein FDR, 0.01; enzyme, trypsin/P; the ‘match between runs’ option was enabled; variable modifications, acetylation of the protein’s N-terminus, carbamidomethylation of cysteine, oxidation of methionine. In addition, for labeling with NHS-acetate or sulfo-NHS-acetate, acetylation of lysine, serine, threonine and tyrosine were set as variable modification (for details see **Table 2.5**). For DEPC labeling, variable CEt-modification of histidine, lysine, cysteine, serine, threonine and tyrosine including neutral losses from histidine, lysine, serine and threonine, formyl-CEt-modified histidine, di-CEt-modified histidine and urethane-CEt-modified histidine including neutral loss were allowed (**Table 2.5**).

**Table 2.5 Labeling induced modifications.** The modifications introduced by S-NHS-aceate, NHS-acetate and DEPC labelling are listed. The individual atomic composition, the introduced mass shift of a modified amino acid within the peptide as well as the composition and mass shift of a neutral loss occurring during HCD fragmentation of labeled peptides is shown.

modification	composition	mass shift	neutral loss composition	neutral loss mass shift
acetylation of lysine, serine, threonine and tyrosine	C <sub>2</sub> H <sub>2</sub> O	42.01 Da	-	-
CEt-modified histidine, lysine,	C <sub>3</sub> H <sub>4</sub> O <sub>2</sub>	72.02 Da	C <sub>2</sub> H <sub>6</sub> O	46.04 Da
CEt-modified serine and threonine	C <sub>3</sub> H <sub>4</sub> O <sub>2</sub>	72.02 Da	C <sub>3</sub> H <sub>6</sub> O <sub>3</sub>	90.03 Da
CEt-modified cysteine and tyrosine	C <sub>3</sub> H <sub>4</sub> O <sub>2</sub>	72.02 Da	-	-
formyl-CEt-modified histidine	C <sub>6</sub> H <sub>10</sub> O <sub>5</sub>	162.05 Da	-	-
di-CEt-modified histidine	C <sub>6</sub> H <sub>10</sub> O <sub>4</sub>	146.06 Da	-	-
urethane-CEt-modified histidine	C <sub>5</sub> H <sub>10</sub> O <sub>4</sub>	134.06 Da	C <sub>3</sub> H <sub>7</sub> O <sub>2</sub> N	89.05 Da

### 2.2.23 Quantification of modified amino acids of model proteins

The intensity values obtained from MaxQuant for modified and unmodified peptides were used for quantification. Data processing and analysis was performed with R (version 3.5.1) [217] using RStudio (version 1.2.1335) [218] and the following R packages: ggforce, openxlsx, reshape and tidyverse. Normalized intensities for each modified site were

calculated. In detail, MaxQuant output files containing information on the individual modifications (e.g. acetylation of lysine or acetylation of serine, threonine, tyrosine) were combined and filtered for the protein of interest. Modified sites with a localization probability < 0.75 and a peptide score < 80 were discarded. For the obtained labeled residues, the intensity was normalized by the sum of all corresponding peptides (i.e., all modified and unmodified peptides) containing the respective site.

The following equation was applied:

$$\frac{(Intensity\ of\ modified\ residue) * 100}{sum\ corresponding\ peptides\ (modified + unmodified)}$$

The mean normalized intensity and standard error for each labeled site were then calculated.

#### **2.2.24 Calculation of the solvent accessible surface area (SASA) and visualization of labeled residues on model proteins**

The tetramer of ADH based on PDB ID 5env [227] and the PK tetramer (PDB ID 1aqf [228]) were used as templates for calculation of SASA of single residues using the webserver GETAREA [210]. For calculation of mean SASA of each residue the two chains of ADH or the four chains of PK were used. PDBePISA (version 1.52) [213] was used to calculate the surface area and the buried area. PyMOL (version 1.3) [216] was used to visualize the tetrameric protein complexes and specific residues within the crystals structure of ADH (PDB ID 5env) and PK (PDB ID 1aqf).

#### **2.2.25 Quantification of labeled residues of synaptic vesicle proteins**

For Quantification of labeled sites of synaptic vesicle proteins, the script described in **Section 2.2.23** was employed. For the two DEPC labeling reagent concentrations, the mean intensity of 2 biological replicates was calculated for each residue. Labeling percentages were calculated for each residue.

For Analysis of acetylated residues, again the described script (**Section 2.2.23**) was then employed to calculate labeling intensities. Significant differences in labeling efficiency were identified using a two-tailed t-test. Residues with a labeling intensity change above 3% and a p-value < 0.05 were considered as changed. In detail, residues with a higher labeling

intensity using S-NHS-acetate compared to NHS-acetate are considered to be located outside and residues with a higher intensity of NHS-acetate compared to S-NHS-acetate were considered to be located in the lumen of synaptic vesicles.

In addition labeled residues of the proton-pumping vacuolar H<sup>+</sup>-ATPase (V-ATPase) were analyzed using UCSF Chimera (version 1.15) [219] and the software tool Xlink Analyzer (version 1.1.4) [220]. For synapsin-1, Rab3a (Ras-related protein Rab-3A) and Cysteine string protein (CSP), each protein's 3D structure was predicted using AlphaFold [229]. PyMOL (version 1.3) [216] was used to visualize the structures.

### **2.2.26 Data availability**

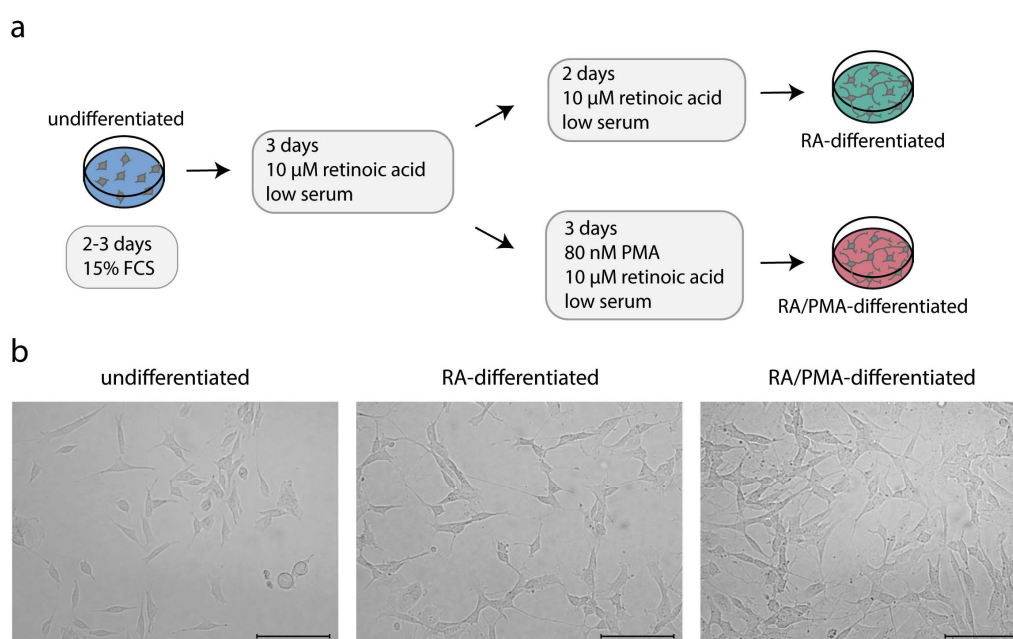
The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository [214]. The dataset identifier for the data from labeling of model proteins is PXD015940 (publicly accessible), for labelled synaptic vesicles PXD020859 (publicly accessible) and for proteomic and cross-linking data of SH-SY5Y cells PXD031054 (Accessible for review with the following log in account: Username: reviewer\_pxd031054@ebi.ac.uk; Password: uLbEpUcP).

### 3. Results

#### 3.1 Proteomic characterization of SH-SY5Y cells<sup>4</sup>

##### 3.1.1 Workflow for proteomic analysis of undifferentiated and differentiated SH-SY5Y cells

The neuroblast-like cell line SH-SY5Y was selected as a neuronal model system. When specific growth conditions are applied, cells with neuronal character can be obtained. Therefore, cells were first grown under standard conditions employing 15% FCS and passaging of the cells every two to three days (undifferentiated cells). Undifferentiated cells were then cultured following a specific protocol to obtain differentiated cells. For differentiation, the cells were, therefore, treated at low serum conditions with RA for five days (RA-differentiation) or with RA for three days followed by three days with RA and PMA (RA/PMA-differentiation) (**Figure 10a** and **Section 2.2.2**). The morphology of the cells changed upon differentiation (**Figure 10b**). Accordingly, undifferentiated cells are characterized by only a few short projections, while differentiated cells have long pronounced projections. However, differentiation did not only result in cells with a neuronal morphology but also in epithelial-like cells (widespread cells without projections).

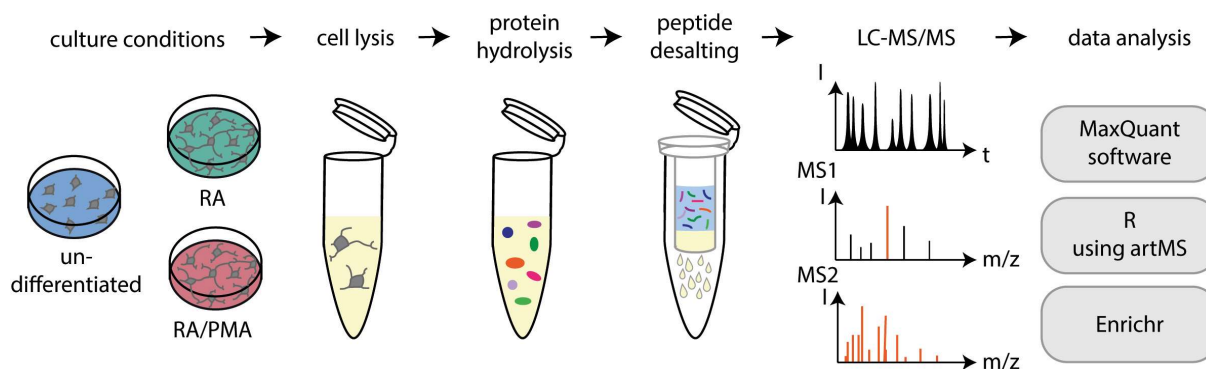


**Figure 10: Differentiation of SH-SY5Y cells.**

**a)** Cells were first grown under standard conditions (undifferentiated). Differentiation was then induced for 5 days with RA (RA-differentiated) or 3 days with RA followed by 3 days of RA/PMA (RA/PMA-differentiated). **b)** The morphology of undifferentiated, RA-differentiated and RA/PMA-differentiated cells was monitored with an inverted Primovert light microscope at 20 × magnification equipped with an 18 MP microscope digital camera. Scale bar: 100 μm. Abbreviations: FCS, fetal calf serum.

<sup>4</sup> The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository [214] with the dataset identifier PXD031054 (Accessible for review with the following log in account: Username: reviewer\_pxd031054@ebi.ac.uk; Password: uLbEpUcP).

Cells of the three different culture conditions were then employed for a proteome comparison. The workflow is shown in **Figure 11**. Six biological replicates of each cell culture condition were used. For this, cells were harvested and lysed using TFA. Protein hydrolysis with trypsin was followed by desalting of the peptide solution. The obtained peptides were then analyzed by LC-MS/MS analysis. Peptides and corresponding proteins were identified using the software MaxQuant [211]. The results were further processed with an in-house developed script in R [217, 218] employing the software package artMS.



**Figure 11: Workflow – Proteome analysis of SH-SY5Y cells**

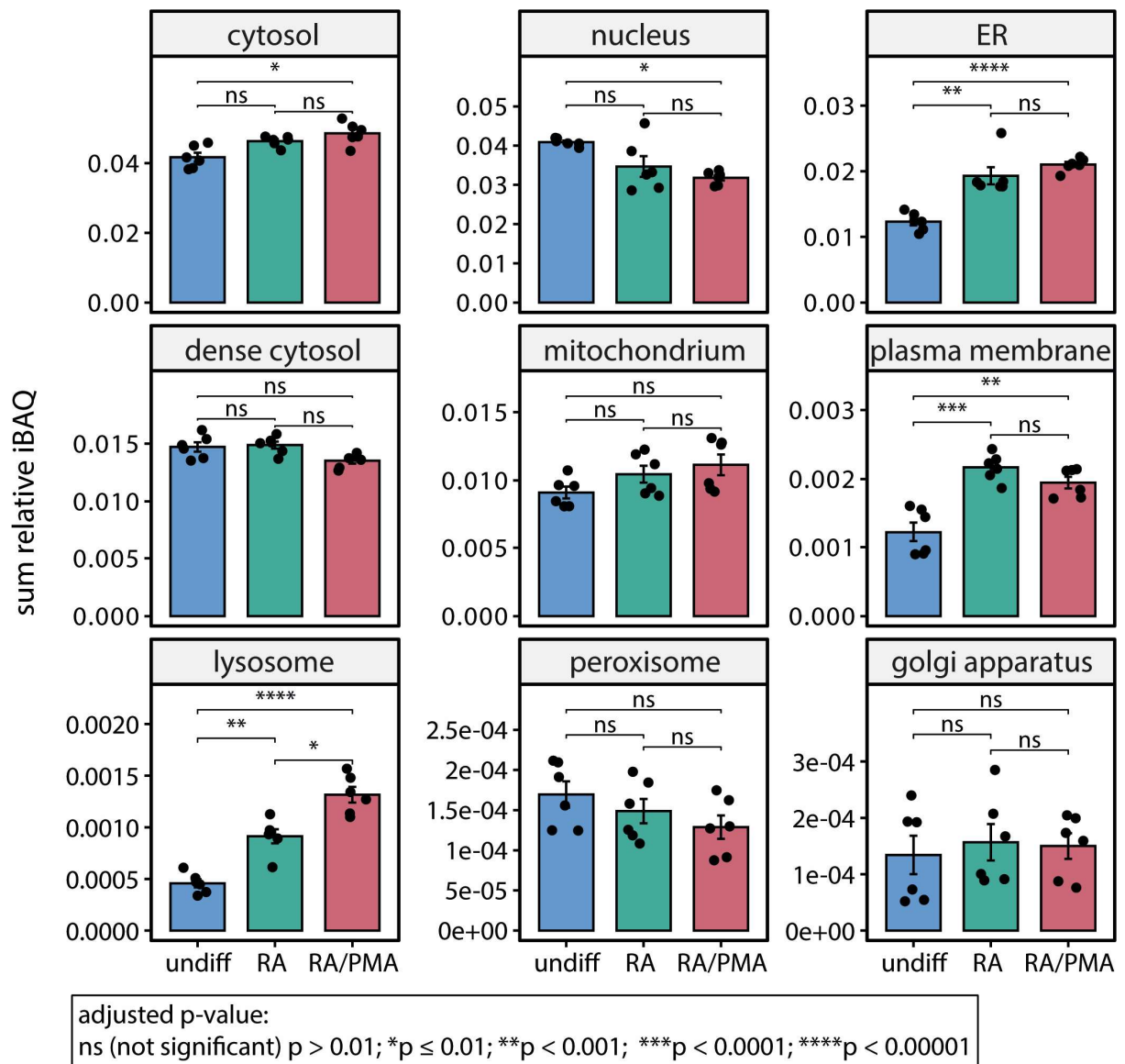
Cells grown under different conditions are lysed and proteins are hydrolyzed with trypsin. Obtained peptides are desalted and analyzed by LC-MS/MS. For data analysis different software is employed.

For this workflow, 5909 different proteins corresponding to approximately 60 % of the expected 10,000 proteins within a cell [230] were identified. Of these proteins, 3661 were identified in all 18 samples (i.e. six replicates for undifferentiated, RA- and RA/PMA-differentiated cells) confirming a good reproducibility of the analysis strategy. Among the identified proteins are typical cellular proteins that occur at high copy numbers [231]; examples are, actin beta, vimentin, enolase 1 as well as ribosomal proteins and histones. Proteins with low copy numbers [231] such as cadherin-2, rotatin or the mitochondrial ribosomal protein S30 were also observed. In conclusion, the employed workflow is well suited for protein identification and the results will be further used for relative quantification of protein abundance.

### 3.1.2 Relative quantification of proteins with a specific subcellular localization

To obtain first insights into changes of protein abundance between the different cell culture conditions, the abundance of proteins with a specific subcellular localization was analyzed. For this, the relative iBAQ [109] for each protein was calculated to identify the abundance

of each protein in the individual samples. Relative iBAQ values of all proteins with a specific subcellular localization [223] were then summed. The resulting protein abundance of each subcellular localization was then compared between the different cell culture conditions. In total, nine different subcellular localizations, namely the cytosol, nucleus, ER, dense cytosol, mitochondria, plasma membrane, lysosome, peroxisome and the golgi apparatus were evaluated. For identification of significantly different protein abundance corresponding to a specific subcellular localization, a two-tailed t-test was applied. An increase in protein abundance in differentiated cells when compared with undifferentiated cells was identified for proteins localized in the ER, plasma membrane and the lysosome (**Figure 12**).



**Figure 12: Relative quantification of proteins with a specific subcellular localization.** The relative iBAQ values of all proteins with a specific subcellular localization identified in undifferentiated (undiff, blue), RA-differentiated (RA, green) or RA/PMA-differentiated (RA/PMA, red) cells was summed. For each cell culture condition, the mean value and the standard error are given (n=6). The adjusted p-value was calculated to determine significant differences between the conditions (see legend for details). Abbreviations: ER, endoplasmic reticulum.

In addition, RA/PMA differentiated cells showed an increasing protein abundance of cytosolic proteins and a decreasing abundance of proteins located in the nucleus compared to undifferentiated cells. Notably, only minor differences were observed between RA- and RA/PMA-differentiated cells. The abundance of proteins located in the lysosome was lower in RA-differentiated cells when compared with cells obtained from RA/PMA-differentiation. The results of this analysis indicate similar changes in protein expression in RA- and RA/PMA-differentiated cells during neuronal differentiation.

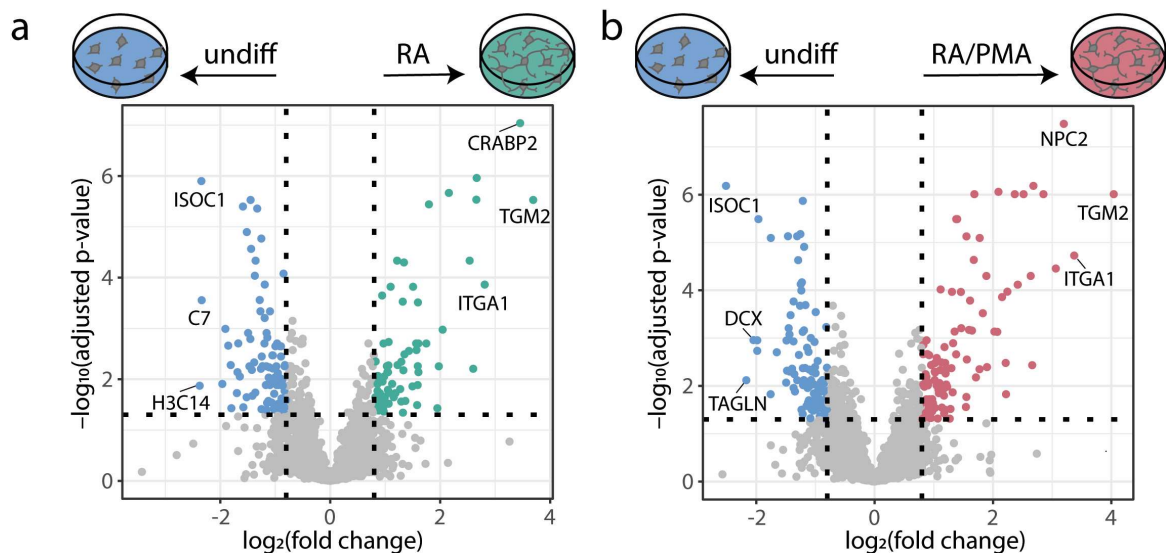
### **3.1.3 Proteome analysis of RA- and RA/PMA-differentiated cells uncovers markers for early neuronal differentiation**

The next step was to compare undifferentiated and differentiated SH-SY5Y cells in more detail by relatively quantifying individual proteins. For this, intensities of the individual peptides were obtained from MaxQuant [107] data analysis and normalized for all proteins in each biological replicate. A two tailed t-test was then applied to identify statistically significant changes in relative protein abundance between the different growth conditions. The false discovery rate was controlled by applying the Benjamini and Hochberg method [232] and adjusted p-values were calculated. Following this approach, only proteins identified in both samples can be relatively compared. When comparing RA-differentiated cells with undifferentiated cells 3787 proteins, when comparing PMA/RA-differentiated cells with undifferentiated cells 3738 proteins and when comparing RA- and RA/PMA-differentiated cells 3837 proteins were quantified. Volcano plots were obtained by plotting adjusted p-values against the fold change of protein intensities (**Figure 13**). The plot enables visualization of the proteomics data and visual identification of proteins with large fold changes that are statistically significant.

For the identification of significantly up- and downregulated proteins, a significance threshold of 5% and a  $\log_2(\text{fold change})$  threshold  $< -0.8$  and  $> 0.8$  was applied. Applying this significance threshold, 71 proteins were identified to be upregulated in RA-differentiated cells and 101 in RA/PMA-differentiated cells when comparing differentiated and undifferentiated cells. The comparison of RA- and RA/PMA-differentiated cells revealed only 6 proteins that were differentially expressed. These results again highlight the similarity of neurons obtained from the two different protocols.



In both, RA- and RA/PMA-differentiated cells, transglutaminase 2 was identified with the highest fold change when compared with the undifferentiated cells. For this protein, a 16- and 13-fold upregulation in RA- and RA/PMA-differentiated neurons was observed (**Figure 13** and **Table 3.1**). An increased expression of transglutaminase 2 following RA treatment was previously described [233] and is therefore expected. In addition to transglutaminase 2, other proteins such as the cellular retinoic acid binding protein are directly linked to RA treatment [234, 235] (**Table 3.1**). This protein binds RA [236], is a co-activator of nuclear RA receptors [237] and by mediating transcriptional activation regulates differentiation [238], neurite initiation and neurite branching [239]. As expected, the cellular retinoic acid binding protein is enriched in RA treated cells. Retinol binding protein 1 and retinal dehydrogenase 10 were upregulated to a lower extent (**Table 3.1**). Of these, the cellular retinol-binding protein 1 is epigenetically regulated by RA receptor alpha [240] and the retinal dehydrogenase 10 protein together with dehydrogenase reductase 3, which was only identified in differentiated cells, form the hetero-oligomeric retinoid oxidoreductase complex essential for controlling RA homeostasis [241]. Treatment with RA, therefore, induced expression of several proteins playing a role in RA metabolism essential for differentiation.



**Figure 13: Relative quantification of undifferentiated versus RA- and RA/PMA-differentiated SH-SY5Y cells.**

The  $\log_2(\text{fold change})$  was plotted against the  $-\log_{10}(\text{adjusted p-value})$  in volcano plots. Quantified proteins with a  $\log_2(\text{fold change}) < -0.8$  or  $> 0.8$  and an adjusted p-value  $< 0.05$  were considered significantly up- or down-regulated, respectively. The gene names of the three top hits that are up- or downregulated in undifferentiated or differentiated cells are labeled in the volcano plots. **a)** Volcano plot of RA-differentiated cells (RA, green) versus undifferentiated cells (undiff, blue). **b)** Volcano plot of RA/PMA differentiated cells (RA/PMA, red) versus undifferentiated cells (undiff, blue). Abbreviations: ISOC1, isochorismatase domain containing 1; C7, complement C7; H3C14, H3 clustered histone 14; ITGA1, integrin subunit alpha 1; CRABP2, cellular retinoic acid binding protein 2; TGM2, transglutaminase 2; DCX, doublecortin; TAGLN, transgelin; NPC2, NPC intracellular cholesterol transporter 2.

Besides proteins that are directly linked with RA-related pathways, a group of proteins that play a role in neuronal differentiation was significantly upregulated in differentiated cells (**Table 3.1**). Among these proteins, annexin A2 and cathepsin B induce neurite outgrowth [242-244], secretogranin II mediates neuronal differentiation [245], vasodilator stimulated phosphoprotein plays a role in filipodia formation [246], doublecortin like kinase 1 mediates structural rearrangements [247] and sequestosome 1 regulates the metabolic shift from aerobic glycolysis to oxidative phosphorylation [248] (**Table 3.1**). The expression level of nestin, a marker protein for early neuronal differentiation [249], was found to be increased in differentiated cells (**Table 3.1**), while doublecortin, a marker for developing neurons [250], was found to be down-regulated.

**Table 3.1 Upregulated proteins in differentiated cells.**

The Uniprot accession number, the gene and protein name, the fold change (FC), the log<sub>2</sub>(FC) and the adjusted p-value (adj. p-value) are given for selected protein groups. FC ratios were calculated for RA-differentiated *versus* undifferentiated cells as well as for RA/PMA-differentiated versus undifferentiated cells.

Uniprot accession number	gene name	protein name	RA vs. undifferentiated			RA/PMA vs. undifferentiated		
			FC	log <sub>2</sub> FC	adj. p-value	FC	log <sub>2</sub> FC	adj-p-value
<b>RA related</b>								
P21980	TGM2	transglutaminase 2	12.93	3.69	2.97E-06	16.46	4.04	9.78E-07
P29373	CRABP2	cellular retinoic acid binding protein 2	10.97	3.46	9.13E-08	5.73	2.52	9.78E-07
P09455	RBP1	retinol binding protein 1	2.45	1.29	4.53E-03	2.13	1.09	9.59E-03
Q8IZV5	RDH10	retinol dehydrogenase 10	3.01	1.59	1.30E-02	3.73	1.9	4.04E-03
<b>neuronal differentiation</b>								
P07355	ANXA2	annexin A2	1.99	0.99	2.23E-02	2.3	1.2	5.49E-03
P07858	CTSB	cathepsin B	2.7	1.43	2.77E-03	3.13	1.65	6.96E-04
P48681	NES	nestin	2.09	1.06	1.85E-03	3.42	1.78	8.06E-06
P13521	SCG2	secretogranin II	3.93	1.98	5.56E-03	2.93	1.55	1.70E-02
P50552	VASP	vasodilator stimulated phosphoprotein	2.09	1.06	3.03E-02	2.34	1.23	1.04E-02
Q13501	SQSTM1	sequestosome 1	5.8	2.54	4.64E-05	4.74	2.24	1.07E-04
O15075	DCLK1	doublecortin like kinase 1	1.85	0.89	2.39E-02	2.28	1.19	3.31E-03
<b>oxidative stress related</b>								
O75874	IDH1	isocitrate dehydrogenase (NADP(+)) 1	1.83	0.87	1.02E-02	1.8	0.85	9.33E-03
P04424	ASL	argininosuccinate lyase	3.02	1.59	5.53E-03	4.21	2.07	7.41E-04
P30043	BLVRB	biliverdin reductase B	2.55	1.35	3.24E-03	3.02	1.6	6.74E-04
P02795	MT2A	metallothionein 2A	3.86	1.95	3.72E-02	4.66	2.22	1.49E-02
P15559	NQO1	NAD(P)H quinone dehydrogenase 1	1.94	0.95	1.11E-02	2.19	1.13	3.23E-03
P07602	PSAP	prosaposin	1.82	0.87	8.59E-03	3.19	1.68	2.32E-05
P04179	SOD2	superoxide dismutase 2	2.99	1.58	3.23E-02	2.91	1.54	2.76E-02
P13521	SCG2	secretogranin II	3.93	1.98	5.56E-03	2.93	1.55	1.70E-02

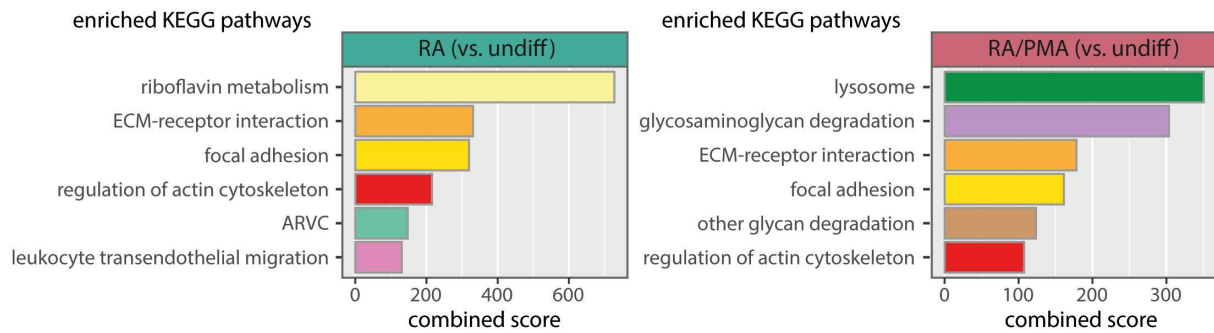
Other proteins such as proteins related to the synaptic vesicle cycle (e.g., synaptophysin, SNAP25, synapsin 1 and synaptotagmin 11, clathrin heavy chain like 1) were only quantified in differentiating cells. Given that these proteins were not quantified in undifferentiated cells, indicates very low abundance or absence in undifferentiated SH-SY5Y cells.

Observed changes suggest that early differentiating neurons are obtained when following our protocols. Interestingly, downregulated of doublecortin in differentiating cells suggests that very early differentiation stages have already passed. Furthermore, the presence of synaptic vesicle proteins indicates the establishment of neuronal synapses.

Neurons have a high energy demand and oxygen consumption and are, therefore, particularly vulnerable to oxidative stress [251]. Accordingly, proteins related to antioxidant defense were found to be upregulated upon differentiation (**Table 3.1**) indicating that antioxidant defense mechanisms are already established during early differentiation. In conclusion, differentiation with RA and RA/PMA leads to an increase in the expression of proteins related to RA administration, proteins required for neuronal development, synaptic proteins and proteins of the antioxidant defense mechanism.

#### **3.1.4 KEGG pathway analysis revealed structural rearrangements during neuronal differentiation**

Identified proteins that were significantly enriched in differentiated cells were further analyzed. For this enriched KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways were determined (**Supplementary Table 1-4**). In detail, corresponding genes of upregulated proteins in differentiated cells were used as input for the gene set search engine Enrichr [207, 208, 252]. This approach resulted in the identification of 21 and 20 enriched KEGG pathways for RA- and RA/PMA-differentiated cells, respectively (**Supplementary Table 1**). Among the top six enriched KEGG pathways in both differentiated cells are ‘regulation of actin cytoskeleton’, ‘extracellular matrix receptor (ECM-receptor) interaction’ and ‘focal adhesion’ (**Figure 14**). These three pathways are closely related to structural rearrangements of the cells.



### Figure 14: Enriched KEGG pathways in differentiated cells

Quantified proteins with a  $\log_2(\text{fold change}) < -0.8$  or  $> 0.8$  and an adjusted  $p\text{-value} < 0.05$  were considered significantly up- or downregulated, respectively. Significantly up- or downregulated proteins were used for KEGG pathway enrichment analysis. Enriched KEGG pathways with high combined scores were determined in RA-differentiated cells vs. undifferentiated cells (green) and RA/PMA differentiated vs. undifferentiated cells (red) using the search engine Enrichr [207, 208, 252]. Abbreviations: ECM, extra cellular matrix; ARVC, Arrhythmogenic right ventricular cardiomyopathy.

The previously described long projections of differentiated SH-SY5Y cells (**Figure 10**) likely require increased focal adhesion and ECM interactions. In agreement with this, proteins such as the vasodilator stimulated phosphoprotein which promotes filopodia formation or paxillin, which plays a role in F-actin assembly and dynamics [253-255], are part of these KEGG pathways.

In RA-differentiated cells the ‘riboflavin metabolism’ pathway was enriched and assigned with the highest score. Two proteins, namely biliverdin reductase B and acid phosphatase 2, are upregulated; they catalyze the reduction of flavin mononucleotide finally yielding fully reduced flavohydroquinone. This catalytic activity might be relevant for oxygen activation during differentiation [256].

In RA/PMA-differentiated cells the pathway ‘lysosome’ was highly enriched. This pathway was also enriched in RA-differentiated cells albeit at a lower score. Proteins within the lysosome are required for macromolecular degradation of proteins, lipids, DNA, RNA and carbohydrates. They further enable recycling of the obtained components. Important roles of axonal and dendritic lysosomes are clearance of dysfunctional synaptic proteins and organelles [16, 17, 257]. The importance of correct lysosomal function is further stressed by linkage of genetical disorders to dysfunction of neuronal lysosomes [258]. Therefore, the enriched KEGG pathway ‘lysosome’ in differentiating SH-SY5Y cells reveals that these cells are challenged with a higher demand for degradation processes during neuronal differentiation.

In summary, KEGG pathway analysis of enriched proteins in differentiated cells, again, confirms the similarity between the cells obtained from RA- or RA/PMA-differentiation.

For instance, enriched KEGG pathways associated with actin regulation and ECM interaction indicate structural and functional changes during differentiation. Differences between the cells obtained using the two differentiation protocols were mainly identified for lysosomal proteins. This confirmed that, in agreement with the results of the quantification of proteins associated with specific organelles, there are only minor differences between RA- and RA/PMA-differentiated neurons.

### **3.1.4 Proliferation is a main characteristic of undifferentiated SH-SY5Y cells**

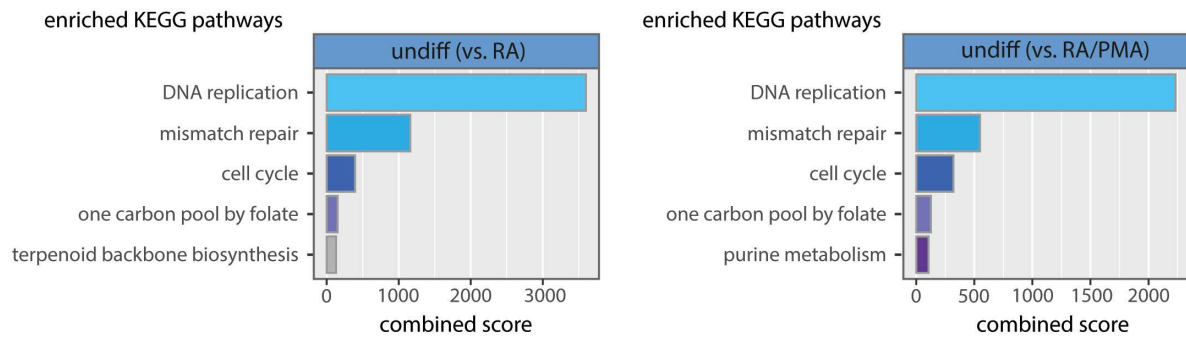
In the next step, protein expression in the undifferentiated cells was compared with the expression levels in differentiated cells in more detail. For identification of significantly upregulated proteins, the same threshold as applied to differentiated cells (see above) was employed, and 86 and 98 proteins were identified as significantly upregulated in undifferentiated cells when compared with RA- and RA/PMA-differentiated cells, respectively (**Figure 13**). Among the highest upregulated proteins was isochorismatase domain containing protein 1 (**Table 3.2**), a protein that is related to ontogenesis [259]. Recently, Cheng et al. observed induced apoptosis and suppressed cell proliferation upon knockdown of isochorismatase domain containing protein 1 in pancreatic cancer [259] suggesting similar proliferation associated functions of this protein in neuroblastoma cells used here. The actin isoform, actin beta like 2 was also enriched in undifferentiated cells compared to differentiated cells. In previous studies, this specific actin isoform was found to be upregulated in colorectal cancer [260]. In addition, proteins regulating cell cycle checkpoints such as cyclin dependent kinase 1 [261] and minichromosome maintenance proteins [262] are enriched in undifferentiated cells (**Table 3.2**). The upregulation of cancer-related proteins and proteins related to the cell cycle reveals cancer characteristics of undifferentiated SH-SY5Y neuroblastoma cells.

**Table 3.2 Upregulated proteins in undifferentiated cells.**

The Uniprot accession number, the gene and protein name, the fold change (FC), the log<sub>2</sub>(FC) and the adjusted p-value (adj. p-value) are given for selected protein groups. FC ratios were calculated for RA-differentiated *versus* undifferentiated cells as well as for RA/PMA-differentiated versus undifferentiated cells.

RA-treatment related								
Uniprot accession number	gene name	protein name	RA vs. undifferentiated			RA/PMA vs. undifferentiated		
			FC	log <sub>2</sub> FC	adj. p-value	FC	log <sub>2</sub> FC	adj.p-value
<b>proliferation</b>								
Q96CN7	ISOC1	isochorismatase domain containing 1	0.2	-2.34	1.26E-06	0.18	-2.51	6.56E-07
P49736	MCM2	minichromosome maintenance complex component 2	0.42	-1.25	1.71E-05	0.4	-1.31	7.41E-06
P25205	MCM3	minichromosome maintenance complex component 3	0.44	-1.2	1.98E-03	0.37	-1.42	3.31E-04
P33991	MCM4	minichromosome maintenance complex component 4	0.44	-1.19	6.22E-04	0.47	-1.08	9.95E-04
P33992	MCM5	minichromosome maintenance complex component 5	0.45	-1.14	1.51E-03	0.44	-1.19	6.89E-04
Q14566	MCM6	minichromosome maintenance complex component 6	0.39	-1.36	4.64E-05	0.42	-1.25	7.46E-05
P33993	MCM7	minichromosome maintenance complex component 7	0.44	-1.19	1.37E-04	0.43	-1.23	6.81E-05
P06493	CDK1	cyclin dependent kinase 1	0.37	-1.45	2.97E-06	0.44	-1.19	1.23E-05
O43602	DCX	doublecortin	0.28	-1.85	2.18E-03	0.24	-2.05	1.10E-03

For the upregulated proteins in undifferentiated, as described above, upregulated KEGG pathways were identified. The pathways ‘DNA replication’, ‘mismatch repair cell cycle’ and ‘one carbon pool by folate’ were found to be upregulated and are associated with the high proliferation rates of undifferentiated cells (**Figure 15, Supplementary Table 1 and 2**). Proteins identified within the ‘DNA replication’ pathway are for instance, the minichromosome maintenance proteins 2 to 7. They act as DNA helicases and regulate cell cycle checkpoints [262]. In the ‘mismatch repair’ pathway, MutS homolog 2 and 6 were observed. MutS homolog 2 and 6 recognize mispaired bases or insertion loops [263-265]. In agreement with our results, a decrease in expression upon differentiation was previously reported for these two proteins [266]. The ‘one carbon metabolism’ plays an important role in maintaining high proliferation rates of cancer cells by providing one-carbon units for biosynthesis and redox reactions. Within this pathway, the proteins dihydrofolate reductase and thymidylate synthase are upregulated (**Table 3.2 and Supplementary Table 1**).



**Figure 15: Enriched KEGG pathways in undifferentiated cells**

Quantified proteins with a  $\log_2(\text{fold change}) < -0.8$  and an adjusted p-value  $< 0.05$  were considered significantly up-regulated in undifferentiated cells compared to differentiated cells. Significantly up-regulated proteins in undifferentiated cells were used for KEGG pathway enrichment analysis. Enriched KEGG pathways with high combined scores in undifferentiated cells were identified using Enrichr [207, 208, 252].

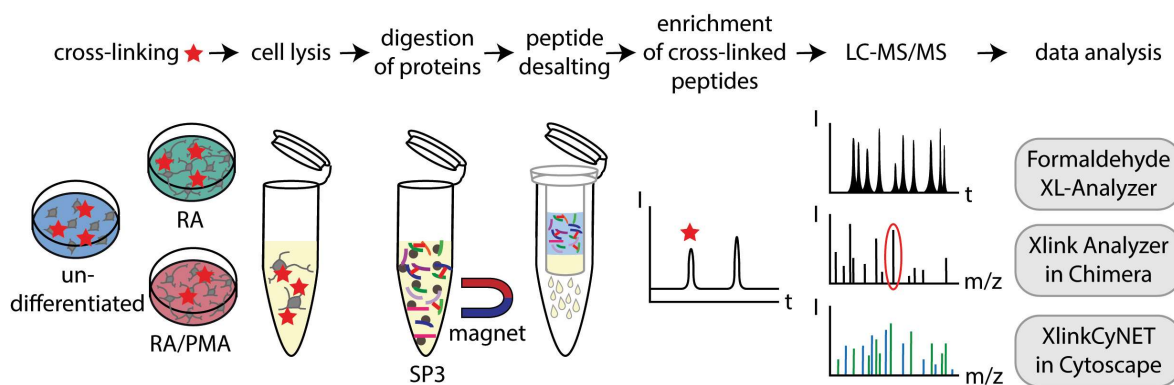
The enriched KEGG pathways identified in undifferentiated cells compared to differentiated cells confirm the identified typical cancer characteristics of undifferentiated cells. Proteins identified in the respective pathways are commonly required for constant proliferation of the cells and cancer progression. Undifferentiated SH-SY5Y cells, therefore, represent neuroblastoma cells which have not yet developed a neuronal character.

## 3.2 Protein cross-linking in SH-SY5Y cells

### 3.2.1 Workflow for in-cell cross-linking

Quantitative proteomics was used to explore changes in protein expression upon differentiation of SH-SY5Y cells. These results suggested structural rearrangements and, therefore, protein interactions within undifferentiated as well as RA- and RA/PMA-differentiated cells were compared by in-cell cross-linking. For this, formaldehyde was used as cross-linking reagent. Formaldehyde is membrane permeable and covalently cross-links proteins in close proximity. Primarily lysine and arginine residues but also asparagine, histidine, aspartic acid, tyrosine and glutamine residues are linked [136].

First, cells were incubated with the short-range cross-linker formaldehyde (**Figure 16**). Upon cross-linking, the cell pellet was lysed using TFA and proteins were hydrolyzed with trypsin on magnetic beads (**Section 2.2.9**). Low-abundant cross-linked peptide pairs were enriched by size exclusion chromatography and corresponding peptide fractions were subsequently analyzed by LC-MS/MS. Cross-linked peptide-pairs were then identified using the ‘Formaldehyde XL Analyzer’ software [136] and a database containing the 800 most abundant proteins identified for the corresponding cell culture condition (see above) used. The results include cross-links within the same protein (intra-molecular protein cross-links) and cross-links between two different proteins (inter-molecular protein cross-links).



### Figure 16: Workflow for in-cell cross-linking of proteins using formaldehyde

SH-SY5Y cells were grown under standard conditions (undifferentiated) or differentiated with RA (RA) or differentiated with RA and PMA (RA/PMA). Proteins were cross-linked using formaldehyde. The cells were pelleted by centrifugation and lysed using TFA. Proteins were enzymatically hydrolyzed using the single-pot, solid-phase-enhanced sample-preparation (SP3) technology. The obtained solution was desalted and peptides were analyzed by LC-MS/MS. Data analysis was performed using different software tools.

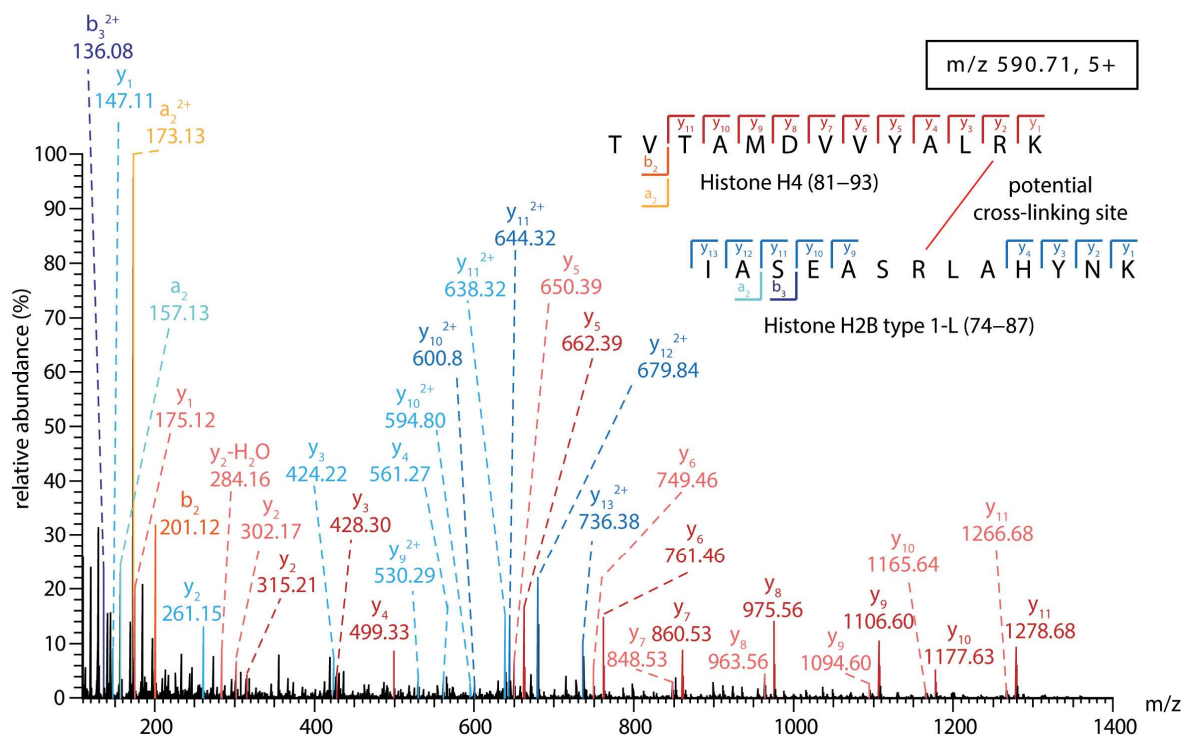
### 3.2.2 Identification of cross-linked peptide pairs

For identification of a cross-linked peptide pair, the precursor mass (recorded as  $m/z$  in the MS1 spectrum) and the fragmentation spectrum of this precursor (MS2) are taken into



account. The mass of a cross-linked peptide pair equals the sum of masses of two peptides plus an additional mass of 12 or 24 Da introduced during formaldehyde cross-linking. Note that, due to fragmentation of the covalent linkage the specific amino acid residues involved in the cross-link cannot always be identified when multiple cross-linkable amino acid residues are present in the peptide sequence.

An example spectrum of an inter-molecular protein cross-link of histone H4 to histone H2B type 1-L is shown in **Figure 17**. The  $m/z$  of the precursor ion corresponds to the mass of both peptides plus an additional mass of 24 Da introduced during formaldehyde cross-linking ( $m/z$  590.71 Da, 5+). Importantly, different charges are acquired during ESI. During fragmentation, peptide bonds as well as the formaldehyde cross-link are cleaved. Therefore,  $y$ -ions including an additional mass of 12 Da and  $y$ -ions without the linker-mass (**Figure 17**) are observed. The fragments of each peptide are used for identification of the corresponding peptide and protein. The cross-linking site is presumably located at R92 (histone H4) and R80 (histone H2B type 1-L) because arginine residues are more reactive towards formaldehyde compared to asparagine, histidine, aspartic acid, tyrosine and glutamine residues [136]. This assumption is supported by the pattern of fragment ions. The  $y_1$ -ion of the peptide 'TVTAMDVVYALRK' was observed without an additional mass of the cross-linker, while the additional mass of 12 Da was observed for the following  $y$ -ions. For the peptide 'IASEASRLAHYNK', the  $y_1$ - to  $y_4$ -ion series does not include the cross-linker modification, while the  $y_9$ - to  $y_{13}$ -ion series includes the additional mass of 12 Da (**Figure 17**). The cross-linking site is, therefore, most likely located between these observed ions.



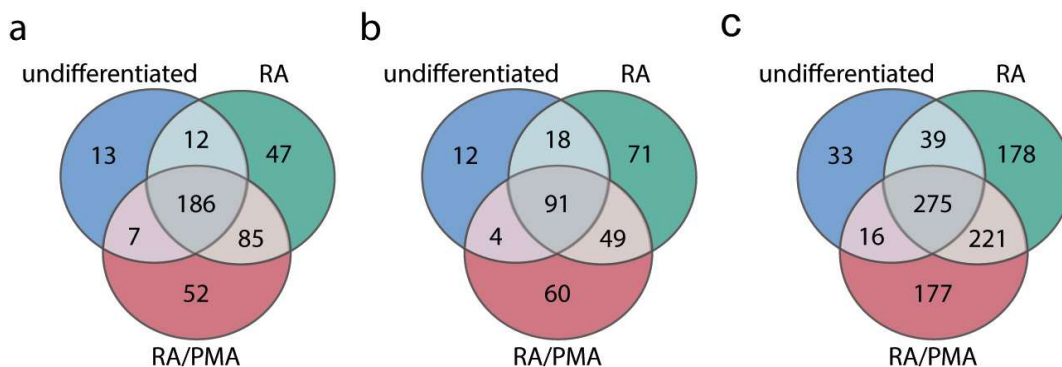
**Figure 17: Example spectrum for the identification of an inter-molecular protein cross-link**

The MS2 spectrum of an inter-protein cross-link between histone H4 and histone H2B including an additional mass of 24 Da is shown. The  $m/z$  and charge of the precursor ion is indicated (see box). For the peptide of histone H4,  $y$ -ions (light red),  $y$ -ions including the additional mass of 12 Da introduced during formaldehyde cross-linking (dark red),  $a$ -ions (yellow) and  $b$ -ions (orange) are assigned. The observed fragment ions of the peptide of histone H2B type 1-L:  $y$ -ions (light blue) as well as  $y$ -ions including an additional mass of 12 Da (blue),  $b$ -ions (dark blue) and  $a$ -ions (turquoise) are assigned. Note that not all observed fragment ions are annotated.

Three replicates of each cell culture condition were used to reproducibly identify cross-linked peptide pairs. The results table obtained using the ‘Formaldehyde XL Analyzer’ software [136] contains information on the cross-linked proteins, the cross-linked peptides and scores enabling filtering for confident cross-links. Only intermediate and high-confident cross-links were used (see **Section 2.2.21** for details). For subsequent analysis, three categories of cross-links were accepted: i) cross-links identified with high confidence in two or three replicates, ii) cross-links identified with high confidence in one replicate and at least intermediate confidence in another replicate and iii) cross-links identified with intermediate confidence in three replicates.

Applying the described confidence threshold, 218 proteins involved in 125 inter-protein and 363 intra-protein cross-links were identified in undifferentiated cells (**Figure 18**). In RA- and RA/PMA-differentiated cells, 330 proteins including 229 inter- and 713 intra-protein cross-links as well as 330 proteins including 204 inter- and 689 intra-protein cross-links, respectively, were identified. Among these identified cross-linked peptide pairs only one

containing a reverse decoy sequence was identified in RA/PMA-differentiated cells indicating a low rate of false positive hits among the identified cross-links. Nonetheless, a lower number of cross-links was identified in undifferentiated cells compared to RA or RA/PMA-differentiated cells.



**Figure 18: Comparison of identified cross-linked proteins, inter-protein cross-links and intra-protein cross-links**

The number of identified proteins and observed inter- and intra-molecular cross-links in undifferentiated (blue), RA-differentiated (green) and RA/PMA-differentiated (RA/PMA) cells is shown. **a)** Identified cross-linked proteins. **b)** Inter-molecular protein cross-links. **c)** Intra-molecular protein cross-links.

186 proteins were cross-linked in undifferentiated and RA- and RA/PMA differentiated cells (**Figure 18a**). The majority of inter-protein cross-links (**Figure 18b**) were observed in all three conditions. This indicates that, independent of the cell culture condition applied, most cross-links are reproducibly identified. The overlap between RA- and RA/PMA-differentiated cells regarding identified proteins involved in cross-links as well as inter- and intra-protein cross-links is higher. Comparing RA- and RA/PMA differentiated cells with undifferentiated cells, a smaller overlap was observed (**Figure 18**). This might be due to a comparable protein expression profile of differentiated cells compared to their undifferentiated origin. This tendency was also observed during proteomic analysis (**Section 3.1**). Again, indicating similarity of the cells obtained by application of the two differentiation protocols.

### 3.2.3 Cross-links identified in within ribosomal proteins

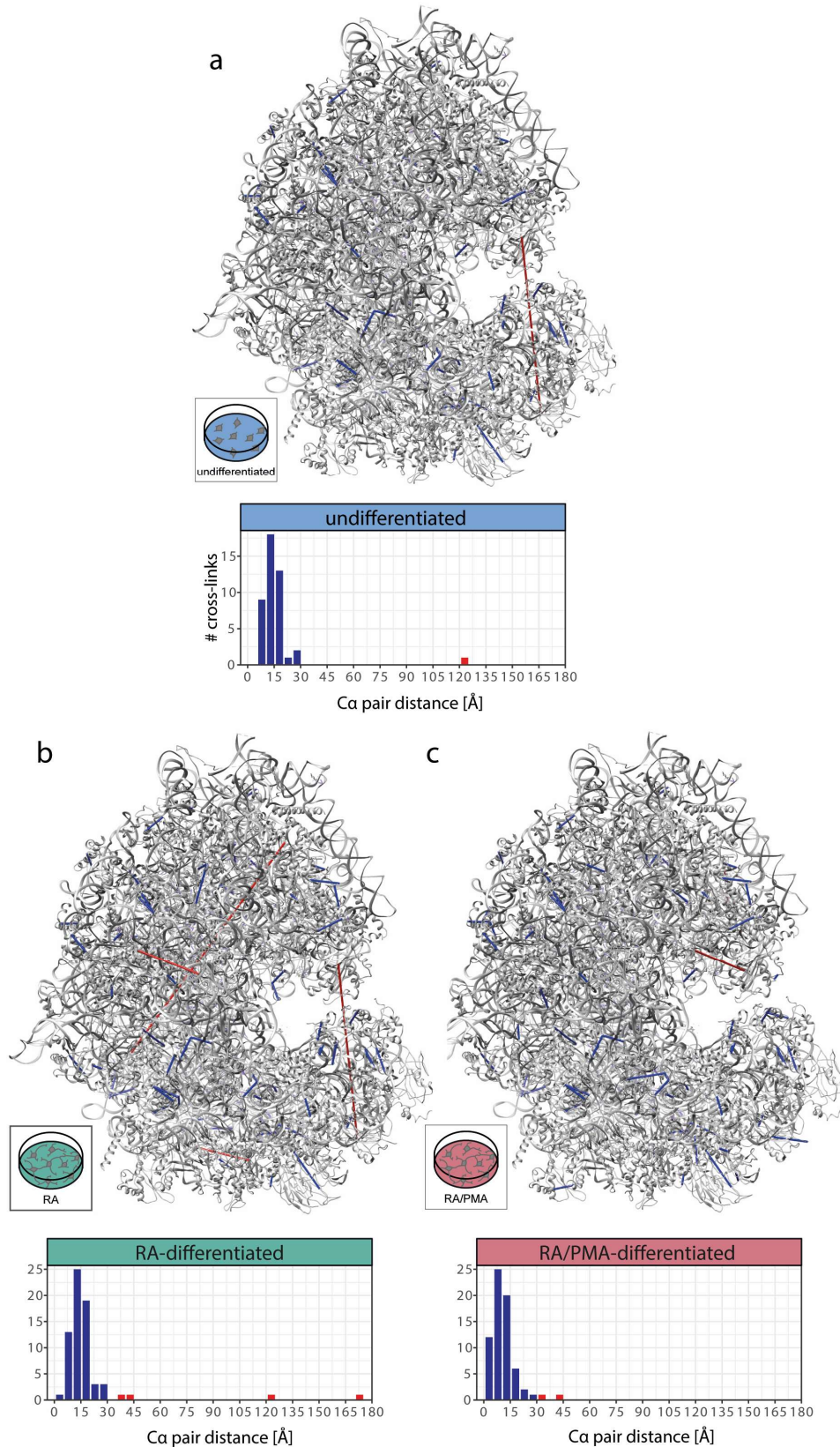
The observed cross-links were validated by plotting on an available high-resolution structure of the ribosome [267]. For this, information on the cross-linked protein subunit and the amino acid involved in the cross-link (i.e., cross-linking site) are required. However, it is impossible to determine the exact cross-linked amino acid within the peptide sequence.

Therefore, the amino acid located in the middle of each cross-linked peptide sequence was determined and employed as cross-linking site.

Identified cross-links of undifferentiated, RA- and RA/PMA-differentiated cells were visualized within the high-resolution structure of the ribosome (**Figure 19**). The  $C\alpha$  distance between two cross-linked amino acids was determined. The expected distance threshold of the identified cross-links was set to a  $C\alpha$  distance  $<30 \text{ \AA}$ . Cross-links with a  $C\alpha$  distance  $<30 \text{ \AA}$  and cross-links with a  $C\alpha$  distance  $>30 \text{ \AA}$  were observed (**Figure 19, Supplementary Table 5**). For undifferentiated cells, 44 of the identified 47 cross-links containing ribosomal proteins present in the crystal structure, were mapped (**Figure 19a**). Among these, for undifferentiated cells, only one cross-link with a  $C\alpha$  distance  $>30 \text{ \AA}$  was determined. For both differentiated cells a higher number of cross-links were identified: 78 and 75 cross-links for RA- and RA/PMA-differentiated cells, respectively. For RA-differentiated cells, 68 cross-links were mapped on the high-resolution structure. Of these four cross-links with a  $C\alpha$  distance  $>30 \text{ \AA}$  were observed (**Figure 19b, Supplementary Table 6**). Again, similar results were obtained for RA/PMA-differentiated cells. Specifically, 67 cross-links including two cross-links with a  $C\alpha$  distance  $>30 \text{ \AA}$  were mapped to the structure (**Figure 19c, Supplementary Table 7**). In all three culture conditions, most of the calculated  $C\alpha$  distances are in good agreement with the high-resolution structure.

Interestingly, for undifferentiated and RA-differentiated cells, the same over-length cross-link between protein subunits RS3(K 60) a part of the small 40S subunit and RL11(T 74) a part of the large 60S subunit was observed. During translation of messenger RNAs (mRNAs) into proteins, the ribosome moves along the mRNA and a new amino acid is incorporated into a nascent peptide [268]. For this, the ribosome undergoes a series of major structural rearrangements [268]. In particular, the rotation of the small 40S subunit relative to the large 60S subunit is essential for translocation (movement of mRNA and tRNA after peptide bond formation) [268]. The visualized structure only represents one conformation; therefore, rotation and flexibility of both subunits might result in a closer proximity of RS3 and RL11 and enable cross-linking of both subunits.

In summary, the majority of the identified cross-links of ribosomal proteins showed a  $C\alpha$  distance  $<30 \text{ \AA}$  and are in good agreement with the high-resolution structure. The validated cross-linking approach, therefore, is further applied for analysis of protein interactions with unknown structures.



**Figure 19: Inter- and intra-molecular cross-links of ribosomal proteins**

Identified inter- and intra-molecular cross-links that contain ribosomal sequences are visualized on the high-resolution structure of the ribosome (PDB: 4ug0 [267]). Cross-links with  $C\alpha$ -distances  $< 30 \text{ \AA}$  (blue) or  $> 30 \text{ \AA}$  (red) are highlighted accordingly. The number of  $C\alpha$ -distances  $< 30 \text{ \AA}$  (blue) or  $> 30 \text{ \AA}$  (red) is given for each cell culture condition. **a)** undifferentiated cells. **b)** RA-differentiated cells. **c)** RA/PMA-differentiated cells.

### 3.2.4 Protein interaction networks of SH-SY5Y cells

Next, protein-protein interactions obtained from intact SH-SY5Y cells were analyzed. For this, the obtained inter-molecular protein cross-links were used to generate a protein-protein interaction network. (**Supplementary Figure 1**). As described above, interactions involving ribosomal proteins were identified. In addition, an interaction cluster including histone proteins was identified. Histones, are abundant proteins that act as spools for DNA to wind around to form nucleosomes. Furthermore, they are abundant in lysine and arginine residues, the main targets during formaldehyde cross-linking. As expected, the facts that histones are abundant, have enough cleavage sites for tryptic hydrolysis and are rich in target residues for cross-linking, result in a high number of identified cross-links. Other protein interactions were observed for instance between proteins of the chaperonin-containing T-complex (TCP-complex), the MICOS complex, a large protein complex of the mitochondrial inner membrane [269] and cytoskeletal proteins including actin.

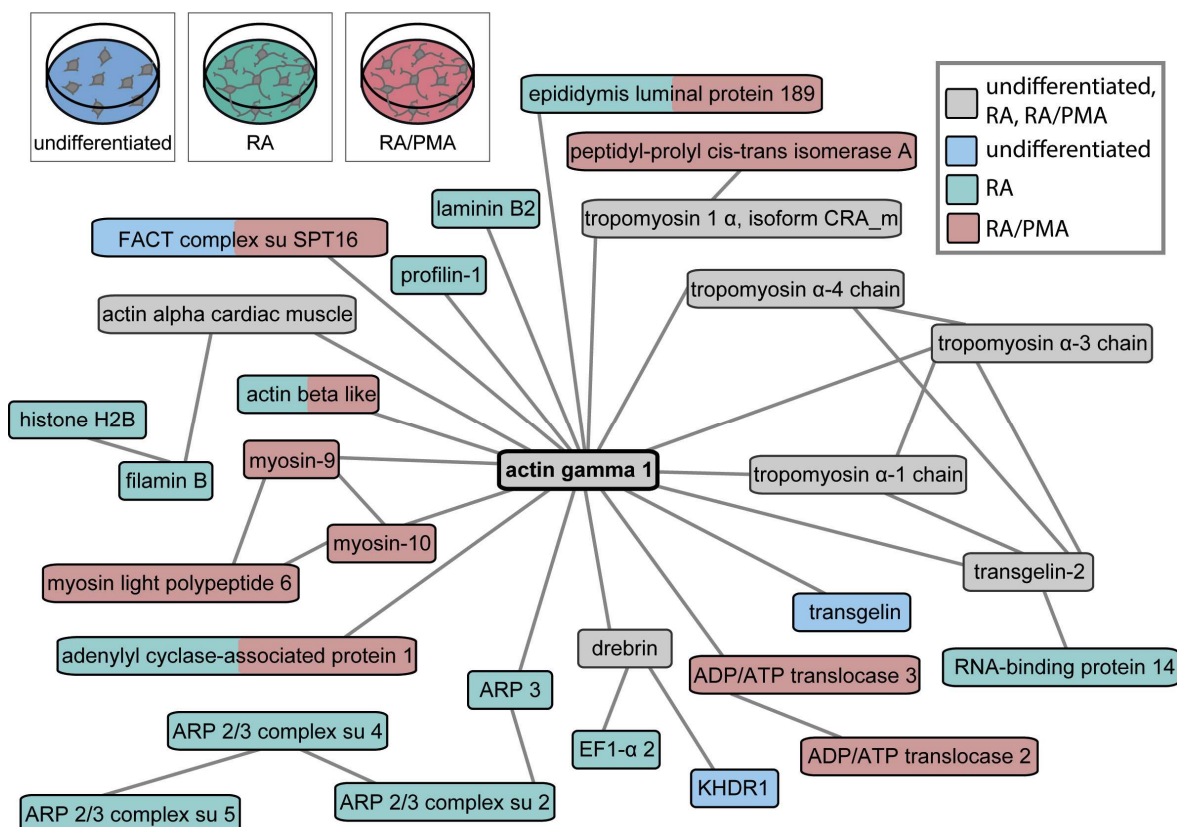
The protein-protein interactions identified, resulted from cross-linked proteins localized in different cellular compartments. For instance, proteins localized in the cytosol, nucleus or mitochondria were found to be cross-linked. These findings highlight the ability of formaldehyde to cross-link proteins in different structural compartments of the cell confirming the applicability for in-cell cross-linking.

#### Actin protein interaction network

The cytoskeleton of cells and importantly of neurons fulfills specific tasks. During differentiation the obtained distinct cell shape depends on a dynamic and changing cytoskeleton [270]. The above-described proteome analysis (**Section 3.1**) revealed enrichment of the KEGG pathway ‘regulation of actin’ in differentiated cells. In particular actin gamma 1, which is a key cytoskeletal protein, was found among the most abundant proteins in SH-SY5Y cells. Therefore, protein interactions of actin gamma 1 were inspected more closely and compared between the three culture conditions (**Figure 20**).

Our analysis revealed cross-links between actin gamma 1 and nine, 13 and 14 other proteins in undifferentiated, RA-differentiated or RA/PMA-differentiated cells, respectively (**Figure 20**). For all three conditions, these included cross-links to actin alpha cardiac muscle, drebrin, transgelin-2 and tropomyosin (**Figure 20**). Interactions between these proteins and actin have been described previously [271-273]. Actin alpha cardiac muscle is an  $\alpha$ - and the

actin gamma 1 an  $\gamma$ -actin isoform. Polymerization of  $\alpha$ - and  $\gamma$ -actin into filamentous actin (F-actin) *in vitro* has been observed [271]. Furthermore, F-actin is stabilized by tropomyosin [272] and drebrin modifies the double helix structure built by two F-actin strands [274]. Transgelin-2 was also found to interact with actin [275]. Therefore, protein interactions between actin gamma 1 and these proteins were expected.

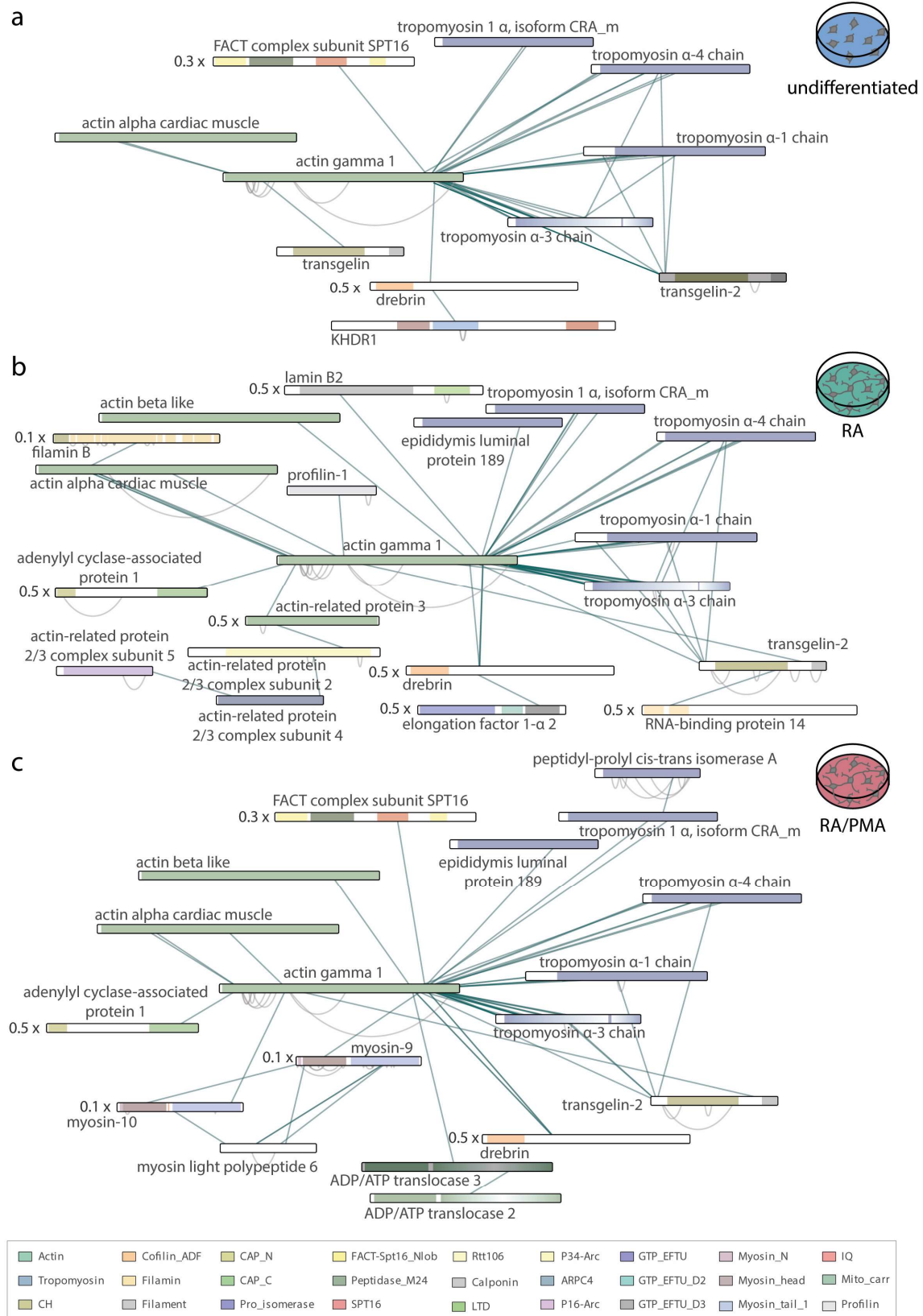


**Figure 20: Protein interaction network of actin gamma 1**

Interactions between actin gamma 1 and other proteins observed in undifferentiated (blue), RA-differentiated (green), RA/PMA-differentiated (red) and in all three cell culture conditions (grey) are shown (see legend for details). The lines between proteins correspond to at least one identified inter-protein cross-link. Note that residue information is not included.

In order to investigate the specific localization of identified cross-links within the protein sequences, a network plot including information on cross-linked amino acids and protein domains was generated for each cell culture condition (**Figure 21, a-c**). In all three conditions, cross-links between actin gamma 1 and the C-terminal peptide of transgelin-2 were observed. In addition, for RA- and RA/PMA-differentiated cells, cross-links between actin gamma 1 and the N-terminal region of transgelin-2 in close proximity to the calponin homology 3 (CH3) domain were observed. Both, C-terminal and CH3 domains of transgelin-2 represent known actin binding sites (**Figure 21, a-c**) [273, 276]. Additional cross-links between the N-terminal part of transgelin-2 and tropomyosin were also observed in all culture conditions of SH-SY5Y cells. Sequence similarity of transgelin-2 with calponin

suggests a tropomyosin binding site within its CH3 domain [277]. The cross-linking experiments, therefore, experimentally confirm the interaction of transgelin-2 with tropomyosin.





### Figure 21: Detailed protein interaction networks of actin gamma 1

Proteins and their domains (see legend for details) are shown as bars and the bar length correlates with the protein's length. The N-termini corresponds to the left side of the bars and the C-termini to the right sides. Some protein bars are scaled (see scaling factor for details). **a)** Detailed actin gamma 1-network observed in undifferentiated cells. **b)** Detailed actin gamma 1-network observed in RA-differentiated cells. **c)** Detailed actin gamma 1-network observed in RA/PMA-differentiated cells.

Abbreviations: CH, calponin homology; ADF, actin depolymerization factor; CAP\_N, adenylate cyclase associated N terminal; CAP\_C, adenylate cyclase associated C terminal; Pro\_isomerase, peptidyl-prolyl cis-trans isomerases; FACT-Spt16\_Nlob, FACT complex subunit SPT16 N-terminal lobe domain; Peptidase\_M24, metallopeptidase family M24; SPT16, FACT complex subunit (SPT16/CDC68); Rtt106, Histone chaperone Rtt106-like; LTD, Lamin Tail Domain; P34-Arc, Arp2/3 complex 34 kDa subunit p34-Arc; ARPC4, ARP2/3 complex 20 kDa subunit (ARPC4); P16-Arc, ARP2/3 complex 16 kDa subunit (p16-Arc); GTP\_EFTU, GTP-binding elongation factor family, EF-Tu/EF-1A subfamily; GTP\_EFTU\_D2, Elongation factor Tu domain 2, GTP\_EFTU\_D3, Elongation factor Tu C-terminal domain; Myosin\_N, Myosin N-terminal SH3-like domain; IQ, Q calmodulin-binding motif; Mito\_carr, Mitochondrial carrier protein.

Cross-links between actin gamma 1 and other proteins than those described above were only observed in one or two of the three cell culture conditions. Among the cross-links only observed in RA- and RA/PMA-differentiated cells, is a cross-link between actin and an actin-dynamic regulating protein named adenylate cyclase-associated protein 1 (**Figure 21**). This protein catalyzes the nucleotide exchange of ADP to ATP upon association with two globular actin monomers [278] and might contribute to the actin reorganization during early neuronal differentiation by acceleration of actin reorganization [279]. The identified interaction suggests reorganization and further establishment of the cytoskeleton during differentiation.

Several unique protein interactions were identified in RA-differentiated cells. These include cross-links between actin gamma 1 and the actin related protein 2/3 (ARP 2/3) complex, a key nucleator of actin branches [280], as well as between actin gamma 1 and profilin-1, a promoter of F-actin assembly [281, 282] (**Figure 21b**). Both, actin filament branch formation and F-actin assembly are required during differentiation and the development and elongation of neurites. Another specific interaction occurring in RA-differentiated cells was observed between actin gamma 1 and laminin B. This interaction might be explained by co-localization of laminin B2 mRNA to actin filaments and the possibly resulting mRNA translation on the cytoskeleton as described previously [283].

In RA/PMA-differentiated cells specific cross-links between actin gamma 1 and actin-dependent molecular motor proteins myosin 9 and myosin 10 were unique [284]. Both are isoforms of non-muscle myosin II and are expressed selectively in the nervous system [285]. They facilitate cell morphological and regulate actin dynamics [285] and might play a role in the changing actin regulation during differentiation.

In conclusion, the majority of protein interactions between actin gamma 1 and other proteins were observed in all three cell culture conditions. Importantly, the highest number of cross-links was observed between actin and tropomyosin which both form stabilized F-actin filaments. Some changes in protein interactions of actin gamma 1 with other proteins were identified between undifferentiated, RA- and RA/PMA-differentiated cells. These changes include interactions of actin gamma 1 with actin-regulating proteins, such as the ARP2/3 complex in RA- or Myosin variants in RA/PMA-differentiated cells and might be associated with changes in actin regulation during differentiation. However, the number of observed cross-links is low and a missing identification in other conditions might result from sample loss or other experimental errors introduced during sample preparation. Nevertheless, in-cell formaldehyde cross-linking proved a valuable tool to study protein interactions within cells and delivers detailed information on protein interactions.

### 3.3 Quantitative labeling of model proteins<sup>5</sup>

#### 3.3.1 Chemical labeling workflow<sup>6</sup>

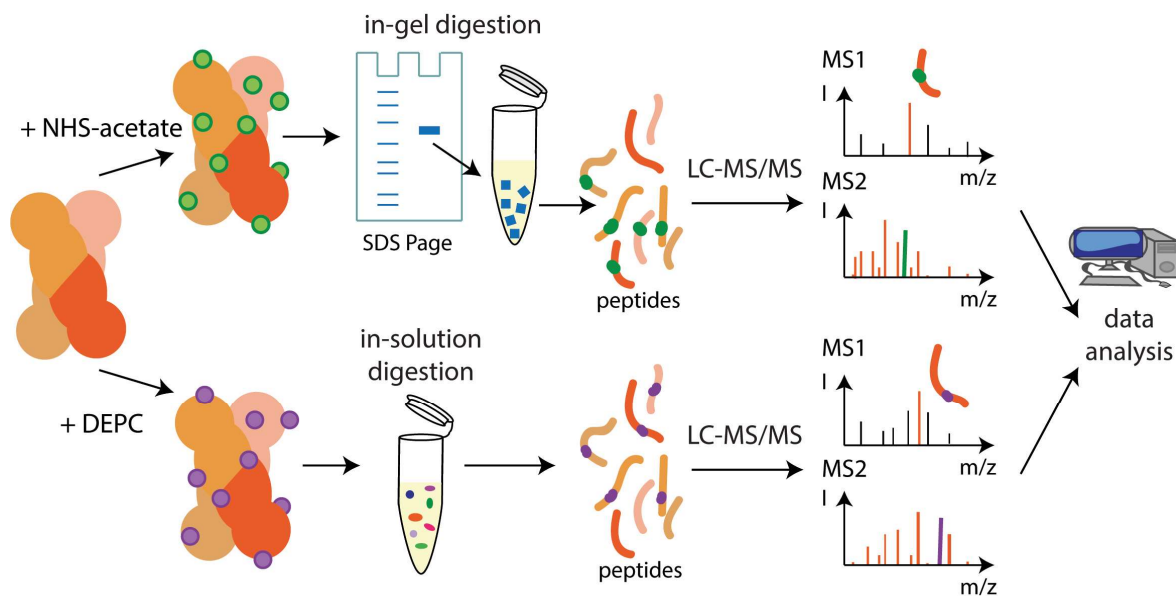
Chemical labeling in combination with MS analysis enables the identification of solvent accessible residues on the surface of proteins or protein complexes (see **Section 1.6.3** for detailed information). For establishing a labeling workflow, two model proteins were used, namely ADH and PK. Both proteins are commercially available and high-resolution structures have been determined [227, 228]. ADH and PK are homo-tetrameric protein complexes and are therefore suited for establishing labeling of protein complexes. Two labeling reagents were used: First, NHS-acetate, which acetylates lysine residues and in side reactions towards serine, threonine and tyrosine residues, and second, DEPC, which modifies histidine, lysine, cysteine, serine, threonine and tyrosine residues. Both reagents specifically modify the named amino acids and in comparison, to other labeling strategies such as HDX, require less complex data analysis (see **Section 1.6.3** for details).

The general labeling protocol was previously established (Master thesis, Marie Barth<sup>2</sup>). The model proteins were incubated either with NHS-acetate or DEPC for labeling of solvent accessible residues (see **Figure 22** for details). ADH or PK were acetylated by NHS-acetate and were separated by gel electrophoresis followed by in-gel hydrolysis using trypsin. Due to the pH-instability of DEPC, DEPC modified proteins were hydrolyzed in-solution in the presence of urea using trypsin at a pH above 7.9 (see **Section 1.1.12** for details). Modified as well as unmodified peptides were obtained using both labeling workflows and were analyzed by LC-MS/MS.

---

<sup>5</sup> The results presented in this chapter were published in **Barth, M.**, Bender, J., Kundlacz, T. & Schmidt, C. Evaluation of NHS-Acetate and DEPC labelling for determination of solvent accessible amino acid residues in protein complexes. *J Proteomics* **222**, 103793 (2020).

<sup>6</sup> The raw data of NHS-acetate labeled ADH and PK of three biological replicates, as well as of two replicates of DEPC labeled ADH and PK were previously obtained (Master thesis: Chemical modification of protein for structure elucidation, Marie Barth, 2018). Additional replicates were performed in this thesis. The raw data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository [214] with the dataset identifier PXD015940.

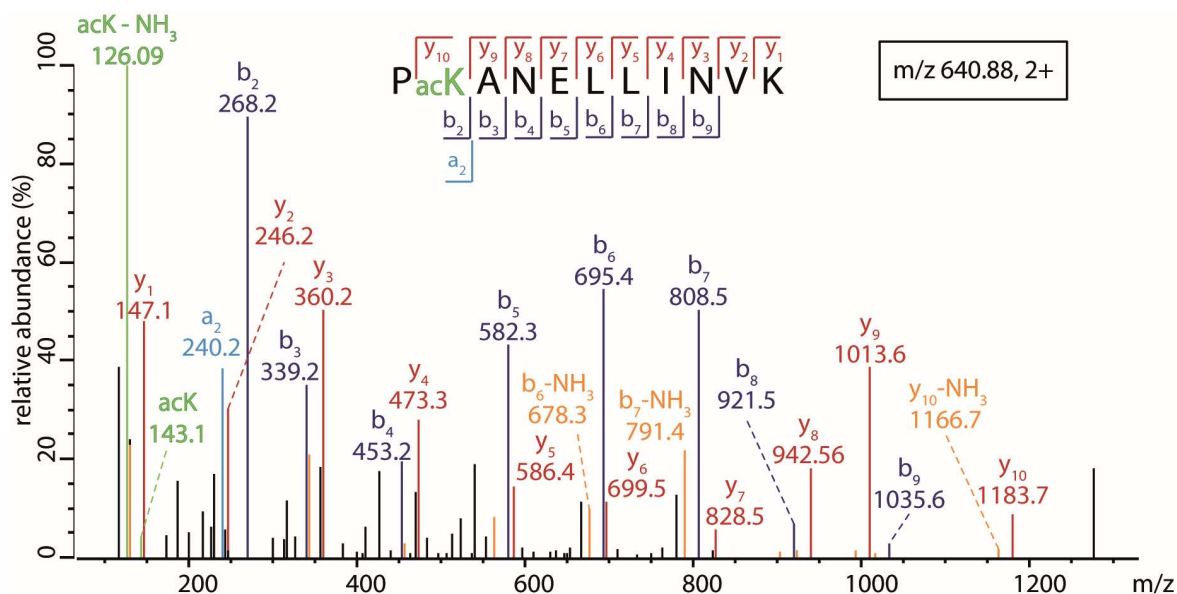


**Figure 22: Workflow for chemical labeling of model proteins using NHS-acetate or DEPC**

Proteins were acetylated using NHS-acetate as labeling reagent (top). Modified proteins were separated by gel electrophoresis. Proteins were enzymatically hydrolyzed in-gel and the obtained peptides were analyzed by LC-MS/MS. DEPC modified proteins (bottom) were enzymatically hydrolyzed in-solution. The peptides were then analyzed by LC-MS/MS.

### 3.3.2 Characterization of fragment spectra of labeled peptides

The identification of modified peptides using MS is based on a characteristic mass shift of the precursor ion as well as specific fragment ions. This mass shift corresponds to the mass of the introduced modification. An example spectrum of an acetylated peptide by NHS-acetate is shown in **Figure 23**. In this spectrum a complete y-ion series as well as a nearly complete b-ion series was observed, confirming the peptide sequence. For the  $y_{10-}$  as well as  $b_{2-}$  to  $b_{9-}$  ions a characteristic mass shift of 42 Da was observed. This allows the assignment of the acetylation to lysine. Furthermore, as described previously for peptides containing acetylated lysine residues [286], marker immonium ions of the acetylated lysine were identified at  $m/z$  143.1 and  $m/z$  126.09 (**Figure 23**).

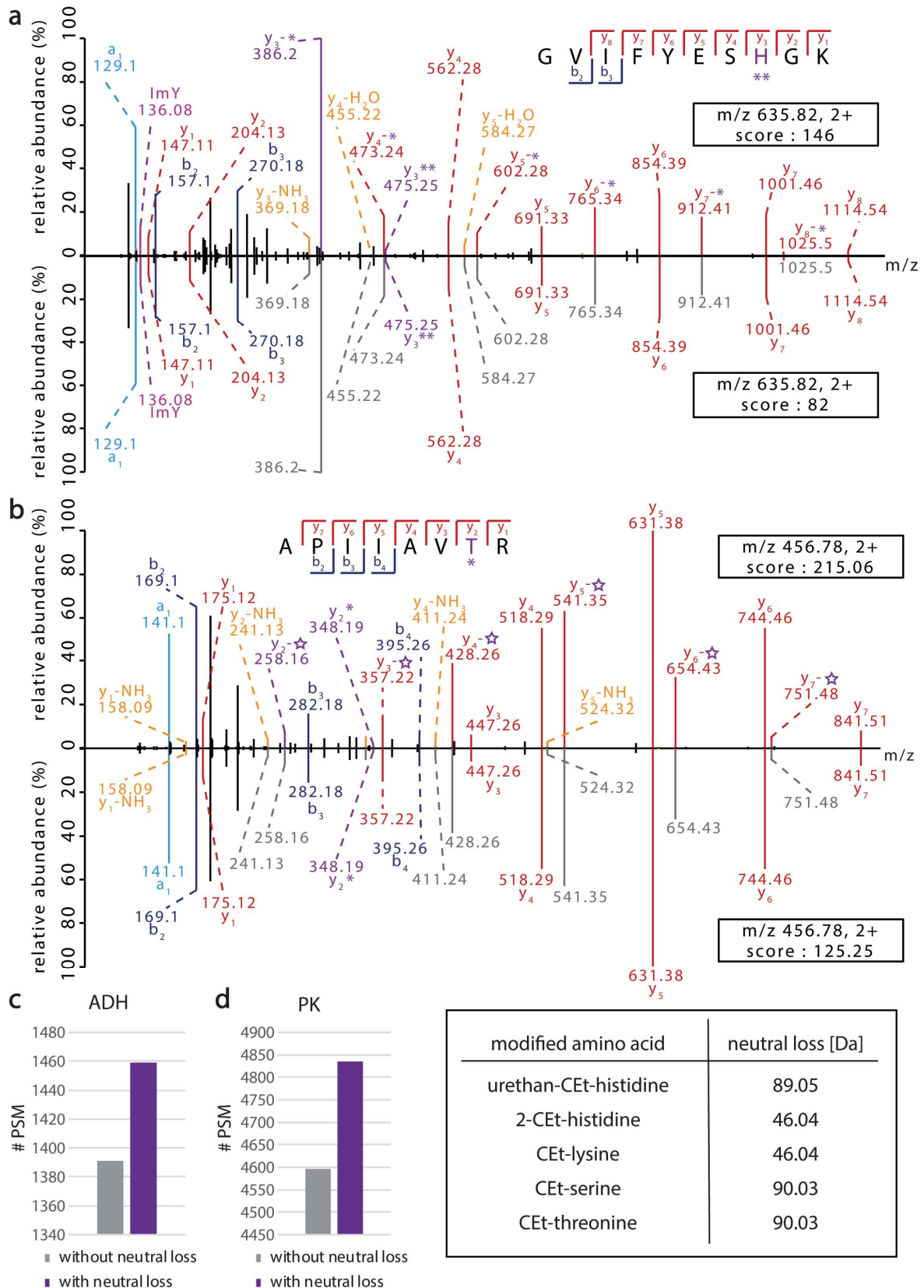


**Figure 23: Example spectrum of an NHS-acetate modified peptide.**

Example spectrum of a peptide containing an acetylated lysine. Y- (red and orange), a- (cyan) and b-ions (blue) are assigned. The  $y_{10}$ - and  $b_{2-}$  to  $b_{9}$ -ions include the 42 Da mass shift introduced by acetylation. In addition, characteristic immonium ions of acetylated lysine (green) are observed.

Example spectra of DEPC modified peptides are shown in mirrored-spectra view in **Figure 24a** and **b** and **Supplementary Figure 2**. In **Figure 24** the mirrored spectra below the x-axis show the identified spectrum with assigned fragment ions. For urethane-CEt-modified histidine, a characteristic mass shift of 134.06 Da is observed. In the spectrum corresponding to the peptide “GVIFYESHGK”, a mass shift of 134.06 Da is assigned to the fragment ions  $y_3$  to  $y_8$  (**Figure 24a**). Close inspection of the mass spectrum revealed many unassigned peaks (**Figure 24a**, mirrored spectrum, below x-axis). For peptides containing urethane-CEt-histidine-modified residues, these unassigned peaks differed from peaks of assigned fragment ions by 89.05 Da (**Figure 24a**). In addition, for CEt-modified histidine and lysine residues, unassigned fragment ions that differ in masses by 46.04 Da and for serine and threonine residues by 90.03 Da were observed (**Figure 24**, **Supplementary Figure 2**). These additional fragment ions might result from a neutral loss of a neutral (i.e. uncharged) fragment of the peptide during HCD fragmentation in the mass spectrometer. Therefore, a neutral loss analysis was performed. When these neutral losses were omitted during database searches, approximately 1390 and 4600 spectra were assigned to ADH and PK (**Figure 24c,d**). This number increased to approximately 1460 and 4830 for ADH and PK, when neutral loss of the CEt-group from lysine, serine, threonine and histidine as well as urethane-CEt-histidine-modified side chains were included (**Figure 24c,d**). Including the observed neutral losses, therefore, increased the number of PSMs by about 5% (**Figure 24c,d**). Of

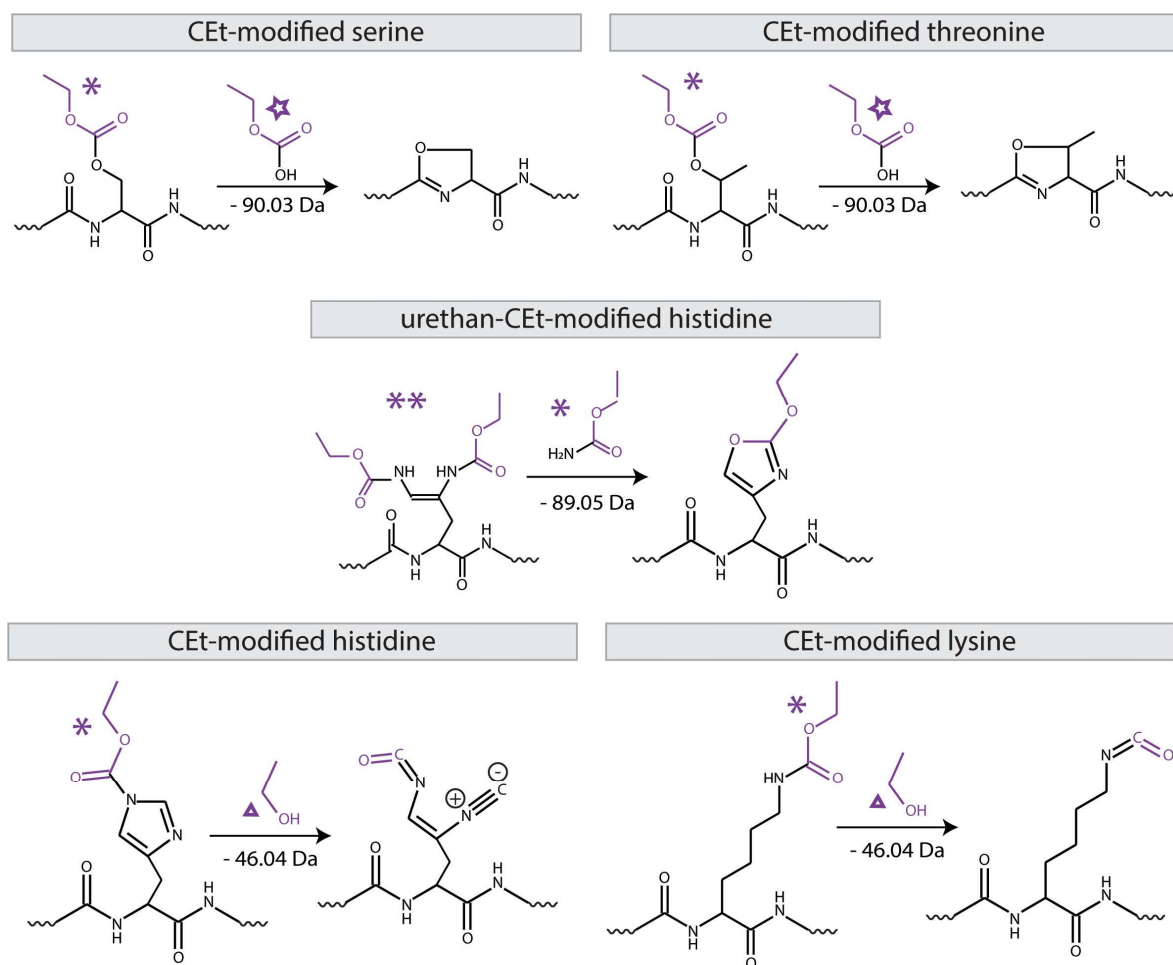
note, the peptide score for most labeled peptides increased significantly (**Figure 24a,b and Supplementary Figure 2**).



### **Figure 24: Example spectra of DEPC labeled peptides.**

The spectra are annotated without (bottom spectrum) and with (top spectrum) neutral loss from DEPC modified residues. Y- (red and orange), a- (cyan) and b-ions (blue) are assigned. The MaxQuant peptide score increases when neutral loss is included in the database search parameters. **a)** Example spectrum of a urethane-CET-histidine labeled peptide. **b)** Example spectrum of a peptide containing a CET-modified threonine residue. Number of identified PSM including or omitting neutral loss (see legend for details) of DEPC labeled peptides during database search. **c)** ADH. **d)** PK. The neutral loss of modified amino acids is given in Da (see box). (Figure adapted from Barth and Schmidt 2020 [153].)

As described above, the observed additional fragment ions might occur due to neutral loss of the CET-group during the applied HCD fragmentation. To-date the mechanism for this in the gas-phase of the mass spectrometer occurring process has not been described. A mechanism similar to phospho-peptides, for which a neutral loss of the phosphate group is routinely observed during CID or HCD fragmentation, is assumed. Accordingly, loss of the CET-group of modified serine and threonine residues suggests the formation of the energetically favored oxazoline ion (**Figure 25**). Presumably, neutral loss from urethane-CET-modified histidine also results in the formation of an oxazoline ion (**Figure 25**). In contrast, neutral loss of 46 Da for CET-modified histidine and CET-modified lysine corresponds in mass to the loss of an ethanol group. This eventually results in formation of isocyanate or related ions (**Figure 25**). However, neutral losses of CET-modified histidine and lysine residues were low abundant, indicating a higher stability of the labeled residue when compared with the neutral losses of urethan-CET-modified histidine and CET-modified serine and threonine side chains (compare **Figure 25** and **Supplementary Figure 2**).



**Figure 25: Possible mechanisms of neutral loss from DEPC labeled residues.**

Neutral losses from CEt-modified serine, CEt-modified threonine, urethane-CEt-modified histidine, CEt-modified histidine and CEt-modified lysine, during HCD fragmentation are shown. (Figure adapted from Barth et al. 2020 [153].)

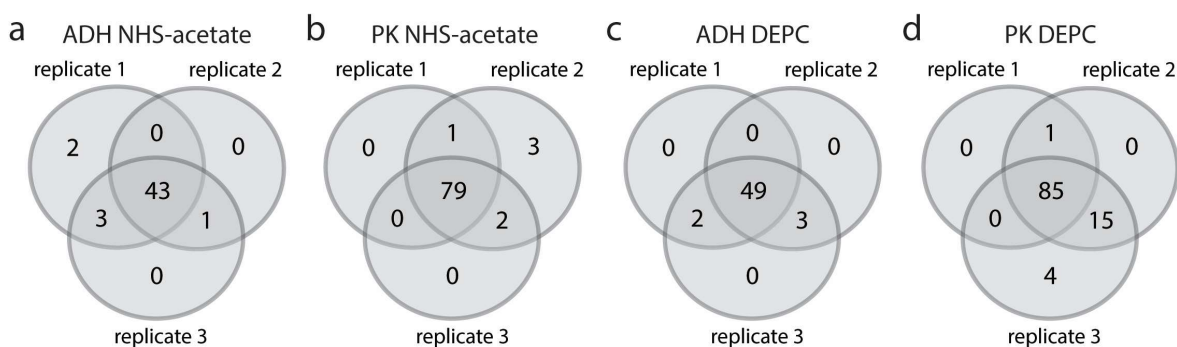
In summary, NHS-acetate or DEPC modified peptide fragmentation results in mass spectra that enable identification and localization of the modification within the peptide. For NHS-acetate modified peptides, the acetylation of a residue results in a 42 Da mass shift and specific fragment and reporter ions are observed. DEPC modified peptides yield complex mass spectra depending on the modified amino acid. Importantly, for CEt-modified histidine, lysine, serine, threonine and urethan-CEt-modified histidine residues, a neutral loss of a part of the modified group is observed. Including this neutral loss in the database search parameters increased the calculated peptide score. Therefore, search parameters including the neutral loss were used for identification of DEPC modified peptides.



### 3.3.3 Labeling efficiency of NHS-acetate and DEPC

For both model proteins and labeling reagents, three independent replicate experiments were performed and analyzed to identify reproducibly modified solvent accessible residues. Data analysis, i.e., peptide spectra annotation, peptide quantification and identification of modified residues was performed using MaxQuant software. For each identified peptide, a score is calculated by comparing the observed fragment ions to the theoretically predicted fragment ions. Within a peptide, several residues can be modified. The localization probability describes the specific probability for an amino acid present in a given peptide to be modified. Peptides with a score  $> 80$  and identified modified sites within this peptide with a localization probability  $> 0.75$  were further analyzed.

Following the described workflow using NHS-acetate, a total of 49 and 85 acetylated lysine, serine, threonine and tyrosine residues were identified in ADH and PK, respectively. The reproducibility of the identification of specific modified sites was assessed by comparing the identified modified residues of three replicates. The overlap of identified NHS-acetate labeled residues in the three replicates showed good reproducibility. In detail, 43 labeled sites of ADH and 79 labeled sites of PK were identified in all three experiments (**Figure 26a,b**).



**Figure 26: Reproducibility of NHS-acetate or DEPC labeling of model proteins**

The overlap of identified labeled residues of ADH (**a** and **c**) and PK (**b** and **d**) using NHS-acetate as labeling reagent (**a** and **b**) or DEPC (**c** and **d**) of three individual replicates was visualized using Venn-diagrams.

Using DEPC as labeling reagent, 54 and 105 modified histidine, lysine, cysteine, serine, threonine and tyrosine residues were identified in ADH and PK, respectively. Again, a high reproducibility was again observed. 49 and 85 residues were modified by DEPC in all three replicates in ADH and PK, respectively (**Figure 26c,d**).

Next, the labeling of different amino acids with NHS-acetate or DEPC was inspected in detail. Using NHS-acetate as labeling reagent, primarily acetylation of lysine residues is

observed; however, side reactions with serine, threonine and tyrosine residues also occur [287, 288]. When these four amino acids are considered as targets for labeling, approximately 15% of all amino acids of ADH and PK can potentially be modified (**Table 3.3**). The acetylation of arginine residues by NHS-acetate was previously described [289], but was not observed even at high concentrations of NHS-acetate. DEPC, in addition to the above-mentioned amino acids lysine, serine, threonine and tyrosine also modifies histidine and cysteine residues. Accordingly, 91 and 103 residues of ADH and PK can potentially be labeled. This is equivalent to 25% of the protein sequence of both proteins.

Nearly all lysine and serine residues of ADH were modified when using NHS-acetate (**Table 3.3**). Threonine and tyrosine residues, however were modified to a lower extent. Similar results were obtained for PK: almost all lysine, serine and threonine residues were labeled. Tyrosine side chains were labeled to a very low degree. DEPC labeling of ADH yielded a high number of modified lysine and serine residues; threonine, tyrosine, cysteine and histidine residues, on the other hand, were labeled to a lower extent. Again, similar results were obtained for PK; lysine, serine, threonine, cysteine and histidine residues were highly modified and while tyrosine was found to be modified to a lower degree. Different DEPC modifications of histidine residues are possible. These are: CEt-histidine, di-CEt-histidine, formyl-CEt-histidine and urethane-CEt-histidine [196, 200, 290]. Of the possible modifications, CEt-histidine and urethane-CEt-histidine were mainly identified.

**Table 3.3. Overview on the number of modified amino acids of ADH and PK when using NHS-acetate and DEPC**

ADH and PK were labeled with NHS-Acetate or DEPC. The number (#) and percentage of lysine, serine, threonine, tyrosine, cysteine and histidine residues that can potentially be labeled (theoretical) as well as the number and percentage of identified labeled residues within the amino acid sequence of ADH and PK (observed) are shown. Abbreviation: n.a., not applicable.

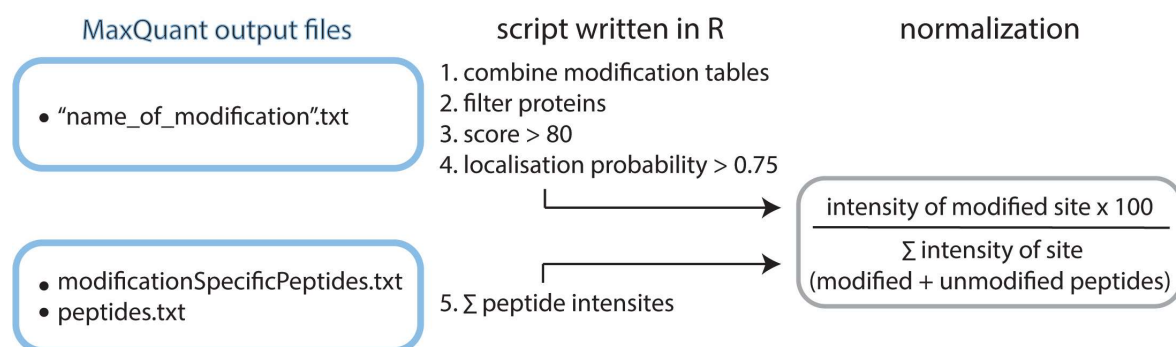
Labeling		Lysine (observed/ theoretical)	Serine (observed/ theoretical)	Threonine (observed/ theoretical)	Tyrosine (observed/ theoretical)	Cysteine (observed/ theoretical)	Histidine (observed/ theoretical)	Total (observed/ theoretical)
<b>alcohol dehydrogenase (ADH)</b>								
<b>NHS- acetate</b>	#	19 / 24	15 / 21	8 / 14	7 / 14	n.a.	n.a.	49 / 73
	%	5.5 / 6.9	4.3 / 6	2.3 / 4	2 / 4	n.a.	n.a.	14 / 20.9
<b>DEPC</b>	#	21 / 24	15 / 21	7 / 14	6 / 14	2 / 8	3 / 10	54 / 91
	%	6 / 6.9	4.3 / 6	2 / 4	1.7 / 4	0.6 / 2.3	0.9 / 2.9	15.5 / 26.1
<b>pyruvate kinase (PK)</b>								
<b>NHS- acetate</b>	#	37 / 37	22 / 31	22 / 26	4 / 9	n.a.	n.a.	85 / 103
	%	7 / 7	4.1 / 5.8	4.1 / 4.9	0.75 / 1.7	n.a.	n.a.	16 / 19.4
<b>DEPC</b>	#	36 / 37	23 / 31	21 / 26	6 / 9	7 / 9	12 / 14	105 / 126
	%	6.8 / 7	4.3 / 5.8	3.95 / 4.9	1.1 / 1.7	1.3 / 1.7	2.2 / 2.6	19.75 / 23.7

In conclusion, both labeling reagents are well-suited to modify specific amino acids and the applied workflow is applicable for identification of labeled residues. As expected, a labeling efficiency of 100% was not achieved. Modification of individual residues strongly depends on their accessibility for the labeling reagent as well as identification of the modified residue during data analysis. Accordingly, residues located in longer peptides or peptides containing several modified residues are in general discriminated during the identification process. Longer peptides are obtained upon enzymatic hydrolysis using trypsin when the number of lysine and arginine residues is low within the protein's amino acid sequence. In addition, ionization and fragmentation of larger peptides is hampered using mass spectrometers and correct assignment of modifications to specific amino acid residues is difficult when multiple residues are labeled and especially when specific reporter ions are absent in the acquired mass spectrum. Nonetheless, the majority of potentially labeled residues were also identified during data analysis. Therefore, NHS-acetate and DEPC are promising labeling reagents for the analysis of solvent accessible amino acid residues on the surface of proteins and protein complexes.

### **3.3.4 Quantification of labeled residues for determination of solvent accessible amino acids on the surface of protein complexes**

Having optimized the identification of labeled amino acid residues, the next step was to analyze differences in solvent accessibility of individual residues within the structured model proteins. For this, the model proteins were labeled with increasing amounts of either NHS-acetate or DEPC. The samples were then prepared and analyzed using LC-MS/MS as described above (see **Section 3.3.1**). Subsequently, the labeling percentage of each labeled site (i.e. the relative proportion of labeled residues compared to their unlabeled counterparts) was calculated for each concentration of labeling reagent. For this, a database search using MaxQuant software was performed. The obtained result tables contain information on the intensity of the observed peptides (unlabeled and labeled) as well as the position of the labeled amino acids within the peptide sequence. The calculated intensities represent the sum of extracted ion chromatograms of the areas of all isotope peaks of the corresponding peptide sequence. For the following analysis, an R script written in house was used (**Figure 27**). In detail, the obtained tables containing information on labeled residues were filtered for ADH or PK excluding identification and modification of typical contaminant proteins such as trypsin. In addition, only modified sites with a score > 80 and a localization

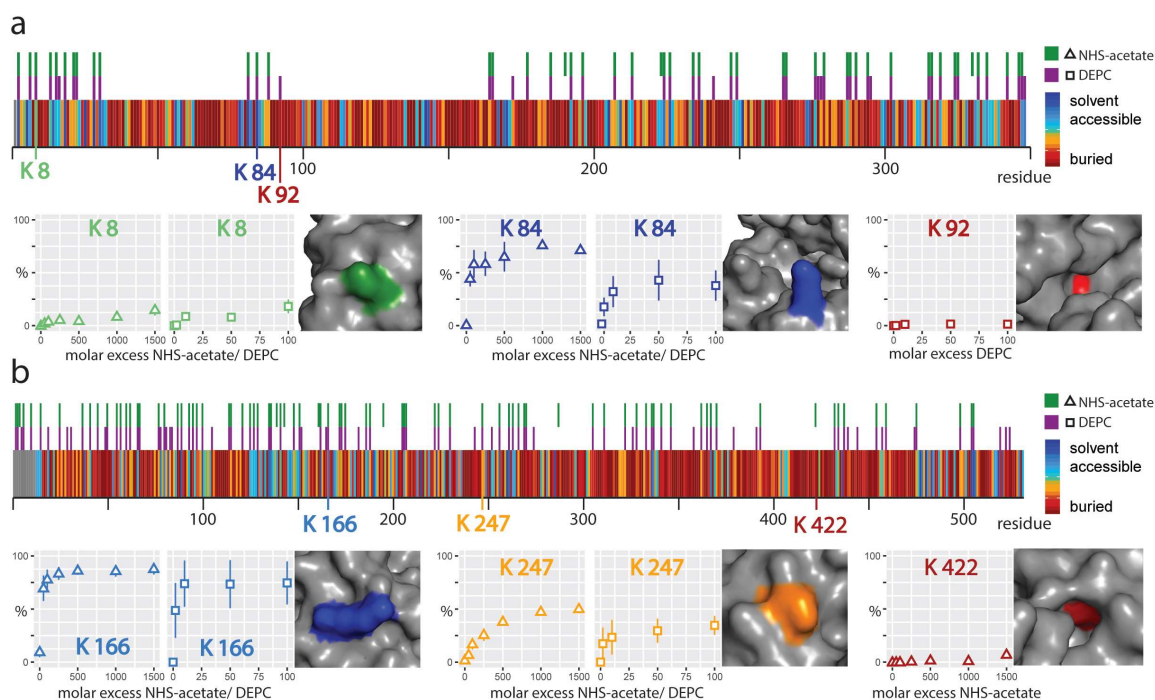
probability > 0.75 were used (see above). The intensity of a modified site was normalized by the sum of intensities of labeled and non-labeled peptides containing this specific residue site (see equation in **Figure 27**). This calculated normalized intensity is equal to the labeling percentage of each modified site. The mean labeling percentage and standard error of each site was calculated from three independent replicates for each concentration of the labeling reagent. The obtained labeling percentage was then plotted against the increasing amount of labeling reagent.



### Figure 27: Data analysis strategy.

Data analysis workflow using MaxQuant software and an R script written in house. The table "name\_of\_modification".txt is among the results files from the software search and is named specifically for each modification. Therefore, these tables are first combined and specific information on the modified amino acid of a protein is used in the following filtering process. This includes filtering for proteins of interest and to identify only valid modifications. For this, modified sites with a score >80 and a localization probability (specific probability for every possibly modified amino acid site present in a given peptide to be modified) >0.75 are selected. In addition, the sum of peptide intensities of all modified and unmodified containing an identified modified residue site are calculated using the modificationSpecificPeptides.txt and peptides.txt tables. The last step is the normalization of the intensity of a modified amino acid site by the sum of intensities of labeled and non-labeled peptides containing this site (see equation for details). This calculated normalized intensity is equal to the labeling percentage of each modified residue. (Figure from Barth et al. 2020 [153].)

For NHS-acetate labeling, a 0-, 50-, 100-, 250-, 500-, 1000- or 1500-fold molar excess, and for DEPC labeling, a 0-, 2-, 10-, 50- or 100-fold molar excess of labeling reagent was used. Predominantly lysine residues showed a significant increase in labeling percentage when higher concentrations of the labeling reagents were employed. When NHS-acetate was used for labeling, the labeling percentage reached a plateau approximately at a 500-fold molar excess and for DEPC at a 50-fold molar excess (**Figure 28**) indicating that lower amounts of DEPC compared to NHS-acetate are required for complete labeling. Other residues did not show this increase in labeling percentage at higher concentrations of the labeling reagent and only increased slightly or maintained the same level of labeling.



**Figure 28: Identification of solvent accessible amino acid residues of ADH and PK.**

**a)** ADH was labeled with increasing amounts of NHS-acetate or DEPC. The bar diagram indicates labeled residues (NHS-acetate in green, DEPC in purple) and the predicted SASA of each residue (see legend for details). The labeling of lysine 8, 84 and 92 including their localization in the crystal structure of ADH (PDB: 5env) is shown in detail. **b)** The labeling results for PK (see **a** for color code) and specific labeling of lysine 166, 247 and 422 and their localization in the crystal structure (PDB: 1aqf) are shown. (Figure from Barth et al. 2020 [153].)

The observations described above, led to the assumption that in particular solvent accessible lysine residues show an increase in labeling percentage, while residues that are less accessible only show a moderate or no increase in labeling.

For validation of this assumption, the relative solvent accessible surface area (SASA) for all amino acid residues of ADH and PK was calculated using the web server GETAREA[210]. The calculated relative SASA of each residue equals the ratio of this residue's surface area to a random coil value. In this equation the random coil value is the average SASA of an amino acid residue X in the tripeptide G-X-G in an ensemble of 30 random conformations [210]. When the obtained ratio is  $>0.5$  residues are considered solvent accessible, while residues with a ratio  $<0.2$  are considered buried. In the next step, the calculated relative SASA based on the crystal structure was compared to the calculated labeling efficiency of each residue. As expected, NHS-acetate and DEPC mainly modify residues located in solvent accessible regions of the two model proteins. In particular, lysine residues with high relative SASA such as K84 of ADH and K166 of PK show a high increase in labeling percentage when using NHS-acetate or DEPC as labeling reagent (**Figure 28a,b**). The

labeling percentage of residues with intermediate SASA show a similar behavior at lower values (see K8 in **Figure 28a** and K247 in **Figure 28b** as an example). For buried residues, an increase in labeling percentage was not observed (see K92 in **Figure 28a** and K422 in **Figure 28b** as an example). Only a few exceptions of this labeling behavior were observed.

In summary, quantitative labeling of proteins with increasing amounts of NHS-acetate or DEPC allows specific identification of the solvent accessibility of surface-exposed residues. In contrast to a simpler labeling workflow employing only one concentration of the labeling reagent, buried residues are reliably distinguished from solvent accessible residues. The labeling procedure and quantitative data analysis established in this thesis is now applicable for the identification of solvent accessibility of residues of other proteins and protein complexes of interest.

### 3.3.5 Ion mobility MS of modified protein complexes

Chemical modifications of proteins are known to influence a protein's function. DEPC for instance is routinely used for inactivation of RNase A by modification of its histidine residues [291]. As function and structure of a protein are closely related, the question whether the native structure of a protein is maintained after labeling arises. Therefore, native MS and ion mobility MS were used to analyze ADH before and after labeling with NHS-acetate and DEPC<sup>7</sup>. For this, a mass spectrometer modified for transmission of high masses maintaining non-covalent interactions of a protein complex in the gas-phase was employed [121] (see **Section 1.6.1**). These results were obtained by Julian Bender and Til Kundlacz and will be discussed in the following paragraphs.

The native mass spectrum of unlabeled ADH revealed that ADH exists predominantly as tetrameric complex, while an ADH monomer as well as the dimeric complex were observed at low intensity (**Figure 29a**). When ADH was labeled with NHS-acetate, the mass of the tetrameric ADH complex shifted to higher  $m/z$  values with increasing amount of labeling reagent (**Figure 29b**). The effect of modified residues on the structure of the tetrameric protein complex was then analyzed by ion mobility enabling separation of analytes by their shape [292]. Before the ions entered the ion mobility cell, increasing collisional voltages were applied to induce unfolding of the protein complex. The drift times of the ions observed

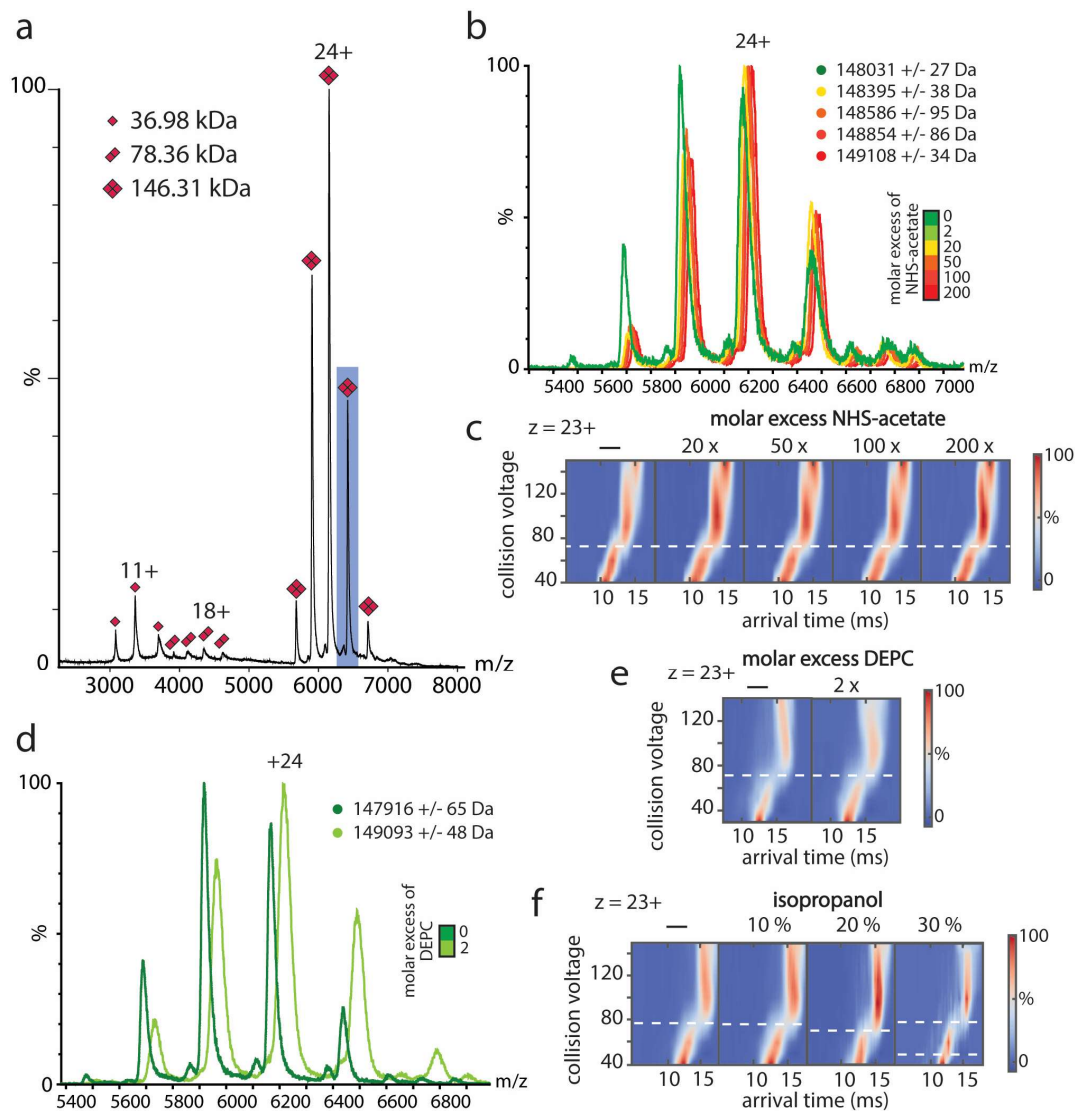
---

<sup>7</sup> Native MS and ion mobility MS of ADH before and after labeling with NHS-Acetate and DEPC were performed by Julian Bender and Til Kundlacz.

upon labeling of ADH at various molar excesses of NHS-acetate were recorded for each setting. The 23<sup>+</sup>-charge state of the ADH tetramer was chosen for detailed examination (**Figure 29c**), as lower charge states of proteins in general resemble more folded states [293]. At low collisional voltages (0 – 60 V) the ions of the 23<sup>+</sup> tetramer required approximately 10 ms to travel from the ion mobility cell to the detector. The arrival time increased to approximately 15 ms at activation voltages between 60 and 80 V indicating a structural transition. At higher collisional voltages (>80 V) unfolding of the protein complex is assumed resulting in further increase of the arrival time. When ADH was labeled with 20-, 50-, 100- and 200-fold molar excess of NHS-acetate, the arrival times and the resulting collisional induced unfolding are comparable with those of the unlabeled protein suggesting that labeling with NHS-acetate does not affect the folded structure and stability of ADH.

For DEPC labeling of ADH, similar results were obtained. However, DEPC is reactive towards several residues and the resulting heterogeneous complexes hamper native MS analysis. Labeling of ADH resulted in an increased mass and multiple modifications caused peak broadening (**Figure 29d**). Therefore, only low amounts of DEPC were applied. Again, unfolding of the DEPC labeled tetramer were monitored as described above and differences in unfolding was not observed between unlabeled and labeled ADH complexes (**Figure 29e**). Therefore, low DEPC concentrations do not alter the protein structure and do not affect complex stability.

As a control experiment, unfolding of ADH was induced by addition of increasing amounts of isopropanol. The required collisional voltage to induce unfolding of the tetramer was reduced at higher concentrations of isopropanol (**Figure 29f**) indicating destabilization of the complex. In summary, the results described above suggest that the structure of tetrameric ADH was preserved upon labeling with NHS-acetate and low concentrations of DEPC enabling the structural analysis of folded proteins and protein complexes.



**Figure 29: Ion mobility MS of NHS-acetate and DEPC labeled ADH.**

**a)** Native mass spectrum of ADH showing peak series corresponding to monomeric, dimeric and tetrameric ADH. **b)** ADH was labeled with increasing amounts of NHS-acetate (see legend for details). Modification of the amino acid residues of the tetrameric complex caused mass shifts to higher  $m/z$ . **c)** The 23+-charge state of ADH tetramer was selected for detailed evaluation of collisional induced unfolding. The recorded arrival time is plotted against the applied collisional voltage. The required collision voltages to unfold the complex does not change for the labeled protein. **d)** ADH was labeled with DEPC (see legend for details). The mass of the modified tetrameric complex increased. **e)** The 23+-charge state of ADH tetramer was selected for detailed evaluation of collisional induced unfolding. The recorded arrival time is plotted against the applied collisional voltage. Similar to NHS-acetate labeling, DEPC labeling did not have an effect on the required collisional voltage to unfold the protein. **f)** As a control, increasing amounts of isopropanol were used to destabilize the ADH tetramer. Unfolding plots were generated. With increasing isopropanol concentrations, lower collisional voltages are required for unfolding. (Figure from Barth et al. 2020 [153].)



## 3.4 Solvent accessibility of synaptic vesicle proteins probed by chemical labeling<sup>8</sup>

### 3.4.1 Labeling of synaptic vesicle proteins using three different chemical labeling reagents

Having established a labeling workflow for the analysis of solvent accessible residues of standard proteins, the next goal was to identify solvent accessible residues of synaptic vesicle proteins. Synaptic vesicles are essential for signal transmission between neurons (see **Section 1.2**). They are densely packed with proteins that mediate, for instance, neurotransmitter loading or vesicle docking to the presynaptic membrane. The explicit function and structure of many proteins during this process is only partially understood; therefore, elucidating solvent accessibility of individual domains of vesicle proteins reveals valuable structural information that might be linked to functional properties of these proteins. To this end, the established labeling strategy (**Section 3.3**) was applied.

Synaptic vesicles were purified from rat brain<sup>9</sup>. For identification of solvent accessible protein regions in synaptic vesicle proteins, intact synaptic vesicles are required. The integrity of the obtained synaptic vesicles was, therefore, confirmed by negative stain electron microscopy [294]. Furthermore, label-free relative quantification of protein abundances showed a typically observed distribution of synaptic vesicle proteins as well as common contaminating proteins, confirming the quality of the obtained synaptic vesicle preparation [294]. For evaluation of the correct orientation of synaptic vesicle proteins, Botulinium neurotoxin B was employed. Botulinium neurotoxin B cleaves the cytosolic domain of synaptobrevin-2, which is the most abundant protein of synaptic vesicles. The complete cleavage without remaining intact synaptobrevin-2 confirmed correct orientation of proteins [294]. Vesicles obtained following the established workflow were then used for labeling experiments.

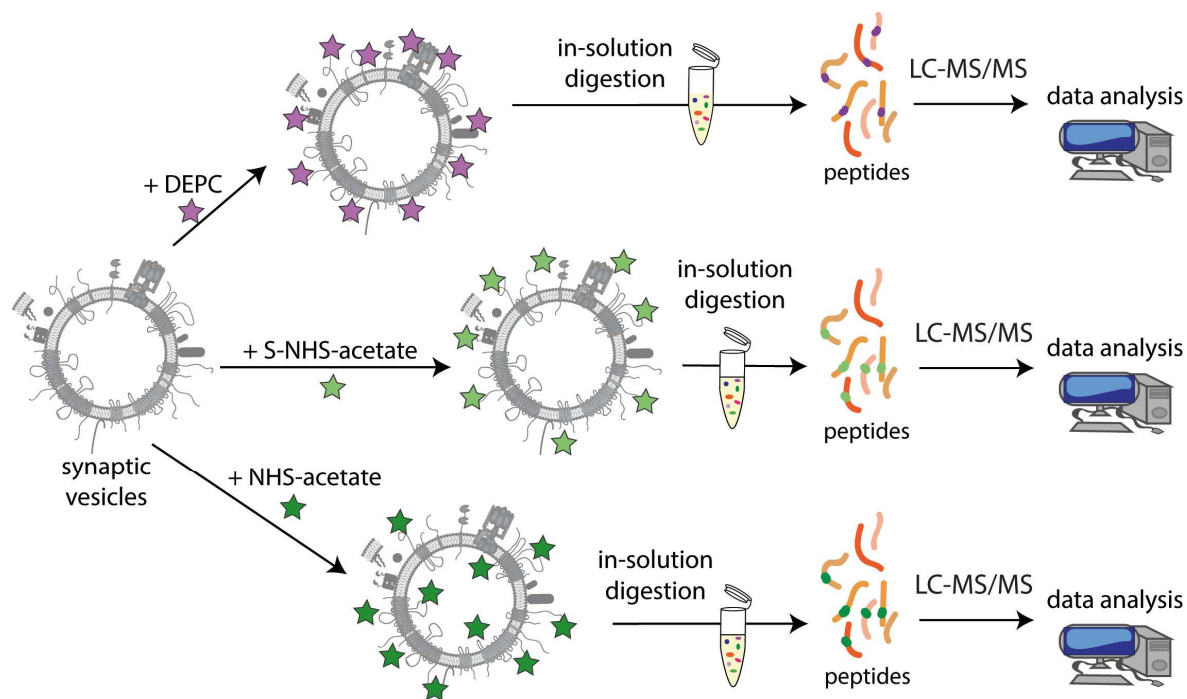
In these experiments, synaptic vesicles were labeled using two concentrations of DEPC, S-NHS-acetate or NHS-acetate (5 and 10 mM). DEPC and S-NHS-acetate are assumed to be

---

<sup>8</sup> Some of the results of this chapter were published in Sabine Wittig, Marcelo Ganzella, **Marie Barth**, Susann Kostmann, Dietmar Riedel, Ángel Pérez-Lara, Reinhard Jahn & Carla Schmidt, Cross-linking mass spectrometry uncovers protein interactions and functional assemblies in synaptic vesicle membranes. *Nature Communications* **12**, 858 (2021).

<sup>9</sup> Purified synaptic vesicles were provided by Prof. Dr. Reinhard Jahn (Max Planck Institute for Biophysical Chemistry, Göttingen). Synaptic vesicles were purified by Dr. Marcelo Ganzella, Susann Kostman or Sabine Wittig. Dr. Dietmar Riedel performed negative stain electron microscopy. Proteomic data analysis of synaptic vesicles was performed by Sabine Wittig.

not membrane permeable and modifications of proteins and protein domains orientated towards the cytoplasmic side of the synaptic vesicle membrane are expected. NHS-acetate is membrane permeable and, in theory, labels cytoplasmic and luminal protein domains. The labeling workflow for synaptic vesicle proteins using the three different labeling reagents is shown in **Figure 30**.



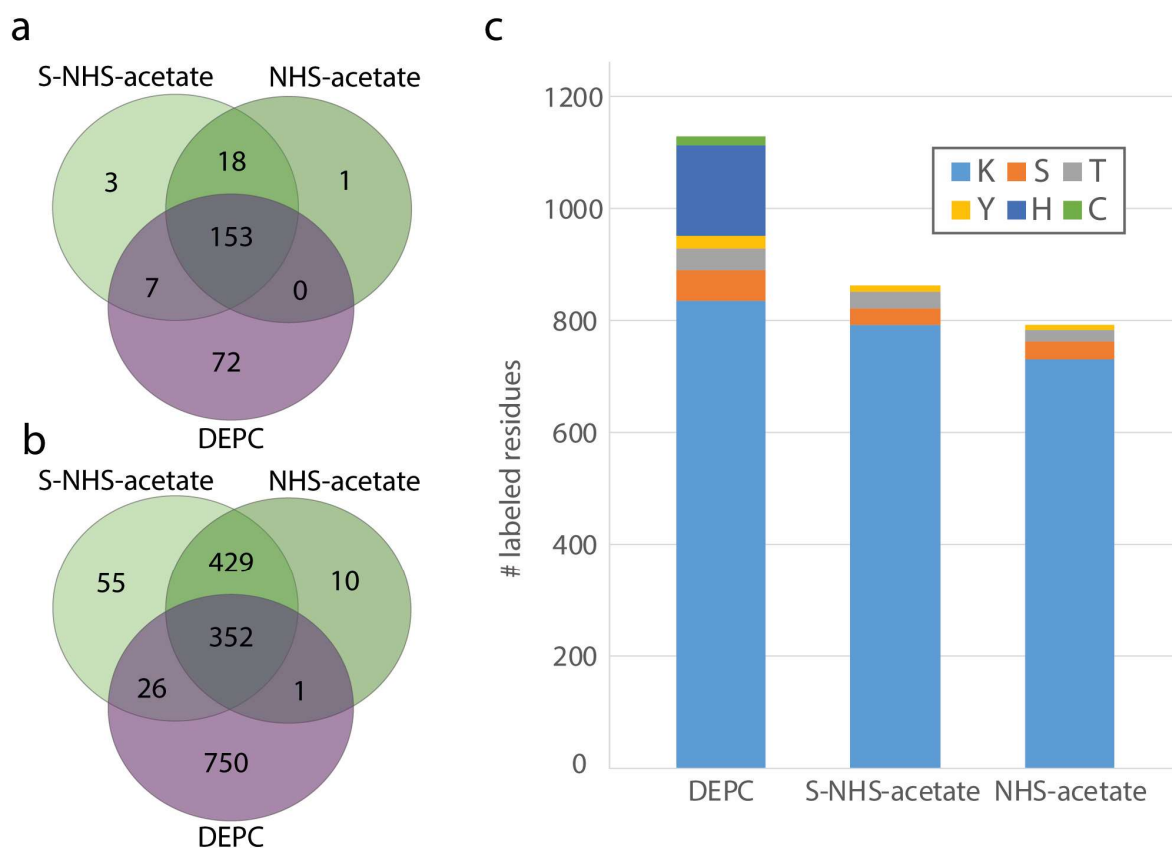
**Figure 30: Labeling workflow for synaptic vesicle proteins**

The proteins of synaptic vesicles, purified from rat brain, were modified using DEPC, S-NHS-acetate or NHS-acetate labeling reagents. Of these, only NHS-acetate is membrane permeable. Labeled proteins were then enzymatically hydrolyzed in-solution and obtained peptides were analyzed by LC-MS/MS. The acquired raw data was analyzed using MaxQuant software.

The workflow includes incubation of synaptic vesicles with DEPC, S-NHS-acetate and NHS-acetate. Following chemical labeling, the proteins were precipitated with ethanol and hydrolyzed using trypsin. Subsequently, the obtained peptides were analyzed by LC-MS/MS. The acquired raw data was further analyzed using MaxQuant software. For this, a database containing the 400 most abundant proteins of synaptic vesicles was used. Acetylations as well as DEPC modifications were included in the database search. The in-house written R script (**Section 3.3**) was used for quantification of the labeling intensity of each modified residue.

Following the described labeling procedure, 254 labeled proteins were identified using either S-NHS-acetate, NHS-acetate or DEPC (**Figure 31a**). Of these, 232 proteins and a total of 1130 amino acid residues were labeled with DEPC (**Figure 31b,c**). The majority of the

modified residues included lysine (835) and histidine (161) residues. Serine, threonine, tyrosine and cysteine residues were labeled to a much lower extent (**Figure 31c**). A lower number of proteins was labeled when using S-NHS-acetate and NHS-acetate (182 proteins): Acetylation of lysine, serine, threonine and tyrosine residues was observed. Again, lysine residues were primarily modified, namely 792 when using S-NHS-acetate and 731 when using NHS-acetate, respectively (**Fig 31c**). Importantly, many residues were modified with both labeling reagents (**Figure 31b**).

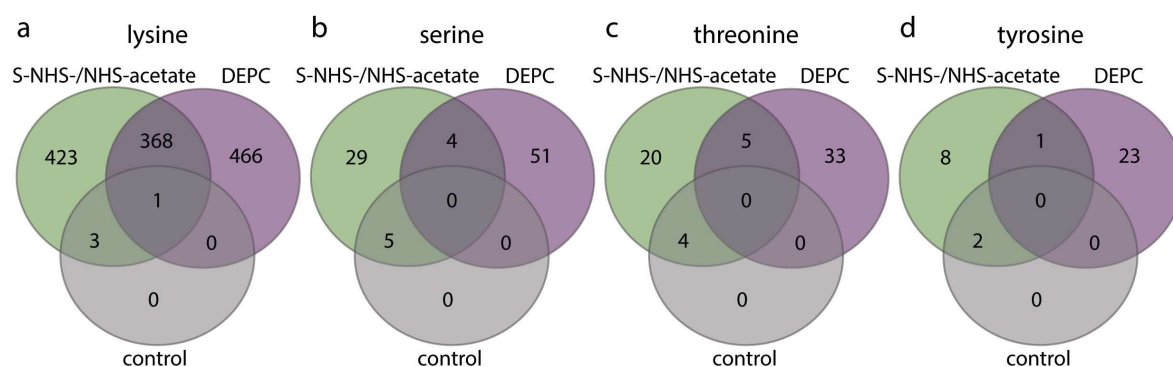


### Figure 31: Overview of labeled proteins and residues

Synaptic vesicle proteins were labeled using S-NHS-acetate, NHS-acetate or DEPC as labeling reagent. **a**) Venn diagram of modified proteins using the different labeling reagents. **b**) The overlap of modified residues using the different labeling reagents is visualized in a Venn diagram. **c**) The number of labeled lysine (K), serine (S), threonine (T), tyrosine (Y), histidine (H) and cysteine (C) residues is indicated for each labeling approach (see legend for details).

Comparing the two labeling reagents introducing acetylating (S-NHS-acetate and NHS-acetate) with DEPC, it becomes apparent that DEPC yields a higher number of modifications. This is due to the ability of DEPC to modify a wider range of amino acids or a higher reactivity. For a detailed assessment of differences between DEPC and S-NHS-acetate and NHS-acetate (S-NHS-/NHS-acetate), only amino acids targeted by all three reagents (lysine, serine, threonine and tyrosine residues) were compared (**Figure 32**). For this, residues modified by S-NHS-/NHS-acetate or DEPC as well as residues that were found

to be acetylated in the control sample without addition of the labeling reagent were compared. Modified lysine residues (1261 residues) were either modified by S-NHS-/NHS-acetate (426 residues), by DEPC (466 residues) or by both labeling strategies (369 residues) (**Figure 32a**). Some modifications were also observed in the control. Modifications identified in the control are either false positive hits or originate from naturally occurring acetylation of lysine residues as PTM. However, only four acetylated lysine residues were detected in the control (**Figure 32a**).



**Figure 32: Comparison of labeled lysine, serine, threonine and tyrosine residues using S-NHS-/NHS-acetate or DEPC**

Acetylated residues identified in the control (grey) and modified residues identified when using S-NHS-/NHS-acetate (green) or DEPC (purple) were compared. The overlap of identified residues between the conditions is visualized in Venn diagrams. The following residues were inspected: lysine (a), serine (b), threonine (c) and tyrosine (d).

More than 50% of the labeled serine, threonine and tyrosine residues were only modified by DEPC (**Figure 32,b-d**). The number of modified serine, threonine and tyrosine residues identified is low when using S-NHS-/NHS-acetate and DEPC (**Figure 32,b-d**). Interestingly, modification of individual residues is either observed when using S-NHS-/NHS-acetate or DEPC indicating differences in labeling reactivity for both labeling strategies. For DEPC labeling of serine, threonine and tyrosine residues, Vachet and co-workers proposed that nearby hydrophobic residues facilitate an increase in local concentration of DEPC, thus making labeling of these residues more likely [295, 296]. Again, few acetylated residues were also identified in the control samples. For serine, threonine and tyrosine residues, these are most likely false positive identifications.

In conclusion, the three different labeling approaches are suited for labeling of synaptic vesicle proteins. Importantly, different residues are modified when using S-NHS-/NHS-acetate or DEPC as chemical labeling reagent. Labeling with S-NHS-/NHS-acetate and DEPC, therefore, results in complementary information on solvent accessibility in particular for serine, threonine and tyrosine residues.

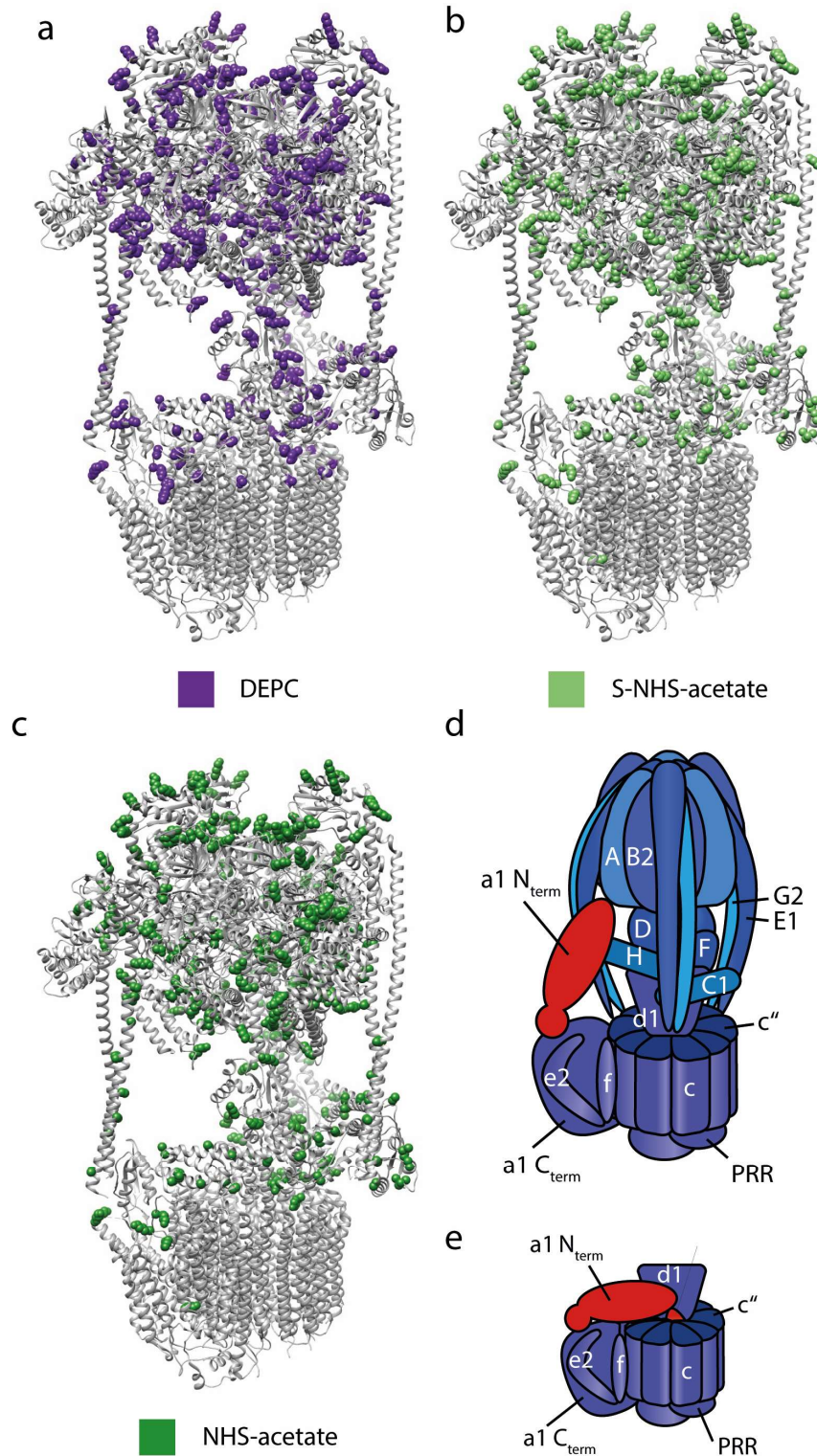
### 3.4.2 Labeling of the synaptic V-ATPase complex

First, the proton-pumping vacuolar H<sup>+</sup>-ATPase (V-ATPase) was inspected in more detail. The V-ATPase is a multi-subunit complex composed of a membrane-embedded domain ('V0') and a cytoplasmic soluble domain ('V1'). In synaptic vesicles, most of the V-ATPase subunits are among the highest abundant proteins [294, 297]. It is generally assumed that one to two copies of the V-ATPase complex are present in one synaptic vesicle [298]. In addition, a high-resolution structure of the V-ATPase from rat brain containing 15 subunits was recently determined [299].

For evaluation of the labeling approach, labeled residues of the V-ATPase were inspected in detail. For this, modified residues were visualized on the available high-resolution structure (**Figure 33**). In total, 179 unique residues were labeled using the three labeling reagents. Using DEPC, 145 residues out of 11 different proteins were found to be modified. Using S-NHS-acetate and NHS-acetate for chemical labeling, 9 different proteins and 133 (S-NHS-acetate) and 122 (NHS-acetate) residues were modified. Of these, all residues that were labeled with NHS-acetate were also labeled when using S-NHS-acetate. Only 80 residues were modified using S-NHS-acetate, NHS-acetate and DEPC again indicating differences in the labeling preferences by specific amino acids for the two labeling strategies. Nonetheless, visualization of the modified residues on the high-resolution structure confirmed solvent accessibility of all residues (**Figure 33**).

Importantly, residues located in the cytosolic 'V1' domain were predominantly labeled. Note that the high-resolution structure only represents one conformation of the highly dynamic V-ATPase and, therefore, modifications correspond to specific conformations of the V-ATPase are not sampled in the structure.

Of the 'V0' domain, residues of subunit 'd1' and 'a1' were mainly modified (**Figure 33,a-c**). Using DEPC labeling, modified residues of the 'c' subunit located on the cytoplasmic side of the c-ring were identified (**Figure 33a**). Importantly, the cytoplasmic domains of subunit 'd1' and of the 'c-ring' are only accessible when the 'V0' and 'V1' domains dissociate. This dissociation has been reported previously and leads to a higher abundance of the 'V0' domain in the obtained synaptic vesicle preparations [294]. The observation that residues within 'd1' and 'c' proteins were modified supporting the assumption, that at least some ATPase complexes dissociated to 'V0' and 'V1' domains. Note that, dissociated 'V1' is lost during the preparation of synaptic vesicles resulting in an excess of 'V0'.

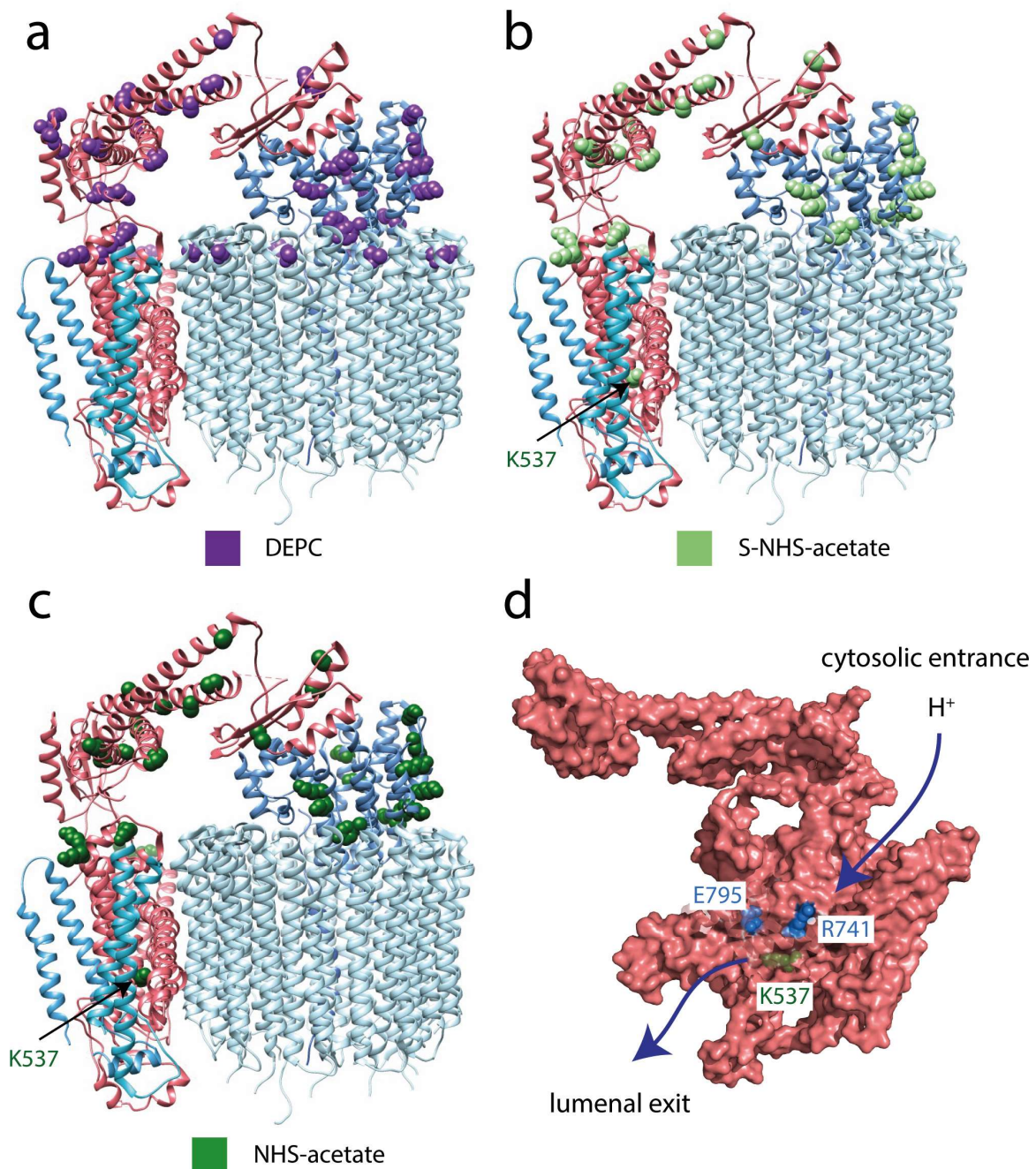


### Figure 33: Labeling of solvent accessible residues of the V-ATPase

The high-resolution structure of rat V-ATPase (PDB: 6vq6 [299]) is shown as cartoon representation (grey, **a-c**). Modified residues are shown as surface representation (**a-c**). **a**) DEPC, **b**) S-NHS-acetate, **c**) and NHS-acetate modified residues are highlighted. **d**) The cartoon model of the V-ATPase includes the indicated subunits of the cytoplasmic 'V1' and membrane-embedded 'V0' domains. **e**) Model of the 'V0' domain is shown. (The cartoon models of the V-ATPase were provided by Sabine Wittig).

In addition, many residues located in the cytosolic domain of subunit 'a1' were modified (**Figure 33, a-c**, subunit 'a1'). This subunit exists in at least two conformations; the open conformation represents the active, fully assembled enzyme (**Figure 33d**) and the closed conformation locks the 'V0' domain (**Figure 33e**) [226, 299]. In the open conformation the N-terminus of subunit a1 interacts with the peripheral stalk built by subunit G2 and E1, while in V0 structures the N-terminus of subunit 'a1' is orientated towards the 'd1' subunit of the central stalk in 'V0' [226, 299]. In addition, in ion mobility and native MS experiments of the ATPase performed by Zhou et. al, a high flexibility of subunit 'a1' was observed [300]. Upon addition of ATP the ion mobility data indicated a more compact and defined structure of 'V0' [300]. These previous findings suggest flexibility of the cytoplasmic domain of subunit 'a1' and presumably result in a high number of modified residues upon chemical labeling.

Interestingly, with the two acetylation reagents, lysine 537 which is located close to the luminal part of synaptic vesicles (**Figure 34b,c**, lysine 537 of subunit 'a1') was found to be labeled. This unexpected labeled residue within the vesicular lumen, is either a false positive identified modified residue or it is indeed accessible for the labeling reagent. Close inspection of this residue, revealed its localization in close proximity to the proton half-channels of subunit 'a1' and the predicted opening site (**Figure 34d**) [299]. Protons enter through the cytosol-facing half-channel and binding of a proton to subunit 'c' occurs. After one full rotation of the c-ring protons exit through the luminal-facing proton half-channel. In detail, the rotation of the c-ring results in abstraction of a proton from the cytoplasmic half-channel to neutralize a charge on a conserved glutamic acid residue in subunit 'c' [299]. The rotation then brings the protonated glutamine residue close to arginine 741 of subunit 'a1' and formation of a salt bridge between arginine 741 and glutamic acid 795 of subunit 'a1' causes release of the proton into the luminal half-channel [299]. In this proposed mechanism, arginine 741 and glutamic acid 795 of subunit 'a1' are located at the predicted opening site of the half-channel [299]. Labeling of lysine 537 of subunit 'a1', therefore, suggests accessibility for the labeling reagent through the proton half-channels.



**Figure 34: Detailed assessment of modified residues in the ‘V0’ domain of the V-ATPase complex.**

The high-resolution structure of the rat V-ATPase subunit ‘V0’ (PDB: 6vq6 [299]) is shown in cartoon representation (a-c). Modified residues are shown in surface representation. Subunit ‘a1’ (red) and the subunits c, d1, e2, PRR and f (blue) are indicated. Residues modified when using different labeling reagents: a) DEPC, b) S-NHS-acetate, c) and NHS-acetate are visualized (see legend for details). d) The cartoon and transparent surface of subunit ‘a1’ is shown (PDB: 6vq6). The two proton half channels are indicated by arrows. Lysine 537 (K537) was modified with S-NHS-acetate and NHS-acetate (green). Glutamic acid 795 (E795) and arginine 741 (R741) are located at the predicted opening of the proton half channels.

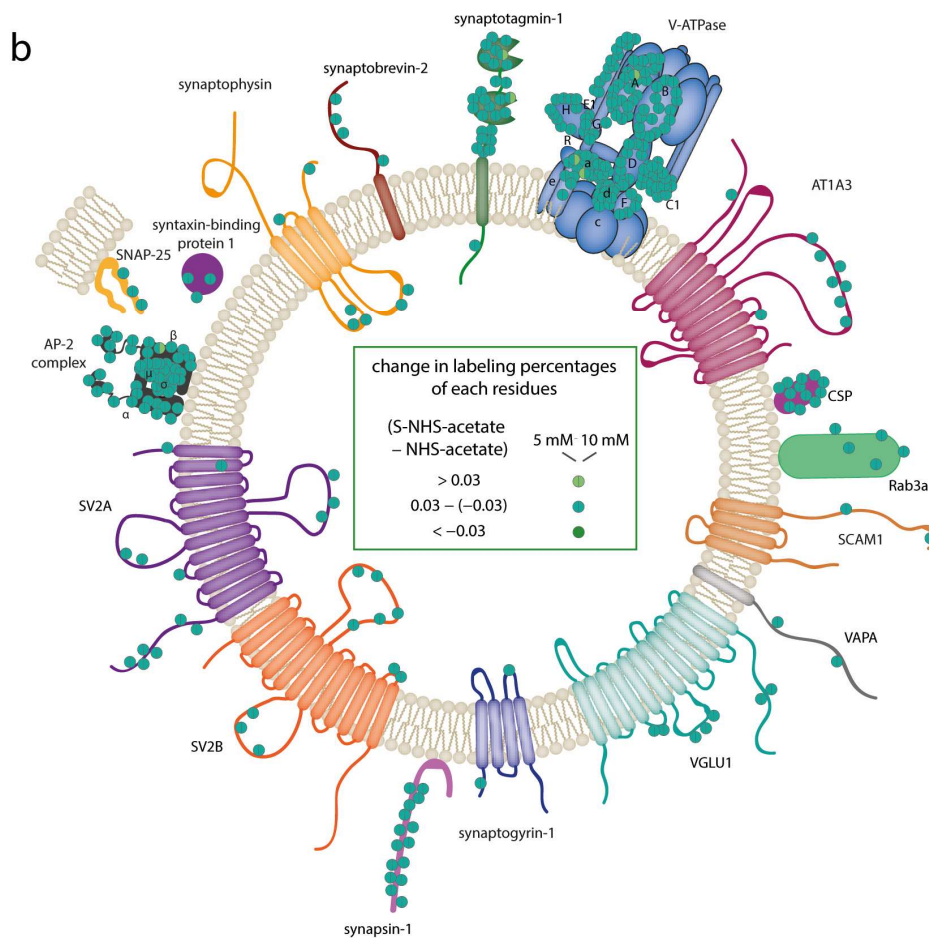
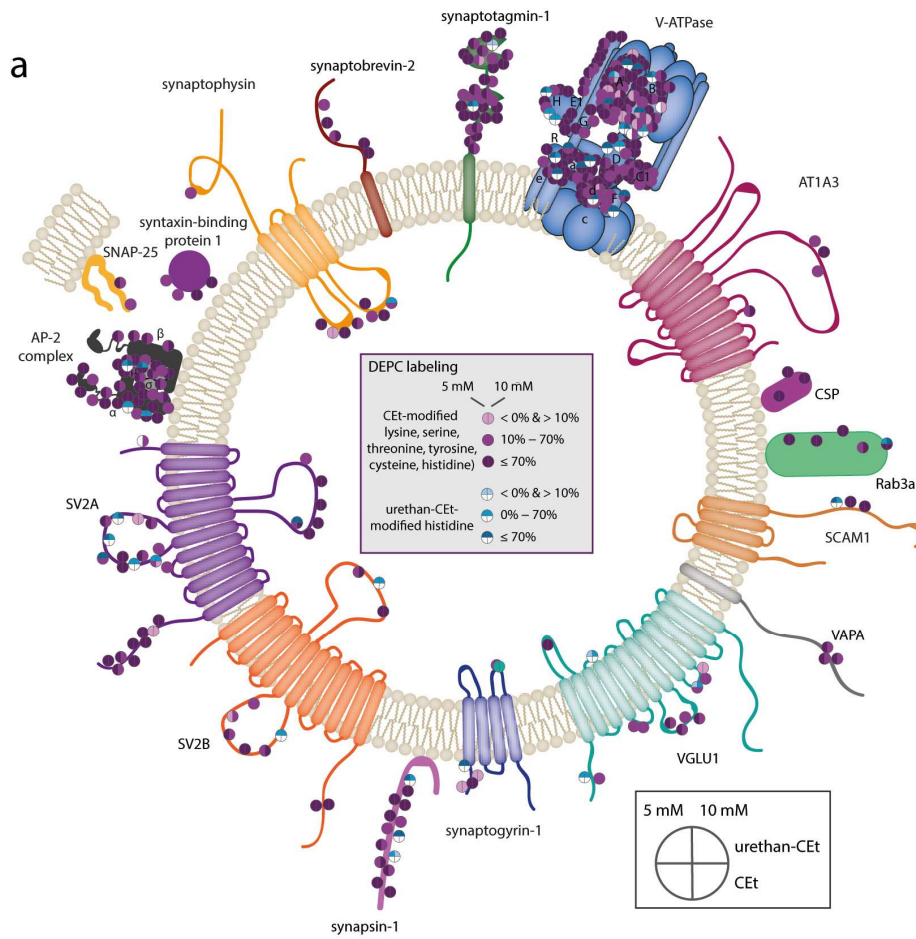
In conclusion, the location of labeled residues of the V-ATPase is in good agreement with solvent accessibility in the high-resolution structure. In particular, labeled residues of the ‘V0’ domain were identified confirming dissociation of the two V-ATPase subunits ‘V0’



and 'V1'. Surprisingly, lysine 537 of subunit 'a1' was found to be modified suggesting accessibility through the proton half-channels of subunit 'a1'. The successful labeling of the V-ATPase protein complex by applying the described labeling approach validates the obtained labeling results.

### **3.4.3 Labeling percentage of individual residues of synaptic vesicle proteins**

Next, information on labeling percentage of solvent exposed residues of several synaptic vesicle proteins was obtained by specifically analyzing the results. Briefly, for each replicate, the labeling percentage of each labeled residue (i.e. the relative proportion of labeled residues with respect to all peptides containing this residue) was calculated for each concentration of the labeling reagent (**Section 3.3**). In the next step, the mean labeling percentage of each modified residue was calculated. For DEPC labeling, the calculated mean labeling percentage was categorized as low ( $< 10\%$ ), intermediate ( $10\% - 70\%$ ) and high ( $\geq 70\%$ ). In the previously performed labeling experiments of model proteins (**Section 3.3**), mainly CEt- and urethane-CEt-modified histidine residues were observed. This was also observed for labeling of synaptic vesicle proteins. Therefore, only CEt- and urethane-CEt-modified histidine residues were considered. Identified DEPC modified residues were then visualized on cartoon models of the most prominent synaptic vesicle proteins (**Figure 35a**). This first overview of labeled synaptic vesicle proteins, revealed solvent exposure of several domains of these proteins.



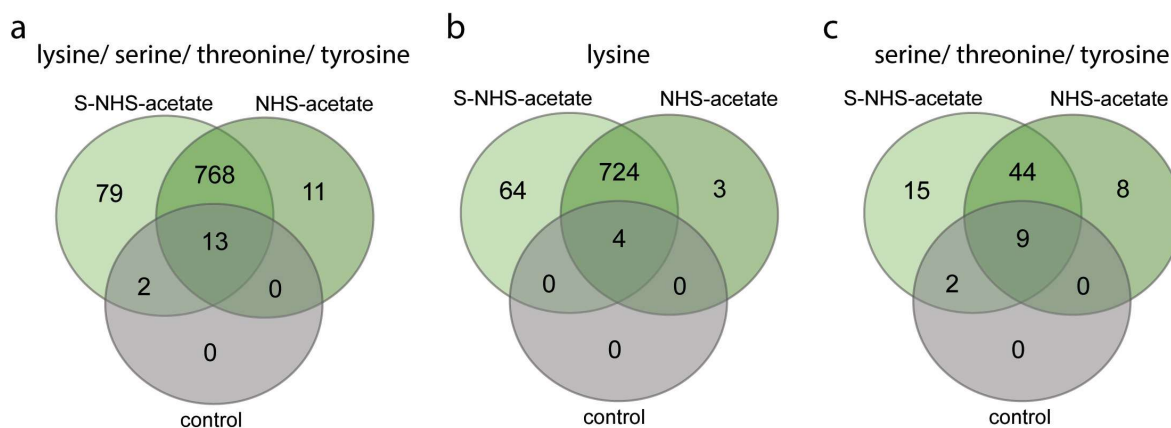
### **Figure 35: Cartoon models of labeled synaptic vesicle proteins**

Prominent synaptic vesicle proteins are shown as cartoon models. **a)** Synaptic vesicles were labeled with 5 mM or 10 mM DEPC. The identified modified residues are colored according to the calculated labeling percentage for CEt-modified residues (histidine, lysine, serine, threonine, tyrosine and cysteine) and urethan-CEt-modified histidine residues (see legend for details). **b)** Synaptic vesicle proteins were labeled with 5 mM or 10 mM using NHS-acetate or S-NHS-acetate. Significant differences in labeling percentages for each residue between S-NHS-acetate and NHS-acetate were determined using a t-test. Intensity changes  $> 0.03$  or  $< -0.03$  with a p-value  $< 0.05$  were considered to be significantly changed (see legend for details).

Abbreviations: AT1A3, sodium/potassium-transporting ATPase subunit alpha-3; CSP, cysteine string protein; Rab3a, Ras-related protein Rab-3A; SCAM1, secretory carrier-associated membrane protein 1; VAPA, vesicle-associated membrane protein-associated protein A; VGLU1, vesicular glutamate transporter 1; SV2B, synaptic vesicle glycoprotein 2B; SV2A, synaptic vesicle glycoprotein 2A; AP-2 complex, adaptor-related protein complex 2; SNAP-25, synaptosomal-associated protein 25; CEt, carboethoxy.

In contrast to DEPC labeling, labeling with S-NHS-acetate and NHS-acetate allows a different analysis strategy. As mentioned above, S-NHS-acetate specifically labels solvent accessible residues of protein domains located at the cytoplasmic side of the synaptic vesicle membrane, while NHS-acetate labels residues located at the cytoplasmic and the luminal side of the synaptic vesicle membrane. The labeling percentage of residues at the cytoplasmic side is, in theory, similar for S-NHS-acetate and NHS-acetate labeled amino acids. In contrast, residues located in the lumen of synaptic vesicles are expected to be exclusively labeled with NHS-acetate. Therefore, the difference in labelling percentage of NHS-acetate and S-NHS-acetate for a specific residue indicates the location of this residues at luminal or cytosolic side of the vesicle. A positive difference in the calculated labeling percentage between S-NHS-acetate and NHS-acetate indicates a cytosolic orientation while a negative difference indicates a luminal orientation.

Accordingly, differences in residue specific labeling of S-NHS-acetate and NHS-acetate were analyzed. Acetylation of 781 lysine, serine, threonine and tyrosine residues was identified after S-NHS-acetate and NHS-acetate labeling (**Figure 36a**). Of these, lysine residues were mainly modified by both reagents. The 728 acetylated lysine residues correspond to a high overlap (91.5%) of both labeling reagents (**Figure 36b**). Side reactions with serine, threonine and tyrosine residues occurred to a lower extent revealing a reduced overlap (53 residues, **Figure 36c**).



**Figure 36: Comparison of identified modified residues identified in synaptic vesicles proteins using S-NHS-acetate or NHS-acetate**

Identified acetylated residues in the control sample without addition of labeling reagent (grey) as well as with addition of S-NHS-acetate (light green) or NHS-acetate (dark green) were compared. The overlap of modified sites between the three conditions is visualized in Venn diagrams. Modifications of different amino acids were analyzed: **a**) lysine, serine, threonine and tyrosine, **b**) lysine and **c**) serine, threonine and tyrosine.

In summary, for S-NHS-acetate labeling, a higher number of modified sites was identified. Although minor differences were observed, the high overlap of modified residues allows a quantitative comparison of labeled residues. For this, a t-test was applied to determine differences for each residue in labeling percentages between S-NHS-acetate and NHS-acetate. Changes in the labeling percentage  $> 0.03$  or  $< -0.03$  with a p-value  $< 0.05$  were considered to be significant. These results were visualized by highlighting identified labeled residues on a cartoon structure of synaptic vesicles proteins (**Figure 35b**). For most modified residues, no significant change in labeling percentage was observed. Some residues were labeled to a greater extent with S-NHS-acetate. This is probably a result of the nature of synaptic vesicles, which are densely packed with proteins resulting in labeling of directly accessible cytosolic domains.

When applying both labeling strategies, modified residues of all prominent synaptic vesicle proteins were identified. For a more specific analysis, specific proteins will be discussed in the following sections.

### 3.4.4 Solvent accessibility of synaptic vesicle proteins

#### Typical cytosolic synaptic vesicle proteins

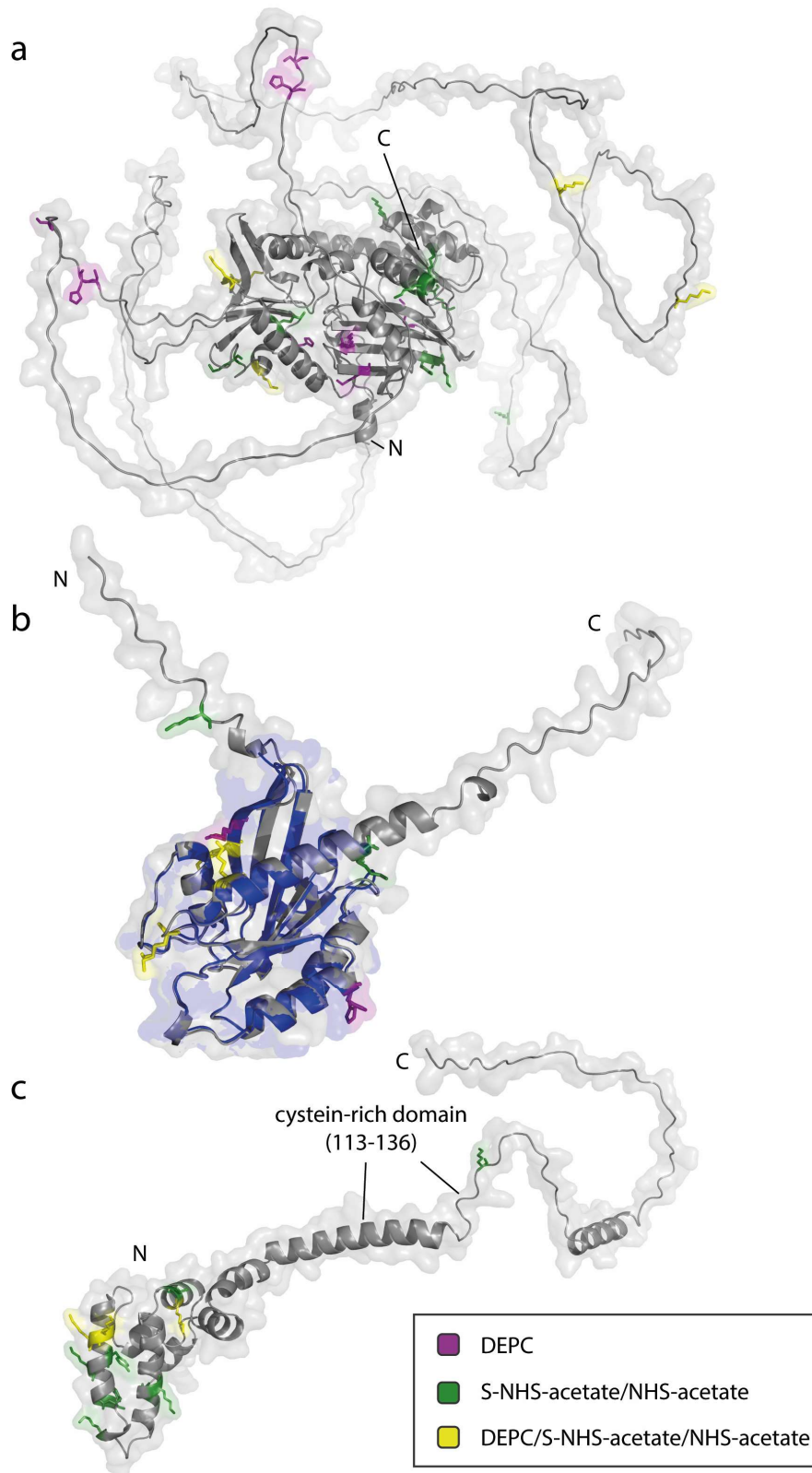
In the following part, synaptic vesicle proteins with predominantly cytosolic orientation were analyzed. As described above, cytosolic proteins and protein complexes of synaptic vesicles were labeled using DEPC and S-NHS-/NHS-acetate. One typical multi subunit

protein complexes in synaptic vesicles is the adaptor-related protein complex 2 (AP-2 complex). The AP-2 complex plays an important role during clathrin-mediated endocytosis (**Figure 2**, step 6) and was highly modified. The proteins of this complex are among the high abundant proteins identified in synaptic vesicle preparations [294, 297, 298]. Labeling of this complex confirmed its localization on the cytoplasmic side of synaptic vesicles.

Other cytosolic proteins such as synapsin-1, Rab3a and cysteine string protein (CSP) were modified by all three labeling reagents. For the three proteins, complete high-resolution structures are not available. Therefore, each protein's 3D structure was predicted using AlphaFold [229] (**Figure 37**).

Synapsin-1 and isoforms of synapsin play a role in the synaptic vesicle cycle by influencing the localization of synaptic vesicles and leading to clustering of synaptic vesicles distal to the active zone in a so-called reserve pool of synaptic vesicles [301]. Approximately, eight copies of synapsin-1 are present per vesicle and binding of synapsin-1 to acidic lipid bilayers has been reported [298, 302]. Upon phosphorylation by numerous protein kinases, synapsin dissociates from synaptic vesicles releasing them from the reserve pool [301]. For synapsin-1, a complete high-resolution structure is not available, presumably due to predicted disordered regions. Modified residues using the three labeling reagents were observed at the surface of the structured part of synapsin-1 and within disordered regions (**Figure 37a**).

Rab3a, is a member of the Rab GTP-binding family and is suggested to play a control role during assembly and disassembly of the SNARE-complex. In its GTP-bound state Rab3a is anchored to the synaptic vesicle membrane. Upon hydrolysis of GTP to GDP, dissociation from the membrane mediated by the protein 'GDP dissociation inhibitor' (GDI) occurs and subsequent synaptic vesicle fusion follows [303, 304]. In the membrane bound state, Rab3a interacts with the synaptic vesicle membrane via its hydrophobic C-terminus [305]. For Rab3a, a high-resolution structure, excluding N- and C-terminal parts, was previously obtained [306]. In the C-terminal region, involved in membrane binding, labeled residues were not identified (**Figure 37b**). Residues within the cytosolic domain, however, were found to be modified. In addition, lysine 12 was modified by S-NHS-/NHS-acetate, indicating surface exposure and probably the short N-terminal domain of Rab3a.



**Figure 37: Labeled residues of Synapsin-1, Rab3a and CSP.**

The protein structure of **a**) synapsin-1 (AF-P09951-F1), **b**) Ras-related protein Rab-3A (Rab3a, AF-P63012-F1) and **c**) Cysteine string protein (CSP, AF-P60905-F1) was predicted using AlphaFold [229]. Proteins are shown in cartoon and transparent surface representation (grey). For Rab3a, a part of the structure was previously determined by x-ray crystallography (blue, PDB: 3RAB [306]). Residues modified by DEPC, S-NHS-acetate or NHS-acetate are shown as sticks and are colored (see legend for details.)

CSP is a member of the DnaJ/Hsp40 family of molecular chaperone proteins. Importantly, it is suggested to prevent misfolding of the SNARE protein SNAP-25, which is part of the fusion machinery of synaptic vesicles [307]. Binding to the synaptic vesicle membrane is mediated by a hydrophobic, cysteine-rich (i.e. 14 cysteine residues) domain including palmitoylated cysteine residues. CSP is 198 amino acid long and the structure of residues one to 100 has been determined by NMR [308]. Modified residues were mainly identified in the cytoplasmic J-domain of CSP (**Figure 37c**). Labeling of the cytoplasmic J-domain indicates its solvent exposure. Only one acetylated residue was identified in the long C-terminal domain and in close proximity to the cysteine-rich membrane binding domain. The absence of modified residues in the cysteine-rich membrane binding region results most likely from embedment of this domain in the membrane. Other modified residues within the C-terminal domain and cysteine-rich membrane binding region were not identified. This might be a result of shielding of this region, for instance, by binding to the membrane or other proteins.

### **Membrane-anchored and membrane-embedded synaptic vesicle proteins**

In addition to cytosolic synaptic vesicle proteins, several transmembrane proteins were modified by DEPC and S-NHS-/NHS-acetate. Among these are prominent synaptic vesicle components such as synaptobrevin-2 and synaptotagmin-1. For synaptobrevin-2, a member of the fusion machinery, 4 and 5 residues were modified with S-NHS-/NHS-acetate and DEPC, respectively (**Figure 35**). These labeled residues are located within the cytoplasmic domain. Although synaptobrevin-2 is the most abundant protein in synaptic vesicles (~ 70 copies per vesicle [298]), only a limited number of sites of this 12.7 kDa protein are amenable to chemical labeling reducing the total number of potentially modified residues.

Similar results were obtained for synaptotagmin-1, which is a highly abundant protein in synaptic vesicles with approximately 15 copies per vesicle. This protein binds calcium and is suggested to play a role in synchronous fusion of synaptic vesicles with the presynaptic membrane [29]. In particular, the two calcium-binding C2 domains, which are located at the cytoplasmic side of synaptic vesicles, were labeled (**Figure 35**, synaptotagmin-1). Residues of the two flexible linker regions were also highly modified. Using DEPC labeling, one modified residue located in the lumen of the vesicle was identified. This is most likely a false positive identification as a modification was not expected.

Other transmembrane proteins present in synaptic vesicles such as sodium/potassium-transporting ATPase subunit alpha-3 (AT1A3), secretory carrier-associated membrane protein 1 (SCAM1) and vesicular glutamate transporter 1 (VGLU1) were modified when using the labeling reagents. These proteins were also predominantly labeled within their cytosolic domains (**Figure 35**).

### **Exposure of luminal loops of tetra-spanning and related proteins**

In the next part, labeling of the transmembrane proteins; i.e. synaptophysin and synaptoporin, an isoform of synaptophysin, synaptogyrin-1 and synaptic vesicle glycoprotein 2A (SV2A), will be analyzed in more detail. The explicit function of these proteins is unknown and, therefore, structural analysis including the identification of solvent exposed domains is of particular interest.

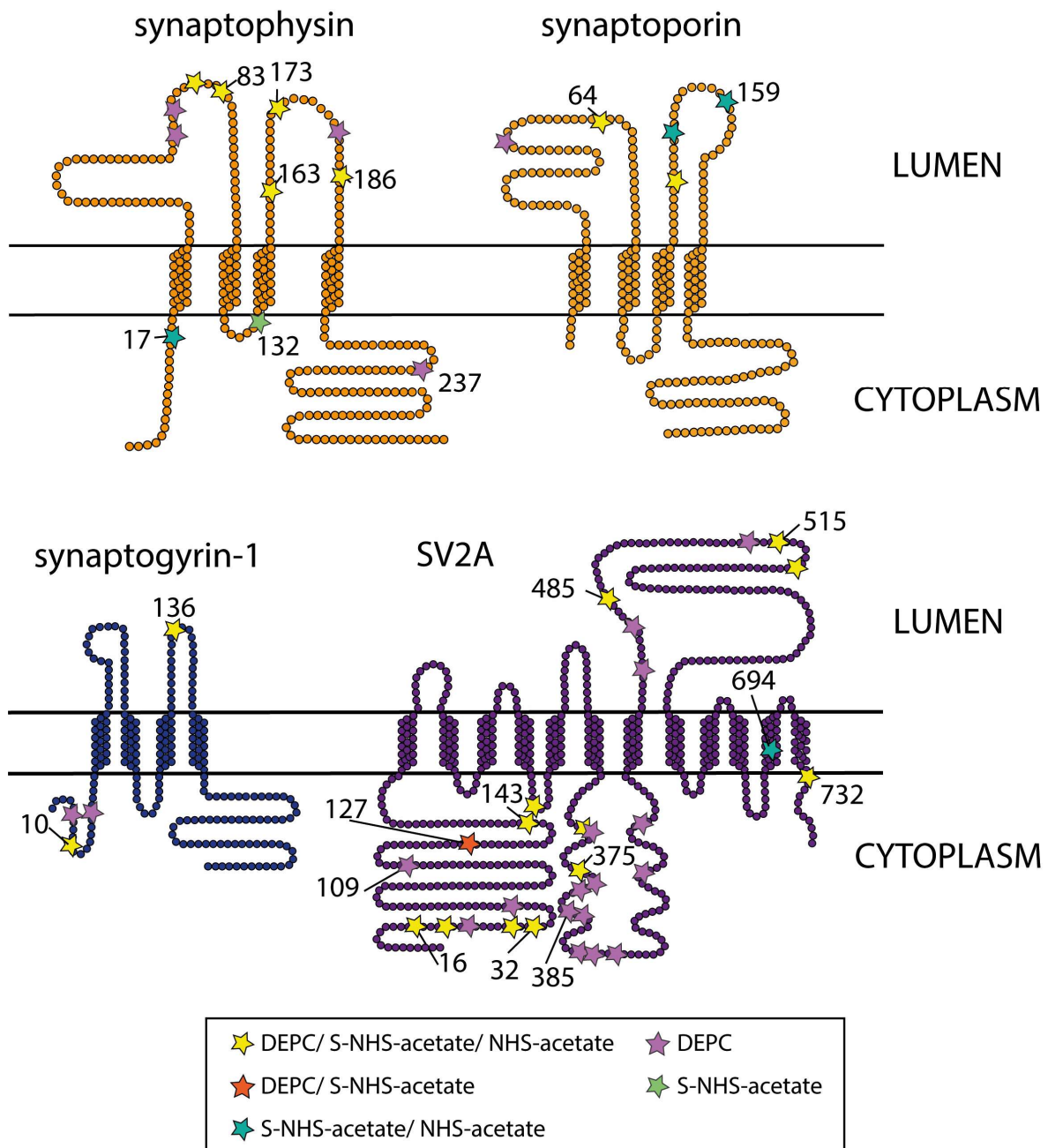
Synaptophysin, synaptoporin and synaptogyrin-1 are tetra-spanning transmembrane proteins. Three cytosolic residues of synaptophysin were found to be labeled. Surprisingly, for both synaptophysin and synaptoporin, modified residues located in both luminal loops were identified (**Figure 38**, synaptophysin and synaptoporin). The modification of residues in the mentioned luminal loops suggests structural changes that result in solvent accessibility of these loops and, consequently, labeling.

For synaptogyrin-1, modified residues located in the cytoplasmic N-terminus and one labeled residue in a luminal loop were identified (**Figure 38**, synaptogyrin-1). This modified luminal residue could be a false positive hit. However, labeling with all three labeling reagents more likely indicates solvent accessibility.

SV2A contains twelve transmembrane domains, six luminal loops, five cytosolic loops and cytosolic N- and C-termini. The large cytosolic N-termini (amino acid 1 – 169) and one large cytosolic loop (amino acid 356 – 447) were found to be heavily labeled (**Figure 38**, SV2A). Again, modified residues located in a large luminal loop (amino acid 469 – 598) were also found to be labeled. Of these, three residues were modified with all three labeling reagents and additional three residues were labeled with DEPC. This high number of modified residues within the luminal loop suggest that these labeled residues are true positive hits. Instead labeling of this luminal domain indicates a functional role of the loop, that involves solvent accessibility.



The described labeling of luminal loops of SV2A, synaptogyrin, synaptophysin and synaptoporin indicates solvent accessibility of these loops and the question of a functional role arises. Previously, glutamatergic vesicles were reported to reversibly increase their size upon filling with glutamate [309]. This increase in size is not mediated by the addition of lipids, but rather conformational changes of proteins resulting in a larger diameter of synaptic vesicles. In particular, SV2A has been suggest playing a role during this process. Large conformational changes of SV2A could be a key component during this process. The labeling results of the luminal loop of SV2A further support this assumption. Specifically, conformational changes resulting in localization of the luminal loop in the membrane might occur. In addition, the labeling results further indicate a similar functional role of synaptophysin, synaptoporin and synaptogyrin-1.



**Figure 38: Modified residues of synaptophysin, synaptoporin, synaptogyrin-1 and SV2A by DEPC and S-NHS-/NHS-acetate**

The predicted localization of the domains of the transmembrane proteins synaptophysin, synaptoporin, synaptogyrin-1 and SV2A is shown. Individual residues are shown as balls. Residues modified by DEPC, S-NHS-acetate and NHS-acetate are indicated with stars (see legend for details). (The figure was provided by Sabine Wittig and adapted for this thesis.)

In summary, labeling experiments of SV2A, synaptophysin, synaptoporin and synaptogyrin-1 in synaptic vesicles suggest a possible functional role of these proteins. Specifically, their luminal loops might play an important structural role, for instance, by enabling the increase in size of synaptic vesicles.

## **4. Discussion and outlook**

### **4.1 Proteome analysis of undifferentiated and differentiated SH-SY5Y cells**

#### **Choosing a neuronal model system**

The neuroblastoma cell-line SH-SY5Y is a prominent neuronal model system. Both, undifferentiated cells and differentiated cells, obtained from various protocols, are commonly employed. In this thesis, the proteomes of undifferentiated as well as RA- and RA/PMA-differentiated SH-SY5Y cells were characterized to determine differences in protein expression and to validate their application as a neuronal model system for in-cell cross-linking experiments.

A systematic review by Xicoy et al. examined the application of SH-SY5Y cells as a neuronal disease system for the analysis of Parkinson's disease [310]. They highlighted that in most studies a differentiation protocol was not applied [310]. When differentiated cells were employed, a variety of differentiation media were applied. RA-induced differentiation was the most prominent differentiation strategy in these studies [310] commonly utilizing a concentration of 10  $\mu$ M RA applied for 3 – 5 days [41]. The second prominent strategy is based on differentiation with 10  $\mu$ M RA and 80 nM PMA [310].

However, the phenotype of RA-differentiated cells obtained in some studies is described as cholinergic, in other studies as dopaminergic [41, 310]. Upon RA/PMA administration, differentiated SH-SY5Y cells show a dopaminergic phenotype [310]. Furthermore, SH-SY5Y cells can be cultured in three-dimensional cultures or on extracellular matrix coated dishes [311, 312]. These protocols require step wise adjustments of a more complex media composition and result in cells with a mature neuronal phenotype [311, 312]. In this thesis, these protocols were neglected, due to limited applicability for in-cell cross-linking experiments as the use of coated dishes would result in heavy cross-linking of the used matrix.

In recent studies, inducible pluripotent stem cells were employed as a model-system for studying neurons *in vitro*. One major advantage is, that they originate from individuals that are healthy or suffer from specific neurological disease [313]. Although, these cells represent a more accurate human neuronal model, their cultivation is complex and obtaining one specific neuronal subtype requires specialized expertise and expensive equipment [313]. The

cultivation of SH-SY5Y cells is, on the contrary, comparably uncomplicated, cost-efficient, and ethical considerations are lower. We, therefore, chose SH-SY5Y cells as neuronal model system.

### **Methodological aspects – mass spectrometry**

Due to, in some cases, contradictory differentiation results and missing proteome studies comparing the different differentiation protocols raised the question, whether the proteomes are different in undifferentiated, RA- and RA/PMA-differentiated SH-SY5Y cells. MS was employed to identify SH-SY5Y proteins and to relatively quantify changes in protein abundance. Although neuronal cell cultures are often limited in the available sample amount, we successfully analyzed the proteome of three different cell-types and identified more than 3,500 proteins in all 18 replicates (i.e. six biological replicates per cell-type). A total of almost 6,000 identified proteins confirms a good coverage of the cellular proteome.

The recently introduced sample preparation strategy termed ‘SPEED’ was employed in this thesis. In contrast to other protocols, neither detergents nor chaotropic agents are used for protein extraction. Instead, cells were lysed in one tube followed by enzymatic hydrolysis in the same reaction tube, thereby reducing sample loss in this thesis. However, upon protein hydrolysis, desalting of the obtained peptide solution is required. In previous experiments, performed by Leonie Jaster during her Bachelor thesis (‘Proteomische Charakterisierung von differenzierten SH-SY5Y-Zellen’, Martin Luther University, 2020), desalting using C18-ZipTips (Merck) and Sep-Pak C18 Plus Short Cartridge desalting columns (Waters) was compared resulting in a higher number of protein identifications when using desalting columns. In addition, SP3 and SPEED protocols were compared in the previous study resulting in similar results. In this thesis, desalting was performed using Pierce<sup>TM</sup> peptide desalting spin columns. In contrast to Sep-Pak C18 Plus Short Cartridge desalting columns, desalting with Pierce<sup>TM</sup> peptide desalting spin columns is less time consuming. Therefore, the sample preparation strategy SPEED and Pierce<sup>TM</sup> peptide desalting spin columns were combined in this thesis.

A three hour gradient LC-MS/MS analysis was employed for peptide separation and identification. With this set-up more than 3,500 proteins were identified in all 18 replicates. To further increase the number of identified peptides and, therefore, protein identification and quantification, prefractionation prior to LC-MS/MS analysis can be performed in future

studies. Prefractionation techniques can be either employed on the protein or peptide level. The most prominent protein fractionation strategy is SDS-Page separating the proteins according to their molecular weight. For peptide prefractionation, strong cation exchange (SCX) [314], hydrophilic interaction chromatography [315], hydrophilic strong anion exchange [316], and high-pH C18-based reversed-phase [317] could be applied. Although prefractionation techniques enable in-depth analysis of the proteome, each additional fractionation step increases the required sample amount and the required analysis time. An alternative, additional separation technique using a FAIMS source, was recently described [318]. Herbert et al. applied label-free quantification (LFQ) experiment with a four hour gradient LC-MS/MS analysis and increased the number of quantified proteins from 6800 to 7800 [318]. Other approaches aim at optimizing LC settings. Accordingly, Bian et al. recently showed that micro-flow LC-MS/MS using a  $1 \times 150$  mm column is advantageous for deep proteome analysis [319]. The authors identified >9000 proteins in 16 h [319]. The addition of a FAIMS source or specific LC adjustments in the LC-MS/MS analysis could, therefore, potentially increase the number of identified proteins even more in the future.

### **Proteome analysis of undifferentiated cells**

The proteome analysis of undifferentiated SH-SY5Y cells uncovered typical cancer characteristics. The cells proliferate constantly, thereby, promoting cancer progression. DNA replication is essential for cell division and proliferation; previous studies showed that DNA replication was inhibited after 24 hours of treatment with RA [320]. Isochorismatase domain containing protein 1 (ISOC1) is among the highly upregulated proteins in undifferentiated cells (when compared with their differentiated counterparts). In other cancer types, such as breast, pancreatic and colon cancer, the knockdown of ISOC1 resulted in upregulation of tumor-suppressive activity [259, 321, 322]. In lung cancer, ISOC1 was found to interact with proteins of DNA damage repair pathways, and in mice xenograft tumor assays, a knockout of ISOC1 significantly inhibited tumor growth [323]. For the isoform of ISOC1, ISOC2, dysregulation of tumor-suppressor p16 (INK4a) was observed [324]. The described findings indicate a similar, however, not fully understood function of ISOC1 in the employed neuroblastoma cells.

Upregulation of minichromosome maintenance complex component (MCM) proteins is linked with poor treatment outcomes and was previously suggested to be a potential prognostic cancer marker [325-333]. In Osteosarcoma, for instance, expression of MSH2 and MSH6 is

associated with shorter survival times for patients [334]. The influence of their high expression in neuroblastoma has not been investigated before.

Upregulation of proteins of the KEGG pathways ‘DNA replication’, ‘mismatch repair’ and ‘cell cycle’ confirm high levels of proliferation. Undifferentiated cells are, therefore characterized by over-expression of proteins necessary for high proliferation rates. The application of undifferentiated SH-SY5Y cells for studying neuronal function and dysfunction, however, remains to be discussed.

### **Proteome analysis of differentiated cells**

SH-SY5Y cells differentiated with RA or RA/PMA showed changes in their morphology and in the expression level of proteins required for structural and functional changes of the cells when compared with their undifferentiated origin. The protein expression profiles of the two differentiated cell-types were, however, very similar. In a previous study, the proteomes of GABAergic and glutamergic synaptic vesicle and synaptic docking complexes were compared [335, 336]. This study revealed only small differences in their protein expression profiles. Differences included their neurotransmitter transporters. In our experiments, we did not apply specific protocols for identification of membrane proteins such as membrane protein enrichment and protein hydrolysis using proteases suited for hydrolysis of membrane proteins. This likely resulted in the absence of the various transporters in our analyses. Nonetheless, the overall goal of our analyses was to uncover differences between undifferentiated and the two differentiated SH-SY5Y cells rather than to obtain a specific neuronal subtype.

Zhang et al. specifically analyzed proteome remodeling during differentiation with RA for 7 days [52]. For this, they employed different TMT labels over a time course [52]. With their setup, the authors relatively quantified 9400 proteins [52]. They further examined the temporal expression pattern of selected proteins [52]. These included three low abundant and downregulated proteins in undifferentiated cells: cytochrome P450 26B1 (CYP26B1), neural cell adhesion molecule 2 (NCAM2) and extracellular leucine-rich repeat and fibronectin type III (ELFN1) [52]. In our analysis, CYP26B1 and NCAM2 were only identified in undifferentiated cells and, therefore, escaped relative quantification. Their identification in undifferentiated cells but not in differentiated cells confirms their low abundance in differentiated cells. ELFN1 was not identified in our analysis. Zhang et al.

further observed downregulation of rabphilin-3A (RPH3A), retrotransposon-like protein 1 (RTL1) and TIMELESS [52]. Again, RPH3A was only identified in undifferentiated cells in our analysis. RTL1 was not identified in RA/PMA-differentiated cells, however, an insignificant decrease was observed for RA-differentiated cells compared to undifferentiated cells. In contrast to Zhang et al., we performed six replicates instead of only two replicates. Therefore, a higher variance might be observed in our experiments. Comparable expression levels of neuron-related proteins were found in this thesis and the study of Zhang et al. Accordingly, PDZ and LIM domain protein 5 (PDLIM5), NAD(P)H dehydrogenase [quinone] 1 (NQO1), integrin alpha1 (ITGA1) and  $\alpha$ -actinin-4 (ACTN4) were upregulated in both RA- and RA/PMA-differentiated cells [52]. Zhang et al. further could not identify a complex time-dependent mechanism, that regulates RA-mediated differentiation of SH-SY5Y cells [52]. In summary, the results of Zhang et al. are comparable to the results obtained in our analysis. In addition, we observed similar changes for RA/PMA differentiated cells as for RA-differentiated cells.

Other important aspects for regulation of neuronal differentiation are changing PTMs of a protein. These might be specific to neuronal subtypes. Among the most frequently analyzed PTMs is phosphorylation of proteins. In SH-SY5Y cells differentiated with a combination of RA and brain-derived neurotrophic factor, phosphorylation levels have already been quantified [42]. Additional PTMs such as acetylation [337] or glycosylation [338] are important for specific neuronal function. This will further allow the characterization of specific neuronal subtypes. In this thesis PTMs were not analyzed but should be addressed in future studies.

In differentiated cells, the expression of mature neuronal markers such as microtubule-associated protein 2 or  $\beta$ -III tubulin did not increase. However, upregulation of nestin and other proteins required for neuronal differentiation suggests, that the cells are in the state of early neuronal differentiation. Downregulation of doublecortin further indicates that the cells are possibly in an intermediate state between early and mature neurons. In this thesis, specifically commonly employed short differentiation protocols were compared. Applying longer differentiation protocols might produce mature neurons and further support differentiation into specific neuronal subtypes (see above). In conclusion, SH-SY5Y cells obtained after 5 days of RA or 6 days of RA/PMA treatment represent early differentiating neurons. These differentiating cells showed overexpression of proteins associated with RA

treatment, neuronal differentiation, establishment of synapses and antioxidant defense as well as proteins regulating the actin cytoskeleton.

## **4.2 In-cell cross-linking of SH-SY5Y cells**

### **In-cell cross-linking – methodological aspects**

The aim of in-cell cross-linking is to capture protein-protein interactions in the native cellular environment. Formaldehyde was used for in-cell cross-linking in this thesis. In addition to formaldehyde, DSS is another prominent membrane permeable in-cell cross-linking reagent. For DSS cross-linking, the cells are often collected and resuspended in cross-linking reagent containing buffer [339, 340]. Cross-linking of the cells, prior to cell collection is probably avoided due to high costs for the cross-linking reagent. Note that, neuronal cells have a distinct shape including long projections. These projections are typically damaged upon cell collection, centrifugation and resuspension. Directly cross-linking intact cells is, therefore, a great advantage of formaldehyde cross-linking. However, cross-linking with formaldehyde has also some draw backs: Formaldehyde does not only cross-link proteins, but also DNA and RNA resulting in a highly complex mixture of cross-linked species. In addition, the introduced linkages are temperature sensitive. These aspects have to be considered during sample preparation and require adjusted protocols at low temperatures. Therefore, the optimized workflow employed for cross-linking SH-SY5Y cells in this thesis avoids high temperatures.

The workflow further includes size exclusion chromatography for enrichment of low abundant cross-links. Alternative enrichment strategies are strong cation exchange chromatography [140] or filtering for higher charge states using FAIMS subsequent to ionization [141]. The latter set-up is particularly beneficial for samples with high complexity such as a cross-linked proteome. In future, it would be of great interest to analyze cross-linked samples with this specialized equipment.

Enrichment strategies are generally employed to increase the abundance of cross-linked peptide pairs and, therefore, the chance for identification of cross-linked peptides. For identification of specific protein interactions, additional purification steps are often included, for instance, a pull-down assay. Recently, Slavin et al. have transfected cells with two nonstructural proteins and one structural protein of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [150]. Upon transfection and protein expression, proteins



were cross-linked within the cell and a viral protein fused to a Strep-tag was used for affinity purification [150]. The authors identified cross-links that allowed integrative modeling to determine the structure of these proteins [150]. This example shows the applicability of in-cell cross-linking for the analysis of specific proteins indicating that in future studies, in-cell cross-linking of SH-SY5Y cells and subsequent pulldown experiments might be valuable for the structural analysis of proteins relevant in neurological diseases. Therefore, the established in-cell cross-linking workflow of SH-SY5Y cells represents a promising neuronal model system to study changing protein interactions in neuronal diseases or at different functional conditions.

Another critical aspect is the data analysis of cross-linked peptides and the required software for their identification. Cross-linked peptide pairs are typically identified by a series of fragment ions corresponding to each peptide. Depending on the length and composition of the peptide, some ions are not or only observed with low intensity hampering identification. In addition, with increasing database sizes the possible peptide pair combinations increase exponentially and thereby challenge search algorithms. In a study from Beveridge et al., a synthetic peptide library was utilized for comparison of routinely used cross-linkers and software [341]. Depending on the analysis strategy the authors observed false discovery rates between 2.4 and 32%. This rate could not be reduced when using MS-cleavable crosslinkers. Furthermore, they found that some software performed better with smaller and other with larger databases. These findings indicate that data analysis improves with careful selection of the analysis software. In this thesis, the recently developed 'Formaldehyde XL Analyzer' software [136] was employed. This software is still under development and has not yet been explored extensively. Future software advancements, including additional parameter settings such as variable oxidation of methionine, will likely increase the identification rates of cross-linked peptides.

### **In-cell cross-linking – biological aspects**

The applied proteome-wide cross-linking strategy resulted in the identification of protein cross-links in undifferentiated, RA- and RA/PMA differentiated cells. For identification of reproducible cross-links, the data analysis included filtering steps for high and intermediate confident cross-links as well as identification of hits in more than one replicate. The results were validated by visualizing cross-links of ribosomal proteins on an available high-

resolution structure of the ribosome. Most of these cross-links were in good agreement with the high-resolution structure confirming applicability of the workflow.

The observed protein interaction network revealed several protein interaction clusters of ribosomal proteins, cytoskeletal proteins or histones (**Supplementary Figure 1**). For detailed assessment, the interaction network of actin gamma 1 was chosen. In all three culture conditions, interactions between actin gamma 1 and different proteins were observed. Interestingly, cross-links between transgelin-2 and actin gamma 1 or tropomyosin were identified, suggesting binding of transgelin-2 to F-actin via its CH3 domain as previously described for other CH3-containing proteins [342].

Upon differentiation, several interactions with actin-regulating proteins were observed (**Figure 21**). The absence of these interactions in undifferentiated cells suggests reorganization of the actin cytoskeleton during differentiation. Interactions between actin gamma 1 to actin dynamic regulating proteins such as the ARP2/3 complex in RA-differentiated cells or Myosin variants in RA/PMA-differentiated cells as well as cyclase-associated protein 1 in both cell types indicate a specific function during early development associated with actin reorganization. However, for most of these protein interactions, only few cross-links were observed suggesting that additional optimization steps are required (see above). For future experimental studies, specific enrichment steps as described above will help increasing the identification of protein interactions. Nonetheless, in-cell protein cross-linking of the complete SH-SY5Y proteome provided several interaction clusters and, therefore, first clues on changes in protein interactions involving cytoskeletal proteins.

### **4.3 Chemical labeling for identification of solvent accessible amino acid residues**

#### **Labeling of model proteins**

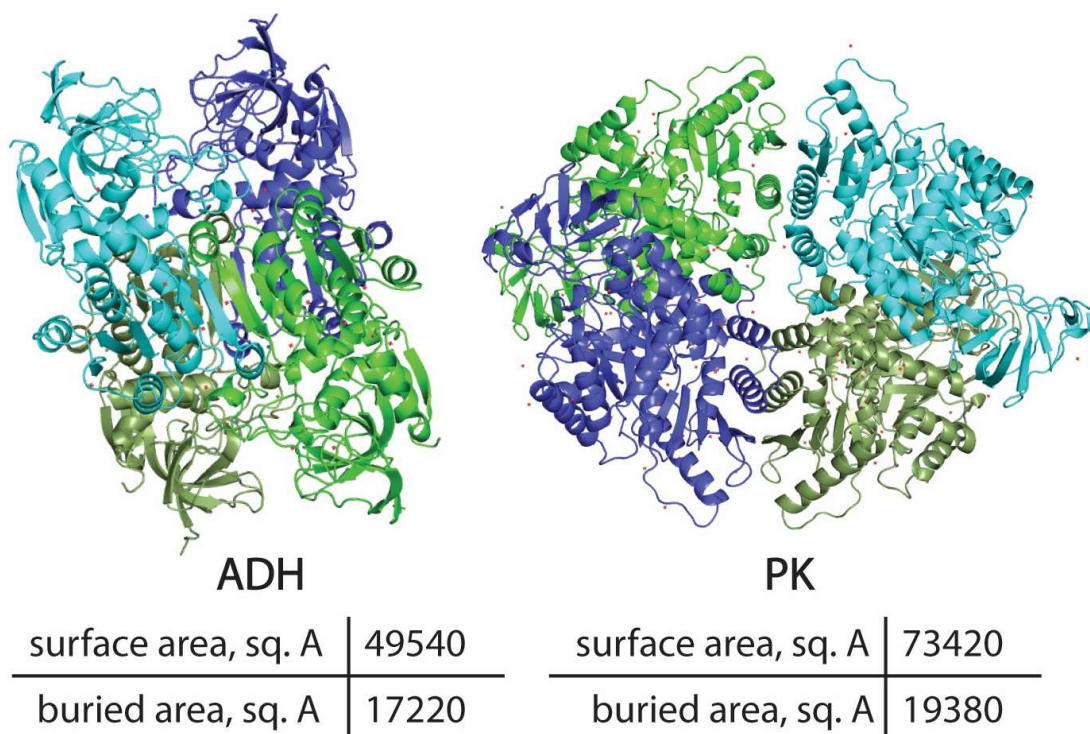
Chemical labeling of proteins provides information on the solvent accessibility of specific residues of a protein or protein complex and adds valuable information for many research questions [290]. The two labeling reagents used in this thesis, i.e., NHS-acetate and DEPC, have previously been applied to study solvent accessibility of several proteins [190, 202, 343]. Most studies focused on the identification of labeled amino acids. However, considering recent advancements in MS instrumentation, low abundant modifications are

also identified and, therefore, a distinction between buried and solvent exposed residues is required. The aim of this thesis was to characterize and compare both labeling approaches and to develop a quantitative workflow for a specific identification of solvent accessible residues compared to buried residues. As expected, labeling of two model proteins resulted in identification of acetylated lysine, serine, threonine and tyrosine for NHS-acetate and in CEt-modified histidine, lysine, serine, threonine, cysteine and tyrosine residues for DEPC. The previously described acetylation of arginine residues by NHS-acetate [289] was not observed. This was the case even at high concentrations of NHS-acetate suggesting that the guanidino group of the arginine side chain is not reactive under the employed experimental conditions.

Labeling with DEPC was proposed to result in four reaction products [290]. Previous studies show that formyl-CEt converts into urethane-CEt under slightly basic conditions (pH 7.5 – 8) [344]. In our experimental setup, basic conditions were employed suggesting that formyl-CEt transformed into the more stable urethane-CEt. Accordingly, mainly CEt- and urethane-CEt-modified histidine residues were identified. These two reaction products were further implemented during the analysis and the following quantification workflow to analyze solvent accessibility.

The reproducibility of NHS-acetate labeling was higher than that observed for DEPC labeling suggesting that the higher variety of residues that can be modified when using DEPC results in more complex labeling products and, therefore, complicates identification. In addition, the labeling efficiency of specific amino acid residues for both model proteins and labeling reagents was found to be different. Nearly all lysine residues were labeled with NHS-acetate and DEPC suggesting a high reactivity of lysine residues in general or a high surface exposure as expected for lysine residues and, therefore, good accessibility for the labeling reagents. For serine residues, similar results as for lysine residues were obtained and as previously reported [196, 345]. Approximately half of the available tyrosine residues were labeled in ADH and PK when using either reagent indicating a lower reactivity of this amino acid or a lower accessibility for the reagents. Differences in the labeling efficiency of threonine and histidine residues were observed between ADH and PK. Histidine residues were expected to be highly labeled using DEPC as labeling reagent [194, 196]. However, labeling of histidine residues was only pronounced for PK while histidine residues of ADH were labeled to a lower extent. This observation was similar for threonine residues. In PK, almost all threonine residues were modified. A comparison of the available high-resolution

structures of ADH and PK revealed a more compact structure of ADH, while the subunits of PK appear to be highly accessible (**Figure 39**). The structure of PK provides a higher accessible surface area, resulting in labeling of most residues targeted by the two labeling reagents. Consequently, the labeling efficiency highly depends on the structure of the protein under investigation.



**Figure 39: High-resolution structures of ADH and PK.**

The tetrameric structures of ADH (left, PDB ID: 5env) and PK (right, PDB ID: 1aqf) are visualized with PyMOL. The four protein chains are shown in different colors in cartoon representation. The accessible surface and buried area of each protein complex was calculated using PDBe PISA [213]. (Figure from Barth et al. 2020 [153].)

Ion mobility MS was employed to observe structural changes upon labeling. For labeling of ADH with NHS-acetate and low concentrations of DEPC the structure was preserved. Although ion mobility MS enabled an overall structural analysis of ADH, small changes in the structure of ADH could not be analyzed by this strategy. The activity of ADH was not assessed. Furthermore, labeling of other proteins might induce even at low concentrations significant structural changes. In future studies, these tests could be included. In summary, the two labeling approaches are well suited for the structural analysis of proteins and their complexes.

### **Quantitative labeling of model proteins**

State-of-the-art mass spectrometers nowadays enable the identification of very low abundant peptides including their modifications raising the question, whether NHS-acetate and DEPC labeled amino acids are not only modified but also solvent accessible and, therefore, suited for structural analysis. To answer this question, a quantitative approach was applied and differences in the labeling percentage of individual residues were monitored. For this, the calculated labeling percentage of each residue was compared to its computed SASA using the available high-resolution structure. This analysis revealed that residues which showed a characteristic increase in labeling percentage at increasing concentration of the labeling reagent correspond to an exposed residue. This agrees well with the computed solvent accessibility. In contrast, this increase was not observed for residues which are less exposed. Labeling of buried residues occurred to a minor degree and is most likely the result of labeling of low amounts of dissociated monomeric and dimeric ADH from tetrameric complexes. The labeling results of both model protein complexes showed the described increase in labeling percentages for labeled residues and their correlation to SASA of the residue.

In contrast to previous results [296], lysine residues appeared to be best suited for a quantitative approach and reliably display solvent accessible residues (**Figure 28**). Interestingly, higher labeling percentages are often observed for NHS-acetate labeling. In contrast to NHS-acetate labeling which results in fixed acetylation of residues, scrambling of the label was described for DEPC [346, 347]. In detail, Zhou et al. observed that free cysteines are able to capture a CEt group from other modified amino acids in solution [347] and Borotto et al. identified scrambling during CID of labeled peptides transferring histidine modifications to other amino acid residues [346]. In addition, the CEt-modification is reversible under various conditions [348]. The described scrambling of the label probably results in the observed lower labeling percentages when using DEPC. However, for few amino acids, a higher labeling percentage was observed suggesting that these residues are better accessible for DEPC than for NHS-acetate. The structure of NHS-acetate is rather bulky while DEPC is smaller resulting in different accessibility of potential labeling sites. Therefore, structural arrangements of the protein might influence labeling by DEPC and NHS-acetate and partially buried residues might be better accessible for DEPC than for NHS-acetate.

### **Comparison of the two labeling approaches**

Labeling of model proteins with NHS-acetate and DEPC resulted in comparable results in most aspects (see above). Nonetheless, there are clear differences between the two labeling strategies. Drawbacks of DEPC labeling are reversible modifications and sensitivity to acids, which are often applied in MS experiments. In addition, heat-induced degradation during sample preparation occurs. Therefore, adjusted sample preparation protocols had to be employed and sample storage was avoided after labeling. Furthermore, DEPC showed a limited solubility in water making the labeling of proteins and protein complexes difficult. NHS-acetate labeling, in comparison, is straightforward and allows the application of various sample preparation protocols including sample storage after labeling. In addition, acetylation of lysine is a commonly identified PTM and, therefore, implemented in most search engines as variable modification. The in nature rarely occurring acetylation of hydroxyl groups of serine, threonine and tyrosine residues can be easily added to the search parameters enabling their identification in several search engines. For identification of DEPC modified residues, the database search setup is more complex. Importantly, the observed neutral loss of DEPC labeled peptides should be included during database searching. In previous studies, the neutral loss was not included [196, 205, 295, 296, 347, 348] and addition might improve assignment of modified residues. When modified residues are identified, the quantification of labeling percentage of each residue can be applied to peptides modified either by DEPC or NHS-acetate.

In comparison to other typically employed labeling strategies such as HRF or HDX, chemical labeling has several advantages: First, covalent chemical labeling does not require specialized equipment and standard proteomic analysis tools are sufficient. In particular, for most labeling reagents, standard workflows and instrumentation are applicable. Second, the variety of existing labeling reagents enables specific targeting of different amino acid side chains. The selection of labeling reagent can be adapted to the protein of interest and its amino acid composition. Third, depending on the labeling reagent, the introduced modification is stable (i.e., no back-exchange) and does not undergo scrambling. Chemical labeling is, therefore, well suited to identify solvent accessible amino acid residues of proteins and protein complexes.

#### **4.4 Labeling of synaptic vesicle proteins**

The identification of solvent exposed protein domains of synaptic vesicle proteins by chemical labeling relies on the identification of proteins and their modifications using MS and the presence of intact synaptic vesicles. The prerequisite of the presence of intact synaptic vesicles was confirmed for the synaptic vesicle purification strategy applied in this study by negative stain electron microscopy, the typical observed synaptic vesicle protein composition, a Botulinium neurotoxin B cleavage assay and a fusion assay [294].

Synaptic vesicles are more complex in their protein composition than the used model protein complexes [294, 297, 298]. They include, for instance, high abundant synaptic vesicle proteins such as synaptobrevin-2 (~70 copies per vesicle) or synaptophysin (~32 copies per vesicle) and at the same time low abundant proteins such as secretory carrier-associated membrane protein 1 (~1 copy per vesicle) [298]. In general, high abundant proteins and protein modifications within a protein mixture are more likely to be identified in MS experiments. Low abundant modified peptides might not be detected and identified and are, therefore, absent in the obtained results. Nonetheless, following the labeling strategy using S-NHS-acetate, NHS-acetate and DEPC modified residues of low, intermediate and high abundant proteins were identified. Modifications of rarely observed proteins (for instance RGD1305455 with ~0.1 copies per vesicle) were, however, not identified. Note that for identification of modifications of rarely observed proteins enrichment strategies should be considered.

#### **Labeling of the ATPase complex**

The V-ATPase represents an essential multi-subunit protein complex of synaptic vesicles. Its main function is the transport of protons across the membrane into the lumen of the synaptic vesicle. This proton gradient is then utilized by specific transporters to refill synaptic vesicles with neurotransmitter molecules [349]. For this active transport of protons across the membrane, the V-ATPase requires the full assembly of the 'V1' and 'V0' domains. The soluble 'V1' domain hydrolyzes ATP resulting in conformational changes and further initiating rotation of the c-ring essential for proton transport. Another central component of the proton transport is subunit 'a1'. This subunit contains a cytosol-facing half-channel for proton entry and another luminal-facing half-channel [299]. Upon entry through the cytosol-facing half-channel, binding of the protons to one 'c' subunit occurs. After one full rotation of the c-ring, exit of protons through the luminal-facing channel into

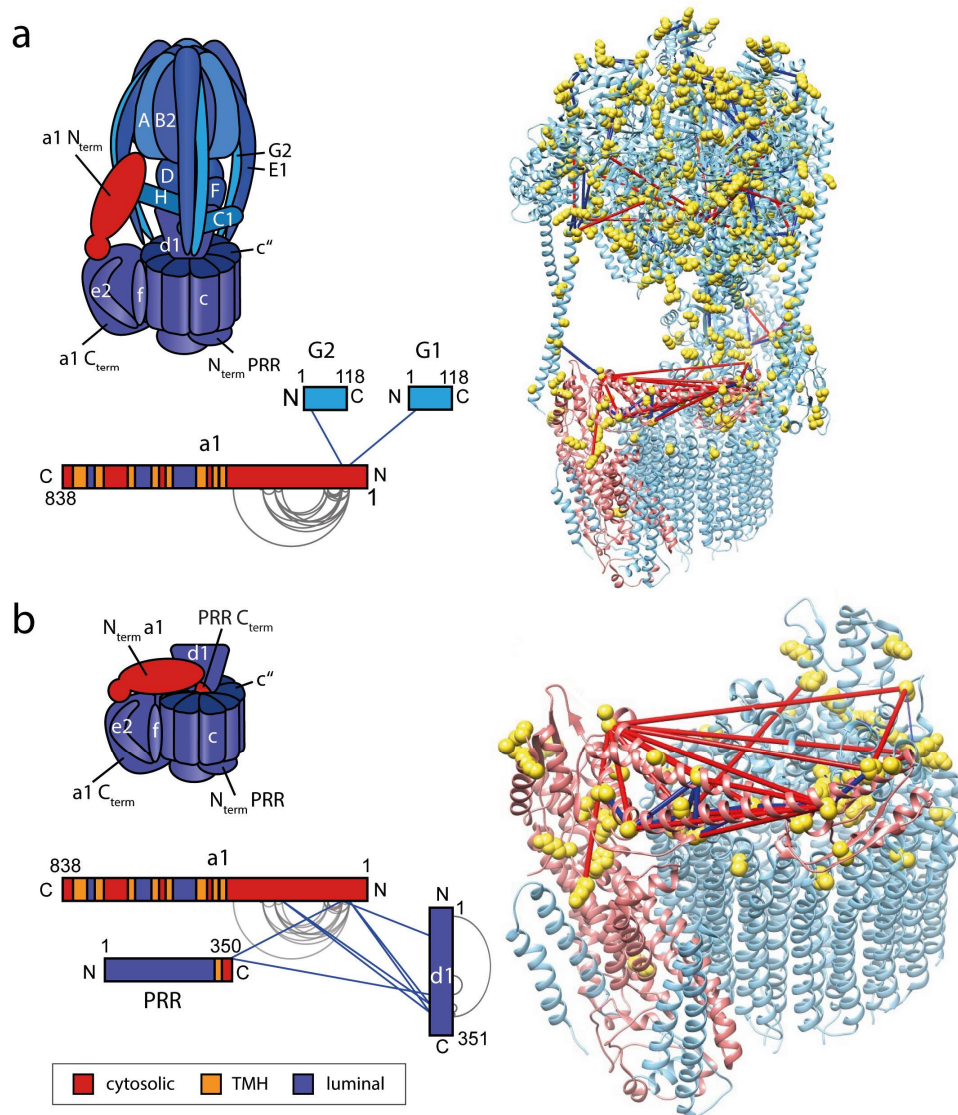
the synaptic vesicle lumen is enabled. Because of its important role for synaptic vesicle function, labeling of the V-ATPase was analyzed in more detail.

The labeling strategy presented here resulted in successful identification of many modified residues using S-NHS-acetate, NHS-acetate and DEPC. As expected, solvent exposed residues of the 'V1' and 'V0' domain were labeled. Previous studies reported dissociation of subunit 'V1' and 'V0' [226, 299]. Vasanthakuma et al. observed conformational changes during reversible dissociation of yeast V-ATPase upon glucose depletion [350]. Accordingly, the 'V1' and 'V0' domains separate and conformational changes of subunit 'H', 'C1', the peripheral stalks of 'V1' and subunit 'a1' of 'V0' prevent spontaneous reassembly [350]. In addition, dissociation of subunit 'C1' was observed. Notably, addition of glucose induced RAVE-mediated reassembly of 'V1', 'V0' as well as subunit 'C1' [350].

In synaptic vesicles, V-ATPase association and dissociation plays an important role. Poëa-Guyon et al. showed that intact V-ATPase complexes prevent vesicle fusion while pharmacologically induced dissociation of 'V0' and 'V1' results in exposure of 'V0' and functional exocytosis [351]. However, the explicit function of 'V0'-mediated, acidification-independent exocytosis is unknown. The results of this thesis revealed a high number of modified residues within the 'V0' domain indicating dissociation of the complex and highlighting solvent exposure. In particular subunit 'a1' was found to be highly modified by the labeling reagents. This subunit undergoes a large conformational change upon disassembly [299, 352]. Furthermore, flexibility of subunit 'a1' has been indicated by ion mobility and native MS experiments [300, 353]. In addition, Zhou et al. investigated the effect of both low pH and low ATP-concentrations and observed dissociation of subunit 'a1' from 'V0' [353]. This dissociation is further supported by cryo-electron microscopy revealing a lack of interactions between subunit 'a1' and the membrane-embedded c-ring [354]. The authors, therefore, assume that subunit 'a1' can move away from the c-ring and might play a regulatory role in sensing both proton and ATP level [353]. The reported flexibility and accessibility of subunit 'a1' was also described using cross-linking experiments of synaptic vesicles (**Figure 40**) [226]. These experiments included cross-links that are in agreement with the fully assembled V-ATPase structure (**Figure 40a**) and cross-links that more likely correspond to the disassembled 'V0' domain (**Figure 40b**) [226]. Specifically, the observed long distance cross-links involving subunit 'a1' indicate structural dynamics [226]. Residues involved in cross-links within subunit 'a1' and in close proximity



to cross-linking sites were modified. Therefore, both labeling and cross-linking experiments suggest flexibility of subunit ‘a1’ in the analyzed synaptic vesicles.



**Figure 40: Solvent accessibility and protein interactions in the V-ATPase complex.**

Synaptic vesicle proteins were cross-linked using BS3 or chemically labeled using S-NHS-acetate, NHS-acetate or DEPC. The cartoon models of the V-ATPase include the indicated subunits. The bar-plots show inter-molecular cross-links identified in one (grey) or at least two (blue) biological replicates. The bar length corresponds to the protein length. N- and C-terminus are indicated. The high-resolution structure of the V-ATPase is shown in cartoon-representation (PDB: 6vq6 [299]). Subunit ‘a1’ is highlighted (red). Residues labeled with at least one chemical labeling reagent are shown in surface representation (yellow). Cross-linked amino acid residues showed distances  $\leq 30$  Å (blue) or  $> 30$  Å (red). **a)** Fully assembled V-ATPase **b)** ‘V0’ domain. Abbreviations: TMH, trans-membrane helix. (Cross-linking of synaptic vesicle proteins with BS3 was performed by Sabine Wittig. Cartoon models of the V-ATPase and bar-plots were provided by Sabine Wittig. (The Figure was adapted from Wittig et al. 2021 [294].)

Importantly, modification of lysine K537 of subunit ‘a1’ located within the transmembrane region was also identified. Distinct binding sites of V-ATPase inhibitors have recently been determined in high-resolution structures of ‘V0’ from *Saccharomyces cerevisiae* [355].

These included bafilomycin A1, concanamycin A and archazolid A which were found to bind subunit 'c' within the transmembrane region suggesting a diversity of available binding sites [355]. The labeling data provides evidence for accessibility of subunit 'c' through the proton-half channel of subunit 'a1' and, furthermore, accessibility of the channel for the inhibitors. The channel might, therefore, represent a target for future drug designs.

The availability of V-ATPase inhibitors suggests that in future studies, labeling and cross-linking experiments of a specific V-ATPase conformation could be applied. For this, locking of V-ATPase by specific inhibitors represents an exciting opportunity. Solvent accessibility of proteins and protein interactions might change and reveal structural changes that are linked to specific functions of the V-ATPase and importantly of the 'V0' domain.

### **Labeling of cytosolic, transmembrane and vesicular domains of synaptic vesicle proteins – general aspects**

For the three chemical labeling reagents S-NHS-acetate, NHS-acetate and DEPC, labeling of solvent exposed residues located in cytoplasmic domains of synaptic vesicle proteins was predominantly identified. Synaptic vesicles are densely packed with proteins and have only a very limited solvent accessible area of their lipid bilayer.

A quantitative labeling approach was applied to compared labeling with S-NHS-acetate to NHS-acetate expecting a higher labeling percentage of luminal residues when using NHS-acetate. However, residues were not found to be labeled to a higher extent with NHS-acetate indicating that most of NHS-acetate already reacted with accessible, cytosolic residues. In addition, reactions of NHS-acetate for instance with the lipid phosphatidylethanolamine, which contains an amine containing head group and contributes to 20% of the lipid composition of synaptic vesicles, are possible [298, 356, 357]. Therefore, quenching of labeling reagents by reactions with lipids might occur and consequently reduce the concentration of labeling reagents resulting in labeling of cytosolic protein domains of synaptic vesicle proteins. If this is the case also for membrane mimetic systems such as proteo-liposomes comprising a lower number of proteins per vesicle and, therefore, providing access to the lipid membrane should be addressed in future studies.

When DEPC was used as labeling reagent, similar protein domains as with S-NHS-acetate and NHS-acetate were modified suggesting that DEPC is not membrane permeable. In a recent study, DEPC has been applied for modification of the transmembrane domain of the vitamin K epoxide reductase [358]. For this, vitamin K epoxide reductase was solubilized in detergent micelles and labeling of residues in the transmembrane domain was enabled by tip

sonication [358]. In contrast to sonication in a water bath, the probe of the ultrasonic processor was immersed into the solution. The findings of Guo et al. support the assumption that DEPC does not pass the membrane when applying the experimental conditions of this thesis. In future experiments, tip-sonication might represent a novel labeling approach to modify transmembrane and luminal residues of transmembrane proteins by providing entry to the lumen of the vesicles. However, exposure to shearing forces during sonication might disrupt protein complexes, as suggested previously [359]. Therefore, integrity of the synaptic vesicles upon sonication should be reevaluated by the methods described above.

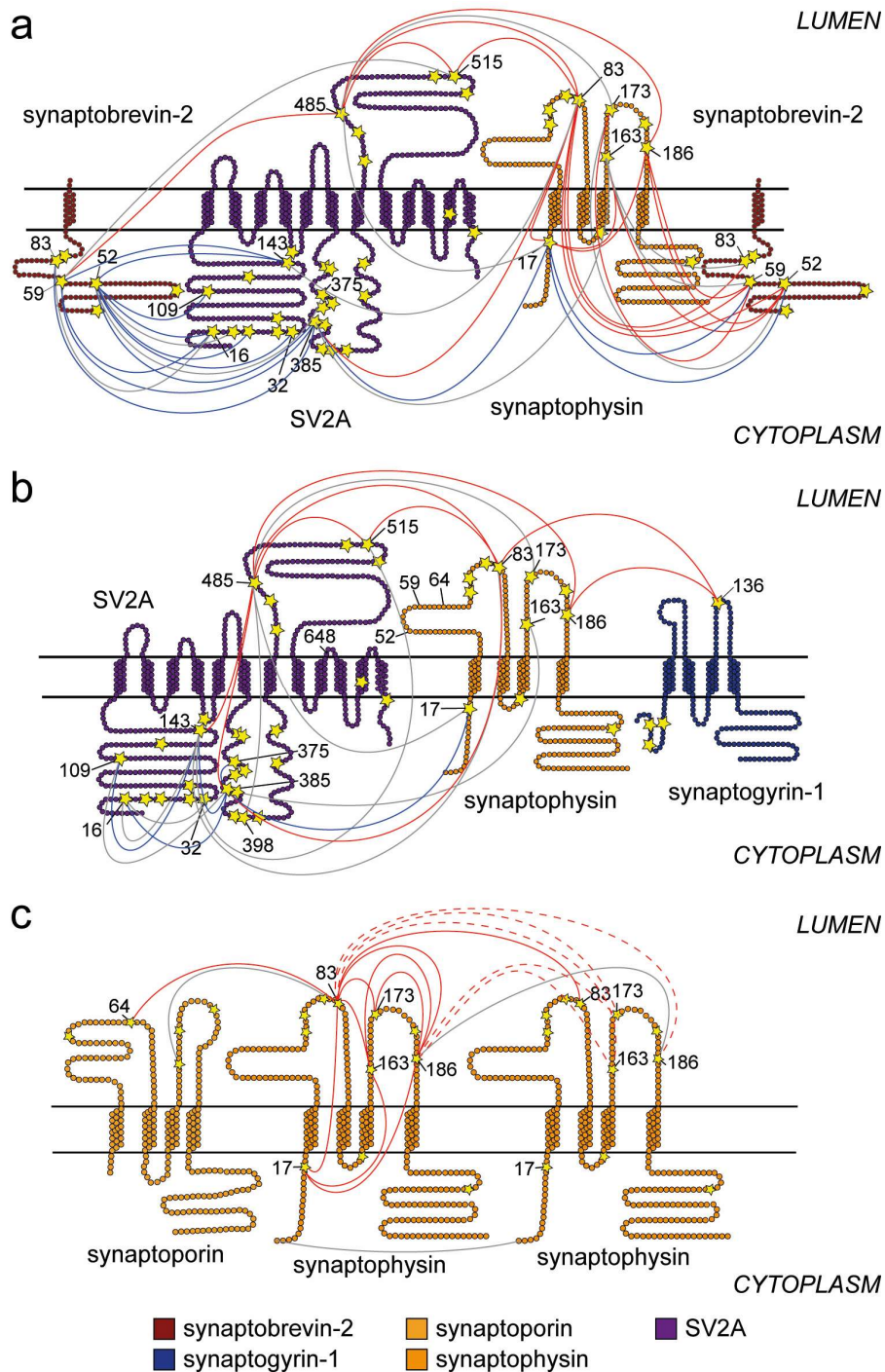
When transmembrane regions are targeted, some properties of membrane proteins should be considered. For identification of proteins and modified residues using MS, peptide generation by enzymatic hydrolysis is required. The protease trypsin, which is cleaving C-terminal of lysine and arginine residues, is in most cases applied. However, transmembrane regions usually only contain a limited number of tryptic cleavage sites. Therefore, alternative proteases such as chymotrypsin, elastase, proteinase 3 or sequential digestion using trypsin and additional alternative proteases might result in a higher number of identified peptides [360, 361]. Another challenge for labeling of transmembrane regions is the availability of target residues for the modification as most labeling reagents primarily target polar or charged amino acid residues; hydrophobic residues are often not targeted. Some unspecific labeling reagents have been described including carbenes that are generated by exposure to UV light of diazo compound or diazirine containing reagents [362]. These could potentially be employed for labeling of transmembrane regions, however the analysis of unspecific labeling is rather complex. In addition, as already indicated above, quenching by lipid molecules as for S-NHS-acetate, NHS-acetate and DEPC is possible (T. Hofmann, L. Steen, M. Frick and C. Schmidt, unpublished work).

The labeling results of cytosolic synaptic vesicle proteins identify solvent exposed residues. The three-dimensional structures of synapsin-1, Rab3a and CSP were predicted using AlphaFold [229]. All three proteins structures include regions with low structure prediction scores. However, the absence of modified residues indicated protein domains shielded from labeling such as lipid binding sites or domains involved in protein interactions.

In summary, the presented labeling workflow of synaptic vesicle proteins is especially suited for identification of solvent accessible cytosolic protein domains.

### **Exposure of luminal loops of tetra-spanning and related proteins**

The tetra-spanning proteins synaptophysin, synaptoporin and synaptogyrin-1 as well as the integral membrane protein SV2A containing 12 transmembrane domains are abundant synaptic vesicle proteins. Surprisingly, close inspection revealed not only modification of cytosolic residues but also of luminal residues (**Figure 41**). In cross-linking experiments using the non-membrane permeable cross-linking reagent BS3, interactions between these proteins were observed [226]. This included extensive cross-linking through both luminal and cytosolic loops (**Figure 41**) [226]. Specifically, interactions of the luminal loop of SV2A (residues 469 – 598) and the two luminal loops of synaptophysin, with each other and with the cytosolic domain of synaptobrevin-2 were identified (**Figure 41a**) [226]. In addition, cross-links between the luminal loops of synaptophysin and lysine 136 located in the luminal loop of synaptogyrin (**Figure 41b**) and luminal cross-linking of the loops of synaptophysin and synaptoporin (**Figure 41c**) were observed [226]. Nearly all residues that were involved in these cross-links were also found to be labeled (**Figure 41**).



**Figure 41: Protein interactions identified in flexible luminal loops.**

Synaptobrevin-2 (dark red), synaptoporin/synaptophysin (light orange/ orange), SV2A (purple) and synaptogyrin-1 (dark blue) are shown as topology models. Luminal (LUMEN) and cytoplasmic (CYTOPLASM) membrane sides are assigned. Amino acid residues labeled with at least one chemical labeling reagent are indicated (yellow stars). Cross-links are indicated by solid and dashed lines (across the membrane and luminal, red; cytosolic, blue; interactions identified in only one replicate, grey). Lysine residues are labeled with residue numbers. **a**) Interactions and modified amino acid residues of synaptobrevin-2, SV2A and synaptophysin. **b**) Interactions and modified amino acid residues of SV2A, synaptophysin and synaptogyrin-1. **c**) Interactions and modified amino acid residues of synaptophysin and synaptoporin. Cross-links that support oligomerisation of synaptophysin are shown as dashed lines. (The figure was provided by Sabine Wittig and is published in Sabine Wittig, Marcelo Ganzella, **Marie Barth**, Susann Kostmann, Dietmar Riedel, Ángel Pérez-Lara, Reinhard Jahn & Carla Schmidt, Cross-linking mass spectrometry uncovers protein interactions and functional assemblies in synaptic vesicle membranes. *Nature Communications* **12**, 858 (2021) [294].)

Possible disruption of synaptic vesicles resulting in accessibility of luminal loops was ruled out (see above). Therefore, the question whether luminal loops are accessible for both labeling and cross-linking reagent is possible remains. There are two hypothesis suggesting structural flexibility of these proteins resulting in integration of luminal loops in the vesicular membrane.

First, the formation of pores by synaptophysin as one hypothesis which will be discussed. For synaptophysin reconstituted into liposomes, the formation of a hexameric homooligomeric channel has been observed [363]. Arthur et al. further resolved its three-dimensional structure using electron microscopy and single-particle 3-D reconstruction [364]. When, in addition to synaptophysin, synaptobrevin-2 was added, an assembly into a hexameric ring in which six synaptophysin molecules bind six synaptobrevin dimers is observed [365]. These findings show the ability of synaptophysin to form pores and suggest that large and dynamic loops might line the inner cavity of the pore and making them solvent accessible. This might also be the case for the similar tetra-spanning proteins. Cross-links between isoforms of tetra-spanning proteins (synptophysin to synpatoporin) further suggests that mixed protein pores occur. However, there is a strong debate whether pores do only occur in reconstituted systems and are not present in synaptic vesicles.

Second, structural changes of proteins resulting in an increase in size of synaptic vesicles is the currently more favored hypothesis. Synaptic vesicles were found to reversibly increase in size upon neurotransmitter loading [309]. This includes an increase in diameter by ~25% corresponding to an increase in surface area of ~50% and in volume of ~100% [309]. Vesicles omitting SV2A did not increase in size [309]. Therefore, the authors suggest a major role of SV2A including conformational changes that enable the increase in size [309]. The conformational flexibility of SV2A is further supported by the observation of two SV2A conformations [366]. However, conformational changes of SV2A alone are not sufficient to describe the mechanism of size increase [309]. The labeling and the cross-linking data support the assumption, that at least some SV2A proteins undergo structural changes resulting in exposure of luminal domains. The tetra-spanning proteins likely contribute by their dynamic structures to the increase in size of synaptic vesicles. In conclusion, the results of this thesis show that the labeling workflow together with cross-linking of synaptic vesicle proteins unravel structural properties of proteins with unknown function.

## References

1. Lodish, H., et al., *Overview of neuron structure and function*. Molecular Cell Biology, 2000. **4**.
2. Ponce, A., et al., *G-protein-gated inward rectifier K<sup>+</sup> channel proteins (GIRK1) are present in the soma and dendrites as well as in nerve terminals of specific neurons in the brain*. J Neurosci, 1996. **16**(6): p. 1990-2001.
3. Panzanelli, P., et al., *Distinct mechanisms regulate GABAA receptor and gephyrin clustering at perisomatic and axo-axonic synapses on CA1 pyramidal cells*. J Physiol, 2011. **589**(Pt 20): p. 4959-80.
4. Dascal, N., *Ion-channel regulation by G proteins*. Trends Endocrinol Metab, 2001. **12**(9): p. 391-8.
5. Climer, L.K., R.D. Hendrix, and V.V. Lupashin, *Conserved Oligomeric Golgi and Neuronal Vesicular Trafficking*. Handb Exp Pharmacol, 2018. **245**: p. 227-247.
6. Grider, M.H., R. Jessu, and R. Kabir, *Physiology, Action Potential*, in StatPearls. 2021, StatPearls Publishing LLC: Treasure Island (FL).
7. Hirokawa, N., et al., *Brain dynein (MAP1C) localizes on both anterogradely and retrogradely transported membranous organelles in vivo*. J Cell Biol, 1990. **111**(3): p. 1027-37.
8. Hirokawa, N., et al., *Kinesin associates with anterogradely transported membranous organelles in vivo*. J Cell Biol, 1991. **114**(2): p. 295-302.
9. Bean, B.P., *The action potential in mammalian central neurons*. Nat Rev Neurosci, 2007. **8**(6): p. 451-65.
10. Südhof, T.C., *Calcium control of neurotransmitter release*. Cold Spring Harb Perspect Biol, 2012. **4**(1): p. a011353.
11. Vos, M., E. Lauwers, and P. Verstreken, *Synaptic mitochondria in synaptic transmission and organization of vesicle pools in health and disease*. Front Synaptic Neurosci, 2010. **2**: p. 139.
12. Edmondson, D.E., et al., *Molecular and mechanistic properties of the membrane-bound mitochondrial monoamine oxidases*. Biochemistry, 2009. **48**(20): p. 4220-30.
13. Becker, N.H., et al., *The localization of enzyme activities in the rat brain*. J Biophys Biochem Cytol, 1960. **8**(3): p. 649-63.
14. Parton, R.G., K. Simons, and C.G. Dotti, *Axonal and dendritic endocytic pathways in cultured neurons*. J Cell Biol, 1992. **119**(1): p. 123-37.
15. Gorenstein, C., et al., *Dendritic transport. I. Colchicine stimulates the transport of lysosomal enzymes from cell bodies to dendrites*. J Neurosci, 1985. **5**(8): p. 2009-17.
16. Goo, M.S., et al., *Activity-dependent trafficking of lysosomes in dendrites and dendritic spines*. J Cell Biol, 2017. **216**(8): p. 2499-2513.
17. Davis, K.M., et al., *Regulated lysosomal trafficking as a mechanism for regulating GABAA receptor abundance at synapses in Caenorhabditis elegans*. Mol Cell Neurosci, 2010. **44**(4): p. 307-17.
18. Padamsey, Z., et al., *Activity-Dependent Exocytosis of Lysosomes Regulates the Structural Plasticity of Dendritic Spines*. Neuron, 2017. **93**(1): p. 132-146.
19. Bott, C.J., et al., *Nestin in immature embryonic neurons affects axon growth cone morphology and Semaphorin3a sensitivity*. Mol Biol Cell, 2019. **30**(10): p. 1214-1229.
20. Gasnier, B., *The SLC32 transporter, a key protein for the synaptic release of inhibitory amino acids*. Pflugers Arch, 2004. **447**(5): p. 756-9.
21. Sagné, C., et al., *Cloning of a functional vesicular GABA and glycine transporter by screening of genome databases*. FEBS Lett, 1997. **417**(2): p. 177-83.
22. Parsons, S.M., *Transport mechanisms in acetylcholine and monoamine storage*. Faseb j, 2000. **14**(15): p. 2423-34.

23. Danbolt, N.C., *Glutamate uptake*. Prog Neurobiol, 2001. **65**(1): p. 1-105.
24. Mukherjee, K., et al., *Piccolo and bassoon maintain synaptic vesicle clustering without directly participating in vesicle exocytosis*. Proc Natl Acad Sci U S A, 2010. **107**(14): p. 6504-9.
25. Coleman, W.L., C.A. Bill, and M. Bykhovskaia, *Rab3a deletion reduces vesicle docking and transmitter release at the mouse diaphragm synapse*. Neuroscience, 2007. **148**(1): p. 1-6.
26. Zarebidaki, F., et al., *Disentangling the Roles of RIM and Munc13 in Synaptic Vesicle Localization and Neurotransmission*. J Neurosci, 2020. **40**(49): p. 9372-9385.
27. Söllner, T., et al., *SNAP receptors implicated in vesicle targeting and fusion*. Nature, 1993. **362**(6418): p. 318-24.
28. Jorquera, R.A., et al., *Complexin controls spontaneous and evoked neurotransmitter release by regulating the timing and properties of synaptotagmin activity*. J Neurosci, 2012. **32**(50): p. 18234-45.
29. Neher, E. and R. Penner, *Mice sans synaptotagmin*. Nature, 1994. **372**(6504): p. 316-7.
30. Barker, L.A., M.J. Dowdall, and V.P. Whittaker, *Choline metabolism in the cerebral cortex of guinea pigs. Stable-bound acetylcholine*. Biochem J, 1972. **130**(4): p. 1063-75.
31. Ceccarelli, B., W.P. Hurlbut, and A. Mauro, *Turnover of transmitter and synaptic vesicles at the frog neuromuscular junction*. J Cell Biol, 1973. **57**(2): p. 499-524.
32. Heuser, J.E. and T.S. Reese, *Evidence for recycling of synaptic vesicle membrane during transmitter release at the frog neuromuscular junction*. J Cell Biol, 1973. **57**(2): p. 315-44.
33. Clayton, E.L., G.J. Evans, and M.A. Cousin, *Bulk synaptic vesicle endocytosis is rapidly triggered during strong stimulation*. J Neurosci, 2008. **28**(26): p. 6627-32.
34. Watanabe, S., et al., *Ultrafast endocytosis at mouse hippocampal synapses*. Nature, 2013. **504**(7479): p. 242-247.
35. Lehoullier, P.F. and M.K. Ticku, *The pharmacological properties of GABA receptor-coupled chloride channels using <sup>36</sup>Cl-influx in cultured spinal cord neurons*. Brain Res, 1989. **487**(2): p. 205-14.
36. Biedler, J.L., et al., *Multiple neurotransmitter synthesis by human neuroblastoma cell lines and clones*. Cancer Res, 1978. **38**(11 Pt 1): p. 3751-7.
37. Biedler, J.L., L. Helson, and B.A. Spengler, *Morphology and growth, tumorigenicity, and cytogenetics of human neuroblastoma cells in continuous culture*. Cancer Res, 1973. **33**(11): p. 2643-52.
38. Barr, E.K. and M.A. Applebaum, *Genetic Predisposition to Neuroblastoma*. Children (Basel), 2018. **5**(9).
39. Berthold, F., et al., *Incidence, Survival, and Treatment of Localized and Metastatic Neuroblastoma in Germany 1979-2015*. Paediatr Drugs, 2017. **19**(6): p. 577-593.
40. Xie, H.R., L.S. Hu, and G.Y. Li, *SH-SY5Y human neuroblastoma cell line: in vitro cell model of dopaminergic neurons in Parkinson's disease*. Chin Med J (Engl), 2010. **123**(8): p. 1086-92.
41. Kovalevich, J. and D. Langford, *Considerations for the use of SH-SY5Y neuroblastoma cells in neurobiology*. Methods Mol Biol, 2013. **1078**: p. 9-21.
42. Murillo, J.R., et al., *Quantitative proteomic analysis identifies proteins and pathways related to neuronal development in differentiated SH-SY5Y neuroblastoma cells*. EuPA Open Proteom, 2017. **16**: p. 1-11.
43. Clagett-Dame, M., E.M. McNeill, and P.D. Muley, *Role of all-trans retinoic acid in neurite outgrowth and axonal elongation*. J Neurobiol, 2006. **66**(7): p. 739-56.
44. Uemura, K., et al., *Presenilin 1 mediates retinoic acid-induced differentiation of SH-SY5Y cells through facilitation of Wnt signaling*. J Neurosci Res, 2003. **73**(2): p. 166-75.
45. Kim, S.N., et al., *Participation of type II protein kinase A in the retinoic acid-induced growth inhibition of SH-SY5Y human neuroblastoma cells*. J Cell Physiol, 2000. **182**(3): p. 421-8.



46. Peinemann, F., et al., *Retinoic acid postconsolidation therapy for high-risk neuroblastoma patients treated with autologous haematopoietic stem cell transplantation*. Cochrane Database Syst Rev, 2017. **8**(8): p. Cd010685.
47. Goldie, B.J., M.M. Barnett, and M.J. Cairns, *BDNF and the maturation of posttranscriptional regulatory networks in human SH-SY5Y neuroblast differentiation*. Front Cell Neurosci, 2014. **8**: p. 325.
48. Kume, T., et al., *Dibutyl cyclic AMP induces differentiation of human neuroblastoma SH-SY5Y cells into a noradrenergic phenotype*. Neurosci Lett, 2008. **443**(3): p. 199-203.
49. Prince, J.A. and L. Orelund, *Staurosporine differentiated human SH-SY5Y neuroblastoma cultures exhibit transient apoptosis and trophic factor independence*. Brain Res Bull, 1997. **43**(6): p. 515-23.
50. Presgraves, S.P., et al., *Terminally differentiated SH-SY5Y cells provide a model system for studying neuroprotective effects of dopamine agonists*. Neurotox Res, 2004. **5**(8): p. 579-98.
51. Katsetos, C.D., M.M. Herman, and S.J. Mörk, *Class III beta-tubulin in human development and cancer*. Cell Motil Cytoskeleton, 2003. **55**(2): p. 77-96.
52. Zhang, T., S.P. Gygi, and J.A. Paulo, *Temporal Proteomic Profiling of SH-SY5Y Differentiation with Retinoic Acid Using FAIMS and Real-Time Searching*. J Proteome Res, 2021. **20**(1): p. 704-714.
53. Yamashita, M. and J.B. Fenn, *Electrospray ion source. Another variation on the free-jet theme*. The Journal of Physical Chemistry, 1984. **88**(20): p. 4451-4459.
54. Smith, R.D., et al., *New developments in biochemical mass spectrometry: electrospray ionization*. Anal Chem, 1990. **62**(9): p. 882-99.
55. Wu, X., R.D. Oleschuk, and N.M. Cann, *Characterization of microstructured fibre emitters: in pursuit of improved nano electrospray ionization performance*. Analyst, 2012. **137**(18): p. 4150-61.
56. Rayleigh, L., *XX. On the equilibrium of liquid conducting masses charged with electricity*. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1882. **14**(87): p. 184-186.
57. Iribarne, J.V. and B.A. Thomson, *On the evaporation of small ions from charged droplets*. The Journal of Chemical Physics, 1976. **64**(6): p. 2287-2294.
58. Michalski, A., et al., *Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer*. Mol Cell Proteomics, 2011. **10**(9): p. M111.011015.
59. Wolff, M.M. and W.E. Stephens, *A Pulsed Mass Spectrometer with Time Dispersion*. Review of Scientific Instruments, 1953. **24**(8): p. 616-617.
60. Makarov, A., *Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis*. Anal Chem, 2000. **72**(6): p. 1156-62.
61. Allen, J.S., *An Improved Electron Multiplier Particle Counter*. Review of Scientific Instruments, 1947. **18**(10): p. 739-749.
62. Tremsin, A.S., et al., *High-Resolution Strain Mapping Through Time-of-Flight Neutron Transmission Diffraction with a Microchannel Plate Neutron Counting Detector*. Strain, 2012. **48**(4): p. 296-305.
63. Glish, G.L. and D.J. Burinsky, *Hybrid mass spectrometers for tandem mass spectrometry*. J Am Soc Mass Spectrom, 2008. **19**(2): p. 161-72.
64. Jennings, K.R., *Collision-induced decompositions of aromatic molecular ions*. International Journal of Mass Spectrometry and Ion Physics, 1968. **1**(3): p. 227-235.
65. Olsen, J.V., et al., *Higher-energy C-trap dissociation for peptide modification analysis*. Nat Methods, 2007. **4**(9): p. 709-12.
66. Syka, J.E., et al., *Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry*. Proc Natl Acad Sci U S A, 2004. **101**(26): p. 9528-33.

67. Zubarev, R.A., N.L. Kelleher, and F.W. McLafferty, *Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process*. Journal of the American Chemical Society, 1998. **120**(13): p. 3265-3266.
68. Miller, P.E. and M.B. Denton, *The quadrupole mass filter: basic operating concepts*. Journal of chemical education, 1986. **63**(7): p. 617.
69. Makarov, A., et al., *Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer*. Anal Chem, 2006. **78**(7): p. 2113-20.
70. Cox, J., et al., *Andromeda: a peptide search engine integrated into the MaxQuant environment*. J Proteome Res, 2011. **10**(4): p. 1794-805.
71. Eng, J.K., A.L. McCormack, and J.R. Yates, *An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database*. J Am Soc Mass Spectrom, 1994. **5**(11): p. 976-89.
72. Craig, R. and R.C. Beavis, *TANDEM: matching proteins with tandem mass spectra*. Bioinformatics, 2004. **20**(9): p. 1466-7.
73. Elias, J.E. and S.P. Gygi, *Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry*. Nat Methods, 2007. **4**(3): p. 207-14.
74. Danilova, Y., et al., *Bias in False Discovery Rate Estimation in Mass-Spectrometry-Based Peptide Identification*. J Proteome Res, 2019. **18**(5): p. 2354-2358.
75. Choi, H. and A.I. Nesvizhskii, *Semisupervised model-based validation of peptide identifications in mass spectrometry-based proteomics*. J Proteome Res, 2008. **7**(1): p. 254-65.
76. Kong, A.T., et al., *MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics*. Nat Methods, 2017. **14**(5): p. 513-520.
77. Ma, B., et al., *PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry*. Rapid Commun Mass Spectrom, 2003. **17**(20): p. 2337-42.
78. Kalb, S.R., et al., *De novo subtype and strain identification of botulinum neurotoxin type B through toxin proteomics*. Anal Bioanal Chem, 2012. **403**(1): p. 215-26.
79. Keller, A., et al., *Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search*. Anal Chem, 2002. **74**(20): p. 5383-92.
80. Medzihradzsky, K.F. and R.J. Chalkley, *Lessons in de novo peptide sequencing by tandem mass spectrometry*. Mass Spectrom Rev, 2015. **34**(1): p. 43-63.
81. Liu, H., R.G. Sadygov, and J.R. Yates, 3rd, *A model for random sampling and estimation of relative protein abundance in shotgun proteomics*. Anal Chem, 2004. **76**(14): p. 4193-201.
82. Shevchenko, A., et al., *In-gel digestion for mass spectrometric characterization of proteins and proteomes*. Nat Protoc, 2006. **1**(6): p. 2856-60.
83. Guilpain, P., et al., *A combined SDS-PAGE and proteomics approach to identify target autoantigens in healthy individuals and patients with autoimmune diseases*. Ann N Y Acad Sci, 2007. **1109**: p. 538-49.
84. Doellinger, J., et al., *Sample Preparation by Easy Extraction and Digestion (SPEED) - A Universal, Rapid, and Detergent-free Protocol for Proteomics Based on Acid Extraction*. Mol Cell Proteomics, 2020. **19**(1): p. 209-222.
85. Wiśniewski, J.R., et al., *Universal sample preparation method for proteome analysis*. Nat Methods, 2009. **6**(5): p. 359-62.
86. Hughes, C.S., et al., *Single-pot, solid-phase-enhanced sample preparation for proteomics experiments*. Nat Protoc, 2019. **14**(1): p. 68-85.
87. Oda, Y., et al., *Accurate quantitation of protein expression and site-specific phosphorylation*. Proc Natl Acad Sci U S A, 1999. **96**(12): p. 6591-6.
88. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics*. Mol Cell Proteomics, 2002. **1**(5): p. 376-86.

89. Milner, E., et al., *The turnover kinetics of major histocompatibility complex peptides of human cancer cells*. Mol Cell Proteomics, 2006. **5**(2): p. 357-65.
90. Lam, Y.W., et al., *Analysis of nucleolar protein dynamics reveals the nuclear degradation of ribosomal proteins*. Curr Biol, 2007. **17**(9): p. 749-60.
91. Geiger, T., et al., *Super-SILAC mix for quantitative proteomics of human tumor tissue*. Nat Methods, 2010. **7**(5): p. 383-5.
92. Thompson, A., et al., *Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS*. Anal Chem, 2003. **75**(8): p. 1895-904.
93. Wiese, S., et al., *Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research*. Proteomics, 2007. **7**(3): p. 340-50.
94. Zhang, R., et al., *Fractionation of isotopically labeled peptides in quantitative proteomics*. Anal Chem, 2001. **73**(21): p. 5142-9.
95. Beynon, R.J., et al., *Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides*. Nat Methods, 2005. **2**(8): p. 587-9.
96. Kito, K., et al., *A synthetic protein approach toward accurate mass spectrometric quantification of component stoichiometry of multiprotein complexes*. J Proteome Res, 2007. **6**(2): p. 792-800.
97. Gerber, S.A., et al., *Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS*. Proc Natl Acad Sci U S A, 2003. **100**(12): p. 6940-5.
98. Powell, D.W., et al., *Cluster analysis of mass spectrometry data reveals a novel component of SAGA*. Molecular and cellular biology, 2004. **24**(16): p. 7249-7259.
99. McAfee, K.J., et al., *Analyzing proteomes and protein function using graphical comparative analysis of tandem mass spectrometry results*. Mol Cell Proteomics, 2006. **5**(8): p. 1497-513.
100. Ishihama, Y., et al., *Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein*. Molecular & cellular proteomics : MCP, 2005. **4**(9): p. 1265-1272.
101. Zybailov, B., et al., *Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae*. Journal of proteome research, 2006. **5**(9): p. 2339-2347.
102. Hoopmann, M.R., et al., *StPeter: Seamless Label-Free Quantification with the Trans-Proteomic Pipeline*. J Proteome Res, 2018. **17**(3): p. 1314-1320.
103. Chelius, D. and P.V. Bondarenko, *Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry*. J Proteome Res, 2002. **1**(4): p. 317-23.
104. Sandra, K., et al., *Highly efficient peptide separations in proteomics. Part 2: bi- and multidimensional liquid-based separation techniques*. J Chromatogr B Analyt Technol Biomed Life Sci, 2009. **877**(11-12): p. 1019-39.
105. Bian, Y., et al., *Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC-MS/MS*. Nat Commun, 2020. **11**(1): p. 157.
106. Sun, W., et al., *An analysis of protein abundance suppression in data dependent liquid chromatography and tandem mass spectrometry with tryptic peptide mixtures of five known proteins*. Eur J Mass Spectrom (Chichester), 2005. **11**(6): p. 575-80.
107. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification*. Nat Biotechnol, 2008. **26**(12): p. 1367-72.
108. Rappsilber, J., et al., *Large-scale proteomic analysis of the human spliceosome*. Genome Res, 2002. **12**(8): p. 1231-45.
109. Schwanhäusser, B., et al., *Global quantification of mammalian gene expression control*. Nature, 2011. **473**(7347): p. 337-42.
110. Kim, M.S., et al., *A draft map of the human proteome*. Nature, 2014. **509**(7502): p. 575-81.

111. Anderson, L. and C.L. Hunter, *Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins*. Mol Cell Proteomics, 2006. **5**(4): p. 573-88.
112. Peterson, A.C., et al., *Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics*. Mol Cell Proteomics, 2012. **11**(11): p. 1475-88.
113. Venable, J.D., et al., *Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra*. Nat Methods, 2004. **1**(1): p. 39-45.
114. Lange, V., et al., *Selected reaction monitoring for quantitative proteomics: a tutorial*. Mol Syst Biol, 2008. **4**: p. 222.
115. Malik, R., et al., *From proteome lists to biological impact--tools and strategies for the analysis of large MS data sets*. Proteomics, 2010. **10**(6): p. 1270-83.
116. Mehmood, S., T.M. Allison, and C.V. Robinson, *Mass spectrometry of protein complexes: from origins to applications*. Annu Rev Phys Chem, 2015. **66**: p. 453-74.
117. Konijnenberg, A., A. Butterer, and F. Sobott, *Native ion mobility-mass spectrometry and related methods in structural biology*. Biochim Biophys Acta, 2013. **1834**(6): p. 1239-56.
118. Konermann, L., *Addressing a Common Misconception: Ammonium Acetate as Neutral pH "Buffer" for Native Electrospray Mass Spectrometry*. Journal of the American Society for Mass Spectrometry, 2017. **28**(9): p. 1827-1835.
119. Dole, M., et al., *Molecular Beams of Macroions*. The Journal of Chemical Physics, 1968. **49**(5): p. 2240-2249.
120. Konermann, L., et al., *Unraveling the mechanism of electrospray ionization*. Anal Chem, 2013. **85**(1): p. 2-9.
121. Sobott, F., et al., *A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies*. Anal Chem, 2002. **74**(6): p. 1402-7.
122. Gault, J., et al., *High-resolution mass spectrometry of small molecules bound to membrane proteins*. Nat Methods, 2016. **13**(4): p. 333-6.
123. Ferrige, A.G., et al., *Disentangling electrospray spectra with maximum entropy*. Rapid Communications in Mass Spectrometry, 1992. **6**(11): p. 707-711.
124. Morgner, N. and C.V. Robinson, *Massign: an assignment strategy for maximizing information from the mass spectra of heterogeneous protein assemblies*. Anal Chem, 2012. **84**(6): p. 2939-48.
125. Marty, M.T., et al., *Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles*. Anal Chem, 2015. **87**(8): p. 4370-6.
126. Bern, M., et al., *Parsimonious Charge Deconvolution for Native Mass Spectrometry*. J Proteome Res, 2018. **17**(3): p. 1216-1226.
127. <https://www.proteinmetrics.com/products/intact-mass/>.
128. Barth, M. and C. Schmidt, *Native mass spectrometry-A valuable tool in structural biology*. J Mass Spectrom, 2020. **55**(10): p. e4578.
129. van de Waterbeemd, M., et al., *High-fidelity mass analysis unveils heterogeneity in intact ribosomal particles*. Nat Methods, 2017. **14**(3): p. 283-286.
130. Chen, Z.A., et al., *Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry*. Embo j, 2010. **29**(4): p. 717-26.
131. Greber, B.J., et al., *The complete structure of the large subunit of the mammalian mitochondrial ribosome*. Nature, 2014. **515**(7526): p. 283-6.
132. Schweppe, D.K., et al., *Mitochondrial protein interactome elucidated by chemical cross-linking mass spectrometry*. Proc Natl Acad Sci U S A, 2017. **114**(7): p. 1732-1737.
133. Staros, J.V., *N-hydroxysulfosuccinimide active esters: bis(N-hydroxysulfosuccinimide) esters of two dicarboxylic acids are hydrophilic, membrane-impermeant, protein cross-linkers*. Biochemistry, 1982. **21**(17): p. 3950-5.
134. Mattson, G., et al., *A practical approach to crosslinking*. Mol Biol Rep, 1993. **17**(3): p. 167-83.

135. Sinz, A., *Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions*. Mass Spectrometry Reviews, 2006. **25**(4): p. 663-682.
136. Tayri-Wilk, T., et al., *Mass spectrometry reveals the chemistry of formaldehyde cross-linking in structured proteins*. Nat Commun, 2020. **11**(1): p. 3128.
137. Metz, B., et al., *Identification of formaldehyde-induced modifications in proteins: reactions with model peptides*. J Biol Chem, 2004. **279**(8): p. 6235-43.
138. Toews, J., et al., *Mass spectrometric identification of formaldehyde-induced peptide modifications under in vivo protein cross-linking conditions*. Anal Chim Acta, 2008. **618**(2): p. 168-83.
139. Leitner, A., et al., *Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography*. Mol Cell Proteomics, 2012. **11**(3): p. M111.014126.
140. Rinner, O., et al., *Identification of cross-linked peptides from large sequence databases*. Nat Methods, 2008. **5**(4): p. 315-8.
141. Schnirch, L., et al., *Expanding the Depth and Sensitivity of Cross-Link Identification by Differential Ion Mobility Using High-Field Asymmetric Waveform Ion Mobility Spectrometry*. Anal Chem, 2020. **92**(15): p. 10495-10503.
142. Hoopmann, M.R., et al., *Kojak: efficient analysis of chemically cross-linked protein complexes*. J Proteome Res, 2015. **14**(5): p. 2190-8.
143. 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): p. 3404.
144. Du, X., et al., *Xlink-identifier: an automated data analysis platform for confident identifications of chemically cross-linked peptides using tandem mass spectrometry*. J Proteome Res, 2011. **10**(3): p. 923-31.
145. Mendes, M.L., et al., *An integrated workflow for crosslinking mass spectrometry*. Mol Syst Biol, 2019. **15**(9): p. e8994.
146. 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): p. 83-97.
147. Sinz, A., *Divide and conquer: cleavable cross-linkers to study protein conformation and protein-protein interactions*. Anal Bioanal Chem, 2017. **409**(1): p. 33-44.
148. Maiolica, A., et al., *Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching*. Mol Cell Proteomics, 2007. **6**(12): p. 2200-11.
149. Yu, C. and L. Huang, *Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology*. Anal Chem, 2018. **90**(1): p. 144-165.
150. Slavin, M., et al., *Targeted in situ cross-linking mass spectrometry and integrative modeling reveal the architectures of three proteins from SARS-CoV-2*. Proceedings of the National Academy of Sciences, 2021. **118**(34): p. e2103554118.
151. Xu, G. and M.R. Chance, *Radiolytic modification and reactivity of amino acid residues serving as structural probes for protein footprinting*. Anal Chem, 2005. **77**(14): p. 4549-55.
152. Xu, G. and M.R. Chance, *Hydroxyl radical-mediated modification of proteins as probes for structural proteomics*. Chem Rev, 2007. **107**(8): p. 3514-43.
153. Barth, M., et al., *Evaluation of NHS-Acetate and DEPC labelling for determination of solvent accessible amino acid residues in protein complexes*. J Proteomics, 2020. **222**: p. 103793.
154. Suckau, D., M. Mak, and M. Przybylski, *Protein surface topology-probing by selective chemical modification and mass spectrometric peptide mapping*. Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(12): p. 5630-5634.
155. Zappacosta, F., et al., *Surface topology of Minibody by selective chemical modifications and mass spectrometry*. Protein Science, 1997. **6**(9): p. 1901-1909.

156. Liu, Y., et al., *Modulation of Replication Protein A Function by Its Hyperphosphorylation-induced Conformational Change Involving DNA Binding Domain B*. The Journal of biological chemistry, 2005. **280**(38): p. 32775-32783.
157. Gao, Y. and Y. Wang, *Site-Selective Modifications of Arginine Residues in Human Hemoglobin Induced by Methylglyoxal*. Biochemistry, 2006. **45**(51): p. 15654-15660.
158. Krell, T., A.R. Pitt, and J.R. Coggins, *The use of electrospray mass spectrometry to identify an essential arginine residue in type II dehydroquinases*. FEBS Letters, 1995. **360**(1): p. 93-96.
159. Leitner, A. and W. Lindner, *Functional Probing of Arginine Residues in Proteins Using Mass Spectrometry and an Arginine-Specific Covalent Tagging Concept*. Analytical Chemistry, 2005. **77**(14): p. 4481-4488.
160. Akinsiku, O.T., E.T. Yu, and D. Fabris, *Mass spectrometric investigation of protein alkylation by the RNA footprinting probe kethoxal*. Journal of Mass Spectrometry, 2005. **40**(10): p. 1372-1381.
161. Akashi, S., et al., *Characterization of the Structural Difference between Active and Inactive Forms of the Ras Protein by Chemical Modification Followed by Mass Spectrometric Peptide Mapping*. Analytical Biochemistry, 1997. **248**(1): p. 15-25.
162. Izumi, S., et al., *Membrane Topology of Guinea Pig Cytochrome P450 17 $\alpha$  Revealed by a Combination of Chemical Modifications and Mass Spectrometry*. Biochemistry, 2003. **42**(49): p. 14663-14669.
163. Liu, H., et al., *Mass spectrometry footprinting reveals the structural rearrangements of cyanobacterial orange carotenoid protein upon light activation*. Biochimica et Biophysica Acta (BBA) - Bioenergetics, 2014. **1837**(12): p. 1955-1963.
164. Yem, A.W., et al., *Chemical modification of Interleukin-18: Biochemical characterization of a carbodiimide-catalyzed intramolecular cross-linked protein*. Journal of Protein Chemistry, 1992. **11**(6): p. 709-722.
165. Apuy, J.L., et al., *Ratiometric Pulsed Alkylation/Mass Spectrometry of the Cysteine Pairs in Individual Zinc Fingers of MRE-Binding Transcription Factor-1 (MTF-1) as a Probe of Zinc Chelate Stability*. Biochemistry, 2001. **40**(50): p. 15164-15175.
166. Maithal, K., et al., *Inhibition of plasmodium falciparum triose-phosphate isomerase by chemical modification of an interface cysteine. Electrospray ionization mass spectrometric analysis of differential cysteine reactivities*. J Biol Chem, 2002. **277**(28): p. 25106-14.
167. Whitehurst, C.B., et al., *Location and role of free cysteinyl residues in the Sindbis virus E1 and E2 glycoproteins*. J Virol, 2007. **81**(12): p. 6231-40.
168. Macpherson, L.J., et al., *Noxious compounds activate TRPA1 ion channels through covalent modification of cysteines*. Nature, 2007. **445**(7127): p. 541-545.
169. Britto, P.J., L. Knipling, and J. Wolff, *The local electrostatic environment determines cysteine reactivity of tubulin*. J Biol Chem, 2002. **277**(32): p. 29018-27.
170. Šantrůček, J., et al., *Tyrosine residues modification studied by MALDI-TOF mass spectrometry*. Biochemical and Biophysical Research Communications, 2004. **323**(4): p. 1151-1156.
171. Knock, S.L., et al., *N-acylation of Aplysia egg-laying hormone with biotin. Characterization of bioactive and inactive derivatives*. J Biol Chem, 1991. **266**(36): p. 24413-9.
172. Suckau, D., M. Mak, and M. Przybylski, *Protein surface topology-probing by selective chemical modification and mass spectrometric peptide mapping*. Proceedings of the National Academy of Sciences, 1992. **89**(12): p. 5630-5634.
173. Ehrhard, B., et al., *Chemical Modification of Recombinant HIV-1 Capsid Protein p24 Leads to the Release of a Hidden Epitope Prior to Changes of the Overall Folding of the Protein*. Biochemistry, 1996. **35**(28): p. 9097-9105.
174. Hassani, O., et al., *Role of lysine and tryptophan residues in the biological activity of toxin VII (Ts  $\gamma$ ) from the scorpion Tityus serrulatus*. European Journal of Biochemistry, 1999. **260**(1): p. 76-86.

175. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification*. *Nature Biotechnology*, 2008. **26**: p. 1367.
176. Cox, J., et al., *Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment*. *Journal of Proteome Research*, 2011. **10**(4): p. 1794-1805.
177. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data*. *ELECTROPHORESIS*, 1999. **20**(18): p. 3551-3567.
178. Ziemianowicz, D.S., V. Sarpe, and D.C. Schriemer, *Quantitative Analysis of Protein Covalent Labeling Mass Spectrometry Data in the Mass Spec Studio*. *Analytical Chemistry*, 2019. **91**(13): p. 8492-8499.
179. Zhang, J., et al., *PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification*. *Molecular & cellular proteomics : MCP*, 2012. **11**(4): p. M111.010587-M111.010587.
180. Shilov, I.V., et al., *The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra*. *Mol Cell Proteomics*, 2007. **6**(9): p. 1638-55.
181. Fu, Y., et al., *Exploiting the kernel trick to correlate fragment ions for peptide identification via tandem mass spectrometry*. *Bioinformatics*, 2004. **20**(12): p. 1948-1954.
182. Wang, L.-h., et al., *pFind 2.0: a software package for peptide and protein identification via tandem mass spectrometry*. *Rapid Communications in Mass Spectrometry*, 2007. **21**(18): p. 2985-2991.
183. Eng, J.K., A.L. McCormack, and J.R. Yates, *An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database*. *Journal of the American Society for Mass Spectrometry*, 1994. **5**(11): p. 976-989.
184. Geer, L.Y., et al., *Open Mass Spectrometry Search Algorithm*. *Journal of Proteome Research*, 2004. **3**(5): p. 958-964.
185. Wenger, C.D., et al., *COMPASS: A suite of pre- and post-search proteomics software tools for OMSSA*. *PROTEOMICS*, 2011. **11**(6): p. 1064-1074.
186. Deutsch, E.W., et al., *A guided tour of the Trans-Proteomic Pipeline*. *PROTEOMICS*, 2010. **10**(6): p. 1150-1159.
187. Lomant, A.J. and G. Fairbanks, *Chemical probes of extended biological structures: Synthesis and properties of the cleavable protein cross-linking reagent [35S]dithiobis(succinimidyl propionate)*. *Journal of Molecular Biology*, 1976. **104**(1): p. 243-261.
188. Yem, A.W., et al., *Biotinylation of reactive amino groups in native recombinant human interleukin-1 beta*. *Journal of Biological Chemistry*, 1989. **264**(30): p. 17691-7.
189. Novak, P., et al., *A Top-down method for the determination of residue-specific solvent accessibility in proteins*. *Journal of Mass Spectrometry*, 2004. **39**(3): p. 322-328.
190. Tollefson, E.J., et al., *Preferential Binding of Cytochrome c to Anionic Ligand-Coated Gold Nanoparticles: A Complementary Computational and Experimental Approach*. *ACS Nano*, 2019. **13**(6): p. 6856-6866.
191. Hassani, O., et al., *Role of lysine and tryptophan residues in the biological activity of toxin VII (Ts gamma) from the scorpion Tityus serrulatus*. *Eur J Biochem*, 1999. **260**(1): p. 76-86.
192. Nuss, J.E., D.J. Sweeney, and G.M. Alter, *Reactivity-based analysis of domain structures in native replication protein A*. *Biochemistry*, 2006. **45**(32): p. 9804-18.
193. Glocker, M.O., et al., *Selective Biochemical Modification of Functional Residues in Recombinant Human Macrophage Colony-Stimulating Factor  $\beta$  (rhM-CSF  $\beta$ ): Identification by Mass Spectrometry*. *Biochemistry*, 1996. **35**(46): p. 14625-14633.
194. Dage, J.L., H. Sun, and H.B. Halsall, *Determination of Diethylpyrocarbonate-Modified Amino Acid Residues in  $\alpha$ 1-Acid Glycoprotein by High-Performance Liquid Chromatography Electro spray Ionization–Mass Spectrometry and Matrix-Assisted Laser*

- Desorption/Ionization Time-of-Flight–Mass Spectrometry*. Analytical Biochemistry, 1998. **257**(2): p. 176-185.
195. Tsubaki, M., et al., *Diethyl Pyrocarbonate Modification Abolishes Fast Electron Accepting Ability of Cytochrome b561 from Ascorbate but Does Not Influence Electron Donation to Monodehydroascorbate Radical: Identification of the Modification Sites by Mass Spectrometric Analysis*. Biochemistry, 2000. **39**(12): p. 3276-3284.
  196. Mendoza, V.L. and R.W. Vachet, *Protein Surface Mapping Using Diethylpyrocarbonate with Mass Spectrometric Detection*. Analytical Chemistry, 2008. **80**(8): p. 2895-2904.
  197. Mendoza, V.L. and R.W. Vachet, *Probing protein structure by amino acid-specific covalent labeling and mass spectrometry*. Mass Spectrom Rev, 2009. **28**(5): p. 785-815.
  198. Nishikawa, K., Y. Kubota, and T. Ooi, *Classification of proteins into groups based on amino acid composition and other characters. II. Grouping into four types*. J Biochem, 1983. **94**(3): p. 997-1007.
  199. Trinquier, G. and Y.H. Sanejouand, *Which effective property of amino acids is best preserved by the genetic code?* Protein Engineering, Design and Selection, 1998. **11**(3): p. 153-169.
  200. Limpikirati, P., T. Liu, and R.W. Vachet, *Covalent labeling-mass spectrometry with non-specific reagents for studying protein structure and interactions*. Methods, 2018. **144**: p. 79-93.
  201. Limpikirati, P., X. Pan, and R.W. Vachet, *Covalent Labeling with Diethylpyrocarbonate: Sensitive to the Residue Microenvironment, Providing Improved Analysis of Protein Higher Order Structure by Mass Spectrometry*. Analytical Chemistry, 2019. **91**(13): p. 8516-8523.
  202. Schmidt, C., et al., *Surface Accessibility and Dynamics of Macromolecular Assemblies Probed by Covalent Labeling Mass Spectrometry and Integrative Modeling*. Analytical chemistry, 2017. **89**(3): p. 1459-1468.
  203. Borotto, N.B., et al., *Investigating Therapeutic Protein Structure with Diethylpyrocarbonate Labeling and Mass Spectrometry*. Anal Chem, 2015. **87**(20): p. 10627-34.
  204. Karmakar, S. and K.P. Das, *Interaction of Cu<sup>+2</sup> with alpha-Crystallin: A Biophysical and Mass Spectrometric Study*. Protein Pept Lett, 2018. **25**(3): p. 275-284.
  205. Liu, T., et al., *Using Covalent Labeling and Mass Spectrometry To Study Protein Binding Sites of Amyloid Inhibiting Molecules*. Analytical Chemistry, 2017. **89**(21): p. 11583-11591.
  206. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks*. Genome Res, 2003. **13**(11): p. 2498-504.
  207. Chen, E.Y., et al., *Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool*. BMC Bioinformatics, 2013. **14**: p. 128.
  208. Kuleshov, M.V., et al., *Enrichr: a comprehensive gene set enrichment analysis web server 2016 update*. Nucleic Acids Res, 2016. **44**(W1): p. W90-7.
  209. Xie, Z., et al., *Gene Set Knowledge Discovery with Enrichr*. Current Protocols, 2021. **1**(3): p. e90.
  210. Fraczekiewicz, R. and W. Braun, *Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules*. Journal of Computational Chemistry, 1998. **19**(3): p. 319-333.
  211. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification*. Nature Biotechnology, 2008. **26**(12): p. 1367-1372.
  212. Choi, M., et al., *MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments*. Bioinformatics, 2014. **30**(17): p. 2524-6.
  213. Krissinel, E. and K. Henrick, *Inference of Macromolecular Assemblies from Crystalline State*. Journal of Molecular Biology, 2007. **372**(3): p. 774-797.
  214. Perez-Riverol, Y., et al., *The PRIDE database and related tools and resources in 2019: improving support for quantification data*. Nucleic Acids Res, 2019. **47**(D1): p. D442-d450.



215. Olivella, R., et al., *QCloud2: An Improved Cloud-based Quality-Control System for Mass-Spectrometry-based Proteomics Laboratories*. Journal of Proteome Research, 2021. **20**(4): p. 2010-2013.
216. Schrödinger, L.L.C., *The PyMOL Molecular Graphics System, Version 1.3*. 2010.
217. R\_Core\_Team, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2020.
218. RStudio\_Team, *R Studio: Integrated Development Environment for R*. R Studio, PBC, Boston, MA. 2020.
219. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis*. J Comput Chem, 2004. **25**(13): p. 1605-12.
220. Kosinski, J., et al., *Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures*. J Struct Biol, 2015. **189**(3): p. 177-83.
221. Lima, D.B., Y. Zhu, and F. Liu, *XlinkCyNET: A Cytoscape Application for Visualization of Protein Interaction Networks Based on Cross-Linking Mass Spectrometry Identifications*. J Proteome Res, 2021. **20**(4): p. 1943-1950.
222. Olsen, J.V., et al., *Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap*. Mol Cell Proteomics, 2005. **4**(12): p. 2010-21.
223. Jean Beltran, P.M., R.A. Mathias, and I.M. Cristea, *A Portrait of the Human Organelle Proteome In Space and Time during Cytomegalovirus Infection*. Cell Syst, 2016. **3**(4): p. 361-373.e6.
224. Dunn, O.J., *Multiple Comparisons among Means*. Journal of the American Statistical Association, 1961. **56**(293): p. 52-64.
225. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. Journal of the Royal Statistical Society: Series B (Methodological), 1995. **57**(1): p. 289-300.
226. Wittig, S., et al., *Cross-linking mass spectrometry uncovers protein interactions and functional assemblies in synaptic vesicle membranes*. Nat Commun, 2021. **12**(1): p. 858.
227. Plapp, B.V., H.A. Charlier, and S. Ramaswamy, *Mechanistic implications from structures of yeast alcohol dehydrogenase complexed with coenzyme and an alcohol*. Archives of Biochemistry and Biophysics, 2016. **591**: p. 35-42.
228. Larsen, T.M., et al., *Ligand-Induced Domain Movement in Pyruvate Kinase: Structure of the Enzyme from Rabbit Muscle with Mg<sup>2+</sup>, K<sup>+</sup>, and D-Phospholactate at 2.7 Å Resolution*. Archives of Biochemistry and Biophysics, 1997. **345**(2): p. 199-206.
229. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold*. Nature, 2021. **596**(7873): p. 583-589.
230. Geiger, T., et al., *Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins*. Mol Cell Proteomics, 2012. **11**(3): p. M111.014050.
231. Beck, M., et al., *The quantitative proteome of a human cell line*. Mol Syst Biol, 2011. **7**: p. 549.
232. Benjamini, Y., et al., *Controlling the false discovery rate in behavior genetics research*. Behav Brain Res, 2001. **125**(1-2): p. 279-84.
233. Jambrovićs, K., et al., *Transglutaminase 2 programs differentiating acute promyelocytic leukemia cells in all-trans retinoic acid treatment to inflammatory stage through NF-κB activation*. Haematologica, 2019. **104**(3): p. 505-515.
234. Collins, C.A. and F.M. Watt, *Dynamic regulation of retinoic acid-binding proteins in developing, adult and neoplastic skin reveals roles for beta-catenin and Notch signalling*. Dev Biol, 2008. **324**(1): p. 55-67.
235. Attoff, K., et al., *Acrylamide alters CREB and retinoic acid signalling pathways during differentiation of the human neuroblastoma SH-SY5Y cell line*. Sci Rep, 2020. **10**(1): p. 16714.

236. Vaezeslami, S., et al., *The structure of Apo-wild-type cellular retinoic acid binding protein II at 1.4 Å and its relationship to ligand binding and nuclear translocation*. J Mol Biol, 2006. **363**(3): p. 687-701.
237. Delva, L., et al., *Physical and functional interactions between cellular retinoic acid binding protein II and the retinoic acid-dependent nuclear complex*. Mol Cell Biol, 1999. **19**(10): p. 7158-67.
238. Delacroix, L., et al., *Cell-specific interaction of retinoic acid receptors with target genes in mouse embryonic fibroblasts and embryonic stem cells*. Mol Cell Biol, 2010. **30**(1): p. 231-44.
239. Dieplinger, B., et al., *The transcriptional corepressor TPA-inducible sequence 7 regulates adult axon growth through cellular retinoic acid binding protein II expression*. Eur J Neurosci, 2007. **26**(12): p. 3358-67.
240. Bistulfi, G., et al., *A repressive epigenetic domino effect confers susceptibility to breast epithelial cell transformation: implications for predicting breast cancer risk*. Cancer Res, 2006. **66**(21): p. 10308-14.
241. Adams, M.K., et al., *Characterization of subunit interactions in the hetero-oligomeric retinoid oxidoreductase complex*. Biochem J, 2021. **478**(19): p. 3597-3611.
242. Han, S., et al., *Effects of annexins II and V on survival of neurons and astrocytes in vitro*. Acta Pharmacol Sin, 2004. **25**(5): p. 602-10.
243. Jacovina, A.T., et al., *Neuritogenesis and the nerve growth factor-induced differentiation of PC-12 cells requires annexin II-mediated plasmin generation*. J Biol Chem, 2001. **276**(52): p. 49350-8.
244. Jiang, M., et al., *Cathepsin B inhibition blocks neurite outgrowth in cultured neurons by regulating lysosomal trafficking and remodeling*. J Neurochem, 2020. **155**(3): p. 300-312.
245. Li, L., A.C. Hung, and A.G. Porter, *Secretogranin II: a key AP-1-regulated protein that mediates neuronal differentiation and protection from nitric oxide-induced apoptosis of neuroblastoma cells*. Cell Death Differ, 2008. **15**(5): p. 879-88.
246. Drees, F. and F.B. Gertler, *Ena/VASP: proteins at the tip of the nervous system*. Curr Opin Neurobiol, 2008. **18**(1): p. 53-9.
247. Koizumi, H., et al., *DCLK1 phosphorylates the microtubule-associated protein MAP7D1 to promote axon elongation in cortical neurons*. Dev Neurobiol, 2017. **77**(4): p. 493-510.
248. Calvo-Garrido, J., et al., *SQSTM1/p62-Directed Metabolic Reprogramming Is Essential for Normal Neurodifferentiation*. Stem Cell Reports, 2019. **12**(4): p. 696-711.
249. Unsicker, C., et al., *SHANK2 mutations impair apoptosis, proliferation and neurite outgrowth during early neuronal differentiation in SH-SY5Y cells*. Sci Rep, 2021. **11**(1): p. 2128.
250. Brown, J.P., et al., *Transient expression of doublecortin during adult neurogenesis*. J Comp Neurol, 2003. **467**(1): p. 1-10.
251. Cobley, J.N., M.L. Fiorello, and D.M. Bailey, *13 reasons why the brain is susceptible to oxidative stress*. Redox Biol, 2018. **15**: p. 490-503.
252. Xie, Z., et al., *Gene Set Knowledge Discovery with Enrichr*. Curr Protoc, 2021. **1**(3): p. e90.
253. Kalinowska, M., et al., *Actinin-4 Governs Dendritic Spine Dynamics and Promotes Their Remodeling by Metabotropic Glutamate Receptors*. J Biol Chem, 2015. **290**(26): p. 15909-20.
254. Rashid, M., et al., *Neural-specific deletion of the focal adhesion adaptor protein paxillin slows migration speed and delays cortical layer formation*. Development, 2017. **144**(21): p. 4002-4014.
255. Huang, L., et al., *Replacement of threonine 558, a critical site of phosphorylation of moesin in vivo, with aspartate activates F-actin binding of moesin. Regulation by conformational change*. J Biol Chem, 1999. **274**(18): p. 12803-10.
256. Müller, F., *The flavin redox-system and its biological function*. Top Curr Chem, 1983. **108**: p. 71-107.

257. Ferguson, S.M., *Neuronal lysosomes*. *Neurosci Lett*, 2019. **697**: p. 1-9.
258. Malik, B.R., et al., *Autophagic and endo-lysosomal dysfunction in neurodegenerative disease*. *Mol Brain*, 2019. **12**(1): p. 100.
259. Cheng, L., et al., *Knockdown of ISOC1 suppresses cell proliferation in pancreatic cancer in vitro*. *Oncol Lett*, 2019. **17**(5): p. 4263-4270.
260. Ghazanfar, S., et al., *Identification of actin beta-like 2 (ACTBL2) as novel, upregulated protein in colorectal cancer*. *J Proteomics*, 2017. **152**: p. 33-40.
261. Kalous, J., D. Jansová, and A. Šušor, *Role of Cyclin-Dependent Kinase 1 in Translational Regulation in the M-Phase*. *Cells*, 2020. **9**(7).
262. Matson, J.P., et al., *Intrinsic checkpoint deficiency during cell cycle re-entry from quiescence*. *J Cell Biol*, 2019. **218**(7): p. 2169-2184.
263. Lamers, M.H., et al., *The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch*. *Nature*, 2000. **407**(6805): p. 711-7.
264. Li, G.M., *The role of mismatch repair in DNA damage-induced apoptosis*. *Oncol Res*, 1999. **11**(9): p. 393-400.
265. Wang, H., et al., *DNA bending and unbending by MutS govern mismatch recognition and specificity*. *Proc Natl Acad Sci U S A*, 2003. **100**(25): p. 14822-7.
266. Belloni, M., et al., *Induction of two DNA mismatch repair proteins, MSH2 and MSH6, in differentiated human neuroblastoma SH-SY5Y cells exposed to doxorubicin*. *J Neurochem*, 1999. **72**(3): p. 974-9.
267. Khatter, H., et al., *Structure of the human 80S ribosome*. *Nature*, 2015. **520**(7549): p. 640-5.
268. Noeske, J. and J.H.D. Cate, *Structural basis for protein synthesis: snapshots of the ribosome in motion*. *Current Opinion in Structural Biology*, 2012. **22**(6): p. 743-749.
269. Harner, M., et al., *The mitochondrial contact site complex, a determinant of mitochondrial architecture*. *Embo j*, 2011. **30**(21): p. 4356-70.
270. Ladtschik, K., A. Ganguly, and S. Roy, *Axonal actin in action: Imaging actin dynamics in neurons*. *Methods Cell Biol*, 2016. **131**: p. 91-106.
271. Müller, M., et al., *Distinct functional interactions between actin isoforms and nonsarcomeric myosins*. *PLoS One*, 2013. **8**(7): p. e70636.
272. von der Ecken, J., et al., *Structure of the F-actin-tropomyosin complex*. *Nature*, 2015. **519**(7541): p. 114-7.
273. Yin, L.M., L. Ulloa, and Y.Q. Yang, *Transgelin-2: Biochemical and Clinical Implications in Cancer and Asthma*. *Trends Biochem Sci*, 2019. **44**(10): p. 885-896.
274. Sharma, S., et al., *Atomic force microscopy reveals drebrin induced remodeling of f-actin with subnanometer resolution*. *Nano Lett*, 2011. **11**(2): p. 825-7.
275. Shapland, C., et al., *Purification and properties of transgelin: a transformation and shape change sensitive actin-gelling protein*. *The Journal of cell biology*, 1993. **121**(5): p. 1065-1073.
276. Yin, L.M., M. Schnoor, and C.D. Jun, *Structural Characteristics, Binding Partners and Related Diseases of the Calponin Homology (CH) Domain*. *Front Cell Dev Biol*, 2020. **8**: p. 342.
277. El-Mezgueldi, M., *Calponin*. *The International Journal of Biochemistry & Cell Biology*, 1996. **28**(11): p. 1185-1189.
278. Kotila, T., et al., *Structural basis of actin monomer re-charging by cyclase-associated protein*. *Nat Commun*, 2018. **9**(1): p. 1892.
279. Kakurina, G.V., E.S. Kolegova, and I.V. Kondakova, *Adenylyl Cyclase-Associated Protein 1: Structure, Regulation, and Participation in Cellular Processes*. *Biochemistry (Mosc)*, 2018. **83**(1): p. 45-53.
280. Pizarro-Cerdá, J., et al., *The Diverse Family of Arp2/3 Complexes*. *Trends Cell Biol*, 2017. **27**(2): p. 93-100.

281. Rouiller, I., et al., *The structural basis of actin filament branching by the Arp2/3 complex*. J Cell Biol, 2008. **180**(5): p. 887-95.
282. Kang, F., D.L. Purich, and F.S. Southwick, *Profilin promotes barbed-end actin filament assembly without lowering the critical concentration*. J Biol Chem, 1999. **274**(52): p. 36963-72.
283. Schaeffer, V., et al., *RNA-binding protein IGF2BP2/IMP2 is required for laminin-β2 mRNA translation and is modulated by glucose concentration*. Am J Physiol Renal Physiol, 2012. **303**(1): p. F75-82.
284. Kovács, M., et al., *Two-headed binding of the unphosphorylated nonmuscle heavy meromyosin-ADP complex to actin*. Biochemistry, 2004. **43**(14): p. 4219-26.
285. Javier-Torrent, M. and C.A. Saura, *Conventional and Non-Conventional Roles of Non-Muscle Myosin II-Actin in Neuronal Development and Degeneration*. Cells, 2020. **9**(9).
286. Kim, J.Y., et al., *Probing Lysine Acetylation with a Modification-Specific Marker Ion Using High-Performance Liquid Chromatography/Electrospray-Mass Spectrometry with Collision-Induced Dissociation*. Analytical Chemistry, 2002. **74**(21): p. 5443-5449.
287. Miller, B.T., *Acylation of Peptide Hydroxyl Groups with the Bolton–Hunter Reagent*. Biochemical and Biophysical Research Communications, 1996. **218**(1): p. 377-382.
288. Miller, B.T., et al., *The occurrence of O-acylation during biotinylation of gonadotropin-releasing hormone and analogs. Evidence for a reactive serine*. Journal of Biological Chemistry, 1992. **267**(8): p. 5060-9.
289. Abello, N., et al., *Selective Acylation of Primary Amines in Peptides and Proteins*. Journal of Proteome Research, 2007. **6**(12): p. 4770-4776.
290. Mendoza, V.L. and R.W. Vachet, *Probing protein structure by amino acid-specific covalent labeling and mass spectrometry*. Mass Spectrometry Reviews, 2009. **28**(5): p. 785-815.
291. Safarian, S., et al., *The structural and functional studies of His119 and His12 in RNase A via chemical modification*. J Protein Chem, 2003. **22**(7-8): p. 643-54.
292. Uetrecht, C., et al., *Ion mobility mass spectrometry of proteins and protein assemblies*. Chem Soc Rev, 2010. **39**(5): p. 1633-55.
293. Jurneczko, E. and P.E. Barran, *How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase*. Analyst, 2011. **136**(1): p. 20-8.
294. Wittig, S., et al., *Cross-linking mass spectrometry uncovers protein interactions and functional assemblies in synaptic vesicle membranes*. Nature Communications, 2021. **12**(1): p. 858.
295. Biehn, S.E., et al., *Utilization of Hydrophobic Microenvironment Sensitivity in Diethylpyrocarbonate Labeling for Protein Structure Prediction*. Anal Chem, 2021. **93**(23): p. 8188-8195.
296. Limpikirati, P., X. Pan, and R.W. Vachet, *Covalent Labeling with Diethylpyrocarbonate: Sensitive to the Residue Microenvironment, Providing Improved Analysis of Protein Higher Order Structure by Mass Spectrometry*. Anal Chem, 2019. **91**(13): p. 8516-8523.
297. Taoufiq, Z., et al., *Hidden proteome of synaptic vesicles in the mammalian brain*. Proc Natl Acad Sci U S A, 2020. **117**(52): p. 33586-33596.
298. Takamori, S., et al., *Molecular anatomy of a trafficking organelle*. Cell, 2006. **127**(4): p. 831-46.
299. Abbas, Y.M., et al., *Structure of V-ATPase from the mammalian brain*. Science, 2020. **367**(6483): p. 1240-1246.
300. Zhou, M., et al., *Ion mobility-mass spectrometry of a rotary ATPase reveals ATP-induced reduction in conformational flexibility*. Nat Chem, 2014. **6**(3): p. 208-215.
301. Zhang, M. and G.J. Augustine, *Synapsins and the Synaptic Vesicle Reserve Pool: Floats or Anchors?* Cells, 2021. **10**(3): p. 658.

302. Benfenati, F., et al., *Electrostatic and hydrophobic interactions of synapsin I and synapsin I fragments with phospholipid bilayers*. The Journal of cell biology, 1989. **108**(5): p. 1851-1862.
303. Fukumoto, Y., et al., *Molecular cloning and characterization of a novel type of regulatory protein (GDI) for the rho proteins, ras p21-like small GTP-binding proteins*. Oncogene, 1990. **5**(9): p. 1321-1328.
304. Macara, I.G., *Role of the Rab3A GTPase in regulated secretion from neuroendocrine cells*. Trends Endocrinol Metab, 1994. **5**(7): p. 267-71.
305. Farnsworth, C.C., et al., *C terminus of the small GTP-binding protein smg p25A contains two geranylgeranylated cysteine residues and a methyl ester*. Proc Natl Acad Sci U S A, 1991. **88**(14): p. 6196-200.
306. Dumas, J.J., et al., *Structural basis of activation and GTP hydrolysis in Rab proteins*. Structure (London, England : 1993), 1999. **7**(4): p. 413-423.
307. Burgoyne, R.D. and A. Morgan, *Chaperoning the SNAREs: a role in preventing neurodegeneration?* Nat Cell Biol, 2011. **13**(1): p. 8-9.
308. Patel, P., et al., *Phosphorylation of Cysteine String Protein Triggers a Major Conformational Switch*. Structure, 2016. **24**(8): p. 1380-1386.
309. Budzinski, K.L., et al., *Large structural change in isolated synaptic vesicles upon loading with neurotransmitter*. Biophys J, 2009. **97**(9): p. 2577-84.
310. Xicoy, H., B. Wieringa, and G.J.M. Martens, *The SH-SY5Y cell line in Parkinson's disease research: a systematic review*. Molecular Neurodegeneration, 2017. **12**(1): p. 10.
311. Innala, M., et al., *3D culturing and differentiation of SH-SY5Y neuroblastoma cells on bacterial nanocellulose scaffolds*. Artif Cells Nanomed Biotechnol, 2014. **42**(5): p. 302-8.
312. Shipley, M.M., C.A. Mangold, and M.L. Szpara, *Differentiation of the SH-SY5Y Human Neuroblastoma Cell Line*. J Vis Exp, 2016(108): p. 53193.
313. Srikanth, P. and T.L. Young-Pearse, *Stem cells on the brain: modeling neurodevelopmental and neurodegenerative diseases using human induced pluripotent stem cells*. J Neurogenet, 2014. **28**(1-2): p. 5-29.
314. Villén, J. and S.P. Gygi, *The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry*. Nat Protoc, 2008. **3**(10): p. 1630-8.
315. McNulty, D.E. and R.S. Annan, *Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection*. Mol Cell Proteomics, 2008. **7**(5): p. 971-80.
316. Ritorto, M.S., et al., *Hydrophilic strong anion exchange (hSAX) chromatography for highly orthogonal peptide separation of complex proteomes*. J Proteome Res, 2013. **12**(6): p. 2449-57.
317. Han, D., et al., *Proteomic analysis of mouse astrocytes and their secretome by a combination of FASP and StageTip-based, high pH, reversed-phase fractionation*. Proteomics, 2014. **14**(13-14): p. 1604-9.
318. Hebert, A.S., et al., *Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer*. Analytical Chemistry, 2018. **90**(15): p. 9529-9537.
319. Bian, Y., et al., *Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC-MS/MS*. Nature Communications, 2020. **11**(1): p. 157.
320. Melino, G., et al., *Retinoids and the control of growth/death decisions in human neuroblastoma cell lines*. J Neurooncol, 1997. **31**(1-2): p. 65-83.
321. Yamaga, R., et al., *Systemic identification of estrogen-regulated genes in breast cancer cells through cap analysis of gene expression mapping*. Biochem Biophys Res Commun, 2014. **447**(3): p. 531-6.
322. Gao, B., et al., *Knockdown of ISOC1 inhibits the proliferation and migration and induces the apoptosis of colon cancer cells through the AKT/GSK-3 $\beta$  pathway*. Carcinogenesis, 2020. **41**(8): p. 1123-1133.

323. Shi, J., et al., *Isochorismatase domain-containing protein 1 (ISOC1) participates in DNA damage repair and inflammation-related pathways to promote lung cancer development*. *Transl Lung Cancer Res*, 2021. **10**(3): p. 1444-1456.
324. Huang, X., et al., *Identification and characterization of a novel protein ISOC2 that interacts with p16INK4a*. *Biochem Biophys Res Commun*, 2007. **361**(2): p. 287-93.
325. Wu, W., et al., *Minichromosome maintenance protein 2 correlates with the malignant status and regulates proliferation and cell cycle in lung squamous cell carcinoma*. *Onco Targets Ther*, 2018. **11**: p. 5025-5034.
326. Zhong, H., et al., *Expression of minichromosome maintenance genes in renal cell carcinoma*. *Cancer Manag Res*, 2017. **9**: p. 637-647.
327. Stewart, P.A., et al., *Upregulation of minichromosome maintenance complex component 3 during epithelial-to-mesenchymal transition in human prostate cancer*. *Oncotarget*, 2017. **8**(24): p. 39209-39217.
328. Cai, H.Q., et al., *Overexpression of MCM6 predicts poor survival in patients with glioma*. *Hum Pathol*, 2018. **78**: p. 182-187.
329. Giaginis, C., et al., *Clinical significance of MCM-2 and MCM-5 expression in colon cancer: association with clinicopathological parameters and tumor proliferative capacity*. *Digestive diseases and sciences*, 2009. **54**(2): p. 282-291.
330. Gu, Y., et al., *MCM6 indicates adverse tumor features and poor outcomes and promotes G1/S cell cycle progression in neuroblastoma*. 2020, Research Square.
331. Marnerides, A., et al., *Immunohistochemical expression and prognostic significance of CCND3, MCM2 and MCM7 in Hodgkin lymphoma*. *Anticancer Res*, 2011. **31**(10): p. 3585-94.
332. Peng, Y.-P., et al., *The Expression and Prognostic Roles of MCMs in Pancreatic Cancer*. *PLoS one*, 2016. **11**(10): p. e0164150-e0164150.
333. Wojnar, A., et al., *Correlation of intensity of MT-I/II expression with Ki-67 and MCM-2 proteins in invasive ductal breast carcinoma*. *Anticancer Res*, 2011. **31**(9): p. 3027-33.
334. Jentsch, T., et al., *Expression of MSH2 and MSH6 on a tissue microarray in patients with osteosarcoma*. *Anticancer Res*, 2014. **34**(12): p. 6961-72.
335. Boyken, J., et al., *Molecular profiling of synaptic vesicle docking sites reveals novel proteins but few differences between glutamatergic and GABAergic synapses*. *Neuron*, 2013. **78**(2): p. 285-97.
336. Grønberg, M., et al., *Quantitative comparison of glutamatergic and GABAergic synaptic vesicles unveils selectivity for few proteins including MAL2, a novel synaptic vesicle protein*. *J Neurosci*, 2010. **30**(1): p. 2-12.
337. Sharma, S.K., *Protein acetylation in synaptic plasticity and memory*. *Neuroscience & Biobehavioral Reviews*, 2010. **34**(8): p. 1234-1240.
338. Scott, H. and V.M. Panin, *The role of protein N-glycosylation in neural transmission*. *Glycobiology*, 2014. **24**(5): p. 407-17.
339. Slavin, M., et al., *Targeted in situ cross-linking mass spectrometry and integrative modeling reveal the architectures of three proteins from SARS-CoV-2*. *Proc Natl Acad Sci U S A*, 2021. **118**(34).
340. Dettmer, U., et al., *In vivo cross-linking reveals principally oligomeric forms of  $\alpha$ -synuclein and  $\beta$ -synuclein in neurons and non-neural cells*. *J Biol Chem*, 2013. **288**(9): p. 6371-85.
341. Beveridge, R., et al., *A synthetic peptide library for benchmarking crosslinking-mass spectrometry search engines for proteins and protein complexes*. *Nature Communications*, 2020. **11**(1): p. 742.
342. Galkin, V.E., et al., *High-resolution cryo-EM structure of the F-actin-fimbrin/plastin ABD2 complex*. *Proc Natl Acad Sci U S A*, 2008. **105**(5): p. 1494-8.
343. Mendoza, V.L., et al., *Structure of the Preamyloid Dimer of  $\beta$ -2-Microglobulin from Covalent Labeling and Mass Spectrometry*. *Biochemistry*, 2010. **49**(7): p. 1522-1532.

344. Kalkum, M., M. Przybylski, and M.O. Glocker, *Structure Characterization of Functional Histidine Residues and Carboxylated Derivatives in Peptides and Proteins by Mass Spectrometry*. *Bioconjugate Chemistry*, 1998. **9**(2): p. 226-235.
345. Melchior, W.B. and D. Fahrney, *Ethoxyformylation of proteins. Reaction of ethoxyformic anhydride with .alpha.-chymotrypsin, pepsin, and pancreatic ribonuclease at pH 4*. *Biochemistry*, 1970. **9**(2): p. 251-258.
346. Borotto, N.B., et al., *Label Scrambling During CID of Covalently Labeled Peptide Ions*. *Journal of The American Society for Mass Spectrometry*, 2014. **25**(10): p. 1739-1746.
347. Zhou, Y. and R.W. Vachet, *Diethylpyrocarbonate labeling for the structural analysis of proteins: label scrambling in solution and how to avoid it*. *J Am Soc Mass Spectrom*, 2012. **23**(5): p. 899-907.
348. Zhou, Y. and R.W. Vachet, *Increased Protein Structural Resolution from Diethylpyrocarbonate-based Covalent Labeling and Mass Spectrometric Detection*. *Journal of The American Society for Mass Spectrometry*, 2012. **23**(4): p. 708-717.
349. Farsi, Z., R. Jahn, and A. Woehler, *Proton electrochemical gradient: Driving and regulating neurotransmitter uptake*. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 2017. **39**(5).
350. Vasanthakumar, T., et al., *Coordinated conformational changes in the V<sub>1</sub> complex during V-ATPase reversible dissociation*. *bioRxiv*, 2021: p. 2021.11.09.467972.
351. Poëa-Guyon, S., et al., *The V-ATPase membrane domain is a sensor of granular pH that controls the exocytotic machinery*. *Journal of Cell Biology*, 2013. **203**(2): p. 283-298.
352. Vasanthakumar, T., et al., *Structural comparison of the vacuolar and Golgi V-ATPases from *Saccharomyces cerevisiae**. *Proc Natl Acad Sci U S A*, 2019. **116**(15): p. 7272-7277.
353. Zhou, M., et al., *Mass spectrometry of intact V-type ATPases reveals bound lipids and the effects of nucleotide binding*. *Science*, 2011. **334**(6054): p. 380-385.
354. Lau, W.C. and J.L. Rubinstein, *Structure of intact *Thermus thermophilus* V-ATPase by cryo-EM reveals organization of the membrane-bound V(O) motor*. *Proc Natl Acad Sci U S A*, 2010. **107**(4): p. 1367-72.
355. Keon, K.A., S. Benlekbir, and J.L. Rubinstein, *Cryo-EM of the yeast V<sub>O</sub> complex reveals distinct binding sites for macrolide V-ATPase inhibitors*. *bioRxiv*, 2021: p. 2021.11.15.468710.
356. Deutsch, J.W. and R.B. Kelly, *Lipids of synaptic vesicles: relevance to the mechanism of membrane fusion*. *Biochemistry*, 1981. **20**(2): p. 378-85.
357. Nagy, A., et al., *The preparation and characterization of synaptic vesicles of high purity*. *Brain Res*, 1976. **109**(2): p. 285-309.
358. Guo, C., et al., *Diethylpyrocarbonate Footprints a Membrane Protein in Micelles*. *J Am Soc Mass Spectrom*, 2021. **32**(11): p. 2636-2643.
359. Chorev, D.S., et al., *Protein assemblies ejected directly from native membranes yield complexes for mass spectrometry*. *Science*, 2018. **362**(6416): p. 829-834.
360. Rietschel, B., et al., *Elastase digests: new ammunition for shotgun membrane proteomics*. *Mol Cell Proteomics*, 2009. **8**(5): p. 1029-43.
361. Min, L., L.H. Choe, and K.H. Lee, *Improved protease digestion conditions for membrane protein detection*. *Electrophoresis*, 2015. **36**(15): p. 1690-8.
362. Jumper, C.C. and D.C. Schriemer, *Mass spectrometry of laser-initiated carbene reactions for protein topographic analysis*. *Anal Chem*, 2011. **83**(8): p. 2913-20.
363. Thomas, L., et al., *Identification of synaptophysin as a hexameric channel protein of the synaptic vesicle membrane*. *Science*, 1988. **242**(4881): p. 1050-3.
364. Arthur, C.P. and M.H. Stowell, *Structure of synaptophysin: a hexameric MARVEL-domain channel protein*. *Structure*, 2007. **15**(6): p. 707-14.
365. Adams, D.J., C.P. Arthur, and M.H. Stowell, *Architecture of the Synaptophysin/Synaptobrevin Complex: Structural Evidence for an Entropic Clustering Function at the Synapse*. *Sci Rep*, 2015. **5**: p. 13659.

366. Lynch, B.A., et al., *Visualization of SV2A conformations in situ by the use of Protein Tomography*. *Biochem Biophys Res Commun*, 2008. **375**(4): p. 491-5.



## 6. Appendix

### 6.1 Supplementary Tables

#### Supplementary Table 1: Enriched KEGG pathways in undifferentiated cells (vs. RA-differentiated cells)

Corresponding gene names of upregulated proteins in each condition were used for gene set search engine Enrichr. Enriched KEGG pathways for each condition with an adjusted P-value < 0.05 are given. Following columns are shown: Term, enriched KEGG term; Overlap, overlapping genes of the input gene list/ genes of the term; P-value, is calculated using the Fisher's exact test; Adjusted P-value, adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypotheses testing; Odds Ratio, odds ratio; Combined Score, multiplication of the odds ratio by the negative natural log of the p-value; Genes, identified genes of the pathway.

undiff (vs. RA)						
Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score	Genes
DNA replication	10/36	2.9E-16	3.2E-14	100.6	3600.2	FEN1; RFC3; MCM7; LIG1; MCM3; RPA2; MCM4; MCM5; MCM6; MCM2
Mismatch repair	5/23	4.1E-08	1.5E-06	68.2	1160.0	MSH6; RFC3; LIG1; MSH2; RPA2
Cell cycle	9/124	3.1E-09	1.7E-07	20.1	394.2	MCM7; RAD21; MCM3; CDK1; MCM4; MCM5; MCM6; SMC3; MCM2
One carbon pool by folate	2/20	3.3E-03	4.5E-02	26.3	150.4	DHFR; TYMS
Terpenoid backbone biosynthesis	2/22	4.0E-03	4.9E-02	23.7	130.8	IDI1; FNTA
Nucleotide excision repair	3/47	1.1E-03	2.4E-02	16.3	111.4	RFC3; LIG1; RPA2
Pyrimidine metabolism	3/57	1.9E-03	3.5E-02	13.3	83.3	DUT; RRM1; TYMS
Spliceosome	5/134	2.8E-04	7.8E-03	9.5	77.4	ISY1; XAB2; SF3B6; SNRPG; BCAS2
Purine metabolism	4/129	2.3E-03	3.7E-02	7.7	46.8	RRM1; IMPDH1; PPAT; PAICS

**Supplementary Table 2: Enriched KEGG pathways in undifferentiated cells (vs. RA/PMA-differentiated cells)**

Corresponding gene names of upregulated proteins in each condition were used for gene set search engine Enrichr. Enriched KEGG pathways for each condition with an adjusted P-value < 0.05 are given. Following columns are shown: Term, enriched KEGG term; Overlap, overlapping genes of the input gene list/ genes of the term; P-value, is calculated using the Fisher's exact test; Adjusted P-value, adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypotheses testing; Odds Ratio, odds ratio; Combined Score, multiplication of the odds ratio by the negative natural log of the p-value; Genes, identified genes of the pathway.

undiff (vs. RA/PMA)						
Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score	Genes
DNA replication	9/36	9.4E-14	8.5E-12	74.4	2232.5	MCM7; POLE3; RPA1; MCM3; RPA2; MCM4; MCM5; MCM6; MCM2
Mismatch repair	4/23	4.5E-06	1.3E-04	44.5	548.6	MSH6; MSH2; RPA1; RPA2
Cell cycle	9/124	1.0E-08	4.5E-07	17.4	320.5	MCM7; RAD21; CDK1; MCM3; MCM4; MCM5; MCM6; SMC3; MCM2
One carbon pool by folate	2/20	4.3E-03	3.8E-02	23.0	125.6	DHFR; TYMS
Purine metabolism	6/129	4.2E-05	9.3E-04	10.5	105.8	PRPS1; RRM1; IMPDH1; PPAT; AK5; PAICS
Nucleotide excision repair	3/47	1.6E-03	2.4E-02	14.3	91.9	POLE3; RPA1; RPA2
Cysteine and methionine metabolism	3/47	1.6E-03	2.4E-02	14.3	91.9	MPST; DNMT1; CBS
Fanconi anemia pathway	3/54	2.4E-03	2.8E-02	12.3	74.3	UBE2T; RPA1; RPA2
Glutathione metabolism	3/56	2.6E-03	2.8E-02	11.8	70.3	RRM1; GPX1; GPX4
Pyrimidine metabolism	3/57	2.8E-03	2.8E-02	11.6	68.4	DUT; RRM1; TYMS

### Supplementary Table 3: Enriched KEGG pathways in RA-differentiated cells (vs. undifferentiated cells)

Corresponding gene names of upregulated proteins in each condition were used for gene set search engine Enrichr. Enriched KEGG pathways for each condition with an adjusted P-value < 0.05 are given. Following columns are shown: Term, enriched KEGG term; Overlap, overlapping genes of the input gene list/ genes of the term; P-value, is calculated using the Fisher's exact test; Adjusted P-value, adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypotheses testing; Odds Ratio, odds ratio; Combined Score, multiplication of the odds ratio by the negative natural log of the p-value; Genes, identified genes of the pathway.

RA (vs. undiff)						
Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score	Genes
Riboflavin metabolism	2/8	3.7E-04	3.9E-03	92.2	728.1	BLVRB; ACP2
ECM-receptor interaction	6/82	5.9E-07	2.4E-05	23.0	330.7	ITGB1; ITGA2; ITGA1; LAMC1; ITGB6; CD44
Focal adhesion	10/199	3.1E-09	3.8E-07	16.3	319.5	ITGB1; VASP; ITGA2; ITGA1; PXN; FLNB; LAMC1; ITGB6; ACTN4; FLNC
Regulation of actin cytoskeleton	9/214	9.4E-08	5.7E-06	13.3	215.5	ITGB1; ITGA2; ITGA1; PXN; MYH9; MSN; ITGB6; ACTN4; GNG12
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	4/72	1.5E-04	2.5E-03	16.7	147.3	ITGB1; ITGA2; ITGA1; ITGB6
Leukocyte transendothelial migration	5/112	6.0E-05	1.2E-03	13.4	130.6	ITGB1; VASP; PXN; MSN; ACTN4
Proteoglycans in cancer	7/201	9.5E-06	2.9E-04	10.6	122.9	ITGB1; ITGA2; PXN; MSN; FLNB; FLNC; CD44
Hypertrophic cardiomyopathy (HCM)	4/85	2.8E-04	3.7E-03	14.0	114.6	ITGB1; ITGA2; ITGA1; ITGB6
Dilated cardiomyopathy (DCM)	4/91	3.6E-04	3.9E-03	13.0	103.3	ITGB1; ITGA2; ITGA1; ITGB6
Nicotinate and nicotinamide metabolism	2/30	5.5E-03	3.5E-02	19.7	102.7	NT5E; NAMPT
Small cell lung cancer	4/93	3.9E-04	3.9E-03	12.7	99.9	ITGB1; CDK6; ITGA2; LAMC1
Hematopoietic cell lineage	4/97	4.6E-04	4.0E-03	12.2	93.6	ITGA2; ITGA1; CD59; CD44
Starch and sucrose metabolism	2/36	7.8E-03	4.5E-02	16.3	78.8	GBE1; PYGL
PI3K-Akt signaling pathway	8/354	4.9E-05	1.2E-03	6.9	68.1	ITGB1; GNG2; CDK6; ITGA2; ITGA1; LAMC1; ITGB6; GNG12
Tight junction	5/170	4.2E-04	3.9E-03	8.7	67.5	ITGB1; VASP; MYH9; MSN; ACTN4
Lysosome	4/123	1.1E-03	8.5E-03	9.5	64.6	NPC2; PSAP; ACP2; CTSB
Peroxisome	3/83	3.6E-03	2.6E-02	10.5	58.9	IDH1; ACSL3; SOD2
Salmonella infection	3/86	4.0E-03	2.7E-02	10.1	55.8	MYH9; FLNB; FLNC
Human papillomavirus infection	7/330	2.2E-04	3.3E-03	6.3	53.5	ITGB1; CDK6; ITGA2; ITGA1; PXN; LAMC1; ITGB6
Pathways in cancer	8/530	7.5E-04	6.0E-03	4.5	32.4	ITGB1; NQO1; GNG2; CDK6; ITGA2; LAMC1; GNG12; NFKB2
Viral carcinogenesis	4/201	6.6E-03	4.0E-02	5.7	28.7	CDK6; PXN; ACTN4; NFKB2

### Supplementary Table 4: Enriched KEGG pathways in RA/PMA-differentiated cells (vs. undifferentiated cells)

Corresponding gene names of upregulated proteins in each condition were used for gene set search engine Enrichr. Enriched KEGG pathways for each condition with an adjusted P-value < 0.05 are given. Following columns are shown: Term, enriched KEGG term; Overlap, overlapping genes of the input gene list/ genes of the term; P-value, is calculated using the Fisher's exact test; Adjusted P-value, adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypotheses testing; Odds Ratio, odds ratio; Combined Score, multiplication of the odds ratio by the negative natural log of the p-value; Genes, identified genes of the pathway.

RA/PMA (vs. Undiff)						
Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Combined Score	Genes
Lysosome	10/123	1.6E-09	2.3E-07	17.3	350.6	CTSA; ARSA; NPC2; GLB1; HEXB; LAMP2; PSAP; M6PR; GNS; CTSB
Glycosaminoglycan degradation	3/19	1.5E-04	3.6E-03	34.5	303.5	GLB1; HEXB; GNS
ECM-receptor interaction	6/82	6.3E-06	2.3E-04	14.9	178.3	ITGB1; ITGA2; ITGA1; LAMC1; ITGB6; CD44
Focal adhesion	10/199	1.6E-07	1.2E-05	10.3	161.3	VASP; ITGB1; PAK1; ITGA2; PXN; ITGA1; FLNB; LAMC1; ITGB6; FLNC
Other glycan degradation	2/18	4.4E-03	3.5E-02	22.8	123.6	GLB1; HEXB
Regulation of actin cytoskeleton	9/214	3.1E-06	1.5E-04	8.5	107.6	ITGB1; PAK1; ARPC1B; ITGA2; PXN; ITGA1; MSN; ITGB6; GNG12
Proteoglycans in cancer	8/201	1.6E-05	4.7E-04	7.9	87.3	ITGB1; PAK1; ITGA2; PXN; MSN; FLNB; FLNC; CD44
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	4/72	6.9E-04	1.4E-02	10.9	79.3	ITGB1; ITGA2; ITGA1; ITGB6
Sphingolipid metabolism	3/47	2.3E-03	2.7E-02	12.5	76.3	ARSA; SGPL1; GLB1
Hypertrophic cardiomyopathy (HCM)	4/85	1.3E-03	2.3E-02	9.1	60.8	ITGB1; ITGA2; ITGA1; ITGB6
Pathogenic Escherichia coli infection	3/55	3.5E-03	3.3E-02	10.6	59.8	ITGB1; TUBB6; ARPC1B
Glutathione metabolism	3/56	3.7E-03	3.3E-02	10.4	58.1	G6PD; RRM2B; IDH1
Pyrimidine metabolism	3/57	3.9E-03	3.3E-02	10.2	56.6	NT5E; RRM2B; NME3
Dilated cardiomyopathy (DCM)	4/91	1.7E-03	2.4E-02	8.5	54.5	ITGB1; ITGA2; ITGA1; ITGB6
Hematopoietic cell lineage	4/97	2.1E-03	2.7E-02	8.0	49.1	ITGA2; ITGA1; CD59; CD44
Shigellosis	3/65	5.7E-03	4.1E-02	8.9	46.0	ITGB1; ARPC1B; CD44
Phagosome	5/152	1.6E-03	2.4E-02	6.3	40.8	ITGB1; TUBB6; ITGA2; LAMP2; M6PR
Leukocyte transendothelial migration	4/112	3.5E-03	3.3E-02	6.8	38.7	VASP; ITGB1; PXN; MSN
Purine metabolism	4/129	5.8E-03	4.1E-02	5.9	30.4	NT5E; RRM2B; AK3; NME3
Human immunodeficiency virus 1 infection	5/212	6.6E-03	4.3E-02	4.5	22.5	PAK1; GNG2; PXN; GNG12; TAPBP
PI3K-Akt signaling pathway	7/354	3.6E-03	3.3E-02	3.8	21.3	ITGB1; GNG2; ITGA2; ITGA1; LAMC1; ITGB6; GNG12
MAPK signaling pathway	6/295	6.1E-03	4.1E-02	3.9	19.8	RPS6KA3; PAK1; FLNB; FLNC; GNG12; NFKB2

### Supplementary Table 5: Inter- and intra-molecular cross-links of ribosomal proteins observed in undifferentiated SH-SY5Y cells

Analysis of cross-links on a high-resolution structure of the ribosome (PDB ID: 4UG0) was performed using UCSF Chimera (version 1.15) and the software tool Xlink Analyzer (version 1.1.4). The Id, consisting of peptide a and peptide b as well as cross-link position within each peptide, the cross-linked protein, the position with the protein (AbsPos1 and AbsPos2) and the C $\alpha$ -pair distances (distance in Å) of the cross-link are shown.

Inter- and intra-molecular cross-links of ribosomal proteins in undifferentiated SH-SY5Y cells					
Id	Protein1	Protein2	AbsPos1	AbsPos2	distance
QHKPRFTTK-EHTALLKIEGVYAR-a5-b7	RL18A	RL35A	167	30	16.33
AVKFQR-DIPGLTDTTVPRR-a3-b7	RL24	RS6	78	127	18.09
KYAVK-IAPRPASGPIRPIVR-a3-b8	RL35	RL13	121	54	10.49
NVSKPR-GFSLEELRVAGIHK-a3-b7	RL36	RL13	23	82	14.46
NVSKPR-GFSLEELRVAGIHK-a3-b8	RL36	RL13	23	83	14.84
SVARIAK-ASGNYATVISHNPETKK-a4-b9	RL7A	RL8	250	138	27.89
YYTRLGNDFHNTK-FSPNSSNPIIVScGWDKLVK-a7-b10	RS17	RACK1	27	166	27.68
ITLTSRNVK-FVADGIFKAELNEFLTR-a5-b9	RS20	RS3	27	20	19.74
NVKGPPVR-QAVDVSPLRR-a4-b5	RS28	RS5	49	142	6.60
TQNVLGEKGR-IAVHcTVRGAK-a5-b6	RS3	RL11	60	74	124.86
EKYAK-EIKDILIQYDR-a3-b6	RS5	RS16	46	113	12.56
RPFEKSR-RFVNVVPTFGK-a4-b6	RS9	RS30	22	47	23.44
QIPRILGPLNK-AVDIPHMDIEALKK-a6-b7	RL10A	RL10A	125	86	15.81
RNPDTQWITKPVHK-FFEVLIDPFHKAIR-a7-b8	RL15	RL15	151	137	7.03
EYVRHFVK-APGTPHSHTKPYVR-a4-b7	RL18	RL18	151	162	9.70
EYKVVGR-cHTPPLYRMR-a4-b5	RL18A	RL18A	13	27	6.65
IEHIKHSK-HGVVPLATYMRIYK-a4-b7	RL21	RL21	97	29	7.58
VVITRLK-GQIQGKVVQVYR-a4-b6	RL26	RL26	108	70	11.27
KKEELLK-VLTVINQTQK-a4-b5	RL35	RL35	17	62	12.05
VTGGAASKLSK-QLDDLKVELSQLR-a6-b7	RL35	RL35	42	27	13.78
AMELLK VSK-KREELSNVLAAMRK-a5-b7	RL36	RL36	62	93	12.15
FcIWTESAFRK-GEESGNVTLPAVFK-a6-b8	RL4	RL4	255	23	13.64
EFNAEVHRK-NSVTPTDMMEMYKK-a5-b7	RL5	RL5	194	236	16.55
IDISNVK-AVDSQILPKIK-a4-b6	RL6	RL6	205	258	13.34
LLARAEK-GDVPTKRPPVLR-a4-b6	RL7A	RL7A	118	132	16.51
VTKAAGTK-DVQIGDIVTGEcRPLSK-a4-b9	RS11	RS11	149	128	19.40
RVLLGETGK-QPTIFQNKK-a5-b5	RS11	RS11	27	18	16.41
KIAFAITAIK-RAGELTEDEVER-a5-b6	RS18	RS18	30	61	14.76
RVLQALEGLK-LKVPEWVDTVK-a5-b6	RS19	RS19	107	34	16.10
ALAAFLKK-HKELAPYDENWFYTR-a4-b8	RS19	RS19	21	50	16.71
ALAAFLKK-ELAPYDENWFYTR-a4-b7	RS19	RS19	21	51	13.12
LGRLVK-SPYQEFTHLVKTHTR-a3-b8	RS2	RS2	69	272	8.22
ATVPKTEIR-LARHGLYEK-a5-b5	RS24	RS24	38	96	14.69
AQVIYTRNTK-EVPNYKLITPAVVSER-a5-b8	RS25	RS25	110	69	8.46
ATYDKLcK-LITPAVVSERLK-a4-b6	RS25	RS25	57	73	10.79
VQPIKLAR-VEFMDDTSRSIIR-a4-b7	RS28	RS28	10	39	11.38
VQPIKLAR-EGDVLTLLESEREAR-a4-b8	RS28	RS28	10	60	8.23
VTKVLGR-TGSQGQcTQVRVEFMDDTSR-a4-b10	RS28	RS28	18	31	8.27
EVQTNDLK-EVVKLIPDSIGKDIEK-a4-b9	RS3A	RS3A	179	192	14.29
VVDPFSKK-TLVTRTQGTK-a4-b5	RS3A	RS3A	25	52	13.70
TLVTRTQGTK-LITEDVQGKNcLTNFHGMDLTR-a5-b11	RS3A	RS3A	52	97	18.53
IDGKVR-GIPHLVTHDARTIR-a3-b7	RS4X	RS4X	75	142	18.81
HIRVR-KQVVNIPSFIVR-a3-b6	RS9	RS9	137	145	12.27
VKFTLAK-SRLDQELK-a4-b4	RS9	RS9	50	27	17.53

### Supplementary Table 6: Inter- and intra-molecular cross-links of ribosomal proteins observed in RA-differentiated SH-SY5Y cells

Analysis of cross-links on a high-resolution structure of the ribosome (PDB ID: 4UG0) was performed using UCSF Chimera (version 1.15) and the software tool Xlink Analyzer (version 1.1.4). The Id, consisting of peptide a and peptide b as well as cross-link position within each peptide, the cross-linked protein, the position with the protein (AbsPos1 and AbsPos2) and the C $\alpha$ -pair distances (distance in Å) of the cross-link are shown.

Inter- and intra-molecular cross-links of ribosomal proteins in RA-differentiated SH-SY5Y cells					
Id	Protein1	Protein2	AbsPos1	AbsPos2	distance
ARVITEEEK-YPMAVGLNKGHK-a5-b6	RL13	RL36	171	11	27.45
QHKPRFTTK-EHTALLKIEGVYAR-a5-b7	RL18A	RL35A	167	30	16.33
ILRR-IVEPYIAWGYPNLKS VNELIYK-a2-b11	RL19	RL7	103	146	170.04
GQSEEIQKK-RMATEVAADALGEEWK-a5-b8	RL24	RS6	67	39	20.20
AVKFQR-DIPGLTDTTVPRR-a3-b7	RL24	RS6	78	127	18.09
QPVIVKAK-ARVITEEEK-a4-b5	RL27A	RL13	124	171	15.58
SQKPVMVK-MPRYYPTEDVPR-a4-b6	RL28	RL6	124	118	17.73
YVLGYKQTLK-VYNYNHLMPTR-a5-b6	RL30	RL27	32	80	13.80
KYAVK-IAPRPASGPPIVR-a3-b8	RL35	RL13	121	54	10.49
NVSKPR-GFSLEELRVAGIHK-a3-b7	RL36	RL13	23	82	14.46
NVSKPR-GFSLEELRVAGIHKK-a3-b8	RL36	RL13	23	83	14.84
LTKHTK-VSRDTLYEAVR-a3-b6	RL36	RL10A	36	11	42.01
SVARIAK-ASGNYATVISHNPETKK-a4-b9	RL7A	RL8	250	138	27.89
KRVLLGETGK-NAKISSLLEE QFQQGK-a5-b8	RS11	RS8	26	163	20.35
IVVNLGRLNK-TYSYLTPDLWKETVFTK-a6-b9	RS15A	RS2	67	256	10.79
YYTRLGND FHTNK-FSPNSSNPIIVScGWDKLVK-a7-b10	RS17	RACK1	27	166	27.68
GYWGNKIGKPHTVPcK-QFGFIVLTTSAGIMDHEEAR-a8-b10	RS2	RS15A	176	108	39.38
ITLTSRNVK-FVADGIFKAELNEFLTR-a5-b9	RS20	RS3	27	20	19.74
KFMTNRLQR-TDITYPAGFMDVISIDKTGENFR-a5-b12	RS24	RS4X	16	90	16.65
NVKG PVR-QAVDV SPLRR-a4-b5	RS28	RS5	49	142	6.60
TQNVLGEKGR-IAVHcTVRGAK-a5-b6	RS3	RL11	60	74	124.86
EKYAK-EIKDILIQYDR-a3-b6	RS5	RS16	46	113	12.56
RPFEKSR-RFVNVVPTFGK-a4-b6	RS9	RS30	22	47	23.44
QIPRILG PGLNK-AVDIPHMDIEALKK-a6-b7	RL10A	RL10A	125	86	15.81
LNKNK-QIPRILG PGLNK-a3-b6	RL10A	RL10A	96	125	13.47
RNPDTQWITKPVHK-FFEVLIDPFHKAIR-a7-b8	RL15	RL15	151	137	7.03
EVYRHF GK-APGTPHSHTKPYVR-a4-b7	RL18	RL18	151	162	9.70
EYKVVGR-cHTPPLYRMR-a4-b5	RL18A	RL18A	13	27	6.65
IEHIKHSK-HGVVPLATYMRIYK-a4-b7	RL21	RL21	97	29	7.58
VVITRLK-GQQIGKVVQVYR-a4-b6	RL26	RL26	108	70	11.27
IKSFVK-NIDDGTS DRPYSHALVAGIDRYPR-a3-b12	RL27	RL27	71	40	16.83
EMGTPDVRIDTR-KRNEDESPNKLYTLVTYVPVTTFK-a6-b13	RL31	RL31	62	104	4.51
KKEELLK-VLTVINQTQK-a4-b5	RL35	RL35	17	62	12.05
AIFAGYKR-NQREHTALLK-a4-b5	RL35A	RL35A	13	25	8.78
FRSNLPAK-VIWKVTR-a4-b4	RL35A	RL35A	92	73	12.94
AMELLK VSK-EELSNVLAAMRK-a5-b6	RL36	RL36	62	94	8.83
FcIWTESAFRK-GESSGKNVTLPAVFK-a6-b8	RL4	RL4	255	23	13.64
LLNRFGMDK-ENPVYEKKPK-a5-b5	RL5	RL5	114	254	17.96
EFNAEVHRK-NSVTPDMMMEEMYKK-a5-b7	RL5	RL5	194	236	16.55
ENPVYEK-IEGMIVcAAYAHHELPKYGVK-a4-b11	RL5	RL5	253	80	12.81
IDISNVK-AVDSQILPKIK-a4-b6	RL6	RL6	205	258	13.34
LLARA EK-GDVPTKRPPVLR-a4-b6	RL7A	RL7A	118	132	16.51
MGVPYcIIK GK-VPPAINQFTQALDRQTATQLLK-a6-b11	RL7A	RL7A	183	87	5.79
VTKAAGTK-DVQIGDIVTVGECRPLSK-a4-b9	RS11	RS11	149	128	19.40
RVLLGETGK-OPTIFQNKK-a5-b5	RS11	RS11	27	18	16.41
KIAFAITAIK-RAGELTEDEVER-a5-b6	RS18	RS18	30	61	14.76
RVLQALEGLK-LKVPEWVDTVK-a5-b6	RS19	RS19	107	34	16.10

<b>Id</b>	<b>Protein1</b>	<b>Protein2</b>	<b>AbsPos1</b>	<b>AbsPos2</b>	<b>distance</b>
ALAAFLKK-HKELAPYDENWIFYTR-a4-b8	RS19	RS19	21	50	16.71
ALAAFLKK-ELAPYDENWIFYTR-a4-b7	RS19	RS19	21	51	13.12
LGRLVK-SPYQEFTDHLVKTHTR-a3-b8	RS2	RS2	69	272	8.22
ATVPKTEIR-LARHGLYEK-a5-b5	RS24	RS24	38	96	14.69
LVSKHR-AALQELLSK-a3-b5	RS25	RS25	102	91	14.53
AQVIYTRNTK-EVPNYKLITPAVVSER-a5-b8	RS25	RS25	110	69	8.46
AQVIYTRNTK-EVPNYKLITPAVVSERLK-a5-b9	RS25	RS25	110	70	11.68
ATYDKLcK-LITPAVVSERLK-a4-b6	RS25	RS25	57	73	10.79
VQPIKLAR-VEFMDDTSRSIIR-a4-b7	RS28	RS28	10	39	11.38
VQPIKLAR-EGDVLTLLESEREAR-a4-b8	RS28	RS28	10	60	8.23
VTKVLGR-TGSQGGcTQVRVEFMDDTSR-a4-b10	RS28	RS28	18	31	8.27
KTSYAQQHQQVR-KWQTMIEAHVDVK-a6-b7	RS3A	RS3A	158	123	13.68
EVQTNDLK-EVVNKLIPDSIGK-a4-b7	RS3A	RS3A	179	190	16.39
EVQTNDLK-EVVNKLIPDSIGKDIEK-a4-b9	RS3A	RS3A	179	192	14.29
VVDPFSKK-TLVTRTQGTK-a4-b5	RS3A	RS3A	25	52	13.70
TLVTRTQGTK-LITEDVQGNcLTNFGMDLTR-a5-b11	RS3A	RS3A	52	97	18.53
LITEDVQGK-FELGKLMELHGEGSSGK-a5-b9	RS3A	RS3A	91	232	12.50
IDGKVR-GIPHLVTHDARTIR-a3-b7	RS4X	RS4X	75	142	18.81
KGAKLTPEEEIILNK-QWYESHYALPLGRK-a8-b7	RS8	RS8	133	118	8.01
HIRVR-KQVVNIPSFIVR-a3-b6	RS9	RS9	137	145	12.27
VKFTLAK-SRLDQELK-a4-b4	RS9	RS9	50	27	17.53

### Supplementary Table 7: Inter- and intra-molecular cross-links of ribosomal proteins observed in RA/PMA-differentiated SH-SY5Y cells

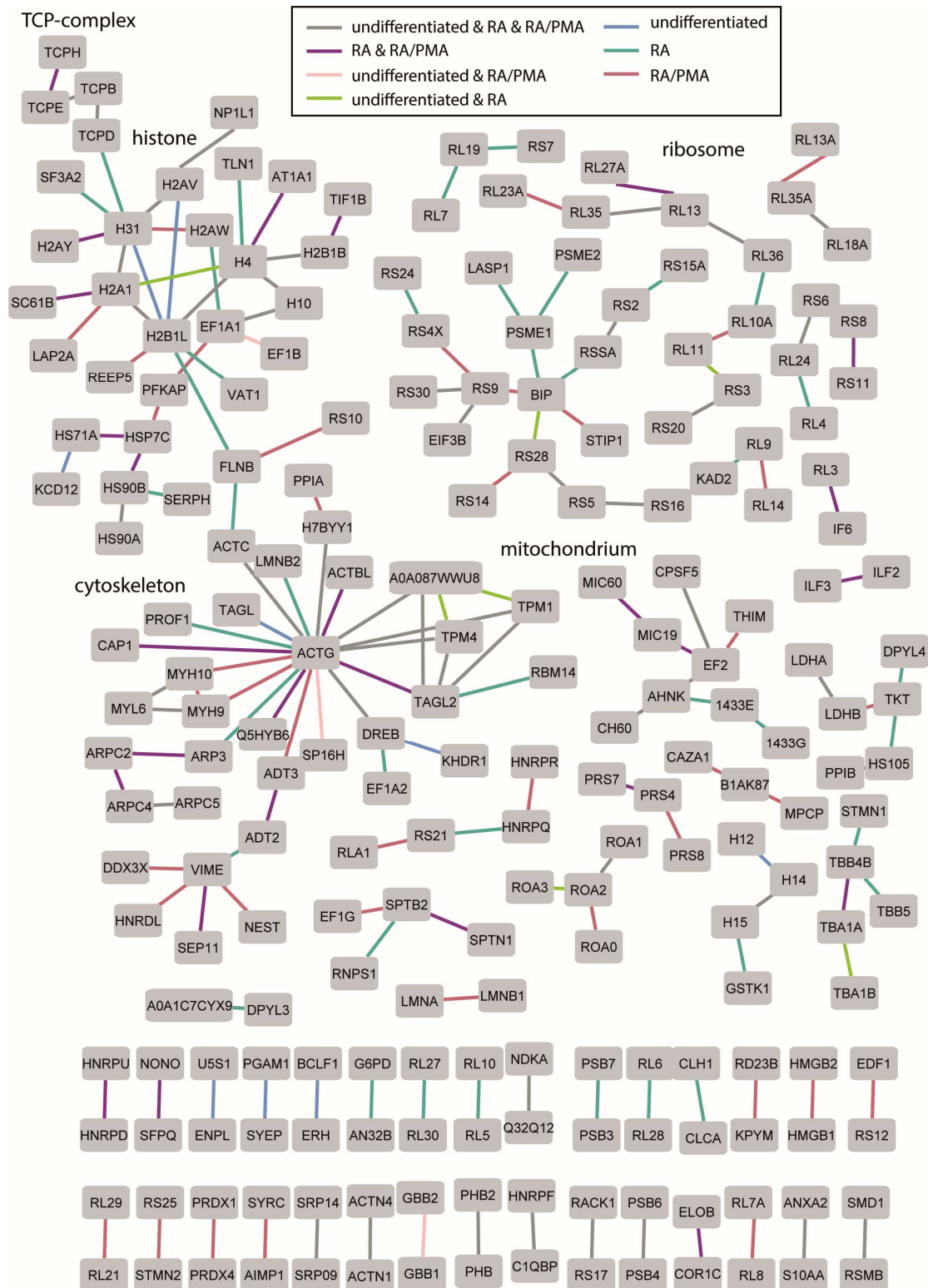
Analysis of cross-links on a high-resolution structure of the ribosome (PDB ID: 4UG0) was performed using UCSF Chimera (version 1.15) and the software tool Xlink Analyzer (version 1.1.4). The Id, consisting of peptide a and peptide b as well as cross-link position within each peptide, the cross-linked protein, the position with the protein (AbsPos1 and AbsPos2) and the C $\alpha$ -pair distances (distance in Å) of the cross-link are shown.

Inter- and intra-molecular cross-links of ribosomal proteins in RA/PMA-differentiated SH-SY5Y cells					
Id	Protein1	Protein2	AbsPos1	AbsPos2	distance
AAKVLEQLTGQTPVFSK-FSVcVLGDQQHcDEAK-a9-b8	RL11	RL10A	45	71	47.84
NVYKK-TGAAPIIDVVRSGYYK-a3-b8	RL13	RL27A	162	103	12.87
QVLLGRK-EHTALLKIEGVYAR-a4-b7	RL13A	RL35A	30	30	13.67
LVAIVDVIDQNR-[MKTILSNQTVDIPENV DITLK-a6-b11	RL14	RL9	30	11	33.51
QHKPRFTTK-EHTALLKIEGVYAR-a5-b7	RL18A	RL35A	167	30	16.33
RINVR-YESLKGVDPK-a3-b5	RL21	RL29	91	34	13.57
GQSEEIQKK-RMATEVAADALGEEWK-a5-b8	RL24	RS6	67	39	20.20
AVKFQR-DIPGLTDTTTPRR-a3-b7	RL24	RS6	78	127	18.09
KYAVK-IAPRPASGPPIVR-a3-b8	RL35	RL13	121	54	10.49
VELSQLRVAK-LDHYAIIKFPLTTESAMK-a5-b9	RL35	RL23A	31	80	11.30
NVSKPR-GFSLEELRVAGIHK-a3-b7	RL36	RL13	23	82	14.46
NVSKPR-GFSLEELRVAGIHKK-a3-b8	RL36	RL13	23	83	14.84
SVARIAK-ASGNYATVISHNPETKK-a4-b9	RL7A	RL8	250	138	27.89
DPYRFK-NFGIGQDIQPKR-a3-b6	RL8	RL7A	68	44	10.67
KRVLLGETGK-NAKISSLLEEQQQK-a5-b8	RS11	RS8	26	163	20.35
SGMKIGR-EGDVLTLLESEREAR-a4-b8	RS14	RS28	126	60	22.62
YYTRLGNDFTNKR-FSPNSSNPIIVScGWDKLVK-a7-b10	RS17	RACK1	27	166	27.68
ITLTSRNVK-FVADGIFKAELNEFLTR-a5-b9	RS20	RS3	27	20	19.74
NVKGpVR-QAVDVSPLRR-a4-b5	RS28	RS5	49	142	6.60
LAAKQSSG]-LFEGNALLRR-a5-b5	RS4X	RS9	261	76	21.70
EKYAK-EIKDILIQYDR-a3-b6	RS5	RS16	46	113	12.56
RPFEKSR-RFVNVVPTFGK-a4-b6	RS9	RS30	22	47	23.44
QIVSGSR-LWNTLGVcKYTVQDESHSEWVScVR-a4-b13	RACK1	RACK1	123	144	12.48
DETNYGIPQR-DKTIIMWKLTR-a5-b6	RACK1	RACK1	53	43	6.54
QIPRILGPLNK-AVDIPHMDIEALKK-a6-b7	RL10A	RL10A	125	86	15.81
YDGIILPGK-VREYELRK-a5-b4	RL11	RL11	175	93	7.00
RNPDTQWITKPVHK-FFEVLIDPFHKAIR-a7-b8	RL15	RL15	151	137	7.03
EYKVVGR-cHTPPLYRMR-a4-b5	RL18A	RL18A	13	27	6.65
IEHIKHSK-HGVVPLATYMRIYK-a4-b7	RL21	RL21	97	29	7.58
VVITRLK-GQQIGKVVQVYR-a4-b6	RL26	RL26	108	70	11.27
IKSFVK-NIDDGTSRDPYSHALVAGIDRYPR-a3-b12	RL27	RL27	71	40	16.83
GKKKEELLK-QLDDLKVELSQLR-a5-b7	RL35	RL35	16	27	17.12
KKEELLK-VLTVINQTQK-a4-b5	RL35	RL35	17	62	12.05
VTGGAASKLSK-QLDDLKVELSQLR-a6-b7	RL35	RL35	42	27	13.78
AIFAGYKR-NQREHTALLK-a4-b5	RL35A	RL35A	13	25	8.78
AMELLKVSKEELS NVLAAMRK-a5-b6	RL36	RL36	62	94	8.83
FcIWTESAFRK-GESSGKNVTLPAVFK-a6-b8	RL4	RL4	255	23	13.64
LLNRFMDK-ENPVYEKKPK-a5-b5	RL5	RL5	114	254	17.96
EFNAEVHRK-NSVTPDMMEEMYKK-a5-b7	RL5	RL5	194	236	16.55
ENPVYEK-IEGDMIVcAAYAHHELPKYGVK-a4-b11	RL5	RL5	253	80	12.81
IDISNVK-AVDSQILPKIK-a4-b6	RL6	RL6	205	258	13.34
LLARA EK-GDVPTKRPPVLR-a4-b6	RL7A	RL7A	118	132	16.51
HWGGNVLPK-LKVPPAINQFTQALDR-a5-b8	RL7A	RL7A	241	82	16.49
VTKAAGTK-DVQIGDIVTGEcRPLSK-a4-b9	RS11	RS11	149	128	19.40
RVLLGETGK-OPTIFQNKK-a5-b5	RS11	RS11	27	18	16.41
KIAFAITAIK-RAGELTEDEVER-a5-b6	RS18	RS18	30	61	14.76
VITIMQNPRQYK-DGKYSQVLANGLDNK-a6-b8	RS18	RS18	73	100	15.04



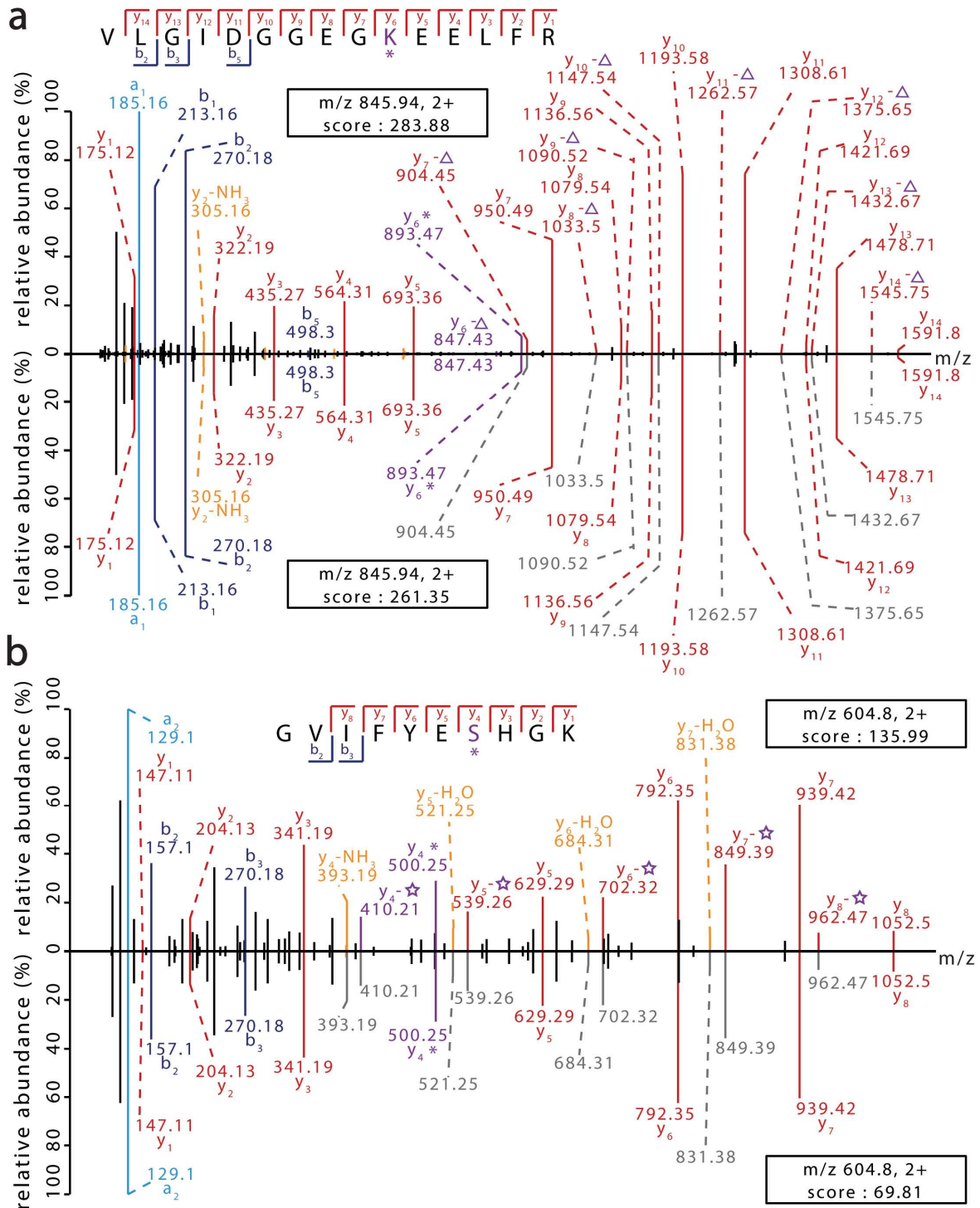
<b>Id</b>	<b>Protein1</b>	<b>Protein2</b>	<b>AbsPos1</b>	<b>AbsPos2</b>	<b>distance</b>
RVLQALEGLK-LKVPEWVDTVK-a5-b6	RS19	RS19	107	34	16.10
ALAAFLKK-HKELAPYDENWFYTR-a4-b8	RS19	RS19	21	50	16.71
ALAAFLKK-ELAPYDENWFYTR-a4-b7	RS19	RS19	21	51	13.12
AEDKEWMPVTK-EVATAIRGAILAK-a6-b7	RS2	RS2	61	153	19.69
LGRLVK-SPYQEFTDHLVKTHTR-a3-b8	RS2	RS2	69	272	8.22
VKGPVR-TPVEPEVAIHRIR-a3-b7	RS20	RS20	53	16	23.74
ATVPKTEIR-LARHGLYEK-a5-b5	RS24	RS24	38	96	14.69
AQVIYTRNTK-EVPNYKLITPAVVSER-a5-b8	RS25	RS25	110	69	8.46
AQVIYTRNTK-EVPNYKLITPAVVSERLK-a5-b9	RS25	RS25	110	70	11.68
ATYDKLcK-LITPAVVSERLK-a4-b6	RS25	RS25	57	73	10.79
VQPIKLAR-VEFMDDTSRSIIR-a4-b7	RS28	RS28	10	39	11.38
VQPIKLAR-EGDVLTLLESEREAR-a4-b8	RS28	RS28	10	60	8.23
VTKVLGR-TGSQGGcTQVRVEFMDDTSR-a4-b10	RS28	RS28	18	31	8.27
EVQTNDLK-EVVKLIPDSIGK-a4-b7	RS3A	RS3A	179	190	16.39
EVQTNDLK-EVVKLIPDSIGKDIEK-a4-b9	RS3A	RS3A	179	192	14.29
VVDPFSKK-TLVTRTQGTK-a4-b5	RS3A	RS3A	25	52	13.70
TLVTRTQGTK-LITEDVQGNcLTNFGMDLTR-a5-b11	RS3A	RS3A	52	97	18.53
LVRELEK-AQQNNVEHKVETFSGVYK-a4-b9	RS7	RS7	83	170	19.37
HIRVR-KQVVNIPSFIVR-a3-b6	RS9	RS9	137	145	12.27
VKFTLAK-SRLDQELK-a4-b4	RS9	RS9	50	27	17.53

## 6.2 Supplementary Figures



**Supplementary Figure 1: Protein interaction network obtained from formaldehyde cross-linking.**

The protein interaction network obtained from in-cell formaldehyde cross-linking is shown. Protein interactions observed in specific culture conditions of SH-SY5Y are colored (see legend for details). The protein names were derived from uniprot identifier names. Note that lines represent at least one observed inter-molecular cross-links and that information on cross-linked residues is not included.



**Supplementary Figure 2: Example spectra of DEPC labeled peptides.**

The spectra are annotated without (bottom spectrum) and with (top spectrum) neutral loss from DEPC modified residues. Y- (red and orange), a- (cyan) and b-ions (blue) are assigned. The MaxQuant peptide score increases when neutral loss is included in the database search parameters **a**) Example spectrum of a peptide containing a CET-modified lysine. **b**) Example spectrum of peptide containing a CET-modified serine. (Figure adapted from Barth and Schmidt 2020 [153].)

## Acknowledgment

First, I would like to thank Carla for giving me the opportunity to work in the exciting field of mass spectrometry, for supervision of this work and importantly my thesis. I am grateful for having her as a great mentor and her valuable feedback. Without her continuous support, I would not have learnt as much as I did during the past years.

Second, I would like to thank the second and third reviewer for reviewing this thesis.

Next, I would like to thank the whole group: Julian, Julia, Melissa, Lea, Tommy, Christian, Til, Jenny and Leonie for fruitful discussions, the nice time and support. Special thank goes to Til, Tommy and Julian for reading this thesis and providing feedback. In addition, I would like to thank the former lab members Sven and Sabine for their encouragement. I would like to thank all for the fun in the breaks with ice cream, the nice retreats, conferences, Christmas parties and more...

Very special thanks go to Fritzi for reading this thesis very carefully and giving spelling and grammar advice.

I would also like to thank my family and friends, for support, encouragement, fun and discussion.

At last, I would like to thank my fiancé Sven for his outstanding support and interest in all steps of my thesis, including technical support, reading this thesis aloud, encouragement and much more.

## Curriculum vitae

**Name:** Marie Bärbel Barth

**Nationality:** German

**Date of birth:** 17.02.1994

**Place of birth:** Dresden

### Education

Since 05/2018

**PhD student**

Institute for Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg, Halle  
Interdisciplinary Research Centre HALOmem, Charles Tanford Protein Center

10/2016 – 03/2018

**Master Biochemistry**

Institute for Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg, Halle  
Title of the master thesis: *Chemical modification of proteins for structure elucidation.*  
Degree: M. Sc., Biochemie

10/2013 – 08/2016

**Bachelor Biochemie**

Institute for Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg, Halle  
Title of the bachelor thesis: *Expression von offenen Leserahmen von humanen endogenen Retroviren in Hodgkinlymphomzellen.*  
Degree: B. Sc. Biochemie

2013

**Abitur**

Thomas-Mann-Gymnasium, Lübeck

## Publication list

8. Sethumadhavan, S., **Barth, M.**, Spaapen, R. M., Schmidt, C., Trowitzsch, S., Tampé, R., (2022) Viral immune evasins impact antigen presentation by allele-specific trapping of MHC I at the peptide-loading complex, **Scientific Reports**, accepted.
7. Hofmann, T., **Barth, M.**, Meister, A., Kastritis, P. L., and Schmidt, C. (2021) Thin-Layer Chromatography and Coomassie Staining of Phospholipids for Fast and Simple Lipidomics Sample Preparation, **Analysis & Sensing** 1, 171-179.
6. Sekulovski, S., Devant, P., Panizza, S., Gogakos, T., Pitiriciu, A., Heitmeier, K., Ramsay, E. P., **Barth, M.**, Schmidt, C., Tuschl, T., Baas, F., Weitzer, S., Martinez, J., and Trowitzsch, S. (2021) Assembly defects of human tRNA splicing endonuclease contribute to impaired pre-tRNA processing in pontocerebellar hypoplasia, **Nature Communications** 12, 5610.
5. Wittig, S., Ganzella, M., **Barth, M.**, Kostmann, S., Riedel, D., Pérez-Lara, Á., Jahn, R., and Schmidt, C. (2021) Cross-linking mass spectrometry uncovers protein interactions and functional assemblies in synaptic vesicle membranes, **Nature Communications** 12, 858.
4. **Barth, M.**, and Schmidt, C. (2021) Quantitative Cross-Linking of Proteins and Protein Complexes, **Methods in Molecular Biology** 2228, 385-400.
3. **Barth, M.**, Bender, J., Kundlacz, T., and Schmidt, C. (2020) Evaluation of NHS-Acetate and DEPC labelling for determination of solvent accessible amino acid residues in protein complexes, **Journal of Proteomics** 222, 103793.
2. **Barth, M.**, and Schmidt, C. (2020) Native mass spectrometry-A valuable tool in structural biology, **Journal of Mass Spectrometry** 55, e4578.
1. **Barth, M.**, Gröger, V., Cynis, H., and Staege, M. S. (2019) Identification of human endogenous retrovirus transcripts in Hodgkin Lymphoma cells, **Molecular Biology Reports** 46, 1885-1893.

## Unpublished work

**Barth, M.**, Schmidt, C., Quantitative proteomics and in-cell cross-linking reveal cellular reorganization during early neuronal differentiation of SH-SY5Y cells induced by retinoic acid or phorbol-12-myristat-13-acetate, **in revision**.

## Oral presentations

7. **HALOmem Retreat**, Lutherstadt Wittenberg (15. – 17.7.2021)  
Quantitative proteomics and in-cell cross-linking reveal cellular reorganization during early neuronal differentiation of SH-SY5Y cells induced by retinoic acid or phorbol-12-myristat-13-acetate
6. **Statusseminar HALOmem**, Halle (12.11.2020)  
Proteomic evaluation of neuronal model systems for cross-linking experiments
5. **Statusseminar HALOmem**, Halle (8.11.2018)  
Covalent labelling of surface accessible amino acids to elucidate the orientation of membrane proteins in a lipid bilayer
4. **Symposium on structural proteomics**, Berlin (10. – 12.10.2018)  
Chemical modification of proteins for structure elucidation
3. **Ringberg Retreat (Max-Planck-Institute Göttingen)**, Kreuth (29.5. – 2.6.2018)  
Chemical modification of proteins to study the architecture of protein complexes
2. **HALOmem Retreat**, Gohrisch (24. – 26.5.2018)  
Protein labelling for structure elucidation
1. **Statusseminar HALOmem**, Halle (2.11.2017)  
Covalent labelling of proteins for structure elucidation



## Poster presentations

7. **53rd annual conference of the DGMS** (Deutsche Gesellschaft für Massenspektrometrie), Münster (1. – 4.3.2020)  
Chemical modification of proteins for structure elucidation and analysis of protein orientation in lipid bilayers
6. **Symposium on structural proteomics**, Göttingen (3. – 6.11.2019)  
Chemical modification of proteins for structure elucidation and analysis of protein orientation in lipid bilayers
5. **Sanibel Conference**, St. Petersburg, Florida (USA) (23. – 27.01.2019)  
Chemical labelling of proteins for structure elucidation and determination of the protein orientation within membranes  
**Kongressreisestipendium**, Deutscher Akademischer Austauschdienst (DAAD)
4. **Statusseminar HALOmem**, Halle (8.11.2018)  
Covalent labelling of proteins for structure elucidation
3. **Symposium on structural proteomics**, Berlin (10. – 12.10.2018)  
Chemical modification of proteins for structure elucidation
2. **Statusseminar HALOmem**, Halle (2.11.2017)  
Covalent labelling of proteins for structure elucidation
1. **Symposium on structural proteomics**, Wien (23. – 25.10.2017)  
Covalent labelling of proteins for structure elucidation

## **Affidavit**

I hereby declare that this thesis has been written independently and with no other sources and aids than quoted. Parts directly or indirectly obtained from foreign sources are indicated. In addition, material provided, or experiments performed by other persons are indicated. The submitted written version of the thesis corresponds to that on the electronic storage medium. Furthermore, I assure that this doctoral thesis has not been submitted as a thesis elsewhere. I have not submitted any other doctoral thesis before this thesis.

Halle (Saale), 20.01.2022

Marie Barth

## **Eidesstattliche Erklärung**

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Die aus fremden Quellen direkt oder indirekt übernommenen Stellen sind als solche kenntlich gemacht. Ebenfalls sind alle von anderen Personen bereitgestellten Materialien oder erbrachten Experimente als solche gekennzeichnet. Die eingereichte schriftliche Fassung der Arbeit entspricht der auf dem elektronischen Speichermedium. Weiterhin versichere ich, dass die vorliegende Promotionsarbeit noch nicht als Abschlussarbeit an anderer Stelle eingereicht wurde. Ich habe vor dieser Arbeit keine andere Promotionsarbeit eingereicht.

Halle (Saale), den 20.01.2022

Marie Barth