Mass Spectrometric Profiling of Lysosomal Protein Interactions, Structures, and Half-lives via Cross-Linking and Pulse Stable Isotope Labeling

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Jasjot Singh

aus

Bremen, Deutschland

Bonn, 2022

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Volkmar Gieselmann

2. Gutachter: Prof. Dr. rer. nat. Christoph Thiele

Tag der mündlichen Prüfung: 13.06.2022
Erscheinungsjahr: 2022

Acknowledgements

I owe special gratitude to Prof. Dr. Volkmar Gieselmann for giving me the opportunity to be part of his working group, for his supervision, time, and interest in my projects, as well as being my first supervisor. I always enjoyed discussing my results with him and sincerely thank him for his time in my grant writings and evaluation letters. Additionally, I would like to thank Prof. Dr. rer. nat. Christoph Thiele for being my second supervisor, and my third and fourth supervisors, PD Dr. Marianne Engeser and Prof. Dr. Pavel Kroupa, for their time and being part of my examination procedure.

Special thanks go to Dr. rer. nat. Dominic Winter for being the best supervisor one can have as a young PhD student and being a friend who supported my ideas throughout the four years. Thanks, Dominic, for your time writing grants and papers with me, patience and efforts teaching me, being my mentor, giving me the opportunities to be part of all the great conferences, and always taking time and listening. Your ideas were the foundation of 4 years of my work I pursued with passion and conviction and would not have been possible without your experience and expertise in the field of mass spectrometry and biology.

This work would not have been possible without the help of numerous people involved in my projects, including my bachelor and master students (Lucas Gudermann, Hang My Dang, and Christine Möller), our technician Norbert Rösel, my senior colleagues who helped me during my first year and our collaboration partners. I consider myself very lucky to be part of an amazing working atmosphere during these years, where colleagues became good friends. I will miss all the discussions and fun we had together in the lab, as well as in our free time doing barbecues, celebrating carnival, or being on a canoeing trip. Thanks a lot guys for the great time!

Last but not least, I would like to thank my scholarship donors “Konrad-Adenauer Stiftung” and “Studienstiftung des deutschen Volkes” for the financial support throughout my Bachelors, Masters, and PhD, but above all for giving me the opportunity to study abroad, be part of all the great interdisciplinary events and international conferences.

List of Contents:

Abbreviations and Acronyms

"Å"	Angstrom
"°C"	Celsius
"β"-hex	Beta-hexosaminidase
ACN	Acetonitrile
AcOH	Acetic acid
AGC	Automatic gain control
AMPK	5’-AMP-activated protein kinase
AP	Adaptor protein complexes
APCI	Atmospheric-pressure chemical ionization
APS	Ammonium persulfate
ATG5-KO	Autophagy related gene 5
Bis-Tris	(2-hydroxyethyl) amino-tris (hydroxymethyl) methane
BORC1	BLOC-1-related complex)
BS3	Bis(sulfosuccinimidyl)suberate
BSA	Bovine serum albumin
BTZ	Bortezomib
BuUrBu	1,3-bis-(4-oxo butyl)-urea group
$CaCl_2$	Calcium chloride
CHEVI	Class C homologs in endosome vesicle interaction
CI	Chemical ionization
CID	Collision-induced dissociation
Co-IP	Co-immunoprecipitation
CO_2	Carbon dioxide
COA	Chicken ovalbumin
CORVET	Class C core vacuole/endosome tethering
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	Clustered regularly interspaced short palindromic repeats ribonucleic acid
Cryo-EM	Cryogenic electron microscopy
CSM	Cross-link spectral matches
CV	Coefficient of variation
Da	Dalton
DDA	Data dependent acquisition
DIA	Data independent acquisition
DMSO	Dimethyl sulfoxide
DPBS	Dulbecco`s phosphate buffered saline
DR	Disrupted
DSBU	Disuccinimidyl dibutyric urea
DSS	Disuccinimidyl suberate
DSSO	Disuccinimidyl sulfoxide
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EEF	Early endosome enriched fraction
EL	Eluate
EM	Electron microscopy
ER	Endoplasmic reticulum
ESCRT	Endosomal sorting complex required for transport
ESI	Electrospray ionization
f.c.	Final concentration
FA	Formic acid
FAB	Fast atom bombardment atmospheric-pressure photoionization
FASP	Filter aided sample preparation
FCS	Fetal calf serum
FDR	False discovery rate
FT	Flow-through
FTICR	Fourier-transform ion cyclotron resonance
G/g	Gravitational constant

GO	Gene ontology
GPCRs	G-protein-coupled receptors
GSK3β	Glycogen synthase kinase 3β
h	Hour
HCD	High collision dissociation
HCl	Hydrogen acid
HEK293	Human embryonic kidney cells 293
HEPES	Hydroxyethyl piperazineethanesulfonic acid acid
HOPS	Homotypic fusion and protein sorting
HPLC	High-performance liquid chromatography
iBAQ	Intensity-based absolute quantification
IN	Input
IP	Immunoprecipitation
ISD	In solution digest
IT	Intact/ion trap
iTRAQ	Isobaric tags for relative and absolute quantitation
K	Known
K_2HPO_4	Potassium hydrogen phosphate
KCl	Potassium chloride
KO	Knock out
LC	Liquid chromatography
LDL	Low-density lipoproteins
LEF	Lysosome enriched fraction
LS	Lysosome/lysosomal
LSD	Lysosomal storage diseases
LYNUS	Lysosomal nutrient sensing
LYS	Lysosome/lysosomal
MA	M (log ratio) and A (mean average) scales
MALDI	Matrix assisted laser desorption ionization
MC	Manually curated
MgAc	Magnesium acetate
$MgCl_2$	Magnesium chloride
MGM-KO	Multiple glycosylation machinery related genes knock out [C1GALT1C1 (C1GALT1-specific chaperone 1), GNPTAB (N-acetylglucosamine-1-phosphotransferase subunits alpha/beta), MGAT1 (alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase), and ALG3 (Dol-P-Man:Man(5)GlcNAc(2)-PP-Dol alpha-1,3-mannosyltransferase)]
min	Minute
mM	Millimolar
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
mTORC1	Mechanistic target of rapamycin complex1
NaCl	Sodium chloride
NH_4HCO_3	Ammonium hydrogen carbonate
NHS	N-hydroxysuccinimide
NK	Not known
nm	Nanometer
NPC1	Niemann-Pick type C1
OEP	Organelle enriched pellet
OT	Orbitrap
PAM	Protospacer adjacent motif
PCA	Principal component analysis
PCR	Polymerase chain reaction
PM	Plasma membrane
PNS	Post nuclear supernatant
PPIs	Protein-protein interactions
PPL	Porcine pancreas lipase
PPM	Parts per million
PSM	Peptide spectrum match

PSTA	Pepstatin A
PTM	Post translational modification
QP	Quadrupole
R	Replicate
rpm	Rounds per minute
RT	Room temperature
SCX	Strong cation exchange chromatography
SDGC	Sucrose density gradient centrifugation
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SILAC	Stable isotope labeling by/with amino acids in cell culture
SMOAC	Sequential enrichment using metal oxide affinity chromatography
SN	Supernatant
SNAREs	Soluble NSF attachment protein receptors
SPIONs	Superparamagnetic iron oxide nanoparticles
STAGE	Stop-and-go-extraction
$T_{1/2}$	Half-life
TBST	Tris-buffered saline Tween 20
TEMED	N,N,N’,N’-Tetramethylethylenediamine
TFA	Trifluoroacetic acid
TFEB	Transcription factor EB
TMEM192	Transmembrane protein
TMT	Tandem mass tag
TOF	Time-of-flight
v-ATPase	Vacuolar-type adenosine triphosphatase
W/WA	Wash
WCL	Whole cell lysate
WT	Wild-type
XL	Cross-linking/cross-linker
z	Charge

Summary

In mammalian cells, the lysosome is the central organelle for the degradation of macromolecules and recycling of their building blocks. Mutations in almost every known lysosomal hydrolase lead to a group of rare but devastating diseases known as lysosomal storage diseases, demonstrating lysosomal protein's pathogenic significance. Besides, in recent years, it was reported that lysosomal proteins and such interacting with the lysosome play key roles in a variety of common diseases such as cancer and neurodegenerative disorders, leading to an increasing interest in the characterization of the lysosomal proteome.

For the investigation of the lysosomal proteome on a large scale, lysosome enriched fractions were analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Systematic comparisons of different lysosome enrichment methods revealed a significant increase of lysosomal proteins by up to 118-fold compared to whole-cell lysate using the superparamagnetic iron oxide nanoparticles method. An approach that was also found to be superior in terms of preserving intact lysosomes and enriching them in large quantities. The established enrichment methodology was then used to study lysosomal structures and their interactions using cross-linking MS. Moreover, it was used to determine protein half-lives by using pulse stable isotope labeling with amino acids in cell culture (pSILAC) in combination with data-independent acquisition (DIA) MS/MS.

To further better characterize and validate cross-linking experiments, antibodies specific against commercially available cross-linkers were raised and affinity-purified from rabbits. Those were successfully applied in western blotting and immunocytochemistry experiments. The results further demonstrated the ability of the antibodies to retain peptides modified with water-quenched dead-links and cross-linker-modified proteins in immunoprecipitation assays (most likely also through water-quenched dead-links).

The presented data cover the first large-scale cross-linking study of lysosome enriched fractions. Analysis of cross-links revealed a highly interconnected network of 847 proteins forming 1,024 protein-protein interactions (PPIs), including 67 % potentially novel PPIs, of which two were confirmed by co-immunoprecipitation (co-IP). Additionally, 161 cross-links mapped on 34 lysosomal proteins confirm known crystal structures in a physiological state, as well as validate predicated AlphaFold models and provide evidence for a novel tetrameric model of palmitoyl-protein thioesterase 1. For the lysosome-interacting proteins flotillin 1 and 2, the first heterodimeric structure, as well as a 38 oligomeric assembly were proposed. Further, the investigation of their putative vesicular cargo led to the identification of 328 proteins including 50 receptors. For instance, all three members of latrophilins (G-protein-coupled receptors) were identified to be endocytosed in a flotillin dependent manner.

In addition, the pSILAC DIA MS/MS resulted in the identification of 33,968 half-lives for 6,915 unique proteins, presenting the most extensive turnover study so far. The study included a wild-type condition (control), representing the steady-state and different perturbational cellular states such as the impairment of autophagosome-lysosome fusion, proteasome activity, cathepsin activity, cholesterol homeostasis, and protein glycosylation. The results revealed a median half-life of 18 h in wild-type cells, with significant half-life differences of up to 30 h for specific proteins among the different perturbational cellular states. For example, a drastic increase of half-lives upon proteasomal inhibition or a decrease upon cathepsin D inhibition was observed. The data suggest that the regulation of cellular degradational processes are the primary driving factors influencing protein half-lives.

1. Introduction

1.1 The Lysosome - Recycling Organelle of the Cell

Proteins, lipids, polysaccharides, and nucleic acids are the essential components of the cell. These molecules are subject to constant turnover, so the cell must be able to degrade them continuously. To accomplish this, cells have several mechanisms, with lysosomes being the central degrading organelle [1, 2]. The word lysosome originated from the Greek words: Lysis (destroy) and Soma (body). To degrade the mentioned macromolecules, lysosomes receive their cargo through various routes, primarily dependent on vesicular transport [3]. While extracellular material reaches lysosomes via phagosomes and endosomes, intracellular substrates are delivered by autophagosomes. In addition, proteins can be transported directly across the lysosomal membrane via chaperone-mediated autophagy [3]. Due to the great structural diversity of these molecules, their degradation requires a set of 60-70 lysosomal hydrolases with different specificities such as proteases, glycosidases, nucleases, and lipases [1, 2] **(Figure 1)**. These reach their maximal activity inside the lumen of lysosomes at an acidic pH (between pH 4.5-5), which is maintained through a vacuolar-type adenosine triphosphatase (v-ATPase) [4]. The end products of the degraded macromolecules are recycled and transported back to the cytosol via more than 30 transporters [1, 2] **(Figure 1)**.

For a long time, the lysosome was considered a purely digestive organelle and often referred to as a simple “waste-disposal system” of the cell [5]. Furthermore, it has been regarded as a “housekeeping” organelle, acting regardless of the cell’s status and performing its degradative function largely isolated from other cellular organelles. However, this restricted view has changed dramatically. Recent results have demonstrated the vital role of lysosomes beyond degradation, including gene regulation, metabolic signaling, repair of the plasma membrane, immunity, migration, and cell adhesion [2]. Notably, the lysosome can sense the cell's nutritional state by integrating information about cytosolic and lysosomal amino acid levels, growth factor signaling, cholesterol, and glucose levels. The complex machinery, which is largely localized within and on the cytosolic surface of the lysosomal membrane, has been termed lysosomal nutrient sensing (LYNUS) machinery [3] **(Figure 1)**. The integration of these multiple metabolic inputs within LYNUS results in either activation or inactivation of one of the most important and best-known complexes, the "mechanistic Target of Rapamycin Complex 1" (mTORC1) on the cytosolic surface of lysosomes, which is critical for cell survival and growth. Apart from mTORC1, also other kinases act as key players of cellular metabolism, such as 5’-AMP-activated protein kinase (AMPK) or the glycogen synthase kinase 3β (GSK3β), can be activated on the lysosomal surface [6, 7]. These signaling events are crucial for overall cellular homeostasis by regulating mechanisms such as cellular differentiation and growth and impact the lysosomal proteome by linking nutrient availability to lysosomal biogenesis. Hereby, dependent on the amino acid abundance, mTORC1 phosphorylates the transcription factor EB (TFEB) [8], also known as the master regulator of lysosomal biogenesis, regulating TFEB translocation to the nucleus and hence its transcriptional activity [9].

1.1.1 The Lysosomal Interactome

Several studies (>60) have already investigated the lysosomal proteome by mass spectrometry [1]. In this process, a large number of putative lysosomal proteins (>300) were identified, and for many of these, localization to the lysosomes has been verified by biological methods [10]. Yet, apart from the qualitative determination of the lysosomal proteome, no studies have so far systematically explored interactions of lysosomal proteins on a large scale. Characterization of individual multimeric protein complexes localized on the lysosomal membrane, such as the mTORC1 complex, demonstrates such interactions exists and have high functional significance [9, 11]. Other examples of protein complexes localizing on the lysosomal membrane include: v-ATPase (vacuolar-type ATPase) complex [12], Ragulator complex [13], BORC1 (BLOC-1-related complex) [14], HOPS (homotypic fusion and protein sorting) [15], CORVET (class C core vacuole/endosome tethering) [16], and AP (adaptor protein) complexes, assemblies of SNAREs (soluble NSF attachment protein receptors), and the ESCRT (endosomal sorting complex required for transport) complex [17]. The mentioned complexes fulfill a variety of functions such as the regulation of lysosomal transport, fusion with other vesicles, repairing lysosomes, and transport of proteins, among others [2]. Further, given the large number of substances degraded by lysosomes and provided to the cell, one must assume that other regulatory processes beyond mTORC1 rely on lysosomal and cytosolic protein interactions. An assumption further supported by the view of lysosome's diverse and vital role beyond degradation and metabolic signaling in other biological processes like repair of the plasma membrane, antigen presentation, exocytosis, cell adhesion, and apoptosis. Apart from this, lysosomes have been shown to build membrane contact sites with other organelles, including endoplasmic reticulum [18], Golgi complex [19], peroxisomes [20], and mitochondria [21] **(Figure 1)**. Based on the diversity of lysosomal functions, it has to be assumed that yet unknown interaction partners of known and unknown complexes, as well as among the lysosome and other organelles are yet to be discovered.

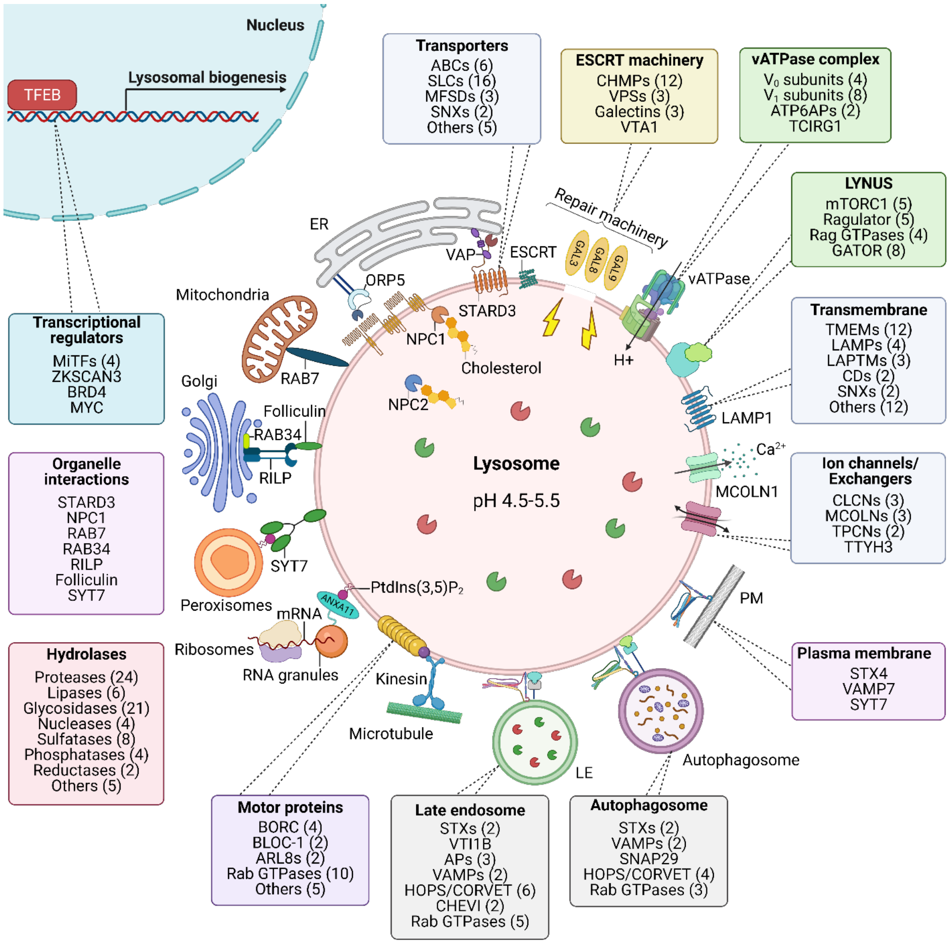


Figure 1: **Draft of the lysosomal proteome.** Classification of the lysosomal proteome into intraluminal, membrane and lysosome associated/interacting proteins. Localization and number of members for each of the protein classes and complexes have been indicated in the boxes. **Abbreviations:** ESCRT: endosomal sorting complex required for transport; HOPS: homotypic fusion and protein sorting; CORVET: class C core vacuole/endosome tethering; CHEVI: class C homologs in endosome–vesicle interaction; LYNUS: lysosomal nutrient sensing machinery; PM: plasma membrane; LE: late endosomes. Figure has been copied from [1].

1.2 Lysosomal Storage Disorders

Genetic mutations have been described for almost all lysosomal hydrolases in humans, leading to the accumulation of the corresponding intermediates of the affected degradation pathways in lysosomes [22]. These diseases are therefore referred to as lysosomal storage diseases (LSDs). Due to the large number of hydrolases, more than ~70 of these diseases are known, with an estimated cumulative prevalence of 1/5000 live births [22]. For example, defects in the enzymes that degrade glycosaminoglycans lead to the accumulation of glycosaminoglycan fragments that cannot be degraded - a subset of LSDs grouped into the mucopolysaccharidoses. Diseases include, for example, "Hurler disease" & "Hunter disease", in which a defect in alpha-L iduronidase or iduronate-2-sulfatase results in accumulation of dermatan sulfate and heparan sulfate, respectively [23]. Another important subgroup of LSDs results from the defect of sphingolipid degrading enzymes, with diseases such as "Gaucher disease", "Fabry disease", and “metachromatic leukodystrophy”. In Gaucher disease, reduced activity of the enzyme beta-glucocerebrosidase results in the accumulation of glucocerebrosides [24]. In Fabry disease, reduced activity of the enzyme α-galactosidase A results in the accumulation of the sphingolipid globotriaosylceramide, affecting a variety of organs [25]. In metachromatic leukodystrophy, the deficiency of the enzyme arylsulfatase A leads to the accumulation of sulfatides within tissues, destroying the nervous system's myelin sheath [26]. The severity of LSDs is based on the level of impairment of lysosomal function, demonstrating the central role of the lysosome in overall cell homeostasis. Therefore, it is not surprising that the lysosome also plays a role in diseases such as Parkinson's and Alzheimer's disease, as well as cancer, infections, and obesity (reviewed in [22]). The severe course of LSDs often leads to the death of patients in childhood. By far, the majority of lysosomal diseases can only be treated symptomatically. However, a few disorders such as Gaucher disease or metachromatic leukodystrophy can be treated by enzyme replacement or gene therapy, respectively [27, 28]. These treatments result in a considerable improvement of the clinical course of these disorders.

1.2.1 Niemann-Pick type C

The autosomal recessive metabolic disease Niemann-Pick type C (NPC) belongs to the group of lysosomal lipid storage disorders. NPC is lethal in the first 5-20 years of life and has a prevalence of 1/120,000 live births [22, 29]. Symptoms of NPC include ataxia, epilepsy, dysphagia, deformity of the hands, visual disturbances, depression, muscle hypotonia, and increasing dementia [30]. NPC disease is currently not treatable, where the cause in 95 % of patients is a defect in the NPC1 protein (5 % NPC2). It has been shown that the accumulation of unesterified cholesterol within lysosomes presents the main feature of the disease [30]. NPC1 is a lysosomal transmembrane protein, while NPC2 localizes within the lumen of lysosomes. The functional interaction of both proteins is essential for the intracellular transport of cholesterol from lysosomes to the endoplasmic reticulum (ER). Low-density lipoproteins (LDLs) particles are transported into the cells via receptor-mediated endocytosis. The cholesterol esters they contain are hydrolyzed by a lysosomal acid lipase to cholesterol that binds to NPC2. NPC2 then transfers cholesterol to the sterol-sensing domain of the N-terminal part of NPC1, which transfers the cholesterol through the lysosomal membrane to the ER [31, 32]. If NPC1 is absent, cholesterol cannot be transported and accumulates, leading to the disease.

1.3 Lysosomal Proteomics

For the investigation of the lysosomal proteome on a large and unbiased scale, proteomics approaches have proven their strength [1]. Proteomics is defined as the large-scale study of proteins originating from different samples, such as tissue extracts, whole-cell lysates, or subcellular enriched fractions [33]. In large scale mass spectrometry based proteomics approaches, lysosomal proteins are inherently difficult to be reproducibly identified/quantified due to their low abundance [34]. For example, it has been estimated that lysosomal proteins contribute only to ~0.2 % of the cellular protein mass in HeLa cells [35]. Furthermore, expression levels of individual lysosomal proteins such as LAPTM4A have been further estimated to be ~3000 copies/cell compared to GAPDH, which is 38,000x higher. This is not unexpected, especially considering the low number of lysosomes (~hundreds/cell) compared to other organelles such as mitochondria with 1000-2000/cell, contributing to 20 % of the cellular volume [35-37]. Hence, for successful comprehensive lysosomal proteomics, either lysosomal enrichment or the development of targeted MS approaches is a prerequisite.

Different techniques have been employed to enrich lysosomes and lysosomal proteins: 1) Density gradient centrifugation, wherein a first step, low-speed centrifugation is applied to separate nuclei, intact cells, and debris from the organelles, followed by the separation of individual organelle fractions via ultracentrifugation. In addition, different types of matrices can be applied to further improve the enrichment of lysosomes such as sucrose, Nycodenz, and Percoll [38]. 2) Superparamagnetic iron oxide nanoparticles (SPIONs) containing an iron oxide core covered with biocompatible substances like dextran, utilized to enrich lysosomes from cells grown in cell culture. SPIONs are unspecifically endocytosed by the cells and reach the lysosomes via the endocytic route, allowing lysosome enrichment within a strong magnetic field [10, 39, 40]. 3) Immunoprecipitation (IP) in which immobilized antibodies on beads bind to surface-exposed antigens of lysosomal tagged overexpressed membrane proteins such as LAMP1-2x/3xFLAG [41, 42], and TMEM192-3xHA tag, which can be used to pull down lysosomes [43].

Each of these methodologies comes with its advantages and disadvantages (reviewed in [1]). One of the advantages of the SPIONs and IP approach is that lysosomal proteins are enriched significantly (up to 100 fold) compared to density gradient centrifugation approaches while contaminating proteins from other organelles are depleted [44]. However, the IP approach has the disadvantage that lysosome enriched fractions contain broken lysosomes, while the SPIONs approach allows for the enrichment of intact lysosomes (up to 80 % intact status). Hence, lysosomes enriched via SPIONs allow for the analyzes of the intact lysosomal proteome, enabling among others, structural analyzes of lysosomal proteins and their interaction partners in a native state.

The efficiency of lysosome enrichment can be assessed by screening marker proteins specific for lysosomes and contaminants arising from other organelles by western blotting. The integrity of lysosomes can be assessed by enzymatic assays, such as the beta-hexosaminidase assay in combination with detergents [45]. The combination of a successfully enriched lysosome sample and its measurement on state-of-the-art mass spectrometers allows to comprehensively study the lysosomal proteome [1, 43, 46, 47].

1.4 Mass Spectrometry

Mass-spectrometry (MS) refers to a method for measuring molecular masses or, more specifically, mass to charge ratios (m/z) of ions, presented as a mass spectrum. MS is used to measure a large variety of molecules, particles, and chemical compounds to study their physical, chemical, or biological properties [48]. In a typical MS experiment, solid, liquid, or gas samples have to be ionized. To do so, various ionization strategies can be applied , such as chemical ionization (CI), fast atom bombardment (FAB), atmospheric-pressure photoionization (APPI), atmospheric-pressure chemical ionization (APCI), electrospray ionization (ESI) [49], or matrix-assisted laser desorption ionization (MALDI) [50]. The latter two ionization strategies are especially beneficial for ionizing macromolecules such as proteins, peptides, lipids, or sugars, as these, for example among others, overcome the hurdle of molecules being fragmented during ionization. Ions are then transferred to different mass analyzers, such as Fourier-transform ion cyclotron resonance (FTICR), sector fields, time-of-flight analyzers (TOFs), ion traps (ITs), quadrupoles (QPs), and Orbitraps (OTs). These analyzers, central to the technology, differ from each other strongly in design and performance and can stand alone as well as combined in tandem to extract their strengths [48, 51].

1.4.1 Proteomics Workflow

Generally, MS-based proteomics experiments can be grouped into “top-down” proteomics analyzing full-length proteins and “bottom-up” proteomics, where proteins are digested into peptides and analysed **(Figure 2)**. While MALDI-TOF, for example, is usually used to analyze samples of low complexity, samples of higher complexity are usually measured with a liquid chromatography (LC) system via ESI coupled to a high-resolution tandem mass spectrometer (MS/MS) [52]. In a typical proteomics experiment, proteins from a sample (e.g., cells, subcellular fractions, or tissues) are digested to peptides applying a protease such as trypsin. Due to the high complexity of samples, usually, those are fractionated via chromatographic or electrophoretic methods, either on protein or peptide level. Samples, for example, can be separated by applying size exclusion chromatography (SEC) or strong cation exchange (SCX) chromatography fractionating proteins or peptides based on their size or either charge. Electrophoresis-based fractionation methods such as sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE), for example, separates proteins based on their size, while OFFGEL electrophoresis, on the other hand, separates proteins or peptides according to their isoelectric points [52, 53].

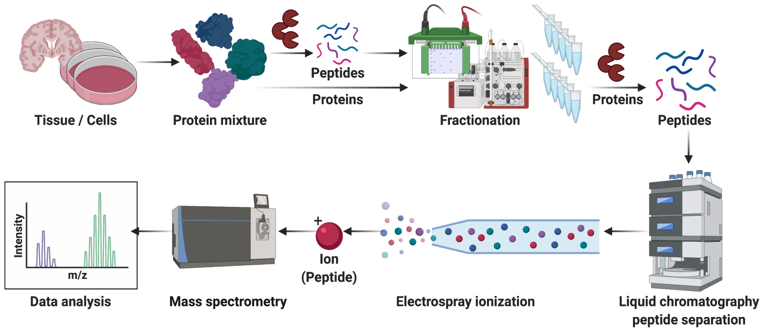


Figure 2: Scheme of a typical mass spectrometry-based proteomics workflow. Samples derived from tissue or cells are lysed, proteolytically digested and protein/peptide mixture fractionated via size exclusion chromatography (SEC), strong cation exchange chromatography (SCX), or electrophoresis-based fractionation methods. Fractionated non-digested proteins are digested with a protease like trypsin, and peptides are separated via a nano-LC chromatography system. Separated peptides are subjected to mass spectrometric analysis, where peptides are ionized and analyzed with specific algorithms for protein identification.

1.4.2 Liquid Chromatography Mass Spectrometry

After fractionation, peptides generated by proteolytic cleavage are usually separated by reversed-phase liquid chromatography, coupled to a tandem mass spectrometer (LC-MS/MS). Applying a solvent gradient with increasing organic content, peptides are eluted from LC columns in order of their hydrophobicity [53, 54]. At the needle of the chromatographic column, the liquid is vaporized in the ion source, and peptides are ionized via a strong electric potential [49]. Subsequently, peptide ions enter the MS through a transfer capillary into the vacuum system, where they are guided and manipulated via electric fields. Dependent on the applied mass analysers, differing in how they determine the mass-to-charge (m/z) ratios of peptides, the mass spectrometer generates a mass spectrum. The mass spectrum records the signal intensity of the ion at each value on a mass/charge (m/z) scale (representing units of Dalton/charge) [51] **(Figure 2)**.

1.4.3 Peptide Fragmentation and Identification

The resulting MS1 scans with their respective m/z values can be matched to peptide masses of *in silico* digested proteins deposited in a database for protein identification, known as "peptide mass fingerprinting" (usually applied in MALDI-TOF experiments). However, the disadvantage of this approach is that it fails to recognize novel peptides [48]. Fragmentation of selected precursor masses (MS1) into fragment ions (MS2), on the other hand, allows for *de novo sequencing*, i.e., that peptide amino acid sequences can be determined. Different fragmentation methods can be applied (reviewed in [55]), such as collision-induced dissociation (CID), where peptide fragmentation is induced by collisions with an inert gas (such as argon or helium) in a collision cell of a mass spectrometer. Other fragmentation methods include higher-energy C-trap dissociation (HCD), a variation of CID using a higher radio frequency (RF) voltage, where ions are fragmented within the HCD cell and stored inside the C-trap before getting injected into the Orbitrap. In contrast, fragmentation approaches such as electron transfer dissociation induce fragmentation by transferring electrons to higher charge state cationic molecules, an approach preferably used to study post-translational modifications (PTMs) [55, 56]. The resulting fragment ions can be used to identify the amino-acid sequence by determining the mass differences between neighboring peaks in a series, differing from its neighbor by one amino acid [51]. Based on the Roepstorff–Fohlmann–Biemann nomenclature, the resulting ions can be classified into a, b, c (charge is retained on the amino-terminal fragment), and x, y, z (charge is retained on the carboxy-terminal fragment) ions **(Figure 3)**. Fragmentation of amide bonds between amino acids resulting in b and y ions present the most common and informative ions.

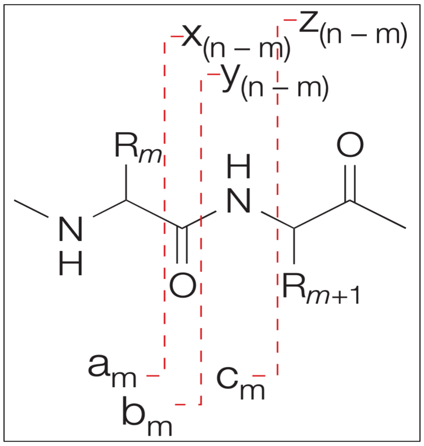


Figure 3: Scheme of a chemical structure of a peptide representing the Roepstorff–Fohlmann–Biemann nomenclature upon fragmentation. Peptide fragmentation is induced by collisions with an inert gas leading to bond breakage, resulting in "a", "b", "c" ions when the charge is retained by the amino-terminal fragment or "x", "y" and "z" ions when it is retained by the carboxy-terminal fragment. Ions are labelled from the original amino terminus a_m, b_m and c_m. Consequently, they are also labeled from the original carboxyl terminus $z_{(n-m)}$, $y_{(n-m)}$, $x_{(n-m)}$. n-m equals to the number of R groups these ions contain, while n is the total number of residues, or R groups in the peptide and m is the number of R groups that the corresponding a-, b- or c-ion would contain [51].

The identification of proteins based on MS2 spectra is more definite than those achieved by peptide mass fingerprinting, as in MS2 based approaches, in addition to the peptide mass, the peak pattern in MS2 spectrum provides information on the peptide sequence. The fact that only an extremely small proportion of the potential peptide amino-acid sequences actually exist in nature makes database searching easier than *de novo* sequencing. A MS2 spectrum may not include enough information to extract the entire amino-acid sequence unambiguously, but it may contain enough information to match it uniquely to a peptide sequence in the database based on the observed and predicted fragment ions [51, 53]. In the most commonly applied database approach, “probability-based matching” calculated *in silico* digested fragments from peptide sequences are compared with observed peaks. A score is calculated for each identified protein from this comparison, reflecting a statistical significance of the match between the observed spectrum and sequence contained in a database [57].

1.4.4 Quantification

Mass spectrometry does not allow to extract quantitative information among different proteins based on the intensity of the signal of a peptide ion itself. This is due to different reasons, such as that upon digestion of a protein, peptide generation is dependent on the accessibility of cleavage sites. Further, peptide solubility, and ionization efficiency of each peptide, contribute significantly to differences in signal intensities [51]. However, as these factors are reproducible among the same peptides, relative amounts of proteins can be determined among experiments. Various MS-based quantification methods have been developed, where a distinction is made between “label-free” and “labeled” approaches [58] **(Figure 4)**. A popular metabolic-based quantification method in this context is SILAC ("stable isotope labeling by amino acids in cell culture") [59]. In such an experiment, one cell population is cultured in SILAC culture medium containing, for example, heavy isotope-labeled versions of lysine and arginine (^{13}C, ^{15}N) (originally developed for leucine). In contrast, the other population is grown in a medium containing non-labeled amino acids. Typically, both of these isotopically labeled amino acids are used to ensure that upon proteolytical digestion of proteins with trypsin, most peptides contain at least one of the labeled amino acids [58]. Labeled and non-labeled cells can then be combined, processed, and analysed. The signal intensities of the heavy and light isotope-labeled peptides, therefore, allow relative quantification of the proteins reducing run to run variability, ion suppression, and saving measurement time [59]. However, the SILAC approach has disadvantages, as it does not allow to compare multiple differentially treated samples (usually up to 3) due to limited labeling combinations available. Further labeling of samples derived from humans is not possible, while high costs are involved for the generation of SILAC *in vivo* animal models [60, 61].

As alternatives, workflows for post-biosynthetic labeling of proteins and peptides are available **(Figure 4)**. In majority of the cases, N-hydroxysuccinimide (NHS) chemistry is applied to modify the N-terminal amino group of proteins and peptides and side chains of lysines. Examples of such reagents are tandem mass tags (TMT) [62] or isobaric tags for relative and absolute quantitation (iTRAQ) [63]. To study absolute amounts of proteins the AQUA (absolute quantification of proteins) [64] approach can be applied. For this, isotope-labeled synthetic standards are mixed into a protein digest to retrieve information on endogenous peptide amounts in the samples. Workflows applying different labeling strategies overall allow more accurate quantification and save measurement time since the samples can be combined and analysed simultaneously after labeling.

Label-free methods, on the other hand, enable the quantification of hundreds or thousands of samples without the expense of labeling reagents **(Figure 4)**. In such an approach, mass spectrometric signal intensity for peptides/fragment ions are directly compared between the samples. In the most frequently used untargeted technique, data-dependent acquisition (DDA), all peptides within a certain mass are fragmented [65]. In DDA mode, fragmentation is based on the selection of the most intense peptide ions in MS1, which are then fragmented via MS/MS. However, as DDA approaches are biased towards the fragmentation of highly abundant peptides, the collected data may contain missing values, especially for the low abundant peptides. On the other hand, targeted acquisition strategies collect peptide and fragment ion data consistently, allowing for precise and sensitive quantification. However, such approaches like single, multiple, or parallel reaction monitoring (SRM, MRM, PRM) are confined to a narrow, preset selection of peptides and limited to a predefined number of proteins per run [66]. Data-independent acquisition (DIA) is another label-free quantificational approach in this context, in which data from all detectable fragment ions in a sample are acquired systematically and consistently [67, 68]. The operation of the mass-spectrometer in a DIA mode circumvents the problem of being biased towards highly abundant peptides. The acquisition strategy methodically covers the entire MS1 scan range collecting comprehensive MS2 spectra, enabling more reproducible and robust quantification covering a wide dynamic range [69].

However, data analysis of DIA data is a challenging task, as the resulting fragment ion spectra are highly complex. Because the fragment ions in DIA spectra may possibly come from numerous precursor ions (any precursor ion present in the m/z range from which the DIA spectrum was produced), the direct relationship between a precursor ion and its fragment ions is lost [68]. Classical DIA data analysis are based on a targeted analysis applying spectral libraries of peptides. A peptide spectral library is a database with the collection of curated, annotated, and non-redundant peptide spectra extracted from LC-MS/MS runs. It serves as consensus templates for peptide/protein identification using the correlation of the template with experimental spectra [68, 70]. Experimentally identified peptides hence can be quantified if present in the spectral library.

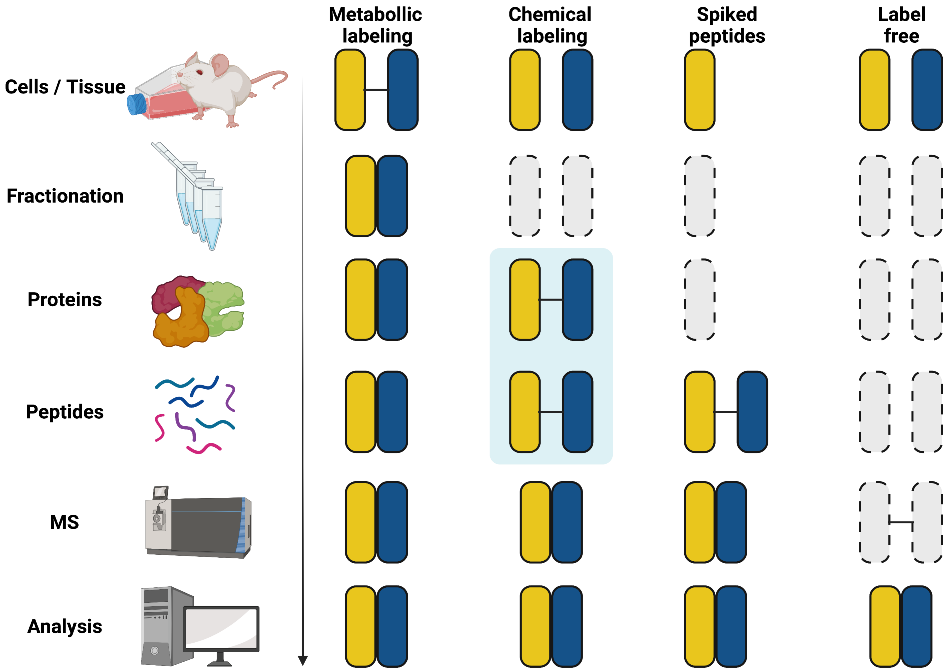


Figure 4: Scheme of common quantitative proteomic approaches. Boxes represent samples, showing at which stage samples are combined. Dashed lines indicate steps at which experimental variation and therefore quantification errors can occur. Modified from [52].

1.5 Cross-linking Mass Spectrometry

One of the promising methods to screen and identify interaction partners on a large, unbiased scale for a given protein, organelle, or whole cell lysate/tissue is cross-linking mass spectrometry (XL-MS) [71-73]. Cross-linking refers to a chemical process to cross-link two or more molecules via a covalent bond [71]. XL-MS was initially mainly applied for protein structure analysis, supplementing well-established techniques such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and electron microscopy (EM) [71]. However, compared to other commonly used approaches like co-immunoprecipitation and proximity labeling (e.g., by biotinylation) - for which often the interactome of one specific protein is investigated in a given study - the main advantage of cross-linking experiments is the unbiased application on a large scale [74]. Furthermore, the covalent bond between two peptides in a cross-linking experiment is physical proof for the direct interaction of the respective proteins. At the same time, co-immunoprecipitation or enrichment after proximity labeling often only indicates the presence in the same cellular compartment.

Numerous chemical strategies for protein cross-linking have been proposed, including targeting carboxyl groups of glutamic acid, aspartic acid, and the protein C-terminus, as well as sulfhydryl groups of cysteine. However, the functional groups targeted by most cross-linkers are primary amines that are present in lysine residues and at the protein N-terminus [74]. For these groups, homobifunctional N-hydroxysuccinimide (NHS)-esters are usually employed. In recent studies, amine-specific cross-linkers are often preferred over other reagents. This is primarily due to the widespread commercial availability of NHS-ester-based cross-linkers and their high reaction specificity, and the relatively even distribution and high prevalence of lysine residues on the surface of most proteins [72, 74]. Dependent on the length of the spacer between the functional groups of a given cross-linker, amino acids in a range of 5-35 Å distance can be linked **(Figure 5)**. For a successful cross-link reaction between two interacting proteins, first and foremost, the functional amino groups of both proteins must be in the range of the cross-linker's length. Furthermore, reaction parameters such as pH, cross-linker/protein ratio, and incubation time and temperature must be determined for individual biological samples [72, 73]. In the ideal case, the successful establishment of the reaction leads to a covalent bond between two different proteins resulting in an inter-link. Alternatively links within the same protein leads to an intra-link, also called loop link, in which two reactive amine groups of one and the same protein are linked together. In the any case, the cross-linker can hydrolyze, and a mono-link is formed, i.e., a chemical modification of a lysine residue corresponding to the cross-linker [75] **(Figure 5)**.

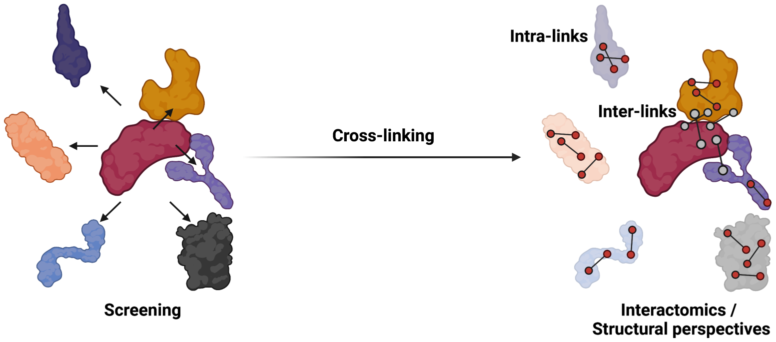


Figure 5: Scheme of the cross-linking approach. Application of cross-linkers for the screening and identification of covalent or transient interactions between different proteins. Inter cross-links are indicated in grey, while intra cross-links originating from lysine residues cross-linked within the same protein in red.

The field of cross-linking faces two major challenges: 1) low abundance of cross-linked peptides in complex samples and 2) data analysis for the reliable identification of cross-links [76]. Current estimations propose that in a typical experiment only 0.1 % of the peptides are cross-linked [75]. This is probably due to a low number of lysine residues present in the right distance to produce a cross-link and partial hydrolyzation of the cross-linker, resulting in undesired modifications of peptides rather than inter-peptide cross-links. Aside from biotin-labeled cross-linkers, which allow for affinity enrichment of cross-linked peptides [77], the currently most common approach to deal with the low abundance of cross-linked peptides is the enrichment by size exclusion chromatography (SEC) [78] or strong cation exchange chromatography (SCX) [76]. These approaches allow for the separation of cross-linked peptide pairs from their unmodified counterparts based on differences in size or charge, increasing the relative abundance of cross-linked species in specific chromatographic fractions.

Analysis of LC-MS/MS data from non-cleavable cross-linkers like bissulfosuccinimidyl suberate (BS3) and disuccinimidyl suberate (DSS), the quadratic expansion of the computational search space (“n-square problem”), presents a major issue [76]. One attempt to overcome the n-square problem is based on the development of MS-cleavable cross-linkers [79, 80] in combination with novel cross-link search algorithms [72, 76]. MS-cleavable cross-linkers allow, after initial fragmentation of the cross-linked peptide pairs, for individual fragmentation of the contained peptides either on the MS2 or MS3 level. This allows for the determination of the precursor mass of the individual peptides reducing the search space significantly [76, 81].

To date, several types of MS2 cleavable cross-linkers have been developed, of which disuccinimidyl sulfoxide (DSSO) [79] and disuccinimidyl dibutyric urea (DSBU) [80] are used frequently in the field of protein-protein interaction analysis **(Figure 6)**. Both compounds include a preferential cleavage site in the spacer arm producing signature fragment ions during fragmentation. Identifying cross-links from such data depends on high-quality MS2 spectra containing signature peaks for the respective MS-cleavable cross-linker. Commonly used algorithms relying on such indicative ions are MeroX [82] and XlinkX [73, 76] **(Figure 6)**. As the identification of cross-links by these algorithms is directly proportional to the presence of the signature peak pairs, and the quality of MS2 spectra, a variety of MS2 methods have been evaluated for samples cross-linked by DSSO and DSBU [81, 83]. Results show fragmentation methods tailored towards a certain cross-linker, in combination with the respective algorithm, strongly enhance the depth and accuracy of cross-link identification in complex samples.

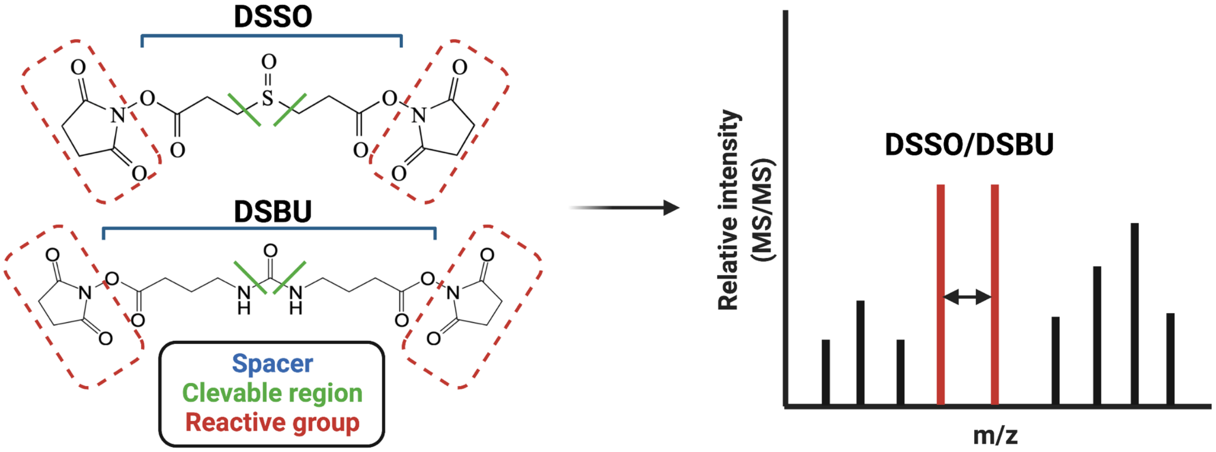


Figure 6: Chemical structure and fragmentation scheme of MS-cleavable cross-linker disuccinimidyl sulfoxide (DSSO) and disuccinimidyl dibutyric urea (DSBU). Both cross-linker belonging to the homobifunctional N-hydroxysuccinimide (NHS)-esters target primary amines of lysine residues and the N-terminus of proteins within a range of 5-35 Å with a MS cleavable spacer in the middle, which is unique to each of the cross-linker resulting in a unique mass difference of signature peaks upon fragmentation of cross-linked peptides.

2. Dissertation Rationale

To study the diverse roles of organelle proteins on a large scale in an unbiased way, MS-based proteomics approaches have proven their strength [84]. To date, >60 studies (reviewed in [1]) have investigated the lysosomal proteome by MS, the majority relying on various lysosomal enrichment procedures due to the low abundance of lysosomal proteins (e.g. [35-37]). However, despite the numerous lysosomal proteomics studies, no systematic comparison of the individual enrichment approaches was performed. Therefore, I compared the four most common lysosome enrichment strategies and differentially evaluated protein abundance, distribution, and contaminant proteins arising from organelles other than the lysosome (Chapter 1 [44]). I identified the application of SPIONs as the optimal enrichment approach. Subsequently, this method was used to optimize lysosomal MS-sample preparation by comparing different proteolytic digestion protocols, fractionation of peptides, desalting, and MS acquisition strategies (Chapter 1 [34, 85]).

Certain lysosomal proteins and complexes (e.g. mTORC1) have been investigated in great detail [13, 86]. However, it is unclear how many complexes involving lysosomal proteins exist and whether more complexes have functional significance similar to the mTORC1 complex. Recent developments in cross-linking mass spectrometry (XL-MS) allow the characterization of protein interactions and structures across different biological systems on a large, unbiased scale [77, 87-90]. One of the approach's strengths is that cross-linker like disuccinimidyl sulfoxide (DSSO) allows to investigate proteins and their interactions in their native state [91, 92], providing an ideal tool for the investigation of lysosomes. A cross-link study's success depends on multiple key steps, many of which have been optimized in the past [71, 72, 74, 81, 83]. Yet, even though the control of a successful cross-link reaction belongs to one of the key steps, it is mainly confirmed by visual inspection of band shift patterns in coomassie/silver-stained gels. However, these approaches have limitations when it comes to the validation of individual cross-linked proteins or limited sample amounts. Further, the question of whether cross-linkers can penetrate the organelle membranes in an active state is actively debated in the field, with no accessible tools to visualize cross-linked proteins in a cell. Hence, to better characterize and validate cross-link reactions, I raised antibodies against the cross-linker DSSO and DSBU in rabbits (Chapter 2 [93]). I further purified, characterized, and applied them for visualizing cross-links *in vivo* and immunoprecipitation of cross-link modified proteins. The results and tools obtained from chapter 1 and 2 were subsequently used to cross-link lysosome enriched fractions presenting the first lysosomal cross-linking interactome draft, which validated existing protein crystal/cryo-EM structures, and allowed to propose novel protein models (Chapter 3 [94]).

The half-life of lysosomal proteins has only been determined by classical biochemical methods for a few selected proteins. Therefore, based on the established lysosome proteomics strategies, I further investigated aspects of protein turnover in a more comprehensive approach determining protein half-lives of lysosome enriched fractions applying the pulse (p) stable isotope labeling by amino acids in cell culture (SILAC) data-independent acquisition (DIA) MS/MS approach [95] (Chapter 4) . Furthermore, to investigate degradational routes of proteins on a large scale, protein half-lives were compared under five different perturbational states. Conditions included the inhibition of proteasomal activity (bortezomib treatment) and conditions affecting lysosomal function, such as the impairment of autophagosome-lysosome fusion (ATG5-KO), cathepsin D activity (pepstatin A treatment), cholesterol homeostasis (NPC1-KO), and protein glycosylation (multiple glycosylation machinery related genes knock out (MGM-KO)). The overall workflow was visualized **(Figure 7)**.

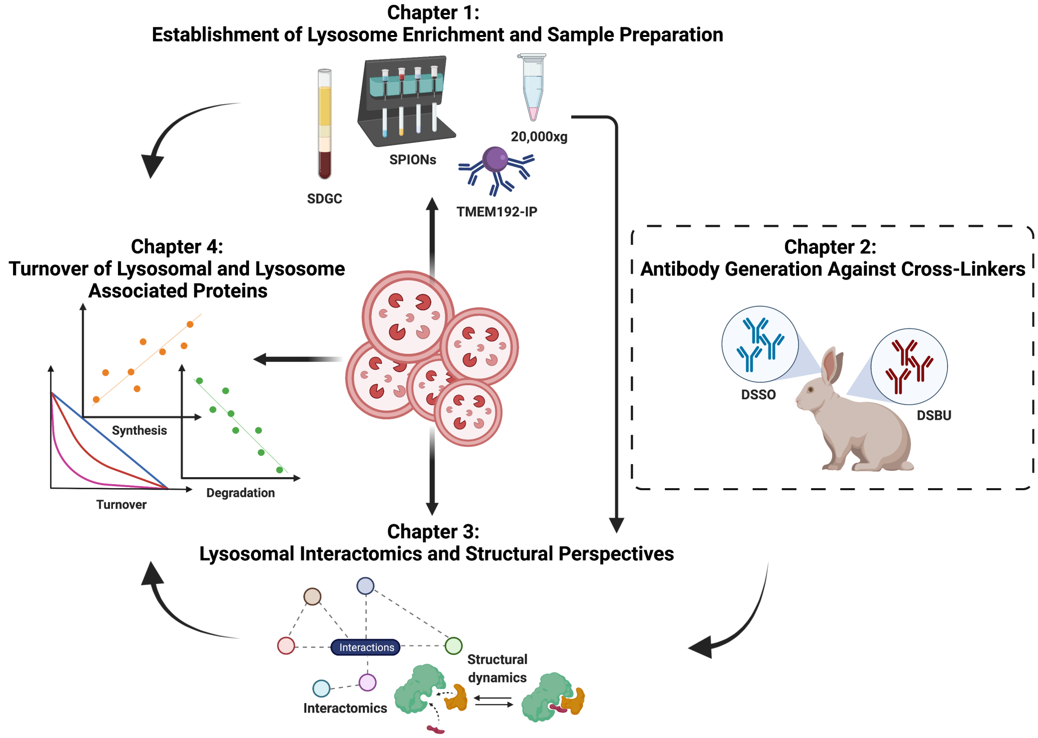


Figure 7: Schematic illustration of the thesis rationale. Four different lysosome enrichment methods were compared to a whole-cell lysate using the data-independent acquisition (DIA) (Chapter 1). Antibodies against the cross-linker DSSO and DSBU were generated, purified, and applied in western blot and immunocytochemistry experiments (Chapter 2). Lysosome enriched fractions were cross-linked to investigate protein interactions and structures (Chapter 3). Half-lives of lysosome enriched fractions were determined applying the pulse stable isotope labeling by amino acids in cell culture (SILAC) DIA approach (Chapter 4). **Abbreviations:** SDGC: Sucrose density gradient centrifugation, SPIONs: Superparamagnetic iron oxide nanoparticles, TMEM192: Transmembrane protein 192, IP: Immunoprecipitation, DSSO: disuccinimidyl sulfoxide, DSBU: and disuccinimidyl dibutyric urea.

2.1 Summary of Publications

1 **Singh, J.**, Kaade, E., Muntel, J., Bruderer, R., Reiter, L., Thelen, M., Winter, D. (2020). Systematic comparison of strategies for the enrichment of lysosomes by data independent acquisition. *Journal of proteome research*, 19(1), 371-381 **(Chapter 1 – pp. 13-34).**

2 Ponnaiyan, S., Akter, F., **Singh, J.**, Winter, D. (2020). Comprehensive draft of the mouse embryonic fibroblast lysosomal proteome by mass spectrometry-based proteomics. *Scientific data*, 7(1), 1-13 **(Chapter 1 – pp. ii-xv) – Supplement.**

3 Mosen, P., Sanner, A., **Singh, J.**, Winter, D. (2021). Targeted Quantification of the Lysosomal Proteome in Complex Samples. *Proteomes*, 9(1), 4 **(Chapter 1 – pp. xvi-xxxiii) – Supplement.**

4 **Singh, J.**, Ponnaiyan, S., Gieselmann, V., Winter, D. (2021). Generation of Antibodies Targeting Cleavable Cross-Linkers. *Analytical Chemistry*, 93(8), 3762-3769 **(Chapter 2 – pp. 35-59).**

5 **Singh, J.**, Elhabashy, H., Muthukottiappan, P., Stepath, M., Eisenacher, M., Kohlbacher, O., Gieselmann, V., Winter, D. (2022). Cross-linking of the Endolysosomal System Reveals Flotillin Structures and Putative Cargo. bioRxiv **(Chapter 3 – pp. 60-116).**

6 **Singh, J.**, Stepath, M., Dang, H.M., Möller, C., Muchamedin, A., Hardt, R., Song, J., Bernhardt, O., Gandhi, T., Bruderer, R., Reiter, L., Eisenacher, M., Becker, T., Vakhrushev, S., Gieselmann, V., Winter, D. The presented results in chapter 4 have not been published yet, as the manuscript is under preparation **(Chapter 4 – pp. 117-145).**

3. Chapter 1: Establishment of Lysosome Enrichment and Sample Preparation

1 Part one of this chapter has been published: [**Singh, J.**, Kaade, E., Muntel, J., Bruderer, R., Reiter, L., Thelen, M., Winter, D. (2020). Systematic comparison of strategies for the enrichment of lysosomes by data independent acquisition. *Journal of proteome research*, 19(1), 371-381].

- For the online version of the paper, visit: https://doi.org/10.1021/acs.jproteome.9b00580
- The supporting information, including supplementary tables, are available at: https://pubs.acs.org/doi/10.1021/acs.jproteome.9b00580
- **Remark:** The first publication within the scope of chapter 1 has been done by me as the only first author and **SHOULD BE CONSIDERED** for this cumulative doctoral dissertation.

2 Part two of this chapter has been published: [Ponnaiyan, S., Akter, F., **Singh, J.**, Winter, D. (2020). Comprehensive draft of the mouse embryonic fibroblast lysosomal proteome by mass spectrometry-based proteomics. Scientific data, 7(1), 1-13].

- For the online version of the paper and supporting information, visit: https://doi.org/10.1038/s41597-020-0399-5
- **Remark:** The second publication within the scope of chapter 1 **SHOULD NOT BE CONSIDERED** for this cumulative doctoral dissertation, as the majority of the work has been done by the first two co-authors. The publication has been attached in the supplement under section 8.1.

3 Part three of this chapter has been published: “[Mosen, P., Sanner, A., **Singh, J.**, Winter, D. (2021). Targeted Quantification of the Lysosomal Proteome in Complex Samples. Proteomes, 9(1), 4].

- For the online version of the paper, visit: https://doi.org/10.3390/proteomes9010004
- The supporting information including supplementary tables are available at: https://www.mdpi.com/2227-7382/9/1/4/s1
- **Remark:** The third publication within the scope of chapter 1 **SHOULD NOT BE CONSIDERED** for this cumulative doctoral dissertation, as the majority of the work has been done by the first two co-authors. The publication has been attached in the supplement under section 8.2.

3.1 Introduction

The success of comprehensive shotgun proteomics approaches depends on multiple factors like sample complexity, adequate sample preparation, and dependent on the goal of a study and experimental design, adapted MS acquisition strategies [96]. While whole-cell proteome analyses containing 12,000 proteins exceed the resolving ability of single-shot proteomics approaches, subcellular proteome approaches decrease the sample complexity to 500-4000 proteins within a typical compartment [96]. Biochemical research has led to the development of highly efficient methods that allow for the enrichment of subcellular compartments up to 90-95 %. However, despite the significant enrichment of organelles such as mitochondria or nuclei, in-depth MS analysis of such fractions still led to the identification of proteins from other cellular compartments [97].

Compared to other organelles such as mitochondria with 1000-2000/cell, lysosomes with ~100/cell are of relatively low abundance, making their enrichment a prerequisite to allow for in-depth proteomic analysis [35, 62]. Although multiple lysosome enrichment approaches have been developed, it has so far not been determined how they affect results of proteomic analysis of lysosomes. Therefore, in chapter 1, I compared four lysosome enrichment approaches:

1) Generation of an organelle-enriched pellet via 20,000×g centrifugation (OEP)
2) A two-step sucrose density gradient centrifugation approach (SDGC)
3) Enrichment by superparamagnetic iron oxide nanoparticles (SPIONs)
4) Immunoprecipitation using a 3xHA tagged version of the lysosomal transmembrane protein 192 (TMEM-IP).

I determined enrichment and depletion efficiencies for lysosomal and contaminant proteins for each enrichment strategy and compared the data to a whole-cell lysate dataset using the data-independent acquisition (DIA) strategy. The DIA approach allowed me to reproducibly identify and quantify proteins across all five conditions [44].

In addition, to assessing different lysosome enrichment approaches, colleagues of our group and I further systematically evaluated MS sample preparation protocols. The comparison included different strategies for the concentration of lysosome enriched fractions, proteolytic digestion, desalting, and fractionation approaches for peptides, and evaluating different LC gradient lengths. We assessed four different strategies, including in-gel digestion, filter aided sample preparation (FASP), and in-solution digestion applying urea and RapiGest [85]. Moreover, colleagues and I further assessed the impact of two different targeted MS acquisition strategies (DIA and parallel reaction monitoring (PRM)) for the analysis/quantification of the lysosomal proteome from samples from different sample complexity [34].

3.2 Publication: Main Article

Reproduced with permission from: **Singh, J.**, Kaade, E., Muntel, J., Bruderer, R., Reiter, L., Thelen, M., Winter, D. (2020). Systematic comparison of strategies for the enrichment of lysosomes by data independent acquisition. *Journal of proteome research*, 19(1), 371-381.
Copyright © 2019 American Chemical Society

Journal of proteome research
Article
Cite This: *J. Proteome Res.* 2020, 19, 371−381
pubs.acs.org/jpr

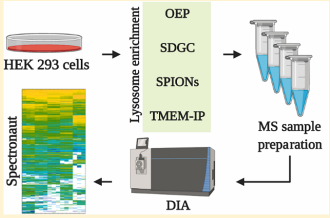
Systematic Comparison of Strategies for the Enrichment of Lysosomes by Data Independent Acquisition

Jasjot Singh,[†] Edgar Kaade,[†] Jan Muntel,[‡] Roland Bruderer,[‡] Lukas Reiter,[‡] Melanie Thelen,[†] and Dominic Winter*,[†]

[†]Institute for Biochemistry and Molecular Biology, University of Bonn, 53115 Bonn, Germany
[‡]Biognosys, 8952 Schlieren, Switzerland

Supporting Information

ABSTRACT: In mammalian cells, the lysosome is the main organelle for the degradation of macromolecules and the recycling of their building blocks. Correct lysosomal function is essential, and mutations in every known lysosomal hydrolase result in so-called lysosomal storage disorders, a group of rare and often fatal inherited diseases. Furthermore, it is becoming more and more apparent that lysosomes play also decisive roles in other diseases, such as cancer and common neurodegenerative disorders. This leads to an increasing interest in the proteomic analysis of lysosomes for which enrichment is a prerequisite. In this study, we compared the four most common strategies for the enrichment of lysosomes using data-independent acquisition. We performed centrifugation at 20,000 × *g* to generate an organelle-enriched pellet, two-step sucrose density gradient centrifugation, enrichment by superparamagnetic iron oxide nanoparticles (SPIONs), and immunoprecipitation using a 3xHA tagged version of the lysosomal membrane protein TMEM192. Our results show that SPIONs and TMEM192 immunoprecipitation outperform the other approaches with enrichment factors of up to 118-fold for certain proteins relative to whole cell lysates. Furthermore, we achieved an increase in identified lysosomal proteins and a higher reproducibility in protein intensities for label-free quantification in comparison to the other strategies.



KEYWORDS: *lysosomes, lysosome enrichment, quantification, mass spectrometry, superparamagnetic iron oxide nanoparticles, sucrose density gradient centrifugation, TMEM192-3xHA immunoprecipitation, data-independent acquisition, organelle, proteomics*

■ INTRODUCTION

The correct function of mammalian cells depends on the ability to recycle macromolecules and organelles on a continuous basis, maintaining a constant turnover.[1] The central cellular compartment for this task is the lysosome, a lytic membrane-enclosed organelle containing a variety of hydrolytic enzymes that catalyze the degradation of a variety of substances such as proteins, lipids, sugars, and nucleic acids.[2] Dysfunction of these hydrolases can result in numerous disorders such as lysosomal storage diseases (LSDs),[2,3] cardiovascular disorders,[4] cancer,[5] and neurological diseases.[6] The group of LSDs comprises ∼50 disorders[3] characterized by the accumulation of undegraded macromolecules in lysosomes. This material results in an impairment of lysosomal function, which, in the majority of cases, leads to a severe phenotype in the affected patients. Due to the direct relationship of lysosomal hydrolases and the respective LSD, their enzymatic activites and roles in the degradation of the respective macromolecules have been investigated in a variety of studies.[2]

For a long time, the lysosome has been regarded as a purely digestive and unregulated organelle. In recent years, however, it is becoming clearer that it functions as a recycling organelle and actively transports the generated molecules such as monosaccharides, amino acids, or lipids into the cytosol, making them available for biosynthetic or oxidative processes.[7] Furthermore, a regulatory network is located at the lysosomal membrane controlling metabolically relevant signaling cascades with mTORC1[8] being its most prominent example.[9]

Due to the emerging role of lysosomes in cellular signaling and a variety of disorders, there has been an increasing interest in the analysis of lysosomal proteins. To this day, ∼160 bona fide lysosomal proteins have been defined in molecular biology studies as well as ∼30 lysosomal-associated proteins from which a significant amount is related to the mTORC1 complex.[10,11] Furthermore, 683 proteins have been associated with lysosomal or late endosomal localization in gene ontology (GO) databases (http://pantherdb.org/).

For the analysis of lysosomal proteins by mass spectrometry, the enrichment of lysosomes is a prerequisite. Common methods applied for this purpose are subcellular density gradient fractionation,[12] delivery of superparamagnetic iron oxide nanoparticles (SPIONs)[13] to lysosomes by unspecific endocytosis, and immunoaffinity enrichment of lysosomes through tagged membrane proteins.[8,14]

Received: August 28, 2019
Published: November 18, 2019

DOI: 10.1021/acs.jproteome.9b00580

In density gradient centrifugation approaches, which are probably the most commonly used method for subcellular fractionation, lysosomes are included in one of the subfractions obtained. A major disadvantage of most gradient centrifugation approaches is the contamination of the lysosome-containing fraction with other organelles, especially with mitochondria and peroxisomes.[15,16] Therefore, numerous approaches have been developed to increase the purity of such lysosome-enriched fractions. This includes, for example, the administration of Triton-WR1339 to rats, which leads to a change in density of lysosomes derived from the liver of these animals,[15] and the combination of this approach with iTRAQ allowing for the quantification of the extent of protein relocalization.[17] Furthermore, different substances for the formation of the gradient have been introduced including Histodenz,[12] Metrizamide,[18] Nycodenz,[19] and Percoll.[20]

The isolation of lysosomes by unspecific fluid-phase endocytosis of SPIONs in combination with magnetic columns presents another method frequently used for the enrichment of lysosomes.[13] This approach has been applied successfully in numerous studies for the investigation of lysosomes by mass spectrometry including the relative quantification of lysosomal proteins with stable isotope labeling by amino acids in cell culture (SILAC) and label-free quantification, as well as lipidomics studies.[10,21]

For the immunoprecipitation of lysosomes with tagged membrane proteins, two constructs have been applied so far. While the initial version of this approach utilized a variant of the lysosomal-associated membrane protein 1 (Lamp1) fused to a red fluorescent protein-FLAG tag construct at its cytosolic tail (LAMP1-RFP-FLAG),[8] the improved version consists of a 3xHA-tagged version of the transmembrane protein 192 (TMEM192-3xHA).[14] This approach has been applied for the proteomic,[14] metabolomic,[22] and lipidomic[23] analysis of lysosomes.

Despite the use of the individual approaches for the enrichment of lysosomes in a multitude of studies, no systematic evaluation of their efficiency has been performed so far. In the current study, we compared the four most common strategies for the enrichment of lysosomes using mass spectrometry-based proteomics. We evaluated the performance of the individual methods relative to each other, comparing the abundance/distribution of lysosomal and total proteins, as determined by data-independent acquisition (DIA).

■ EXPERIMENTAL SECTION

Cell Culture and Enrichment of Lysosomes

Tissue culture plates were coated by poly-L-lysine (0.5 mg/ml in phosphate-buffered saline (PBS)) for 10 min at 37 °C. Human embryonic kidney cells (HEK 293) were seeded on these plates and cultured in Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal calf serum (FCS), 100 IU/mL penicillin, 100 μg/mL streptomycin, 2 mM L-glutamine at 37 °C, and 5% CO_2. Dependent on the sample preparation method, different numbers of cells/plates were used.

Whole Cell Lysates

One 10 cm plate of 6×10^6 HEK 293 cells/plate was seeded in DMEM and incubated for 48 h. The cells were washed twice with ice-cold phosphate-buffered saline (PBS), and the plate was scraped off in 1 mL of ice-cold PBS, transferred to a micro tube, and centrifuged at 600 × *g* for 10 min at 4 °C. The supernatant was discarded, the pellet was resuspended in lysis buffer (20 mM HEPES, 150 mM NaCl, 1.5 mM $MgCl_2$, 0.5 mM dithiothreitol (DTT) (Sigma-Aldrich), 0.4% NP-40, containing complete EDTA-free protease-inhibitor cocktail (Roche Diagnostics, Mannheim, Germany), pH 7.4), and the cells were lyzed at an amplitude of 40 with three cycles of 30 s each using a sonicator (Bioruptor plus, Seraing, Belgium) at 4 °C. Subsequently, proteins were extracted by acetone precipitation (4:1 v/v), and the pellet was stored at −20 °C until further use.

Organelle-Enriched 20,000 × *g* Pellets (OEP)

Three 10 cm plates of 6×10^6 HEK 293 cells/plate were seeded in DMEM and incubated for 48 h. The cells were washed twice with ice-cold PBS, and the plate was scraped off in 2 mL of isolation buffer (250 mM sucrose, 10 mM HEPES, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 1.5 mM MgAc, 1 mM DTT, 1 × PIC). Afterward, the cells were pooled, disrupted using a 15 mL douncer with 25 strokes, and centrifuged at 600 × *g* for 10 min at 4 °C. The supernatant was transferred to another tube, and the pellet was resuspended in 3 mL of isolation buffer, dounced with 25 strokes, and centrifuged at 600 × *g* for 10 min at 4 °C. The post nuclear supernatants (PNS) were pooled, centrifuged at 20,000 × *g* for 30 min at 4 °C, and the pellet was stored at −20 °C.

Sucrose Density Gradient Centrifugation (SDGC)

Ten 10 cm plates of 6×10^6 HEK 293 cells/plate were seeded in DMEM and incubated for 48 h. Cells were washed thrice (twice with ice-cold PBS and once with 250 mM sucrose buffer), and the plate was scraped off in 1 mL of 250 mM sucrose buffer. Subsequently, the cells were pooled and centrifuged at 600 × *g* for 5 min at 4 °C. The pellet was resuspended in 3 mL of 250 mM sucrose buffer, dounced in a 5 mL douncer with 30 strokes and centrifuged at 800 × *g* for 10 min at 4 °C. The PNS was recovered and loaded on top of a sucrose density gradient which was prepared as follows (from bottom to top): 2 mL of 35% Histodenz in 250 mM sucrose, 2 mL of 17% Histodenz in 250 mM sucrose, and 3 mL of 6% Percoll in 250 mM sucrose. The gradient was centrifuged at 70,000 × *g* for 35 min at 4 °C; the second interphase was collected and used for a subsequent gradient centrifugation. For this, the recovered interphase was mixed with 0.879 mL of 80% Histodenz forming the first layer of the new density gradient. On top of this solution, 2 mL of 17% Histodenz in 250 mM sucrose, 2 mL of 5% Histodenz in 250 mM sucrose, and 1 mL of 250 mM sucrose were added. The gradient was then centrifuged at 70,000 × *g* for 35 min at 4 °C resulting in one lysosome containing an interphase band, which was collected. Subsequently, proteins were extracted from this sample by acetone precipitation (4:1 v/v), and the pellet was stored at −20 °C until further use.

Superparamagnetic Iron Oxide Nanoparticles (SPIONs)

Three 10 cm plates of 6×10^6 HEK cells/plate were seeded in DMEM supplemented with 10% (v/v) magnetite dextran solution (EndoMAG40, Liquids Research, North Wales, UK) and incubated for 24 h. Afterward, the cells were washed thrice with PBS to remove the magnetite dextran and resupplemented with DMEM followed by incubation for 24 h. Subsequently, the cells were washed twice with ice-cold PBS, scraped in 2 mL of isolation buffer, pooled, and dounced with 25 strokes in a 15 mL douncer, followed by centrifugation at 600 × *g* for 10 min at 4 °C. The pellet was resuspended in 3 mL of isolation buffer, dounced with 25 strokes, and centrifuged at 600 × *g*, and the PNS was combined with the first PNS. Lysosome isolation was performed at room temperature using Miltenyi LS columns (Miltenyi Biotech, Auburn, CA) in combination with the Quadro MACS magnet (Miltenyi Biotech). The columns were equilibrated with 1 mL of 0.5% (w/v) BSA in PBS by gravity flow. Then, the PNS was passed through the column, and the column was washed three times with 1 mL of isolation buffer and removed from the magnet; lysosomes were eluted two times in 1 mL of isolation buffer using a plunger. The eluates were centrifuged at 20,000 × *g* for 30 min at 4 °C, the supernatant was discarded, and the pellet was stored at −20 °C.

Immunoprecipitation (TMEM-IP)

For the generation of HEK 293 cells stably expressing TMEM192-3xHA, cells were transfected with a plasmid containing TMEM192-3xHA and a puromycin resistance gene using the Turbofect reagent (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions. Following puromycin selection with a previously determined concentration of 2 μg/mL culture medium, single cell clones were selected and propagated, and expression of TMEM192-3xHA was verified by Western blot. pLJC6-3XHA-TMEM192 was a gift from David Sabatini (Addgene plasmid no. 104434; http://n2t.net/addgene:104434; RRID: Addgene 104,434).

Two 15 cm plates of 11×10^6 HEK 293 cells/plate stably expressing TMEM192-3xHA were seeded in DMEM and incubated for 48 h. The cells were washed twice, scraped in 1 mL of ice-cold PBS, and centrifuged at 1000 × *g* for 2 min at 4 °C, and the cell pellet was resuspended in 1 mL of isolation buffer (136 mM KCl, 10 mM KH_2PO_4, pH 7.25)/plate. Subsequently, the cells were dounced with 20 strokes in a 2 mL douncer and the PNS was incubated with prewashed 100 μL of anti HA magnetic beads (Thermo Fisher Scientific) for 5 min on an overhead rotator. The supernatant was removed, and the beads were washed gently three times with 500 μL of isolation buffer using a DynaMag spin magnet (Thermo Fisher Scientific). Finally, proteins were extracted by incubating the beads for 10 min with ice-cold lysis buffer (40 mM HEPES, 1% Triton X-100, 10 mM β-glycerol phosphate, 10 mM pyrophosphate, 2.5 mM $MgCl_2$, pH 7.4) and precipitated by acetone precipitation (4:1). The pellet was then stored at −20 °C.

The efficiency of lysosome enrichment and lysosomal integrity was assessed using the β-hexosaminidase assay[24] with/without addition of 1% Triton X-100 (final concentration). Furthermore, the protein concentration of the individual samples was determined using the DC protein assay (BioRad, Hercules, CA, USA).[25]

MS Sample Preparation

For each sample, 100 μg of protein was denatured with a freshly prepared solution of 8 M urea/100 mM triethylammonium bicarbonate (TEAB) (Sigma-Aldrich, St. Louis, MO, USA), reduced with 10 mM DTT at 56 °C for 30 min, and alkylated with 20 mM chloroacetamide (Merck KGaA, Darmstadt, Germany) at RT for 30 min. The reaction was quenched by adding 10 mM DTT. The samples were diluted 1:2 with 100 mM TEAB, rLysC (Promega, Mannheim, Germany) was added at an enzyme-to-protein ratio of 1:100 (w/w), and the samples were incubated at 37 °C overnight. Subsequently, the urea concentration was diluted to 1.6 M, trypsin (Promega) was added at an enzyme-to-protein ratio of 1:100 (w/w), and the samples were incubated at 37 °C for 8 h. The resulting peptides were desalted using STAGE-tips[26] assembled with Empore C18 reversed-phase material (3 M, St. Paul, MN), dried using a vacuum centrifuge, and finally stored at −20 °C until further use.

LC-ESI-MS/MS

Peptides were reconstituted in 5% ACN (AppliChem)/5% FA (Sigma-Aldrich) and analyzed using a nano-UHPLC-Dionex Ultimate 3000 system connected to an Orbitrap Fusion Lumos mass spectrometer (both Thermo Fisher Scientific). For DDA analyses, each sample material from all three biological replicates was pooled before the analysis. For all analyses, 1 μg of the sample was loaded directly on a 50 cm reversed-phase analytical column at a flow rate of 600 nL/min using 100% solvent A (0.1% FA in water). The columns were produced in-house as follows: spray tips were generated from 360 μm outer diameter/100 μm inner diameter fused silica capillaries using a P-2000 laser puller (Sutter Instruments, Novato, CA) and packed with 1.9 μm ReprosilPur AQ C_{18} particles (Dr. Maisch, Ammerbuch-Entringen, Germany). Peptide separation was performed with 240 min (DDA) and 120 min (DIA) linear gradients from 5−35% solvent B (95% ACN, 0.1% FA) at a flow rate of 300 nL/min. MS1 spectra were acquired in the Orbitrap mass analyzer from m/z 375−1200 at a resolution of 60,000 and with a dynamic exclusion of 120 s fragmenting charge states between 2+ to 4+. Peptides were fragmented using higher energy collisional dissociation (30%) fragmentation, and MS2 scans were acquired at a resolution of 30,000 in the Orbitrap analyzer. The precursor isolation width was set to m/z 1.6 with an automatic gain control of 4×10^5 and 5×10^5 and maximum injection times of 50 and 54 ms for the MS1 and MS2 scans, respectively. For DIA, MS1 scans were acquired with a resolution of 120,000, an AGC target setting of 5×10^5, and a maximum injection time of 20 ms covering a mass range of 375 to 1200 m/z. The MS1 scan was followed by 36 static MS2 DIA segments. The DIA segment isolation width was adjusted to 24.1 m/z to cover the same mass range as the MS1 scan including a 0.5 m/z overlap. The MS2 scans were acquired with a resolution of 30,000, an AGC target setting of 1×10,[6] and a maximum injection time of 60 ms. The HCD collision energy was set to 27%. This window design resulted in a total cycle time of 3.44 s for the 120 min gradient length.

Data Analysis

For library generation, DDA MS*.raw files were analyzed using the Pulsar search engine available in the Library perspective of Spectronaut[27] (version: 13.2.19). The human Uniprot database was used (UniProt database proteome ID: UP000005640, release date: 11.06.2019 with 21,155 entries). DDA and DIA MS*.raw files were analyzed using the default parameters[27] (enzyme: trypsin, mass tolerances were defined dynamically by the Pulsar algorithm, fixed/variable modifications: carbamidomethylation at cysteine/oxidation at methionine; number of allowed missed cleavages: 2; 3 to 6 fragment ions per peptide were selected based on intensity to build the library). An initial search was performed to assess unspecific carbamylation of lysine residues and peptide N-termini, which was found to be <3%. In subsequent searches, the modification was not considered. For retention time alignment, the high-precision iRT concept (dynamic) was applied.[27] Mass tolerances for matching of precursor and fragment ions as well as peak extraction windows were determined automatically by

DOI: 10.1021/acs.jproteome.9b00580
J. Proteome Res. 2020, 19, 371−381

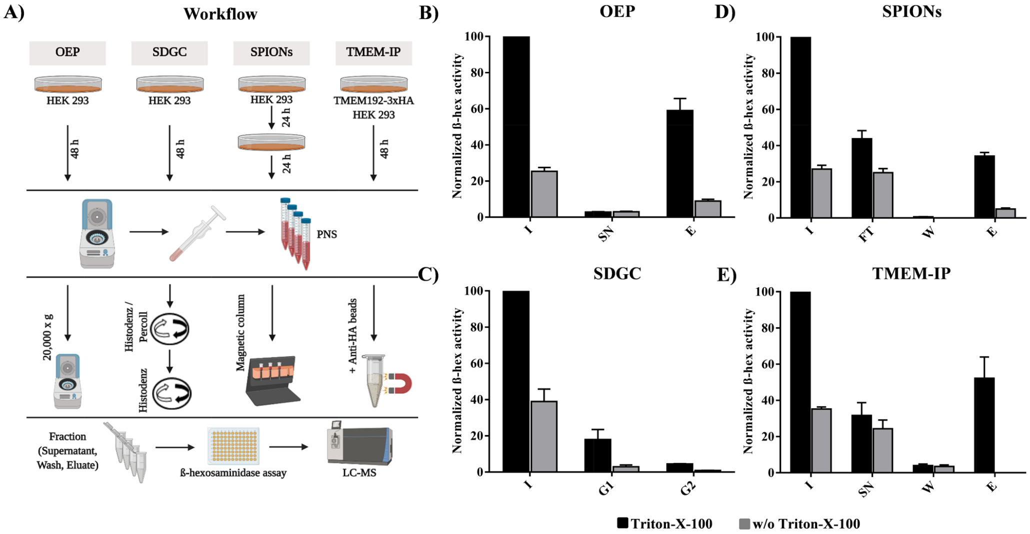


Figure 1. (A) Schematic representation of the workflows for the four lysosome enrichment methods. (B−E) Normalized β-hexosaminidase activity assays for each of the four enrichment methods and the respective fractions. Samples treated with Triton X-100 are shown in black and without in gray. Shown are average values (n = 3) and standard deviation. Lysate: whole cell lysate; OEP: organelle-enriched pellet; SDGC: sucrose density gradient centrifugation; SPIONs: superparamagnetic iron oxide nanoparticles; TMEM-IP: TMEM192-3xHA immunoprecipitation; I: input; SN: supernatant; FT: flow through; W: wash; E: eluate; G1: interphase gradient #1; G2: interphase gradient #2; β-hex: β-hexosaminidase; and PNS: postnuclear supernatant.

Spectronaut. The precursor information was only utilized in peak detection. Interference correction was enabled. Data were normalized using local regression normalization, and the results were filtered with 1% FDR on the precursor and protein level (q-value <0.01).[28] The Spectronaut project was uploaded to Pride (PXD015098) and can be viewed using the freely available Spectronaut Viewer. For unsupervised clustering and determination of p-values, the post analysis pipeline of Spectronaut was used with its default parameters (distance metric: Manhattan Distance; linkage strategy: Ward's method; multiple testing correction: Storey's method).

■ RESULTS AND DISCUSSION

For the analysis of lysosomal composition and function, it is important to recover as many lysosomes as possible, that they stay intact during the process, and that the enriched fractions contain as few contaminating proteins as possible originating from other organelles. To determine which of the currently available methods meets these requirements in the most complete way, we compared four different enrichment strategies that are frequently used for cells grown in tissue culture (Figure 1A). As a benchmark, we chose a crude enrichment strategy, which is the centrifugation of the PNS at 20,000 × g termed in the following as the organelle-enriched pellet (OEP), as this should result in the most complete recovery and intact ratio of lysosomes. We compared this method to a two-step sucrose density gradient centrifugation (SDGC) approach employing Histodenz and Percoll.[12] Furthermore, we included two enrichment strategies developed specifically for lysosomes. In the SPIONs[13] approach, which is based on the endocytosis of dextran-coated nanoparticles and their delivery to the lysosomal compartment, lysosomes containing these particles are isolated using magnetic columns. In the TMEM192-3xHA immunoprecipitation[14] strategy (TMEM-IP), cells stably expressing a 3xHA-tagged version of the lysosomal membrane protein TMEM192 are used for lysosome isolation by anti-HA antibodies immobilized on magnetic beads.

Assessment of Lysosomal Yield

To allow for the reproducible enrichment of lysosomes by TMEM-IP, we initially generated HEK 293 cells stably expressing TMEM192-3xHA and confirmed expression of the protein by Western blotting analysis (Figure S1). We then performed lysosome enrichment with each of the individual methods in three independent biological replicates and assessed the recovery and intact ratio of lysosomes using enzymatic assays for β-hexosaminidase activity,[10] an enzyme residing in the lysosomal lumen (Figure 1B−E). In these assays, values of enzymatic activity without the addition of Triton X-100 (lysosomal membrane is intact) present ruptured lysosomes, and values after addition of Triton X-100 present the total lysosomal population (β-hexosaminidase is released from all lysosomes in the sample). The difference between both values represents the fraction of intact organelles. Due to the different amounts of starting material for the individual approaches (1.8×10^7, 6×10^7, 1.8×10^7, and 3.5×10^7 cells for OEP, SDGC, SPIONs, and TMEM-IP, respectively), we normalized the values for each approach on the input fraction (PNS) treated with Triton X-100 (Table S1). We then calculated the percentage of intact lysosomes contained in the input fraction we recovered for the individual approaches.

DOI: 10.1021/acs.jproteome.9b00580
J. Proteome Res. 2020, 19, 371−381

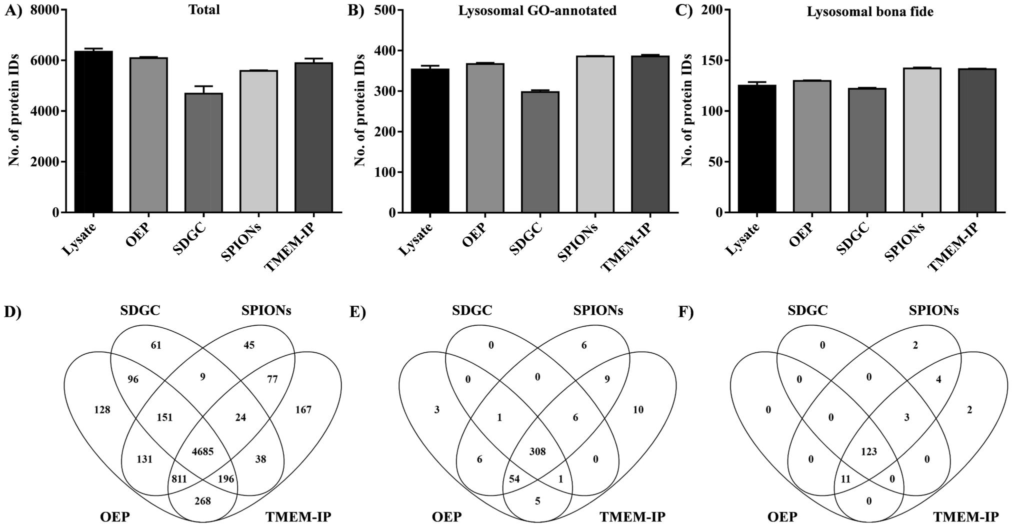


Figure 2. (A–C) Number of identified (A) total, (B) lysosomal GO-annotated, and (C) lysosomal bona fide proteins for the whole cell lysate and the four enrichment strategies. Shown are average values ($n = 3$) and standard deviation. (D–F) Overlap of identified (D) total, (E) lysosomal GO-annotated, and (F) lysosomal bona fide proteins for the four enrichment strategies. Lysate: whole cell lysate; OEP: organelle-enriched pellet; SDGC: sucrose density gradient centrifugation; SPIONs: superparamagnetic iron oxide nanoparticles; and TMEM-IP: TMEM192-3xHA immunoprecipitation.

On average, we were able to retrieve 79% (OEP, Figure 1B), 7% (SDGC, Figure 1C), 47% (SPIONs, Figure 1D), and 81% (TMEM-IP, Figure 1E) of the intact lysosomes contained in the starting material in the lysosome-enriched fraction. Due to the variations in isolation efficiency, different amounts of starting material have to be used for the individual approaches to yield similar amounts of lysosomes. The highest number of cells is needed for SDGC, followed by TMEM-IP and SPIONs, while the OEP can be generated with the lowest amount of input material. As we only needed 1 μg of protein for each analysis, the different yields did not affect our performance as all methods resulted in a sufficient amount of material. However, if higher amounts of material are desired for further studies, for example, addressing the distribution of proteins across lysosomal subcompartments (lumen, membrane, and lysosome-associated), this may present a limitation for the individual approaches. Furthermore, the price for lysosome isolation varies substantially based on the method employed. When only taking the material used for lysosome isolation into account, the OEP and SDGC samples are virtually for free while one isolation by TMEM-IP (~$100 excluding VAT) is approximately 3 times as expensive as the SPION-based isolation (~$33 excluding VAT).

In the lysosome-enriched fractions, 85, 80, and 85% of lysosomes were intact for OEP, SDGC, and SPIONs, respectively, indicating that the lysosomal damage resulting from the individual approaches was comparable. No conclusion could be made for the TMEM-IP samples as elution from the anti-HA beads is performed by using Triton X-100, which disrupts the lysosomal membrane. Since only the enrichment of intact lysosomes results in β-hexosaminidase activity, also for the TMEM-IP, a comparable or higher integrity can be assumed. While the OEP sample contained virtually no lysosomes in the supernatant, the SPION and TMEM-IP samples showed losses of 25 and 11%, respectively. These data indicate that the OEP and the TMEM-IP are the most efficient methods for the recovery of lysosomes from the starting material. When further considering the numbers of cells used (by normalizing the obtained β-hexosaminidase activity in the eluate fraction on the cell number, Table S1); however, it becomes apparent that SPIONs and OEP (which resulted in similar values) outperform the TMEM-IP and SDGC by 1.8-, and 6-fold, respectively. This is due to the fact that for both the TMEM-IP and the SDGC, the douncing of cells has to be performed in relatively small volumes at high cell densities as a highly concentrated sample is used as the starting material. This results in a reduced yield of lysosomes as the douncing efficiency of cells is lower at higher cell densities, representing a drawback of these approaches in their current version. For the SPION and OEP approaches, this limitation does not apply since lysosomes are concentrated from a high-volume input sample. However, also for these approaches, a minimum density of cells has to be used for douncing as the intactness of lysosomes negatively correlates with increasing sample volume. Therefore, if small amounts of cells or tissue are to be analyzed, a compatible low-volume douncer has to be utilized and the conditions must be optimized for the individual sample. This also presents a limitation for the applicability of all these approaches to low-abundant samples as alternative approaches would have to be applied to release intact lysosomes from cells. Furthermore, we observed higher amounts of intact lysosomes in the flow through of the SPION-enriched samples. This is probably related to lysosomal populations that do not receive SPIONs within the pulse-chase conditions used for our experimental setup.

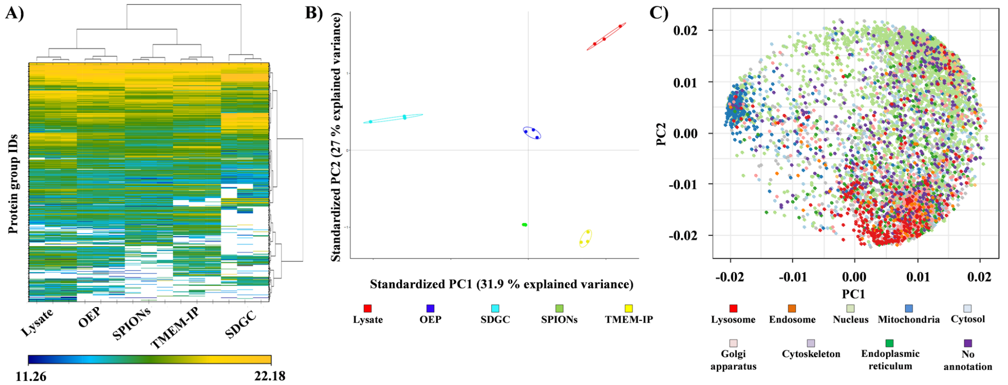


Figure 3. (A) Unsupervised clustering of the quantitative values for the whole cell lysate and the lysosome enrichment methods. Each column represents one individual replicate for the respective method, the color code depicts the normalized intensity of the individual proteins. (B) PCA plot with two defined standardized principal components PC1 and PC2 for the whole cell lysate and the four enrichment methods. (C) PCA loading plot visualizing the individual protein IDs and their GO category. Lysate: whole cell lysate; OEP: organelle-enriched pellet; SDGC: sucrose density gradient centrifugation; SPIONs: superparamagnetic iron oxide nanoparticles; and TMEM-IP: TMEM192-3xHA immunoprecipitation.

DIA Analysis of Lysosome-Enriched Fractions

We subsequently performed in-solution digestion for the individual eluate fractions of the different enrichment methods as well as a whole cell digest, which we included as a reference sample. For the generation of a spectral library, we analyzed all samples initially by 240 min LC−MS/MS DDA runs. The library, generated by the Pulsar algorithm integrated in the Spectronaut software, covered 6966 proteins, 96,268 peptides, and 122,259 precursors. Subsequently, we analyzed three independent biological replicates for each of the five approaches by 120 min gradients and DIA followed by data analysis with Spectronaut (Table S2).

The whole cell lysate resulted, as expected, in the highest number of protein identifications (on average 6329 proteins), followed by the OEP (6086 proteins), TMEM-IP (5890 proteins), SPIONs (5586 proteins), and SDGC (4689 proteins) (Figure 2A). While it was not surprising to us that the whole cell lysate resulted in the highest number of identified proteins, we did not expect to find higher numbers of proteins in the TMEM-IP and SPION samples compared to SDGC as one would assume that the first two should be the more specific approaches. Among the proteins covered with the individual enrichment approaches, we observed an overlap of 68.4%, with the majority being contributed by the SPION and TMEM-IP samples, which shared 85%.

As the analysis of lysosomal proteins was the main goal of our efforts, we further focused on proteins with known lysosomal localization in the datasets (Figure 2B,C). First, we compared the proteins to such which contained the term "lysosome" in their GO annotation (www.pantherdb.org, 683 proteins in total, Table S3). In the individual samples, we were able to identify on average between 353 and 391 of these proteins (Figure 2B). As this list also contains a high number of proteins that are annotated to be lysosomal based on similarities to other proteins, we further used a short list of bona fide lysosomal and lysosome-associated proteins, which have all been shown to be of lysosomal origin in molecular biology experiments (187 proteins in total, Table S3).[10] For this list, relative to the whole cell lysate, we observed an average increase of 12% for the identification of lysosomal proteins for SPIONs and TMEM-IP, whereas no significant increase could be observed for the OEP and SDGC methods (Figure 2C). The identified lysosomal proteins for the TMEM-IP are in agreement with a previously published study employing a similar experimental setup (Figure S2).[14]

We further validated the overlap of identified total, GO-annotated, and bona fide lysosomal proteins for the four enrichment methods (Figure 2D−F). As expected, we observed the highest discrepancy on the level of identified total proteins and with a decreasing size of the population analyzed, also the differences decreased. While for the GO-annotated list, the OEP sample still was able to contribute a few proteins not covered by the other approaches, the SDGC sample failed to provide any orthogonality irrespective of which list was used. The highest number of unique identifications for both lysosomal protein lists was contributed by SPIONs and the TMEM-IP. To assess the reproducibility of the approaches, we furthermore investigated the variability of signal intensities for the identified protein groups in the biological replicates of the respective samples (Figure S3A, gray bars). With 4568 and 3783 identifications with CVs < 20% and < 10%, respectively, the SPION method outperformed the other enrichment strategies significantly. Especially for protein identifications with a CV < 10%, a strong increase was observed. SPIONs identified 5.9-, 5.9-, 2,6-, and 1.7-fold more proteins with CVs < 10% compared to the whole cell lysate, SDGC, OEP, and TMEM-IP, respectively. This shows that despite the use of DIA, the enrichment of lysosomes has a beneficial effect on the generated datasets. This is probably due to the increase in signal intensity, especially for low-abundant lysosomal proteins. This is also confirmed by the average coefficient of variation as well as the dynamic range of variation among the enrichment strategies (Figure S3B), which is lowest for the SPION-enriched samples.

Based on these results, we conclude that the TMEM-IP and SPIONs result in the highest number of identified total and lysosomal proteins, while the data generated by the latter are

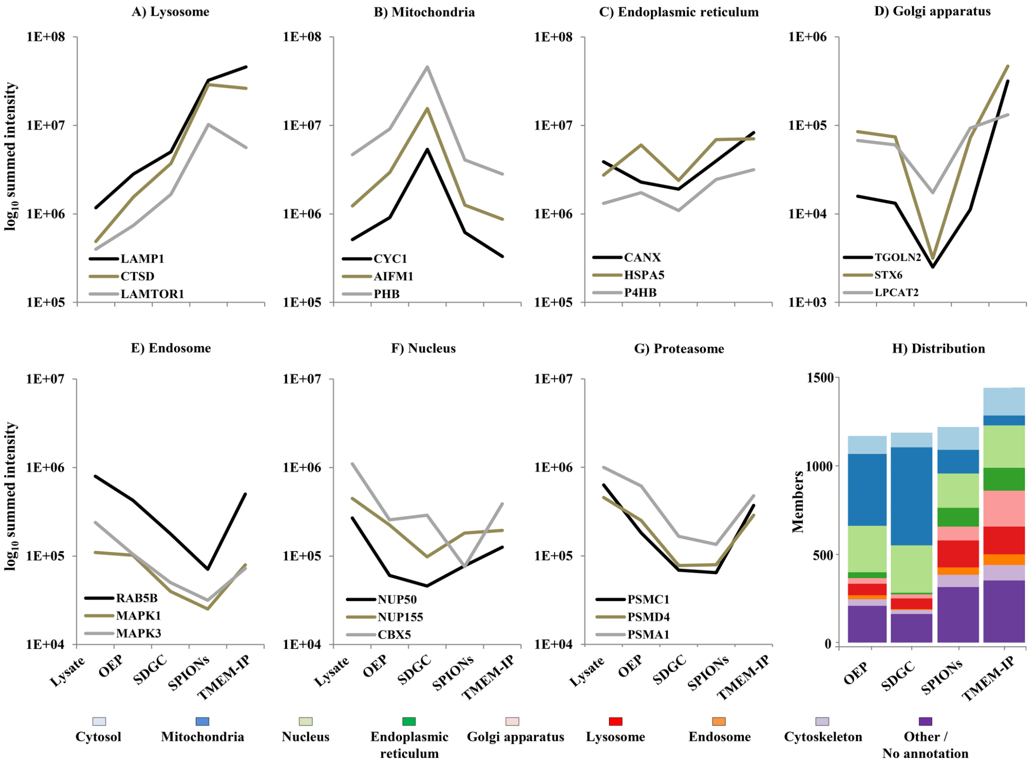


Figure 4. (A–G) Change in protein abundance for three distinct marker proteins of seven individual organelles and large cellular structures for the whole cell lysate and the different enrichment methods. The y axis shows the average summed intensity for the three replicates. H) Overview of enriched proteins (log 2 ratio > 0.58,q < 0.05 compared to the lysate) assigned to organelles. Lysate: whole cell lysate; OEP: organelle-enriched pellet; SDGC: sucrose density gradient centrifugation; SPIONs: superparamagnetic iron oxide nanoparticles; and TMEM-IP: TMEM192-3xHA immunoprecipitation.

more reproducible. Therefore, if high-quality label-free quantification data are of importance, SPIONs or TMEM-IP is the method of choice. If only identification of lysosomal proteins is desired and high CV values are of no concern, the OEP sample presents the best compromise between effort and result quality as it is more cost-effective than SPIONs and TMEM-IP and no stable cell lines have to be generated. Furthermore, for samples that cannot be specifically enriched for lysosomes, the DIA analysis of OEP samples allows an increased performance compared to whole cell lysates for the majority of lysosomal proteins. These are, for example, low abundant cells, tissue samples from animal models or patients, cells that do not actively deliver SPIONs to the lysosomal compartment, or such that cannot be stably transfected with TMEM192-3xHA.

Individual Enrichment Methods Yield Distinct Protein Populations

Initially, to assess the overall quality of the quantitative dataset, we generated a heat map clustered in a row- and column-wise manner for all data points with high confidence (Figure 3A). For the individual methods, we observed a highly reproducible clustering of intensities among the independent biological replicates. Furthermore, we were able to identify for each method protein populations showing a unique pattern indicating that functionally connected protein groups were enriched in the individual datasets.

To further assess these differences and determine how many variables are contained in the dataset, we performed a principal component analysis (PCA). The PCA plot (Figure 3B) shows that the two principal components PC1 and PC2 explained 31.8 and 26.9% of the variance in the dataset. As the replicates of the individual enrichment strategies are clustered closely together, the main variance in the dataset (PC1) can be explained based on the different enrichment strategies themselves. At this point, we did not consider more principal components as PC1 and PC2 already allowed for a good separation of the samples. Considering the TMEM-IP, for example, PC1 explains the difference to the SDGC method, whereas PC2 accounts for the difference to the whole cell lysate.

Subsequently, we investigated which proteins (classes) drove the separation between the enrichment strategies. For this analysis, we extracted the PC1 and PC2 loadings for each protein (individual contribution of each protein to the

DOI: 10.1021/acs.jproteome.9b00580
J. Proteome Res. 2020, 19, 371−381

separation in the respective principal component), plotted them, and colored the proteins according to their GO annotation (Figure 3C). This analysis revealed that nuclear proteins were depleted from the sample in all approaches and that SDGC was especially efficient in the enrichment of mitochondria. PC2 separated the TMEM-IP and SPIONs from the whole cell lysate, and the analysis of protein loadings (Figure 3C) showed that this separation was primarily driven by lysosomal proteins and to a minor extent by proteins originating from the Golgi apparatus. By correlating the organelle-specific protein population distribution (Figure 3C) with the different enrichment approaches (Figure 3B), it further is becoming apparent that the highest enrichment of lysosomal proteins was achieved by the SPION and the TMEM-IP methods, which show a similar profile.

Quantification of Organelle Abundance

To further investigate the efficiency of the individual approaches, we quantified the extent of organelle enrichment based on the label-free quantification values for the individual proteins. As base line, we applied the intensities obtained for the whole cell lysate. Initially, we plotted the summed intensities of three marker proteins for a specific subcellular population (Figure 4A−G and Table S1). For the lysosomal proteins LAMP1 (membrane), CTSD (luminal), and LAMTOR1 (membrane-anchored), we observed an increase in signal intensity for all enrichment methods with the highest values for SPION and TMEM-IP samples ranging from 14- to 59-fold. When considering the summed intensities of bona fide lysosomal proteins that were identified in all samples analyzed, this increase ranged from 2.4- to 28-fold dependent on the enrichment method (Figure S4). While the representative nuclear proteins decreased for all enrichment methods, only SDGC was able to reduce the amount of some of the selected ER proteins, while at the same time, it was the only approach resulting in a strong increase for mitochondrial markers. This is probably due to similar densities of lysosomes and mitochondria, which therefore cannot be separated by the applied gradient. Endosomes and the proteasome were efficiently depleted by all approaches except the TMEM-IP, which furthermore resulted in an increase in signal intensity for the Golgi apparatus marker proteins.

We then extended these analyses to the whole protein population and generated volcano plots for the proteins assigned with the respective GO categories (Figure S5 and Table S4). Shown is the average q-value and log 2 fold-change (relative to the whole cell lysate) across three biological replicates with a significance cutoff of 0.05 for the q-value (multiple testing corrected p-value) and 0.58 for the absolute log 2 fold-change. In these analyses, SPIONs and TMEM-IP show a bigger proportion of significantly enriched lysosomal proteins compared to the OEP and SDGC samples. Furthermore, the enrichment of mitochondrial proteins was further confirmed for the SDGC samples. In all four methods, we found a paucity for the majority of nuclear proteins.

Based on these data, we calculated how many proteins assigned to a specific organelle were enriched (log 2 ratio > 0.58, $q < 0.05$) compared to the lysate by each method to obtain a global picture of the enrichment for individual organelles (Figure 4H and Table S4). The size of the organelle-specific section of each stacked bar represents the number of proteins assigned to the specific organelle for the respective approach. For the OEP and the SDGC samples, we observed roughly similar values for all organelles, with the exception of mitochondria, matching the previous results. For the SPION and TMEM-IP samples, we obtained similar results for the majority of organelles with an equal performance for the enrichment of lysosomal proteins. The only outliers were proteins originating from the Golgi apparatus, which were overrepresented for the TMEM-IP and for such originating from mitochondria, which were higher for SPIONs. While the overabundance of Golgi apparatus-related proteins is in well accordance with the single protein intensities (Figure 4D), the discrepancies in mitochondrial proteins were not apparent for our selected marker proteins (Figure 4B). The higher abundance of mitochondrial proteins in the SPION dataset could be due to the different isolation methods utilized for the individual approaches: SPION-containing lysosomes are isolated using columns, which are packed with a dense bed of metal beads. This structure could retain the mitochondria due to interactions with its surface or due to the fluidic properties of the packing bed, which should not be the case for the TMEM-IP anti-HA beads, for which microtubes in combination with a magnet are used. The increased abundance of Golgi apparatus-related proteins for the TMEM-IP could be either due to the presence of the protein in this location or in vesicles derived from it. In total, the results obtained for SPIONs and TMEM-IP were surprisingly similar, which is also reflected in the >85% overlap in identified proteins for both methods (Figure 2D). This indicates that the majority of unspecifically enriched proteins are associated to the lysosomes and are not due to the columns/beads used as the two approaches utilize different workflows for enrichment.

Analysis of Lysosomal Integrity

Based on the β-hexosaminidase assays, a significant number of lysosomes were already ruptured during the process of cell lysis by douncing. Furthermore, the relatively broad distribution of fold-change values for individual lysosomal proteins in the volcano plots (Figure S5), especially for the SPION and TMEM-IP samples, indicates different enrichment factors for distinct proteins. This could be due to (i) the presence of certain lysosomal proteins at cellular compartments other than the lysosome, (ii) the selective enrichment of lysosomal (sub) fractions, or (iii) the loss of lysosomal proteins during the enrichment procedure, possibly due to a dissociation of associated complexes or by the release of luminal proteins caused by the rupture of the lysosomal membrane.

As such discrepancies may result in data that do not reflect the lysosomal composition properly, we further investigated this by plotting the intensity ratios of a subset of lysosomal proteins for each of the enrichment methods relative to the whole cell lysate (Figure 5A−C). For this analysis, we utilized the list of bona fide lysosomal proteins (Table S3)[10] and identified all proteins in our datasets, which were detected in each of the enrichment methods (70 proteins). Since we were able to recover the majority of intact lysosomes in the OEP, we used the ratio of OEP/whole cell lysate as a benchmark and compared the other three isolation methods in relation to it.

For the SDGC samples, the enrichment factor was for most proteins slightly higher than the OEP, which is in agreement with our previous analyses (Figure 4). We observed, however, a higher variability for the intensity of lysosomal proteins and for several a lower enrichment factor (Figure 5A). Compared to the OEP, both the SPIONs (Figure 5B) and the TMEM-IP (Figure 5C) resulted in significantly higher enrichment factors

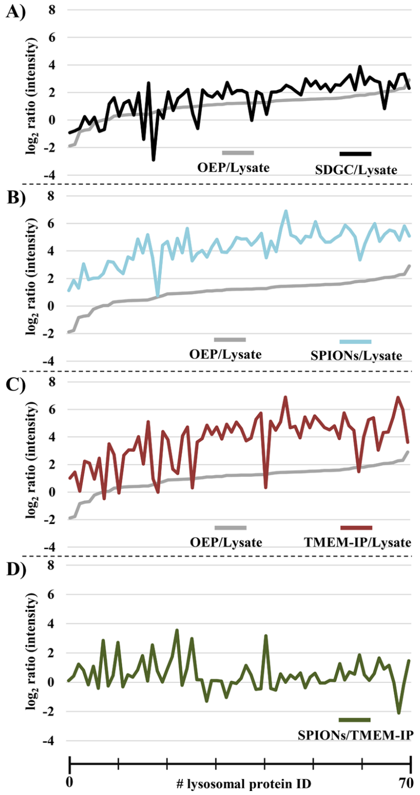


Figure 5. (A−C) Ratio of the summed protein abundances for the individual enrichment methods and the whole cell lysate for bona fide lysosomal proteins detected in all analyses. Lysosomal proteins were plotted based on their OEP/lysate fold-change ratio. (D) Ratio of the summed protein abundances for SPIONs divided by TMEM-IP. Lysate: whole cell lysate; OEP: organelle-enriched pellet; SDGC: sucrose density gradient centrifugation; SPIONs: superparamagnetic iron oxide nanoparticles; and TMEM-IP: TMEM192-3xHA immunoprecipitation.

(up to 118-fold) for the majority of lysosomal proteins. The TMEM-IP, however, resulted in a higher variability and, in certain instances, even in lower enrichment values than the OEP samples. We hypothesized that this may be either due to the dissociation of associated protein complexes or because of the recovery of broken lysosomes by the TMEM-IP as immunoprecipitation of the tagged membrane protein does not require the lysosome to be intact, while for SPIONs, only intact lysosomes containing the magnetic nanoparticles should be retained on the column. We therefore further compared the signal intensities of lysosomal proteins identified in SPION and TMEM-IP samples (Figure 5D). These data show that, with a few exceptions, all proteins showing a differential abundance are overrepresented in the SPION samples.

To further follow up on this effect, we extracted entries from our list of bona fide lysosomal proteins that were identified by both approaches (137 in total) and subsequently performed a binary comparison (Figure S6A). We applied an arbitrary cutoff of 2-fold for up- or downregulation and investigated which proteins were affected. We identified 36 proteins that were overrepresented in the SPION samples and seven proteins that were more abundant in the TMEM-IP (Table S5). The proteins with higher intensities in the TMEM-IP did not follow a clear pattern including five membrane proteins and two luminal hydrolases. The group of proteins with a higher abundance in the SPION sample, however, included 18 proteins belonging to the lysosome-associated complexes vATPase and mTORC1, 10 luminal hydrolases, and only 8 membrane proteins. This included TMEM-192, which is probably retained on the HA beads upon elution of lysosomes with Triton X-100. Strikingly, while the members of the mTORC1 complex associated with the membrane by a lipid anchor (Rheb and LAMTOR1) were similar in abundance for both approaches, all cytosolic parts of the complex (with the exception of LAMTOR 2/3) were overrepresented in the SPION sample (Figure S6B). Also, for the vATPase complex, all cytosolic subunits were more intense for the SPION sample as well as one of the inter-membrane subunits with a large cytosolic domain (Figure S6C). These data suggest that the isolation of lysosomes by TMEM-IP results in a loss of lysosome-associated as well as luminal proteins. This could either be due to rupture of the organelle during the isolation process, a change in lysosomal properties as a result of TMEM192-3xHA expression, or dissociation of the associated proteins because of the buffers employed in the isolation procedure. Aside from the associated complexes, we did not observe major differences in the abundance of known lysosomal proteins between the individual approaches. Therefore, based on these experiments, we have the impression that neither the use of SPIONs nor of TMEM-192-3xHA strongly influences the composition of lysosomes in the HEK 293 cells used in this study.

■ CONCLUSIONS

In the current study, we evaluated four different approaches for the generation of lysosome-enriched fractions and their analysis by DIA. While OEP and SDGC allow for an increase in signal intensity for lysosomal proteins, they do not result in higher identification numbers compared to the whole cell lysate. This is probably due to the DIA analysis of the sample as even in the highly complex whole cell lysate, a sufficient amount of fragment ions is identified to allow for the identification of lysosomal proteins. For SPIONs and TMEM-IP, we achieved a significant enrichment of lysosomal proteins (up to >100-fold increase in signal intensity for selected proteins), a better coverage of the lysosomal proteome, and a more robust quantification with markedly lower CV values compared to the more complex samples. This shows that, despite the application of DIA, the enrichment of lysosomes with these approaches is beneficial for the generation of highly reproducible data and to maximize coverage for lysosomal proteins. While both methods seem to achieve similar results, the TMEM-IP results in a higher recovery of intact lysosomes from the starting material but overall in a lower yield. Furthermore, it exhibits a higher variability in protein abundance compared to the whole cell lysate. This seems mainly to be due to the loss of proteins in lysosome-associated complexes as well as luminal hydrolases. Therefore, the isolation of lysosomes by SPIONs is able to yield the highest number of lysosomes per starting material and is able to preserve their integrity in the most efficient way.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.9b00580.

Figure S1, Western blot verification for the stable expression of TMEM192-3xHA in HEK 293 cells; Figure S2, comparison of TMEM-IP data with a published dataset; Figure S3, variability of signal intensities for the individual datasets; Figure S4, summed abundances of 70 bona fide lysosomal proteins detected in all samples; Figure S5, enrichment factors for individual proteins detected with the four methods of lysosome enrichment relative to the whole cell lysate; Figure S6, variability in protein abundance between SPIONs and TMEM-IP (PDF)

Table S1, β-hexosaminidase assays (XLSX)

Table S2, protein IDs (XLSX)

Table S3, list of lysosomal proteins (XLSX)

Table S4, candidate list (XLSX)

Table S5, absolute abundance and abundance ratios for 70 lysosomal proteins (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: dominic.winter@uni-bonn.de. Fax: +49 228 732416.

ORCID

Jan Muntel: 0000-0003-2320-5829
Roland Bruderer: 0000-0002-1049-5524
Lukas Reiter: 0000-0001-5751-3139
Dominic Winter: 0000-0001-6788-6641

Notes

The authors declare the following competing financial interest(s): JM, RB and LR are employees of Biognosys AG. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository[29] with the dataset identifier PXD015098.

■ ACKNOWLEDGMENTS

The authors appreciate financial support from the Studienstiftung des Deutschen Volkes and the DFG Research Unit FOR2625.

■ REFERENCES

(1) Balogi, Z.; Multhoff, G.; Jensen, T. K.; Lloyd-Evans, E.; Yamashima, T.; Jäättelä, M.; Harwood, J. L.; Vígh, L. Hsp70 interactions with membrane lipids regulate cellular functions in health and disease. *Prog. Lipid Res.* **2019**, *74*, 18−30.

(2) Ballabio, A.; Gieselmann, V. Lysosomal disorders: from storage to cellular damage. *Biochim. Biophys. Acta, Mol. Cell Res.* **2009**, *1793*, 684−696.

(3) Ballabio, A. The awesome lysosome. *EMBO Mol. Med.* **2016**, *8*, 73−76.

(4) Wildenthal, K.; Decker, R. S. The role of lysosomes in the heart. *Adv. Myocardiol.* **1980**, *2*, 349−358.

(5) Davidson, S. M.; Vander Heiden, M. G. Critical Functions of the Lysosome in Cancer Biology. *Annu. Rev. Pharmacol. Toxicol.* **2017**, *57*, 481−507.

(6) Fraldi, A.; Klein, A. D.; Medina, D. L.; Settembre, C. Brain Disorders Due to Lysosomal Dysfunction. *Annu. Rev. Neurosci.* **2016**, *39*, 277−295.

(7) Settembre, C.; Fraldi, A.; Medina, D. L.; Ballabio, A. Signals from the lysosome: a control centre for cellular clearance and energy metabolism. *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 283−296.

(8) Zoncu, R.; Bar-Peled, L.; Efeyan, A.; Wang, S.; Sancak, Y.; Sabatini, D. M. mTORC1 senses lysosomal amino acids through an inside-out mechanism that requires the vacuolar H(+)-ATPase. *Science* **2011**, *334*, 678−683.

(9) Carroll, B.; Dunlop, E. A. The lysosome: a crucial hub for AMPK and mTORC1 signalling. *Biochem. J.* **2017**, *474*, 1453−1466.

(10) Thelen, M.; Winter, D.; Braulke, T.; Gieselmann, V. SILAC-Based Comparative Proteomic Analysis of Lysosomes from Mammalian Cells Using LC-MS/MS. *Methods Mol. Biol.* **2017**, *1594*, 1−18.

(11) Rabanal-Ruiz, Y.; Korolchuk, V. I. mTORC1 and Nutrient Homeostasis: The Central Role of the Lysosome. *Int. J. Mol. Sci.* **2018**, *19*, 818.

(12) Aguado, C.; Pérez-Jiménez, E.; Lahuerta, M.; Knecht, E. Isolation of Lysosomes from Mammalian Tissues and Cultured Cells. *Methods Mol. Biol.* **2016**, *1449*, 299−311.

(13) Walker, M. W.; Lloyd-Evans, E. A rapid method for the preparation of ultrapure, functional lysosomes using functionalized superparamagnetic iron oxide nanoparticles. *Methods Cell Biol.* **2015**, *126*, 21−43.

(14) Wyant, G. A.; Abu-Remaileh, M.; Frenkel, E. M.; Laqtom, N. N.; Dharamdasani, V.; Lewis, C. A.; Chan, S. H.; Heinze, I.; Ori, A.; Sabatini, D. M. NUFIP1 is a ribosome receptor for starvation-induced ribophagy. *Science* **2018**, *360*, 751−758.

(15) Leighton, F.; Poole, B.; Beaufay, H.; Baudhuin, P.; Coffey, J. W.; Fowler, S.; de Duve, C. The large-scale separation of peroxisomes, mitochondria, and lysosomes from the livers of rats injected with triton wr-1339: Improved Isolation Procedures, Automated Analysis, Biochemical and Morphological Properties of Fractions. *J. Cell. Biol.* **1968**, *37*, 482−513.

(16) Graham, J. M. Purification of a crude mitochondrial fraction by density-gradient centrifugation. *Curr. Protoc. Cell Biol.* **2001**, *4*, 3.4.1−3.4.22.

(17) Della Valle, M. C.; Sleat, D. E.; Zheng, H.; Moore, D. F.; Jadot, M.; Lobel, P. Classification of Subcellular Location by Comparative Proteomic Analysis of Native and Density-shifted Lysosomes. *Mol. Cell. Proteomics* **2011**, *10*, M110-006403.

(18) Isolation of rat liver lysosomes by isopycnic centrifugation in a metrizamide gradient. *J. Cell Biol.* **1978**, *78*, 349−368, DOI: 10.1083/jcb.78.2.349.

(19) Chapel, A.; Kieffer-Jaquinod, S.; Sagné, C.; Verdon, Q.; Ivaldi, C.; Mellal, M.; Thirion, J.; Jadot, M.; Bruley, C.; Garin, J.; Gasnier, B.; Journet, A. An Extended Proteome Map of the Lysosomal Membrane Reveals Novel Potential Transporters. *Mol. Cell. Proteomics* **2013**, *12*, 1572−1588.

(20) (A) Pertoft, H.; Rubin, K.; Kjellén, L.; Laurent, T. C.; Klingeborn, B. The viability of cells grown or centrifuged in a new density gradient medium, Percoll(TM). *Exp. Cell Res.* **1977**, *110*, 449−457. (B) Schröder, B.; Wrocklage, C.; Pan, C.; Jäger, R.; Kösters, B.; Schäfer, H.; Elsässer, H.-P.; Mann, M.; Hasilik, A. Integral and associated lysosomal membrane proteins. *Traffic* **2007**, *8*, 1676−1686.

(21) Tharkeshwar, A. K.; Trekker, J.; Vermeire, W.; Pauwels, J.; Sannerud, R.; Priestman, D. A.; Te Vruchte, D.; Vints, K.; Baatsen, P.; Decuypere, J.-P.; Lu, H.; Martin, S.; Vangheluwe, P.; Swinnen, J. V.; Lagae, F.; Impens, F.; Platt, F. M.; Gevaert, K.; Annaert, W. A novel approach to analyze lysosomal dysfunctions through subcellular proteomics and lipidomics: the case of NPC1 deficiency. *Sci. Rep.* **2017**, *7*, 41408.

(22) Abu-Remaileh, M.; Wyant, G. A.; Kim, C.; Laqtom, N. N.; Abbasi, M.; Chan, S. H.; Freinkman, E.; Sabatini, D. M. Lysosomal metabolomics reveals V-ATPase- and mTOR-dependent regulation of amino acid efflux from lysosomes. *Science* **2017**, *358*, 807−813.

(23) Castellano, B. M.; Thelen, A. M.; Moldavski, O.; Feltes, M.; van der Welle, R. E. N.; Mydock-McGrane, L.; Jiang, X.; van Eijkeren, R. J.; Davis, O. B.; Louie, S. M.; Perera, R. M.; Covey, D. F.; Nomura, D. K.; Ory, D. S.; Zoncu, R. Lysosomal cholesterol activates mTORC1 via an SLC38A9-Niemann-Pick C1 signaling complex. *Science* **2017**, *355*, 1306−1311.

(24) Wendeler, M.; Sandhoff, K. Hexosaminidase assays. *Glycoconjugate J.* **2009**, *26*, 945−952.

(25) Lowry, O. H.; Rosebrough, N. J.; Farr, A. L.; Randall, R. J. Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* **1951**, *193*, 265−275.

(26) Rappsilber, J.; Ishihama, Y.; Mann, M. Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. *Anal. Chem.* **2003**, *75*, 663−670.

(27) Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Reiter, L. High-precision iRT prediction in the targeted analysis of data-independent acquisition and its impact on identification and quantitation. *Proteomics* **2016**, *16*, 2246−2256.

(28) Rosenberger, G.; Bludau, I.; Schmitt, U.; Heusel, M.; Hunter, C. L.; Liu, Y.; MacCoss, M. J.; MacLean, B. X.; Nesvizhskii, A. I.; Pedrioli, P. G. A.; Reiter, L.; Röst, H. L.; Tate, S.; Ting, Y. S.; Collins, B. C.; Aebersold, R. Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses. *Nat. Methods* **2017**, *14*, 921−927.

(29) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, Ş.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **2019**, *47*, D442−D450.

3.3 Publication: Supporting Information

Systematic comparison of strategies for the enrichment of lysosomes by data independent acquisition

Jasjot Singh[†], Edgar Kaade[†], Jan Muntel[§], Roland Bruderer[§], Lukas Reiter[§], Melanie Thelen[†], Dominic Winter[*, †]

[†]Institute for Biochemistry and Molecular Biology, University of Bonn, 53115 Bonn, Germany
[§]Biognosys, 8952 Zürich-Schlieren, Switzerland

***Corresponding author**
Dr. Dominic Winter
Institute for Biochemistry and Molecular Biology
University of Bonn
Nussallee 11, 53115 Bonn, Germany
Email: dominic.winter@uni-bonn.de
Fax: +49 228 732416

List of tables:

List of figures:

Supplementary references:

(1) Wyant, G. A.; Abu-Remaileh, M.; Frenkel, E. M.; Laqtom, N. N.; Dharamdasani, V.; Lewis, C. A.; Chan, S. H.; Heinze, I.; Ori, A.; Sabatini, D. M. NUFIP1 is a ribosome receptor for starvation-induced ribophagy. *Science* **2018**, *360*, 751–758.

(2) Hulsen, T.; Vlieg, J. de; Alkema, W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC genomics* **2008**, 9, 488.

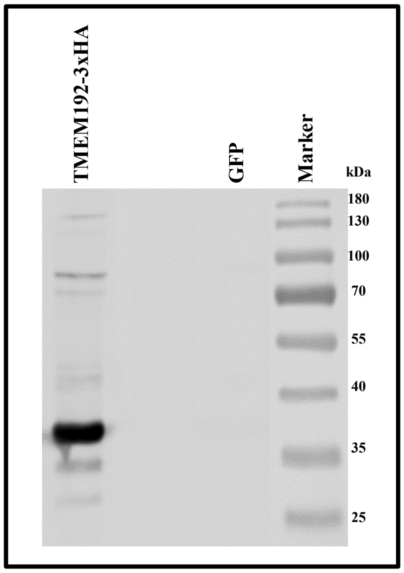


Figure S1: Western-blot verification for the stable expression of TMEM192-3xHA in HEK 293 cells. 20 µg of whole cell lysate for HEK 293 cells stably transfected with TMEM192-3xHA and HEK 293 cells transiently transfected with GFP were separated by a 10 % SDS-PAGE for 1.5 h at 120 V. Proteins separated on the gels were transferred to a nitrocellulose membrane using semi-dry electro blotter. The membrane was incubated in 5 % nonfat dry milk in tris buffer saline containing 0.05 % tween (TBS-T) for 1 h at RT. The membrane was incubated o/n at 4 °C with a primary anti HA antibody (1:2000) (self-made), washed 3x for 10 min at RT with TBS-T and incubated with a secondary goat anti mouse antibody (1:5000) (Millipore, AP180P) for 60 min at RT. Membrane was again washed 3x for 10 min at RT with TBS-T before developing the blot. Protein expression signal was detected using the enhanced chemiluminescence (ECL) kit (Bio-Rad Laboratories, Inc. (Hercules, CA, USA)), visualized with the FUSION SOLO 4M system, and illustrated/analyzed by the FusionCapt advance software.

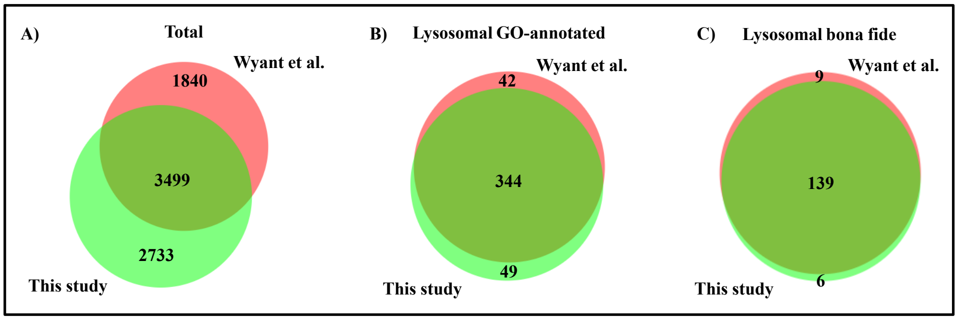


Figure S2: Comparison of TMEM-IP data with a published dataset. Comparison of data generated in this study from lysosomes isolated by TMEM-IP with a published dataset generated by Wyant et al.[1] for A) total B) lysosomal GO-annotated and C) lysosomal bona fide IDs. Venn diagram were generated with the BioVenn tool by Hulsen et al. 2008.[2]

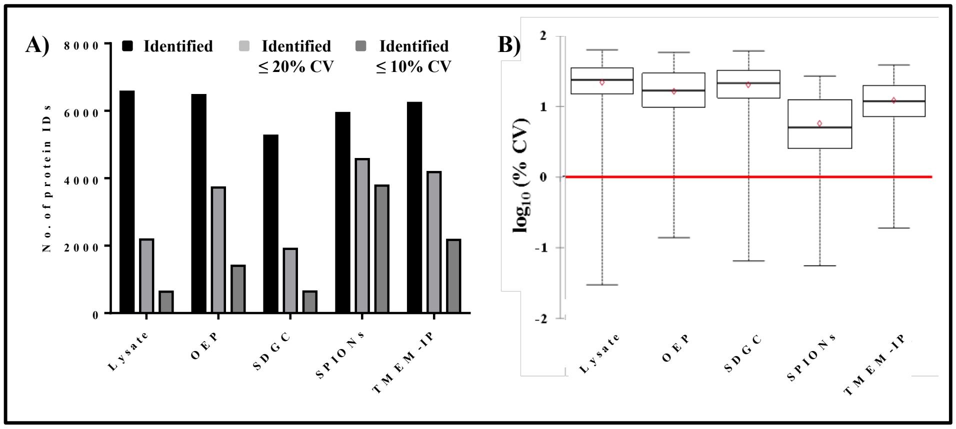


Figure S3: Variability of signal intensities for the individual datasets. Total numbers of identified proteins (black), proteins identified with a CV ≤ 20% (light grey) and a CV ≤ 10% (dark grey) for the individual datasets. B) Average CVs for the datasets of the individual methods. Lysate: whole cell lysate; OEP: organelle enriched pellet; SDGC: sucrose density gradient centrifugation; SPIONs: superparamagnetic iron oxide nanoparticles; TMEM-IP: TMEM192-3xHA immunoprecipitation.

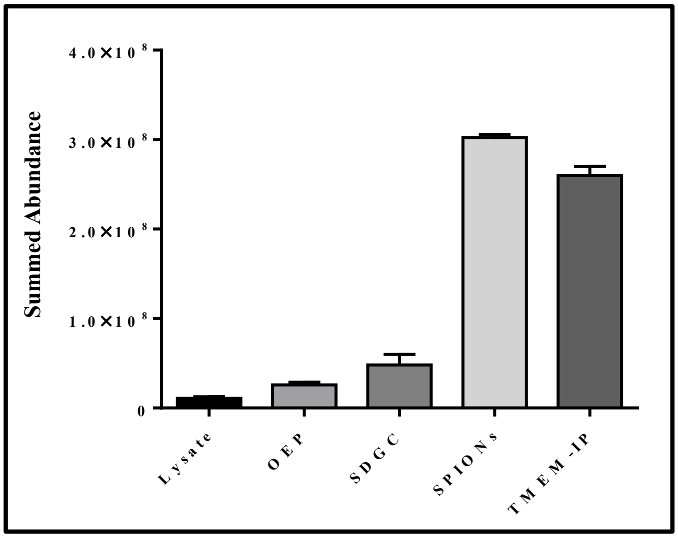


Figure S4: Summed abundances of 70 bona fide lysosomal proteins detected in all samples. Shown are the average values (n=3) plus standard deviation for the summed abundance of signal intensities of bona fide lysosomal proteins for the individual samples. Only proteins which were quantified in each of the conditions were used for this analysis. Lysate: whole cell lysate; OEP: organelle enriched pellet; SDGC: sucrose density gradient centrifugation; SPIONs: superparamagnetic iron oxide nanoparticles; TMEM-IP: TMEM192-3xHA immunoprecipitation.

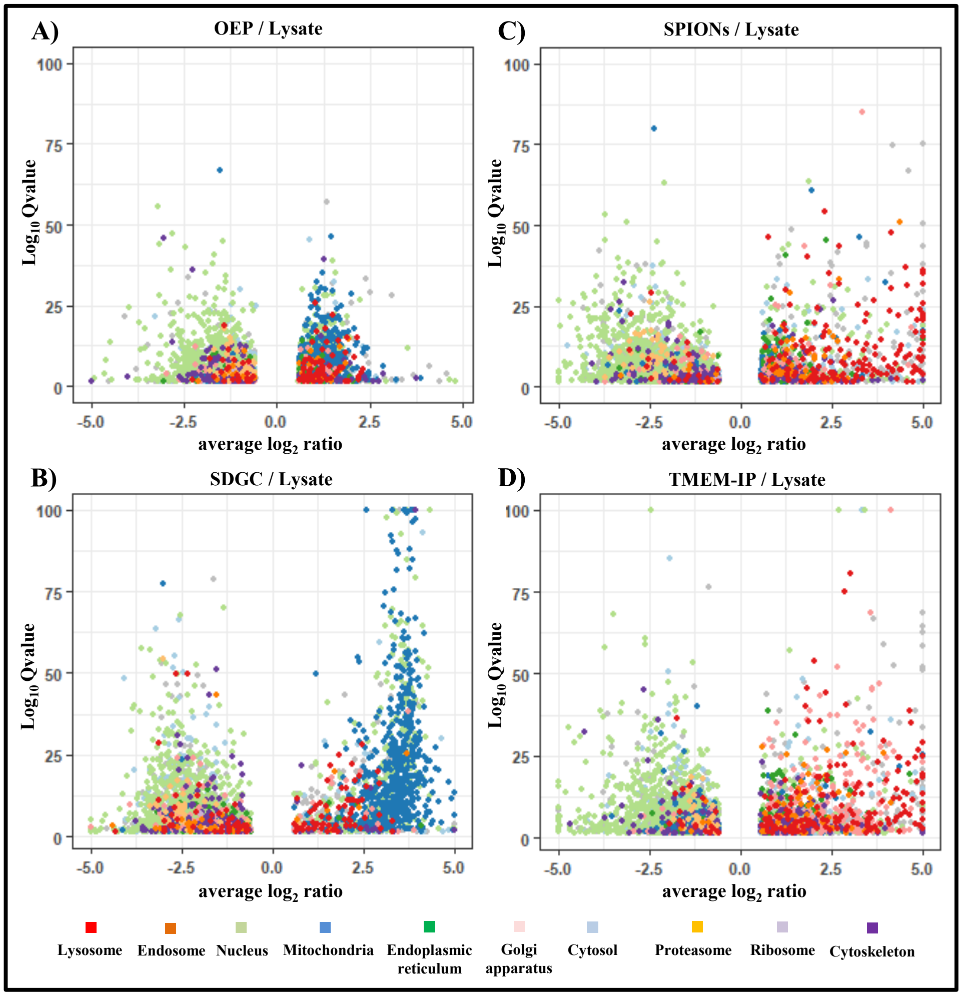
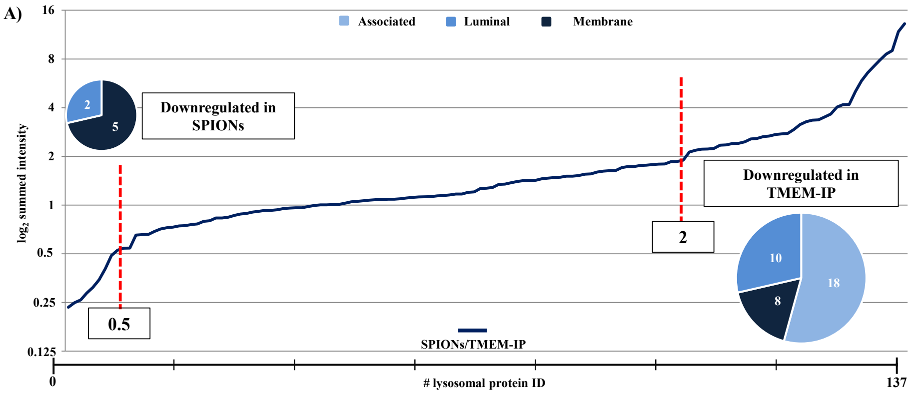


Figure S5: Enrichment factors for individual proteins detected with the four methods of lysosome enrichment relative to the whole cell lysate. Shown are the fold-change ratios of proteins passing the threshold for significant regulation with a q-value (multiple tested corrected p-value) > 0.05 and an absolute $\log_2$ ratio > 0.58 for A) OEP, B) SDGC, C) SPIONs and D) TMEM-IP. Candidates were classified based on their GO subcellular localization annotation among to the following categories: lysosome, endosome, nucleus, mitochondria, endoplasmic reticulum, Golgi apparatus, cytosol, proteasome, ribosome and cytoskeleton. OEP: organelle enriched pellet; SDGC: sucrose density gradient centrifugation; SPIONs: superparamagnetic iron oxide nanoparticles; TMEM-IP: TMEM192-3xHA immunoprecipitation.

A)



B)

Gene	Regulation	A	M
MTOR	3,65	X	
RPTOR	5,04	X	
DEPTOR	4,04	X	
MLST8	5,87	X	
RHEB	1,70		X
RRAGA	2,58	X	
RRAGC	3,51	X	
LAMTOR1	1,86		X
LAMTOR2	1,45	X	
LAMTOR3	1,48	X	
LAMTOR4	2,22	X	
LAMTOR5	11,80	X	
AKT1S1	3,28	X	
TSC2	2,93	X	

C)

Gene	Regulation	A	M
ATP6AP1	1,75		X
ATP6V0A1	2,56		X
ATP6V0C	0,26		X
ATP6V0D1	1,62		X
ATP6V1A	7,88	X	
ATP6V1B2	9,01	X	
ATP6V1C1	2,13	X	
ATP6V1D	7,21	X	
ATP6V1E1	6,57	X	
ATP6V1F	4,19	X	
ATP6V1G1	3,36	X	
ATP6V1H	8,58	X	

Figure S6: Variability in protein abundance between SPIONs and TMEM-IP. A) Abundance ratio values (log2) calculated for the summed intensities of bona fide lysosomal proteins (Table S3) quantified in the SPIONs and the TMEM-IP samples. Based on a 2-fold change cut off, 7 proteins were up- and 35 proteins downregulated in the TMEM-IP relative to SPIONs. The proteins were grouped based on their properties classifying them in: lysosomal associated, luminal or membrane bound. The majority of downregulated proteins consisted of members of the mTORC1 complex (B) and the vATPase complex (C). For the mTORC1 and vATPase complex, proteins were further classified in such being lysosomal membrane associated (A) or membrane bound (M) (Table S5).

3.4 Conclusion

For the analysis of lysosomal composition and function, it is important that as many lysosomes as possible are recovered in an intact state and that the enriched fractions contain as few contaminating proteins derived from other organelles as possible. Hence, I compared four frequently applied lysosome enrichment methods. As a reference, I chose a crude enrichment strategy centrifuging the post-nuclear supernatant (PNS) at 20,000×g, referred to as organelle-enriched pellet (OEP), because this should result in the most complete recovery and intact ratio of lysosomes. I further compared this method with a two-step sucrose density gradient centrifugation (SDGC) approach using Histodenz and Percoll. Additionally, I included two enrichment strategies specifically developed for lysosomes. In the superparamagnetic iron oxide nanoparticles (SPIONs) procedure, based on the endocytosis of dextran-coated nanoparticles and their delivery to the lysosomal compartment, lysosomes containing these particles are isolated using magnetic columns. In the TMEM192-3xHA immunoprecipitation (TMEM-IP) strategy, cells stably expressing a 3xHA-tagged version of the lysosomal membrane protein TMEM192 were used for lysosome isolation by anti-HA antibodies immobilized on magnetic beads **(Figure 1A)**.

Applying the β-hexosaminidase enzymatic activity assay, I assessed the recovery and intact ratio of lysosomes. On average, I was able to recover 79 % (OEP), 7 % (SDGC), 47 % (SPIONs), and 81 % (TMEM-IP) of the intact lysosomes contained in the starting material **(Figure 1B-E)**. In the lysosome-enriched final fractions used for further quantitative analysis, 85, 80, and 85 % of the lysosomes were intact for OEP, SDGC, and SPIONs, respectively, indicating that lysosomal damage by each approach was comparable **(Figure 1B-E)**. Due to the application of the detergent Triton-X-100 during the elution process of TMEM-IP samples, no conclusion could be drawn on the intact status. However, since only the enrichment of intact lysosomes leads to β-hexosaminidase activity, comparable or higher integrity can also be assumed for TMEM-IP.

Utilizing the label-free quantification DIA-MS/MS approach, I further investigated to what extent the different methods cause differences in the proteomic composition of lysosomal proteins. As a reference sample for the calculation of enrichment factors, a whole cell lysate (WCL) sample was included. As expected, the WCL samples resulted on average in the highest number of protein identifications (6,329) followed by the OEP (6,086), TMEM-IP (5,890), SPIONs (5,586), and SDGC (4,689) **(Figure 2A)**. However, as the analysis of the lysosomal proteome was my main goal, I further compared the number of lysosomal proteins based on an extended (683) and a list of biologically validated (187) proteins. A comparison to the first extended list for the individual samples resulted on average between 353 and 391 of these proteins, while applying the second more stringent list, I observed an increase of 12 % for the identification of lysosomal proteins for SPIONs and TMEM-IP when compared to WCL **(Figure 2B, C)**. In contrast, no significant increase could be observed for the OEP and SDGC methods. SPIONs and the TMEM-IP contributed the highest number of unique identifications for both lysosomal protein lists **(Figure 2 E, F)**.

To further assess the reproducibility of the enrichment approaches, I further examined the variability of signal intensities (protein abundance) for the identified protein groups in the biological replicates of the respective samples. With 4,568 and 3,783 identifications with a CV < 20 % and < 10 %, respectively, the SPION method was superior to the other enrichment strategies **(Figure S3 A)**. In particular, a significant increase was observed in protein identifications with a CV < 10 %. SPIONs identified 5.9-, 5.9-, 2.6-, and 1.7-fold more proteins with CVs < 10 % than WCL, SDGC, OEP, and TMEM-IP, respectively. This demonstrates that despite the widespread belief that enrichment approaches are unnecessary while using DIA, my results show that lysosome enrichment is beneficial for generating highly reproducible data and maximizing coverage for lysosomal proteins **(Figure S3 A, B)**. This is most likely due to the increase in signal intensity, especially for low abundant lysosomal proteins.

Based on the conclusion that TMEM-IP and SPIONs yield the highest number of identified lysosomal proteins, while the data generated with the latter are more reproducible, I additionally identified individual proteomic differences among the enrichment strategies. By applying a heat map and principal component analysis (PCA), I showed that the two principal components PC1 and PC2 explain 31.9 and 27 % of the variance in the data set **(Figure 3A, B)**. Since the replicates of the individual enrichment strategies showed a close clustering, the main variance in the dataset (PC1) could be explained by the different enrichment strategies themselves. To define driving factors causing the separation between the enrichment strategies, PC1 and PC2 loadings for each protein (individual contribution of each protein to the separation in each PC) were extracted and plotted according to their subcellular annotation. The analysis revealed an efficient depletion of nuclear proteins for all enrichment approaches, while in particular SDGC yielded samples in which contaminating mitochondrial proteins were enriched **(Figure 3C)**.

Further, the analysis showed an efficient enrichment (abundance) of lysosomal proteins for the TMEM-IP and SPIONs with enrichment factors of 14-59-fold for selected lysosomal proteins such as LAMP1 and CTSD compared to the total cell lysate **(Figure 4A)**. Besides that, I was able to show an overrepresentation of Golgi apparatus-related proteins for the TMEM-IP, which is probably a result due to the presence of the protein in this location or in vesicles derived from it **(Figure 4D)**.

In a final analysis, I further assessed the unique differences of the lysosomal proteome among the enrichment strategies, as differences might arise due to 1) the presence of certain lysosomal proteins in cellular compartments other than the lysosome, 2) selective enrichment of lysosomal (sub)fractions, or 3) loss of lysosomal proteins during the enrichment procedure. For the analysis, intensity ratios of 70 lysosomal proteins for each of the enrichment methods relative to the WCL sample were plotted, in which the abundance ratio of OEP/WCL served as a benchmark. Most importantly, the analysis revealed significantly higher enrichment factors (up to 118-fold) for the majority of lysosomal proteins for the SPIONs and TMEM-IP **(Figure 5A-C)**. However, for the latter a higher variability and for some lysosomal proteins, even lower enrichment values compared to the OEP samples could be observed. To further investigate the finding, the protein abundance of 137 lysosomal proteins identified in both methods were compared, showing an underrepresentation of lysosome-associated complexes in the TMEM-IP samples **(Figure 5D)**. This might be a result of several factors such as rupture of the organelle during the enrichment process, alteration of lysosomal features due to TMEM192-3xHA expression, or the applied buffers leading to the dissociation of associated proteins during the isolation procedure **(Figure S6 A-C)**. However, aside from proteins associated with lysosomal complexes, no major differences could be observed among both methodologies. As a result, it was concluded that enrichment applying SPIONs yields the maximum number of lysosomes with intact ratios of 80 % in the enriched fractions and an efficient and reproducible coverage of the lysosomal proteome.

4. Chapter 2: Antibody Generation Against Cross-Linkers

4 This chapter has been published: [**Singh, J.**, Ponnaiyan, S., Gieselmann, V., Winter, D. (2021). Generation of Antibodies Targeting Cleavable Cross-Linkers. *Analytical Chemistry*, 93(8), 3762-3769].

- For online version of the paper, visit: https://doi.org/10.1021/acs.analchem.0c04043
- The supporting information including supplementary tables is available at: https://pubs.acs.org/doi/10.1021/acs.analchem.0c04043
- **Remark:** The publication within the scope of chapter 2 has been done by me as the only first author and **SHOULD BE CONSIDERED** for this cumulative doctoral dissertation.

4.1 Introduction

Protein complexes and protein-protein interactions (PPIs) belong to one of the key functional modules of cells [74]. Numerous biological processes in cells involve tightly regulated transient, non-covalent, and highly stable covalent interactions controlling cellular events. Numerous approaches have been developed to study PPIs and structures, one of them being cross-linking mass-spectrometry (XL-MS), a method, which proved its potential by supplementing well-established techniques such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and electron microscopy (EM) [71]. Functional groups targeted by most cross-linkers like the commercially available cross-linker disuccinimidyl sulfoxide (DSSO) [79] and disuccinimidyl dibutyric urea (DSBU) [80] are primary amines, which are present in lysine residues and at the protein/peptide N-terminus [53]. Dependent on the length of the spacer between the functional groups of a cross-linker, amino acids in various distances can be linked [74, 98]. Although the XL-MS approach has been successfully applied to provide structural insights of individual protein complexes [90, 99], as well as on a large scale targeting interactomes of different organelles [76, 87, 98, 100], evaluation of cross-linking sample preparation still presents a major issue.

Especially for *in vivo* cross-linking studies, including tissue extracts, or intact organelles, the membrane permeability of a cross-linker is a key factor. Although the hydrophobic nature of the commercially available cross-linkers DSSO and DSBU should allow them to penetrate the membranes efficiently, a low penetration efficiency has been reported [81]. Additionally, the validation of whether a cross-link reaction was successful or not might be further aggravated by limited amounts of samples for testing purposes.

Compared to cross-linking studies on nuclei, where cross-linkers diffuse through the nuclear pore complex into the nucleus (=efficient cross-linking of nuclear proteins), lysosomes are membrane-enclosed organelles, which hamper cross-linking efficiency. Hence, to establish and validate the cross-linking of lysosome enriched fractions, I generated cross-linker specific antibodies presenting a novel approach for identifying, validating, and visualizing cross-linker modified proteins. The antibodies developed were further used to establish lysosomal cross-linking.

4.2 Publication: Main Article

Reproduced with permission from: Singh, J., Ponnaiyan, S., Gieselmann, V., Winter, D. (2021). Generation of Antibodies Targeting Cleavable Cross-Linkers. Analytical Chemistry, 93(8), 3762-3769.

analytical chemistry

pubs.acs.org/ac Article

Generation of Antibodies Targeting Cleavable Cross-Linkers

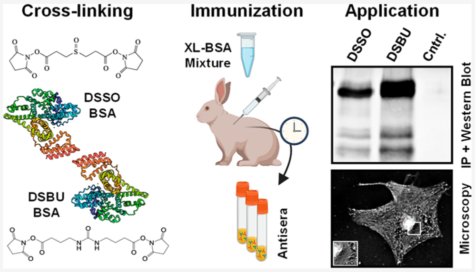
Jasjot Singh, Srigayatri Ponnaiyan, Volkmar Gieselmann, and Dominic Winter*

Cite This: *Anal. Chem.* 2021, 93, 3762−3769

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Chemical cross-linking has become a powerful tool for the analysis of protein structures and interactions by mass spectrometry. A particular strength of this approach is the ability to investigate native states in vivo, investigating intact organelles, cells, or tissues. For such applications, the cleavable cross-linkers disuccinimidyl sulfoxide (DSSO) and disuccinimidyl dibutyric urea (DSBU) are gaining increasing popularity, as they allow for the analysis of complex mixtures. It is inherently difficult to follow the reaction of cross-linkers with proteins in intact biological structures, stalling the optimization of in vivo cross-linking experiments. We generated polyclonal antibodies targeting DSSO- and DSBU-modified proteins, by injection of cross-linked bovine serum albumin (BSA) in rabbits. We show that the cross-linker-modified BSA successfully triggered an immune response, and that DSSO- and DSBU-specific antibodies were generated by the animals. Using affinity-purified antibodies specific for the individual cross-linkers, we demonstrate their application to the detection of cross-linker-modified proteins in Western blot and immunocytochemistry experiments of intact and permeabilized cells. Furthermore, we show their ability to immunoprecipitate DSSO/DSBU-modified proteins and provide evidence for their affinity toward water-quenched dead-links. These antibodies provide a valuable tool for the investigation of proteins modified with the cross-linkers DSSO and DSBU.



Protein complexes are considered as key functional modules in cells, being elementary for a plethora of biological processes.[1] Identification of such complexes, and the underlying protein−protein interactions (PPIs), is the basis for a better understanding of cellular function. The current method of choice for the investigation of PPIs on a large scale is mass spectrometry-based proteomics, as it allows for the unbiased identification and quantification of thousands of proteins from a given sample.[2] Numerous approaches have been developed to study PPIs in cells and tissues, with co-immunoprecipitation[3] and proximity labeling[4] being currently used most frequently. In recent years, cross-linking mass spectrometry (XL-MS), which was initially mainly applied to structural studies of purified proteins and protein complexes, is increasingly utilized for the identification of PPIs. Major factors for this development are novel cross-linkers[5−8] and algorithms for the analysis of XL-MS data,[2,9,10] facilitating the analysis of more complex samples. The applicability of these tools has been shown in several recent studies addressing samples of increasing complexity, ranging from isolated organelles to whole cells or tissues.[11−15] Currently, the commercially available cleavable cross-linkers disuccinimidyl sulfoxide (DSSO)[5] and disuccinimidyl dibutyric urea (BuUrBu/DSBU),[6] which are targeting primary amines present in lysine residues and at protein N-termini, are among those used most widely in the field.[16]

For the analysis of complex samples by XL-MS, several key steps are decisive for the success of the experiment and have therefore been optimized in various studies. This includes the cross-linking of proteins, their proteolytic digestion, the enrichment of cross-linked peptides, their analysis by LC-MS/MS, and the identification/assignment of cross-links.[17−19] Among these steps, the control of the cross-linking reaction is a key element, as reaction efficiencies can vary between samples, and the amount of cross-linked peptides is decisive for the success of the experiment. Currently, the efficiency of cross-linking reactions is mainly assessed by the visual interpretation of band shift patterns in coomassie/silver-stained SDS-PAGE gels. This strategy comes with certain limitations related to the abundance and complexity of samples, as the ability to visualize and identify individual protein bands is a prerequisite. Furthermore, for labeling of intact organelles, cells, or tissues, it cannot be distinguished if the cross-linker passed the membrane in an active state and reacted with intracellular

Received: September 24, 2020
Accepted: February 5, 2021
Published: February 16, 2021

proteins, as the identity of proteins which reacted with the cross-linkers is only revealed in the final MS experiment.

In comparison to other approaches, however, the unique feature of XL-MS lies in the ability to label proteins in their natural environment in vivo, facilitating the analysis of PPIs and protein structures in the native state.[20] For such in vivo XL experiments, it is elementary to be able to cross-link proteins in intact organelles, cells, or tissues, for which membrane permeability of the cross-linker is a decisive feature.[20] In principle, due to the hydrophobic nature of DSSO and DSBU, a certain membrane permeability should be given, but a low penetration efficiency has been reported in a recent study, based on the number and localization of identified cross-links.[21] Further analysis of the membrane permeability of cross-linkers, and optimization of protocols toward a better cross-linking efficiency in vivo, is inherently difficult. This is due to the fact, that the assessment of the subcellular localization of DSSO/DSBU labeling is currently limited to the final readout after mass-spectrometric analysis. As such mass-spectrometric experiments are highly laborious and usually present the final stage of the experiment, the possibilities for optimization of cross-linking parameters in vivo are currently strongly limited. To allow for a better development of in vivo cross-linking methods, additional tools allowing for direct detection/visualization of proteins which reacted with the cross-linkers like Western blotting and immunocytochemistry would be a valuable addition to the toolbox. For several protein modifications, such as methylated/acetylated/β-hydroxybutyrated lysine, phosphorylated tyrosine, or biotinylated amino acids, specific antibodies are available enabling such experiments.[22−26] For cross-linking experiments, however, only the protein interaction reporter (PIR) cross-linker, which includes a biotin molecule, and can be detected by Western blot and immunocytochemistry, presents to our knowledge such options orthogonal to mass spectrometric analyses.[27,28]

In the current study, we raised antibodies, which specifically recognize DSSO- and DSBU-modified proteins, in rabbits. We demonstrate the specificity of these antibodies toward the individual cross-linkers and their purification and characterization and apply them for the visualization of cross-linker-modified proteins in Western blotting and immunocytochemistry experiments. Finally, we show that the antibodies can be utilized for the immunoaffinity enrichment of DSSO/DSBU-modified proteins and that they show a preference toward water-quenched dead-links.

■ EXPERIMENTAL SECTION

Antigen Generation and Antibody Production. Bovine serum albumin (BSA, Sigma-Aldrich, Seelze, Germany) was cross-linked at a concentration of 2 mg/mL in two independent reactions with 9.5 mM (500× molar excess) of either DSSO or DSBU (both Thermo Fisher Scientific, Darmstadt, Germany, 50 mM stock solutions were prepared in DMSO) in 100 mM sodium phosphate (pH 7.5), 150 mM NaCl for 30 min at room temperature (RT). The reactions were quenched with 20 mM Tris-HCl (pH 8.0), and DSSO/DSBU-linked BSA was combined in a 1:1 (w/w) ratio. Two rabbits were immunized five times (on day 1, 20, 30, 40, and 61) with 500 μg of the mixture (application by subcutaneous injection, performed by Pineda Antibody Service, Berlin, Germany). Antisera of the immunized animals were collected prior to immunization (pre-immune serum) and 60, 90, 120, 150, and 180 days (final bleeding) after immunization. Antisera were supplemented with 0.05% NaN_3 (final concentration) and stored at 4 °C.

Affinity Column Generation and Antibody Purification. Porcine pancreas lipase (PPL, Sigma-Aldrich), was cross-linked in two independent reactions at a concentration of 1 mg/mL with 1 mM (50× molar excess) of either DSSO or DSBU (50 mM stock solutions were prepared in DMSO) in 100 mM sodium phosphate (pH 7.5), 150 mM NaCl for 30 min at room temperature (RT). The reactions were quenched with 20 mM Tris-HCl (pH 8.0), concentrated with 30 kDa molecular weight cutoff spin filters (Amicon-YM30, Merck Millipore, Darmstadt, Germany), and buffered to 0.1 M MES [2-(*N*-morpholino)ethansulfonic acid], 150 mM NaCl, pH 4.7 (coupling buffer). Cross-linked PPL was coupled to CarboxyLink resin (Thermo Fisher Scientific, Darmstadt, Germany) using 1-ethyl-3-(3-(dimethylamino)propyl)carbodiimide hydrochloride according to the manufacturer's protocol (EDC, BioRad, Germany). Briefly, 20 mg of DSSO or DSBU cross-linked PPL in 4 mL of coupling buffer containing 480 mg of EDC was incubated with 8 mL of CarboxyLink resin (16 mL of slurry) by end-over-end rotation at 25 rpm for 3 h, at RT. Subsequently, the columns were washed with 50 mL 1 M NaCl, and stored in 1× phosphate buffer saline (PBS), 0.05% NaN_3 at 4 °C until further use.

Antibody Purification, Storage, and Biotinylation. Affinity columns were equilibrated with 50 mL of 100 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES) pH 7.5 and 20 mL of 10 mM glycine/HCl pH 2.5, and the pH value of the column was adjusted by 10 mM HEPES/OH pH 8.8. The diluted antiserum (50 mL in total, diluted 1:5 with 10 mM HEPES/OH pH 7.5) was applied to the DSSO column and the flow-through was passed over the DSBU column. Individual columns were washed with 50 mL of 10 mM HEPES/OH pH 7.5, 20 mL of 500 mM NaCl, and 50 mL of 10 mM HEPES/OH pH 7.5. Antibodies were eluted with 10 mL of 100 mM glycine/HCl pH 2.5 and collected in 15 mL tubes containing 1 mL of 500 mM NaCl, 1 M HEPES/OH pH 8.0. Columns were re-equilibrated by 10 mM HEPES/OH pH 7.5. Antibody-containing eluate fractions were concentrated using 30 kDa molecular weight cutoff spin filters (Amicon-YM30, Merck Millipore), rebuffered to 100 mM sodium phosphate pH 7.5, 150 mM NaCl, 30% glycerol, and stored at −80 °C until further use. Affinity-purified antibodies were modified using EZ-Link Sulfo-NHS-LC-biotin (Thermo Fisher Scientific) at a 100-fold molar excess, or with bissulfosuccinimidyl suberate (BS3, Thermo Fisher Scientific) at a 100-/500-fold molar excess, at 500 rpm in a thermomixer for 45 min at RT and the reaction quenched with 20 mM Tris-HCl pH 8.0.

Cell Culture, Digitonin Treatment, and Cross-Linking. Human embryonic kidney (HEK293) and NIH/3T3 cells were cultivated in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal calf serum (FCS), 100 IU/mL penicillin, 100 μg/mL streptomycin, and 2 mM L-glutamine at 37 °C and 5% CO_2. For in vivo cross-linking experiments, 5000 NIH/3T3 cells/well were seeded in 24-well plates containing 12 mm coverslips, cultivated for 24 h, and washed twice with ice-cold PBS. Digitonin treatment was performed with 10 μg/mL digitonin (Sigma-Aldrich) in PBS for 2 min on ice. Subsequently, cells were washed with PBS at RT and incubated with 5 mM DSSO or DSBU in PBS for 10 min at RT. The cross-link reaction was quenched with 20 mM Tris-

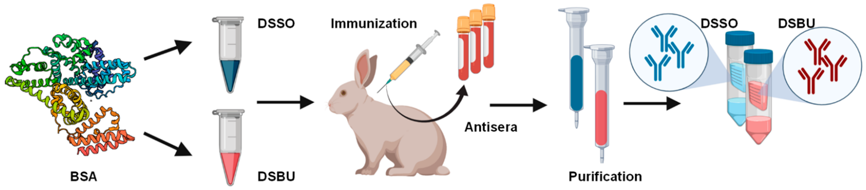


Figure 1. Generation of cross-linker-specific antibodies in rabbits. Animals were immunized with an equimolar mixture of disuccinimidyl sulfoxide (DSSO) and disuccinimidyl dibutyric urea (DSBU)-cross-linked BSA and antisera were collected at different time points. Individual antibody populations were affinity purified and concentrated.

HCl pH 8.0, and the cells were washed with PBS and fixed with 4% paraformaldehyde for 20 min at RT. For in vitro cross-linking experiments, 6 × 10^6 HEK293 cells were seeded per 10 cm plate, cultivated for 48 h, washed twice with ice-cold PBS, and harvested by scraping in 1 mL of ice-cold PBS. The cell suspension was transferred to a microtube and centrifuged at 600*g* for 10 min at 4 °C, the supernatant was discarded, and the cell pellet was resuspended in 2 mL of cross-linking buffer [10 mM HEPES/OH pH 7.5, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 1.5 mM $Mg(CH_3COO)_2$, containing complete EDTA-free protease-inhibitor cocktail (Roche, Mannheim, Germany)]. Cells were mechanically disrupted using a sonicator (Bioruptor Plus, Seraing, Belgium) with three cycles of 30 s each (amplitude: 40) at 4 °C. The cell lysate, BSA, PPL, and chicken ovalbumin (COA, Sigma-Aldrich) were cross-linked at a protein concentration of 1 mg/mL with 1 mM (cell lysate) or 10 mM (BSA, PPL, COA) DSSO or DSBU, respectively, for 30 min at RT. The reactions were quenched with 20 mM Tris-HCl pH 8.0 and the samples concentrated with 10 kDa molecular weight cutoff spin filters (Amicon-YM10, Merck Millipore), rebuffered to 100 mM sodium phosphate (pH 7.5), 150 mM NaCl, and stored at 4 °C until further use.

Co-immunoprecipitation of Cross-Linked Proteins via Streptavidin. Biotinylated antibodies were incubated with the cross-linked samples at a ratio of 1:2.5 (w/w) overnight with end-over-end rotation at 25 rpm, 4 °C. The next day, streptavidin sepharose high performance beads (GE Healthcare) were added to the samples at a ratio of 1:5 (slurry to sample, v/v) and incubated for 3 h with end-over-end rotation at 25 rpm, 4 °C. Samples were centrifuged at 1000*g*, 4 °C, and supernatants transferred to a new tube. Subsequently, the beads were washed three times with 200 μL of ice-cold PBS each, centrifuged at 1000*g*, 4 °C, and transferred to new tubes. Antibody–protein complexes were eluted from the beads by incubation for 10 min at 95 °C in 20 μL of 2% β-mercaptoethanol, 4% SDS, 20% glycerol, 2% bromophenol blue, and 120 mM Tris-HCl (pH 7.4).

Western Blotting. Proteins were separated by SDS-PAGE on 10% polyacrylamide gels and transferred to a nitrocellulose membrane (0.45 μm pore size) using a semidry electroblotter (Bio-Rad, Hercules, CA). The membrane was blocked in 5% nonfat dry milk in 1× Tris-buffered saline pH 7.4 containing 0.05% Tween-20 (TBS-T) for 1 h at RT. The membrane was incubated overnight at 4 °C with the anti-DSSO or anti-DSBU primary antibody diluted 1:2000 in 5% nonfat dry milk in 1×TBS-T, washed three times for 10 min at RT with TBS-T, and incubated with the secondary HRP-coupled goat anti-rabbit antibody (1:5,000, #111-035-003, Dianova, Hamburg, Germany) diluted in 5% nonfat dry milk in 1×TBS-T for 60 min at RT. The membrane was washed again three times for 10 min at RT with TBS-T, and protein signals were detected using the enhanced chemiluminescence (ECL) kit (Bio-Rad) and visualized with a FUSION SOLO 4 M system followed by analysis by FusionCapt advance software (both).

For the investigation of purified DSSO and DSBU antibodies, membranes were incubated 1:5000 with the goat anti-rabbit antibody (IgG heavy chain, #111-035-046, Dianova) and 1:5,000 with the mouse anti-rabbit HRP conjugate (IgG light chain, #211-032-171, Dianova). Biotinylation efficiency of antibody fractions was assessed using streptavidin HRP, 1:10 000 (#21126, Thermo Fisher Scientific).

Immunocytochemistry. For immunocytochemistry, all steps were carried out at RT. Coverslips containing fixed cells were washed three times with PBS, permeabilized with 0.1% Triton X-100 in PBS for 10 min, and blocked using 10% normal goat serum (NGS) in PBS for 1 h, followed by incubation with anti-DSSO or anti-DSBU antibodies (1:50 in 2% NGS) overnight at 4 °C. Subsequently, cells were further washed three times with PBS, incubated with Cy3-conjugated goat-anti rabbit secondary antibody, (1:400, #111-165-003, Dianova) diluted in 2% NGS for 1 h in the dark at RT, and coverslips were washed twice with PBS and once with water. Subsequently, coverslips were mounted on glass slides using DAPI Fluoromount G (Thermo Fischer Scientific) and kept for drying at RT. Images were taken using an Axiovert 200 M microscope equipped with an AxioCamMR3 (both Zeiss, Oberkochen, Germany).

■ RESULTS AND DISCUSSION

Cross-Linking and Antibody Generation. For a variety of protein modifications, antibodies serve as valuable tools for detection and enrichment.[25,29,30] For the generation of such antibodies, animals typically are injected with a peptide or protein containing the respective modification, which then triggers an immune response.[31] Subsequently, these (polyclonal) antibodies are purified from the animal's serum (Figure 1). We argued that proteins modified with cross-linkers could exhibit a sufficient immunogenic potential for the generation of specific antibodies. Along this line, we attempted the generation of antibodies against the commercially available cross-linkers DSSO and DSBU, which are frequently used in the community for complex mixtures, as they provide a simple approach for such samples. We modified bovine serum albumin (BSA) with either DSSO or DSBU, analyzed the cross-linking efficiency by SDS-PAGE (Figure S1A), and characterized the modified proteins by in-solution digestion and LC-MS/MS (Supporting Information). For DSSO/DSBU-modified BSA, we identified 46/32 cross-links, 113/86 water-quenched dead-links, and 12/34 Tris-quenched dead-

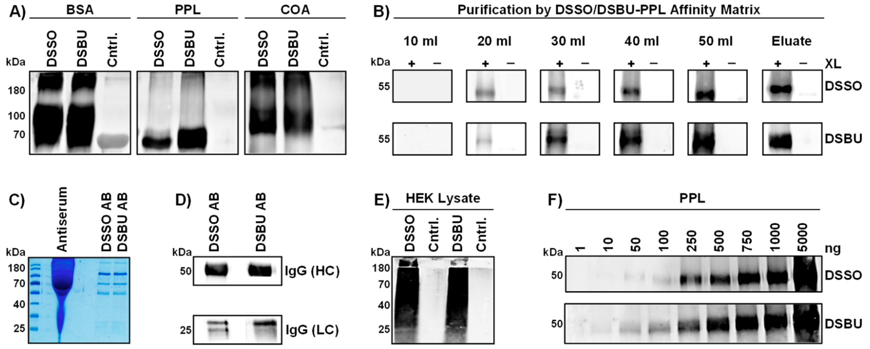


Figure 2. Performance evaluation and affinity purification of cross-linker-specific antibodies. (A) Western blot analysis of 10 μg of bovine serum albumin (BSA), porcine pancreas lipase (PPL), and chicken ovalbumin (COA), modified with DSSO or DSBU, respectively, utilizing rabbit antisera. Non-cross-linker-modified proteins served as control sample (Cntrl.). For full blots, see Figure S1C. (B) Affinity purification of DSSO/DSBU-specific antibodies from rabbit antisera. DSSO-/DSBU-modified PPL was immobilized via its carboxyl groups to the stationary material, followed by purification of antisera. Aliquots were withdrawn from the flow-through of the affinity column after every 10 mL (at 10, 20, 30, 40, and 50 mL) and applied in Western blot experiments for the detection of 10 μg of DSSO-/DSBU-modified PPL. Bound antibodies were eluted by acidic pH. For full blots, see Figure S2. (C) Analysis of 10 μg of purified antibody fractions (anti-DSSO/DSBU) by SDS-PAGE. (D) Confirmation of antibody class by Western blotting with an IgG-specific antibody (heavy chain: HC, light chain: LC). For full blots, see Figure S3A. (E) Detection of cross-links in 10 μg of DSSO-/DSBU-modified HEK293 whole cell lysates using purified antibodies. Non-cross-linker-modified cell lysates served as control sample (Cntrl.). (F) Investigation of antibody sensitivity using cross-linker-modified PPL. Numbers indicate the total protein amount utilized for the Western blots. For full blots, see Figure S3B.

links, respectively, at 1% FDR indicating a successful cross-link reaction (Figure S1B, Table S1). After collection of preimmune sera, we immunized three rabbits with an equimolar mixture of DSSO- and DSBU-modified BSA. We collected antisera from these animals 60, 90, 120, 150, and 180 day postimmunization and tested their reactivity against DSSO/DSBU-modified BSA, chicken ovalbumin (COA), and porcine pancreas lipase (PPL) in Western blot experiments. For all combinations of proteins and cross-linkers, the antisera showed a strong reactivity (Figure 2 A, Figure S1C), while the preimmune serum did not result in any detectable signals (Figure S1D), confirming the successful immunization of the animals against both DSSO and DSBU.

While we observed a signal in the non-cross-linker-modified control samples of BSA and COA, the antiserum did not recognize unmodified PPL. This was probably due to the use of cross-linker-modified BSA as carrier protein for immunization, as, in addition to the cross-linker-specific modifications, the unmodified BSA protein itself acted immunogenic, resulting in antibodies which also recognize the unmodified protein. Due to the sequence homology of BSA and COA (42%), BSA-specific antibodies are likely to recognize both proteins. For all three proteins, we detected signals at and above the molecular weight of the unmodified protein, which is in line with the assumption that the incubation with the respective cross-linker should result either in modified monomers (intra-links and dead-links) or multimers (additional interlinks) of the proteins.

Antibody Purification. Because we immunized the animals with a mixture of DSSO/DSBU-modified BSA, and therefore both antibodies were present in their serum, we subsequently purified the respective antibody populations by affinity chromatography. As we determined no cross-reactivity of our antisera with unmodified PPL, we generated two individual affinity columns utilizing PPL modified with either DSSO or DSBU, respectively. Initially, we determined the column binding capacity with 50 mL of 1:5 diluted antiserum. We collected the flow-through fractions (every 10 mL) as well as the eluate and tested their reactivity in Western blot experiments against PPL which was modified with either DSSO or DSBU (Figure 2 B, Figure S2A/B). For both cross-linkers, we started to observe antibodies in the flow-through fraction after loading of 20 mL with intensities reaching their maximum after 40 mL. Also, for the eluate fraction, we observed a strong signal, indicating that the respective antibodies were successfully affinity-purified from the antiserum, and that they retained their activity. Utilizing 40 mL of input for each purification step, we generated enriched fractions of the individual antibodies. From each individual animal, we obtained 11−12 mg for each the anti-DSSO and -DSBU antibody, respectively, indicating a comparable immune reaction toward both cross-linkers. We further investigated the possibility of cross-reactivity between both antibody species. For this purpose, we performed Western blot experiments detecting PPL modified with DSBU with anti-DSSO antibodies and vice versa (Figure S2C/D). For both populations of purified antibodies, we did not observe any cross-reactivity, indicating that the respective cross-linker-specific groups (sulfoxide group of DSSO and urea group of DSBU) are sufficiently different to be recognized exclusively by the respective antibody.

Characterization of Purified Antibodies. Subsequently, we determined the purity of the enriched antibodies by SDS-PAGE, confirming the depletion of the majority of serum proteins (Figure 2C). We identified proteins at ∼160, 110, 55, and 25 kDa for both the anti-DSSO and -DSBU eluate fractions, corresponding to the molecular weights of the intact antibody as well as dimers and monomers of the heavy and light chains, respectively, confirming the purity of our antibody-containing fractions. Subsequently, we confirmed

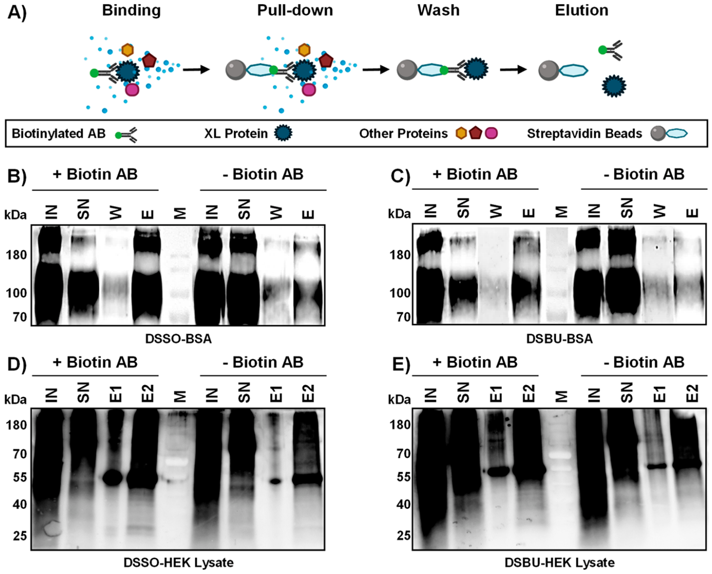


Figure 3. Immunoprecipitation of cross-linker-modified proteins. (A) Workflow for the immunoprecipitation of cross-linker-modified proteins utilizing biotinylated antibodies. (B/C/D/E) Western blot analysis for the immunoprecipitation of cross-linker-modified BSA and HEK293 whole cell lysates. Biotin ± indicates biotinylation of antibodies by a monovalent biotin-NHS ester. The left panels show the results using biotinylated antibodies, and the right panels the results from a control experiment using nonbiotinylated antibodies. (B) BSA modified with DSSO; (C) BSA modified with DSBU; (D) HEK293 whole cell lysate modified with DSSO; (E) HEK293 whole cell lysate modified with DSBU. Samples shown are input (IN), supernatant after immunoprecipitation (SN), wash fraction (W), eluate (E1), streptavidin beads (E2), and protein molecular weight marker (M). For each of the input fractions, 10 μg of protein was loaded on the gel. All other fractions were normalized to the input fractions to yield the same percentage of sample loading. Wash fractions were excluded due to their high salt content. Full blots and images recorded with different exposure times are included in Figures S5 and S6.

both antibodies to be of immunoglobulin G (IgG) type (Figure 2D, Figure S3A), which are the result of somatic hypermutation and affinity maturation due to repeated antigen exposure and usually exhibit an affinity and maturation higher than that of IgMs.[32] Finally, we tested the ability of the antibodies to react with different types of cross-linker protein conjugates and their limit of detection. Using DSSO- or DSBU-modified HEK293 whole cell lysates, respectively, we observed a strong signal over the whole molecular weight range for both antibodies, indicating that antibody binding is irrespective of protein sequence and solely dependent on the cross-linker (Figure 2E). Using a dilution series of DSSO- or DSBU-modified PPL, we determined a limit of detection of ~10 ng/mL for the anti-DSBU antibody while it was in the range of ~50 ng/mL for the anti-DSSO antibody (Figure 2F, Figure S3B).

Application to Immunoprecipitation. Cross-linking reactions result typically in low yields, and the enrichment of cross-linked proteins or peptides presents a promising avenue to increase the number of identified cross-links in mass spectrometry experiments.[7] We therefore tested the applicability of our cross-linker-specific antibodies in immunoprecipitation experiments. Initially, we tried several commonly used strategies for the covalent immobilization of proteins to different types of beads based on the targeting of primary amines (CnBr- and NHS-activated agarose beads) as well as protein A/G beads in combination with bissulfosuccinimidyl suberate (BS3) and performed immunoprecipitation experiments using either cross-linker-modified protein- or peptide-samples. All protocols, however, resulted in the inactivation of our antibodies, as we were not able to detect cross-linker-modified proteins by Western blot with these antibodies anymore (Figure S4A/B). Furthermore, we were unable to immunoprecipitate cross-linker-modified proteins (data not shown) or peptides (Supporting Information, Figure S4C, Table S2) with antibodies immobilized on these beads. We currently hypothesize that this could be due to a certain degree of affinity of our antibodies' antigen binding sites to the reagents/reactive groups used for the coupling of proteins to the beads, as most cross-linkers exhibit a certain resemblance to DSSO/DSBU in their structure. This would result in preferential cross-linking of the antigen binding site (either within the domain or to the beads), therefore quenching the antibodies' affinity. The only strategy with which we were able to retain a sufficient amount of antibody activity, as determined by Western blot experiments, was their biotinylation with a monovalent biotin-NHS ester (Figure S4A/B), facilitating their enrichment with streptavidin beads. After confirmation of antibody biotinylation by Western blot (Figure S4D), we performed immunoprecipitation experiments (Figure 3A) of cross-linker-modified proteins (Figure 3B–E, Figures S5 and S6) and peptides (Figure S7A, Table S3) by streptavidin affinity enrichment of the biotinylated antibodies. We were able to immunoprecipitate both BSA (Figure 3B/C, Figure S5)

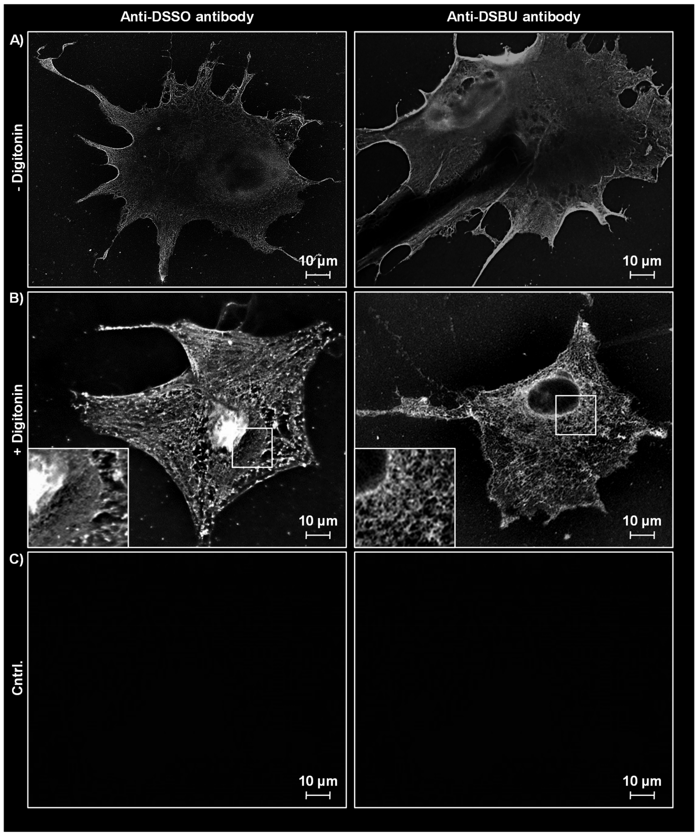


Figure 4. Immunocytochemistry analyses of cross-linker-modified cells. (A) NIH/3T3 cells modified with DSSO or DSBU and stained with the respective antibody. (B) NIH/3T3 cells treated with digitonin and modified with DSSO and DSBU followed by staining with the respective antibody. (C) NIH/3T3 cells modified with DSSO or DSBU and incubated only with the secondary antibody.

and proteins from HEK293 whole cell lysates (Figure 3C/D, Figure S6) which were modified with DSSO or DSBU. In comparison to a control sample containing nonbiotinylated antibodies, we observed both reduced abundance in the supernatant of the input fraction and increased abundance in the eluate fraction, indicating a cross-linker-specific enrichment of proteins. Surprisingly, despite its higher sensitivity in Western blot experiments (Figure 2F), the DSBU antibody performed less well in comparison to the DSSO antibody in these experiments. The enrichment of these proteins can either be related to residue to residue cross-links (inter- or intra-links) or dead-links. To further discriminate between these possibilities, we performed immunoprecipitation experiments with a library of cross-linked synthetic peptides[33] and analyzed the eluate fractions by LC-MS/MS. In addition to the biotinylated antibodies bound to streptavidin beads (Figure S7A, Table S3), we furthermore utilized our antibodies in combination with molecular weight cut-off spin filters (Figure S7B, Table S4), as this allows to use the antibodies in their unmodified form. While we were not able to immunoprecipitate any cross-linked peptides in these analyses, we observed for both the biotin- and spin filter-immobilized antibodies almost exclusively water quenched dead-links (only one Tris-quenched dead-link was identified) in the eluate fraction, while no cross-linker-modified peptides were found in control samples (Figures S7A/B, Tables S3/S4). These results show that the antibodies' main affinity is directed toward water-quenched dead-links, which also presented the main fraction of cross-linker-modified peptides detected in the analysis of the BSA used for immunization of the rabbits (Figure S1B, Table S1). A likely reason for this specificity is that the antibodies generated by the rabbit's immune system toward water-quenched dead-links were most efficient in targeting the cross-linker-modified BSA while those which bound interlinks were less effective (as also less interlinks were present per BSA molecule and they are less exposed). As the main purpose of the immune system is the removal of the antigen, antibodies with a more efficient binding behavior were produced in highest quantities resulting in a natural selection for the production of antibodies against water-quenched dead-links. Therefore, to produce antibodies which are specific for interlinks, another dead-link free immunization material would have to be utilized such as, for example, exclusively interlinked synthetic peptides.

Application to Immunocytochemistry. The majority of cross-link experiments published to date were performed with disrupted cells. While this is certainly due to practical reasons (it is easier to adjust experimental parameters in cell lysates rather than intact cells) another likely factor is that intact membranes present a natural barrier for the diffusion of many reactive cross-linkers. Even if the cross-linker itself is membrane permeable, it is highly likely that its reactive groups interact with membrane proteins, lipids, and sugars. Furthermore, current protocols for the assessment of DSSO/DSBU reactions upstream of mass spectrometric analysis do not allow investigation of the cellular localization of the cross-linking reaction, as they usually solely rely on molecular weight patterns in SDS-PAGE. We therefore evaluated the use of our antibodies for the investigation of cross-linking reactions in living NIH/3T3 cells by immunocytochemistry (Figure 4). Initially, we incubated intact cells with DSSO or DSBU and processed them for immunocytochemistry with the affinity-purified antibodies at different cross-linker concentrations (Figure S8). We observed for both DSSO and DSBU predominantly labeling of proteins located at the plasma membrane, as almost no other labeled cellular structures were observable. This confirms a limited diffusion of reactive DSSO/DSBU across the plasma membrane, which is probably due to their interaction with primary amines in lipids or proteins which are localized at the outside/inside of the plasma membrane (Figure 4A).[21] To increase membrane permeability, we treated cells with digitonin, a steroidal saponin which results in reversible permeabilization of the plasma membrane. Cells which were incubated with DSSO or DSBU after digitonin treatment showed a striking difference in subcellular structures which could be visualized by our DSSO/DSBU-specific antibodies (Figure 4B). Incubation of unmodified cells with the anti-DSSO/DSBU antibodies did not result in any signal at the same exposure time, confirming that they are specific for cross-linker-modified proteins (Figure 4C). These experiments show that permeabilization of the plasma membrane allows for diffusion of reactive cross-linkers into the cell, as only cross-linkers which are covalently bound to cellular structures are retained during the sample preparation steps for microscopy. Therefore, permeabilization reagents present an attractive way to facilitate efficient cross-linking of cellular proteins in vivo. Furthermore, these data demonstrate that our antibodies can be utilized for the monitoring of the reaction of DSSO and DSBU with living cells, providing a tool for the optimization of conditions for in vivo experiments with these cross-linkers.

CONCLUSION

Our results demonstrate the feasibility of DSSO/DSBU-specific antibody generation in rabbits. The obtained antibodies show an excellent performance in Western blotting and immunocytochemistry experiments. For immunoprecipitation experiments, the antibodies are able to retain peptides which are modified with water-quenched dead-links, as well as cross-linker-modified proteins (most likely also through water-quenched dead-links). These results imply that the antibodies generated by cross-linked BSA are specific for water-quenched dead-links. To obtain interlink specific antibodies, a different type of material would probably have to be utilized for immunization. However, especially for the cross-linking of intact cells or tissues, these antibodies provide a valuable tool, as they allow for the investigation of reactivity and (sub-) cellular distribution of DSSO and DSBU. As cross-linker-accessible proteins will always receive a mixture of cross-linked reaction products (intra-, inter-, and dead-links), the information for which (sub)cellular locations receive water-quenched dead-links is of high interest, as it allows for the characterization of reaction conditions. Therefore, these antibodies present a valuable tool for the optimization of in vivo cross-linking experiments.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.0c04043.

Supporting experimental procedures and figures (PDF)

LC-MS/MS results from tryptic digests of DSSO/DSBU-modified BSA (XLSX)

LC-MS/MS results for the immunoprecipitation of a DSSO-modified synthetic peptide library with antibodies immobilized on CnBr and ProtA beads (XLSX)

LC-MS/MS results for the immunoprecipitation of a DSSO-modified synthetic peptide library with antibodies immobilized on streptavidin beads (XLSX)

LC-MS/MS results for the immunoprecipitation of a DSSO-modified synthetic peptide library with antibodies in combination with molecular weight cut off spin filters (XLSX)

AUTHOR INFORMATION

Corresponding Author

Dominic Winter – *Institute for Biochemistry and Molecular Biology, Medical Faculty, University of Bonn, 53115 Bonn, Germany*; orcid.org/0000-0001-6788-6641; Phone: +49 228 73 7081; Email: dominic.winter@uni-bonn.de

Authors

Jasjot Singh – *Institute for Biochemistry and Molecular Biology, Medical Faculty, University of Bonn, 53115 Bonn, Germany*

Srigayatri Ponnaiyan – *Institute for Biochemistry and Molecular Biology, Medical Faculty, University of Bonn, 53115 Bonn, Germany*

Volkmar Gieselmann – *Institute for Biochemistry and Molecular Biology, Medical Faculty, University of Bonn, 53115 Bonn, Germany*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.analchem.0c04043

Author Contributions

J.S. performed experiments, analyzed the data, and wrote the manuscript. S.G.P. performed microscopy experiments. V.G. provided assistance for the experimental setup and edited the manuscript. D.W. designed the study and wrote the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors appreciate financial support from the Studienstiftung des Deutschen Volkes for J.S. and the National Overseas Scholarship by the Indian Ministry of Social Justice and Empowerment for S.G.P. We further thank Karl Mechtler for providing us with the cross-link peptide library. Workflows presented in Figure 1 and the Table of Contents graphic were created with BioRender.com.

■ REFERENCES

(1) Chavez, J. D.; Bruce, J. E. *Curr. Opin. Chem. Biol.* **2019**, *48*, 8−18.
(2) Liu, F.; Rijkers, D. T. S.; Post, H.; Heck, A. J. R. *Nat. Methods* **2015**, *12*, 1179−1184.
(3) Markham, K.; Bai, Y.; Schmitt-Ulms, G. *Anal. Bioanal. Chem.* **2007**, *389*, 461−473.
(4) Kim, D. I.; Roux, K. J. *Trends Cell Biol.* **2016**, *26*, 804−817.
(5) Kao, A.; Chiu, C.-L.; Vellucci, D.; Yang, Y.; Patel, V. R.; Guan, S.; Randall, A.; Baldi, P.; Rychnovsky, S. D.; Huang, L. *Mol. Cell. Proteomics* **2011**, *10*, 1.
(6) Müller, M. Q.; Dreiocker, F.; Ihling, C. H.; Schäfer, M.; Sinz, A. *Anal. Chem.* **2010**, *82*, 6958−6968.
(7) Steigenberger, B.; Pieters, R. J.; Heck, A. J. R.; Scheltema, R. A. *ACS Cent. Sci.* **2019**, *5*, 1514−1522.
(8) Tang, X.; Munske, G. R.; Siems, W. F.; Bruce, J. E. *Anal. Chem.* **2005**, *77*, 311−318.
(9) Götze, M.; Pettelkau, J.; Fritzsche, R.; Ihling, C. H.; Schäfer, M.; Sinz, A. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 83−97.
(10) Mendes, M. L.; Fischer, L.; Chen, Z. A.; Barbon, M.; O'Reilly, F. J.; Giese, S. H.; Bohlke-Schneider, M.; Belsom, A.; Dau, T.; Combe, C. W.; et al. *Mol. Syst. Biol.* **2019**, *15*, 8994−9107.
(11) Chavez, J. D.; Keller, A.; Zhou, B.; Tian, R.; Bruce, J. E. *Cell Rep.* **2019**, *29*, 2371−2383.
(12) Fasci, D.; van Ingen, H.; Scheltema, R. A.; Heck, A. J. R. *Mol. Cell. Proteomics* **2018**, *17*, 2018−2033.
(13) Götze, M.; Iacobucci, C.; Ihling, C. H.; Sinz, A. *Anal. Chem.* **2019**, *91*, 10236−10244.
(14) Liu, F.; Lössl, P.; Rabbitts, B. M.; Balaban, R. S.; Heck, A. J. R. *Mol. Cell. Proteomics* **2018**, *17*, 216−232.
(15) Schweppe, D. K.; Chavez, J. D.; Lee, C. F.; Caudal, A.; Kruse, S. E.; Stuppard, R.; Marcinek, D. J.; Shadel, G. S.; Tian, R.; Bruce, J. E. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 1732−1737.
(16) Iacobucci, C.; Piotrowski, C.; Aebersold, R.; Amaral, B. C.; Andrews, P.; Bernfur, K.; Borchers, C.; Brodie, N. I.; Bruce, J. E.; Cao, Y.; et al. *Anal. Chem.* **2019**, *91*, 6953−6961.
(17) Iacobucci, C.; Götze, M.; Ihling, C. H.; Piotrowski, C.; Arlt, C.; Schäfer, M.; Hage, C.; Schmidt, R.; Sinz, A. *Nat. Protoc.* **2018**, *13*, 2864−2889.
(18) Klykov, O.; Steigenberger, B.; Pektaş, S.; Fasci, D.; Heck, A. J. R.; Scheltema, R. A. *Nat. Protoc.* **2018**, *13*, 2964−2990.
(19) Chen, Z. A.; Rappsilber, J. *Nat. Protoc.* **2019**, *14*, 171−201.
(20) Chavez, J. D.; Mohr, J. P.; Mathay, M.; Zhong, X.; Keller, A.; Bruce, J. E. *Nat. Protoc.* **2019**, *14*, 2318−2343.
(21) Liu, F.; Lössl, P.; Scheltema, R.; Viner, R.; Heck, A. J. R. *Nat. Commun.* **2017**, *8*, 15473.
(22) Gratzner, H. G. *Science* **1982**, *218*, 474−475.
(23) Kim, S.-Y.; Sim, C. K.; Zhang, Q.; Tang, H.; Brunmeir, R.; Pan, H.; Karnani, N.; Han, W.; Zhang, K.; Xu, F. *PLoS One* **2016**, *11*, No. e0162528.
(24) Perez-Burgos, L.; Peters, A. H.F.M.; Opravil, S.; Kauer, M.; Mechtler, K.; Jenuwein, T. *Methods Enzymol.* **2003**, *376*, 234−254.
(25) Udeshi, N. D.; Pedram, K.; Svinkina, T.; Fereshetian, S.; Myers, S. A.; Aygun, O.; Krug, K.; Clauser, K.; Ryan, D.; Ast, T.; et al. *Nat. Methods* **2017**, *14*, 1167−1170.
(26) Xie, Z.; Zhang, D.; Chung, D.; Tang, Z.; Huang, H.; Dai, L.; Qi, S.; Li, J.; Colak, G.; Chen, Y.; et al. *Mol. Cell* **2016**, *62*, 194−206.
(27) Chavez, J. D.; Weisbrod, C. R.; Zheng, C.; Eng, J. K.; Bruce, J. E. *Mol. Cell. Proteomics* **2013**, *12*, 1451−1467.
(28) Tang, X.; Yi, W.; Munske, G. R.; Adhikari, D. P.; Zakharova, N. L.; Bruce, J. E. *J. Proteome Res.* **2007**, *6*, 724−734.
(29) Rush, J.; Moritz, A.; Lee, K. A.; Guo, A.; Goss, V. L.; Spek, E. J.; Zhang, H.; Zha, X.-M.; Polakiewicz, R. D.; Comb, M. J. *Nat. Biotechnol.* **2005**, *23*, 94−101.
(30) Xu, G.; Paige, J. S.; Jaffrey, S. R. *Nat. Biotechnol.* **2010**, *28*, 868−873.
(31) Clementi, M. E.; Marini, S.; Condò, S. G.; Giardina, B. *Ann. Ist. Super. Sanita* **1991**, *27*, 139−143.
(32) Sathe, A.; Cusick, J. K. Biochemistry, Immunoglobulin M (IgM). In *StatPearls*; Sathe, A., Cusick, J. K., Eds.; StatPearls Publishing, 2020.
(33) Beveridge, R.; Stadlmann, J.; Penninger, J. M.; Mechtler, K. *Nat. Commun.* **2020**, *11*, 1−9.

4.3 Publication: Supporting Information

Supporting information

Generation of Antibodies Targeting Cleavable Cross-Linkers

Jasjot Singh, Srigayatri Ponnaiyan, Volkmar Gieselmann, Dominic Winter*

Institute for Biochemistry and Molecular Biology, Medical Faculty, University of Bonn, 53115 Bonn, Germany

***Corresponding author**
Dr. Dominic Winter
Institute for Biochemistry and Molecular Biology
Medical Faculty
Rheinische Friedrich-Wilhelms-University of Bonn
Nussallee 11, 53115 Bonn, Germany
Email: dominic.winter@uni-bonn.de
Tel: +49 228 73 7081

Contents

Figure S5: Complete western-blots for immunoprecipitation of cross-linked BSA with different exposure times.
Figure S6: Complete western-blots for immunoprecipitation of cross-linked HEK-lysate with different exposure times.
Figure S7: Immunoprecipitation of DSSO and DSBU cross-linked peptide libraries with streptavidin beads and a molecular with cut off approach.
Figure S8: Immunocytochemistry of DSSO cross-linked cells at different concentrations.

References:

(1) Beveridge, R.; Stadlmann, J.; Penninger, J.M.; Mechtler, K. A synthetic peptide library for benchmarking crosslinking-mass spectrometry search engines for proteins and protein complexes. *Nat. Commun.* **2020**, *11*, 1–9.
(2) Rappsilber, J.; Ishihama, Y.; & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem*. **2003** 75(3), 663-670.

Proteolytic digestion of DSSO and DSBU cross-linked BSA

Cross-linked BSA was precipitated with acetone (1:4 v/v) overnight at -20 °C and the protein pellet denatured in freshly prepared 8 M urea/100 mM triethylammonium bicarbonate (TEAB) for 30 min at 37 °C. Proteins were reduced with 10 mM dithiothreitol (DTT, final concentration) at 56 °C for 30 min, and alkylated with 20 mM chloroacetamide (final concentration) at room temperature for 30 min, followed by quenching of the reaction with 10 mM DTT. Subsequently, samples were diluted with 100 mM TEAB to a concentration of 4 M urea and rLysC (Promega, Mannheim, Germany) was added at an enzyme to protein ratio of 1:100 (w/w). Proteins were digested overnight at 37 °C, samples further diluted with 100 mM TEAB to 1.6 M urea, and trypsin (Promega, Mannheim, Germany) was added at a ratio of 1:50 (w/w). After digestion for 10 h, peptides were desalted using Oasis HLB cartridges (Waters, Milford, MA), the eluate fraction dried in a vacuum centrifuge, and stored at -20 °C until further use.

Peptide immunoprecipitation of DSSO cross-linked library with streptavidin beads

DSSO biotinylated antibodies were incubated with the DSSO cross-linked peptide library[1] at a molar ratio of 1:2 overnight with end-over-end rotation at 25 rpm, 4 °C. The next day, streptavidin sepharose high performance beads (GE Healthcare) were added to the samples at a ratio of 1:5 (slurry to sample, v/v), and incubated for 3 h with end-over-end rotation at 25 rpm, 4 °C. Samples were centrifuged at 1,000 x g, 4 °C and supernatants transferred to a new tube. Subsequently, the beads were washed three times with 200 µl of ice-cold PBS each, centrifuged at 1,000 x g, 4 °C, and transferred to new tubes. Antibody-peptide complexes were eluted from the beads by applying 0.1 % TFA, followed by desalting of the fractions with stage-tips.[2]

Peptide immunoprecipitation of DSBU cross-linked library with a molecular weight cut off filter approach

DSBU antibodies were incubated with the DSBU cross-linked peptide library[1] at a molar ratio of 1:3 overnight with end-over-end rotation at 25 rpm, 4 °C. The next day, samples were passed through a 30 kDa molecular weight cut off filter (Merck Millipore), the membrane washed 3x with 200 µl of ice-cold PBS at 14,000 x g and the antibody-peptide complexes were eluted from the filter by applying 0.1 % TFA, followed by desalting of the fractions with stage-tips.

LC-MS/MS and data analysis

LC-MS/MS analyses were conducted with a nano-UHPLC-Dionex Ultimate 3000 system coupled to an Orbitrap Fusion Lumos mass spectrometer (both Thermo Fisher Scientific, Bremen, Germany). Analytical columns were produced in-house as follows: spray tips were generated from 360 µm outer diameter and 100 µm inner diameter fused silica capillaries using a P-2000 laser puller (Sutter Instruments, Novato, CA) and packed with 1.9 µm Reprosil AQ C_{18} particles (Dr. Maisch, Ammerbuch, Germany) to a length of 50 cm. Peptides were dissolved in 5 % ACN/5 % FA and 1 µg was loaded directly on the analytical column at a flow rate of 600 nl/min in 95 % solvent A (0.1 % FA in water). Peptides were separated with a 120 min linear gradient from 5 % - 35 % solvent B (90 % ACN/ 0.1 % FA), eluting peptides ionized in the positive ion mode at a cone voltage of 2 kV, and survey scans acquired in the Orbitrap mass analyzer at m/z 375-1600 and a resolution of 120,000. Automatic gain control (AGC) and maximum injection time settings were set to standard (system setting the recommended target in an automated fashion per scan type) and auto, respectively, dynamic exclusion was set to 45 sec. Peptides were fragmented using stepped HCD with normalized collision energies of 26 % ± 5 % with an AGC setting of standard and a maximum injection time of 70 ms with a cycle time set to 5 sec (top speed). MS2 scans were acquired in the

Orbitrap mass analyzer with a dynamic mass range set to normal and a resolution of 30,000 fragmenting peptides with charge states ≥ 3+.

Thermo *.raw files were analyzed by Proteome Discoverer 2.4 (Thermo Fisher Scientific, Bremen, Germany) utilizing the XlinkX node (Proteome Discoverer Version 2.4) with the following settings: precursor ion mass tolerance: 10 ppm; Orbitrap fragment ion mass tolerance: 20 ppm, fixed modification: carbamidomethyl (cysteine); variable modification: oxidation (methionine); number of allowed missed cleavage sites: 2; minimum peptide length: 5 amino acids; cross-linking site: lysine. Data were searched against a cRAP database containing the BSA sequence and common contaminants (262 entries, Max Planck Institute of Biochemistry, Martinsried URL: https://lotus1.gwdg.de/mpg/mmbc/maxquant_input.nsf/7994124a4298328fc125748d0048fee2/$FILE/contaminants.fasta) and the published Cas9 library of synthetic peptides.[1] Data were searched with a false discovery rate of 1 % at cross-linked peptide spectra level matching a minimum score threshold of 40. The Proteome Discoverer XlinkX workflow was divided by the “XlinkX filter” node into two branches including a standard peptide search and a cross-link search. Spectra containing indicative reporter ions were searched by the node “XlinkX search” employing a MS2 search option. Spectra which did not contain reporter ions were searched using Sequest HT for the identification of unmodified peptides, dead-end links (H_2O hydrolyzed, as well as Tris quenched) applied as dynamic modifications at lysines.

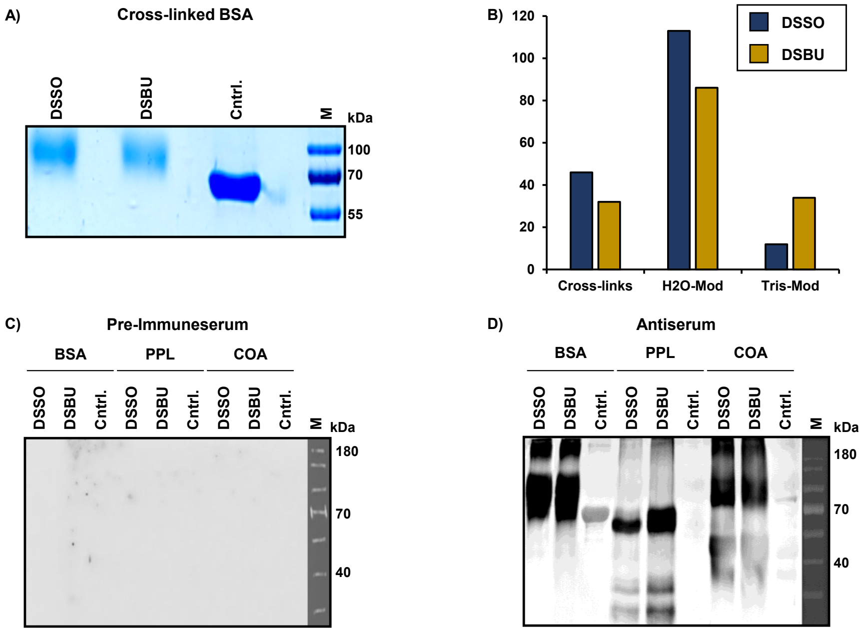


Figure S1: Verification of BSA cross-linking with DSSO/DSBU followed by LC-MS/MS analysis of antigen, and initial rabbit antisera tests. A) Verification of cross-linking by DSSO and DSBU for BSA by SDS-PAGE. BSA was cross-linked at a concentration of 2 mg/ml with 9.5 mM (500x fold molar excess) of either DSSO or DSBU for 30 min at RT, and quenched with 20 mM Tris-HCl. For each sample, 10 µg were separated by SDS-PAGE and stained by coomassie brilliant blue. The band shift of cross-linked relative to the native BSA (control sample) confirms the successful reaction. **B)** LC-MS/MS analysis of cross-linked BSA depicting cross-links, and dead-links (modified with H_2O and Tris) for DSSO and DSBU modified BSA, respectively. **C/D)** Western-blot analysis of 10 µg of DSBO/DSBU cross-linked and unmodified (Cntrl.) standard proteins (BSA, PPL, COA). Membranes were incubated overnight at 4 °C with the pre- or antiserum diluted 1:2000 in 5 % non-fat dry milk in Tris buffer saline containing 0.05 % tween. **C)** Detection of cross-linked proteins with pre-serum obtained prior to immunization of animals. **D)** Detection of cross-linked proteins with antiserum obtained after immunization of animals. DSSO: disuccinimidyl sulfoxide; DSBU: 1,3-bis-(4-oxo-butyl)-urea group; BSA: bovine serum albumin; PPL: porcine pancreas lipase; COA: chicken ovalbumin; Cntrl.: control.

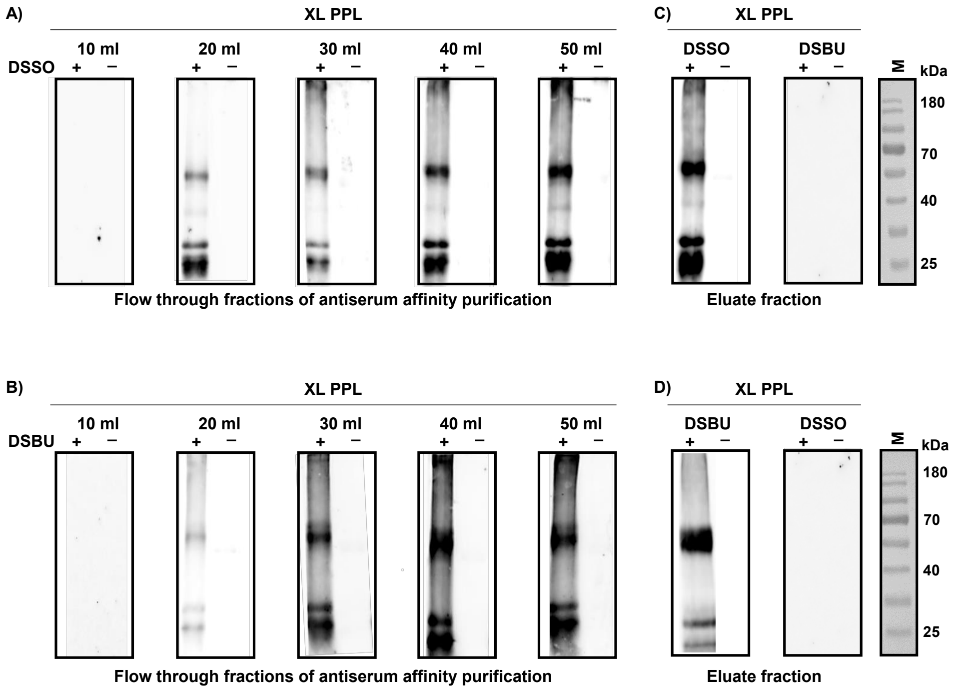


Figure S2: Evaluation of binding capacity for antibody affinity column purification and cross-reactivity of purified antibodies. A/B) Western blot analysis of flow-through and eluate fraction for the affinity purification of antisera obtained from rabbits immunized with DSSO/DSSU cross-linked BSA. Initially, antisera were passed through the DSSO affinity column. Subsequently, the flow-through fractions of this column were pooled and passed through the DSBU affinity column. **A)** Fractions obtained from the application of antisera to an affinity column containing DSSO cross-linked PPL coupled to the stationary phase through carboxyl group specific linkage. **B)** Fractions obtained from the application of the pooled flow-through fractions to the affinity column containing DSBU cross-linked PPL. **C)** Investigation of the specificity of antibodies purified by the DSSO specific column. **D)** Investigation of the specificity of antibodies purified by the DSBU specific column. DSSO: disuccinimidyl sulfoxide; DSBU: 1,3-bis-(4-oxo-butyl)-urea group; PPL: porcine pancreas lipase; FT: flow through; AB: Antibody.

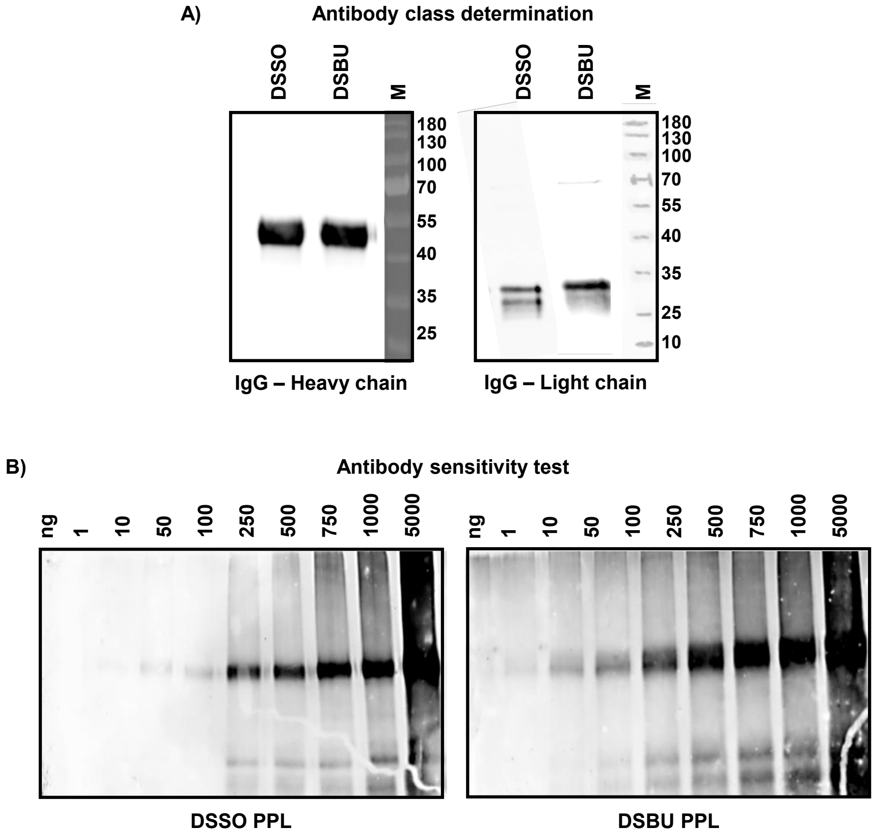


Figure S3: A) Confirmation of antibody class by western-blotting with an IgG specific antibody (left panel antibody specific for IgG heavy chains, right panel antibody detecting light chains). **B)** Investigation of antibody sensitivity using cross-linked PPL. Numbers indicate the total protein amount loaded on the respective lane of the SDS-PAGE gels used for the western-blots in ng. DSSO: disuccinimidyl sulfoxide; DSBU: 1,3-bis-(4-oxo-butyl)-urea group; PPL: porcine pancreas lipase.

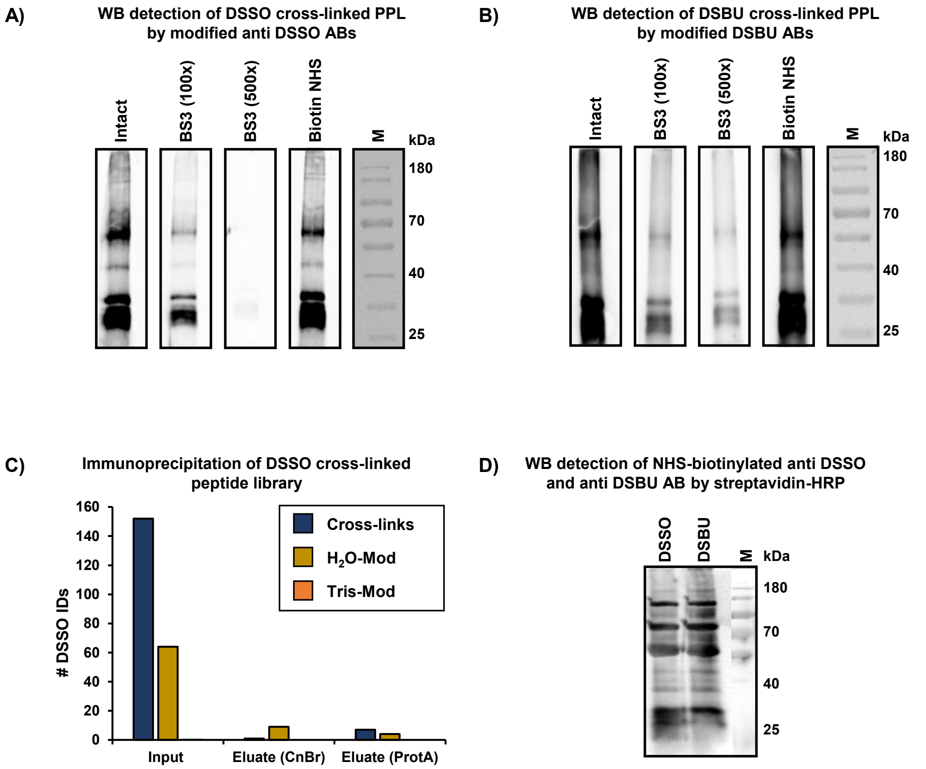


Figure S4: Determination of antibody activity following treatment by BS3 and NHS-biotin and peptide immunoprecipitation via CnBr and ProtA beads. A/B) Investigation of antibody activity following chemical modification. BS3 treatment of antibodies resembles the BS3-mediated coupling of antibodies to beads with primary amines on their surface. **A)** Individual western-blot analyses of DSSO cross-linked PPL by purified DSSO antibodies, which were unmodified, treated with either a 100x or 500x molar excess of BS3, or with a monovalent biotin NHS-ester. **B)** Individual western-blot analyses of DSBU cross-linked PPL by purified DSSO antibodies, which were unmodified, treated with either a 100x or 500x molar excess of BS3, or with a monovalent biotin NHS-ester. **C)** DSSO cross-linked peptide library was incubated with CnBr and protein A DSSO antibody coupled beads in a 2:1 molar ratio overnight. Beads were washed and eluted applying acidic conditions on the subsequent day, followed by LC-MS/MS analysis of the input and eluate fractions for each of the coupling strategies for cross-links, water and Tris quenched dead-links. **D)** Western-blot analysis for the investigation of antibody biotinylation by a monovalent biotin NHS-ester using a streptavidin-HRP conjugate. BS3: Bissulfosuccinimidyl suberate, CnBr: Cyanogen bromide, ProtA: protein A, DSSO: disuccinimidyl sulfoxide; DSBU: 1,3-bis-(4-oxo-butyl)-urea group; PPL: porcine pancreas lipase; AB: antibody; WB: western blot.

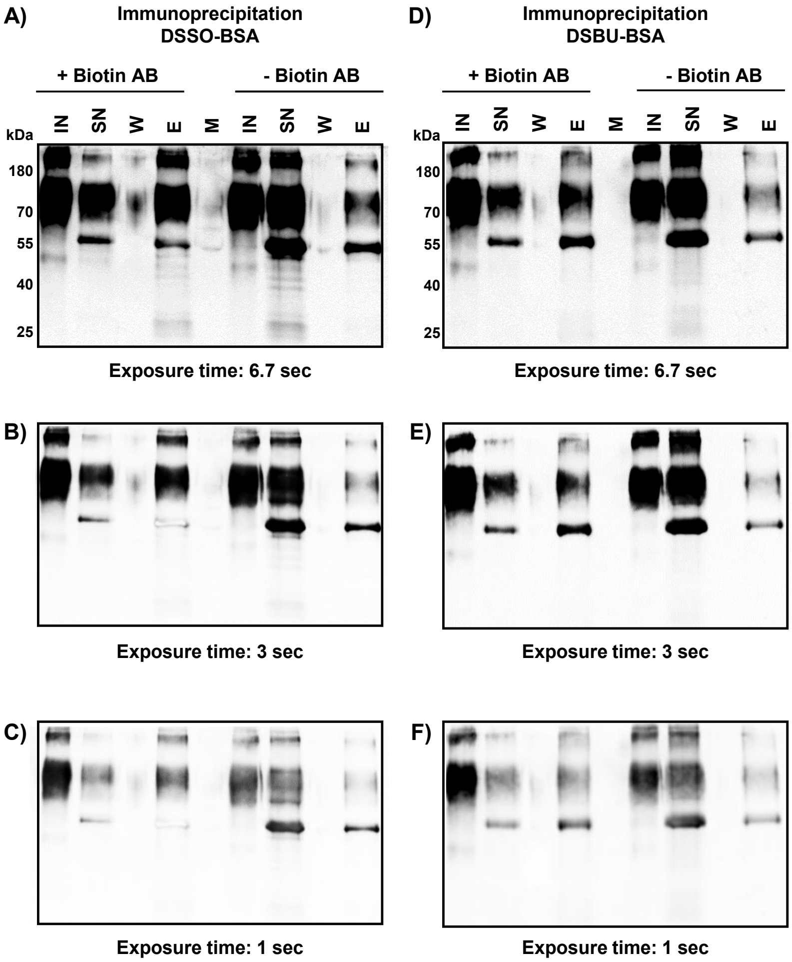
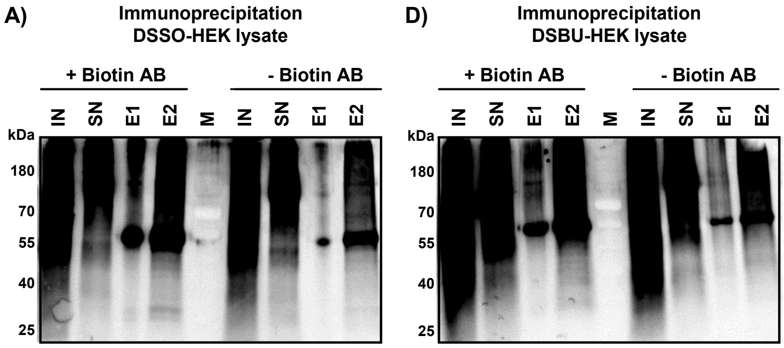
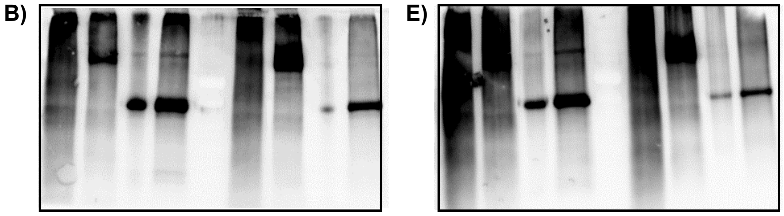


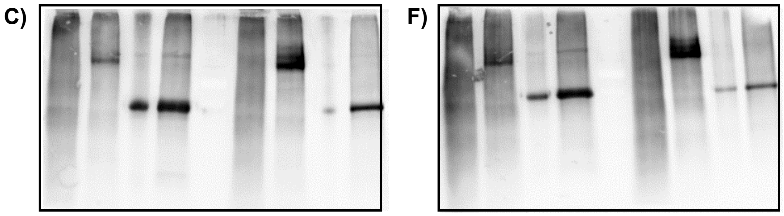
Figure S5: Complete western-blots for immunoprecipitation of cross-linked BSA with different exposure times. A-F). Biotin +/- indicates biotinylation of antibodies by a monovalent biotin-NHS ester. The left panels show the results using biotinylated antibodies, the right panels the results from a control experiment using non-biotinylated antibodies. **A-C)** BSA cross-linked with DSSO depicting different exposure times. **D-F)** BSA cross-linked with DSBU depicting different exposure times. Samples shown are: Input (IN), supernatant (SN) after immunoprecipitation, wash (W) fraction, eluate (E), and protein molecular weight marker (M).



Exposure time: 7.5 sec Exposure time: 7.5 sec



Exposure time: 3 sec Exposure time: 3 sec



Exposure time: 1 sec Exposure time: 1 sec

Figure S6: Complete western-blots for immunoprecipitation of cross-linked HEK293-lysate with different exposure times. A-F). Biotin +/- indicates biotinylation of antibodies by a monovalent biotin-NHS ester. The left panels show the results using biotinylated antibodies, the right panels the results from a control experiment using non-biotinylated antibodies. **A-C)** HEK293-lysate cross-linked with DSSO depicting different exposure times. **D-F)** HEK293-lysate cross-linked with DSBU depicting different exposure times. Samples shown are: Input (IN), supernatant (SN) after immunoprecipitation, wash (W) fraction, eluate (E1), streptavidin beads (E2) and protein molecular weight marker (M).

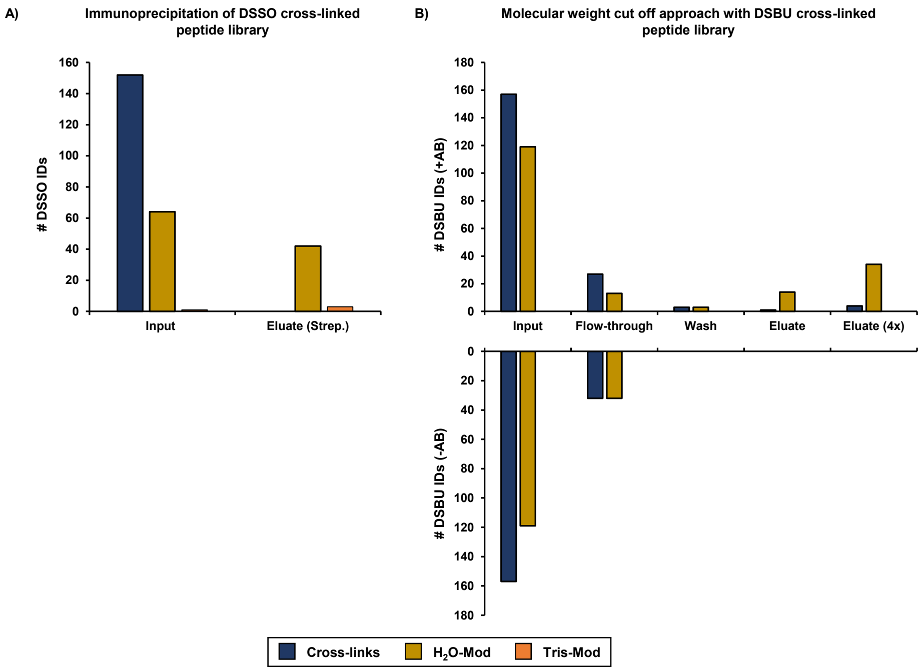


Figure S7: Immunoprecipitation of DSSO and DSBU cross-linked peptide libraries with streptavidin beads and a molecular with cut off approach. A) The DSSO cross-linked peptide library was incubated with biotinylated DSSO antibodies in a 2:1 molar ratio overnight. Streptavidin beads were added on the subsequent day, washed and eluted applying acidic conditions, followed by LC-MS/MS analysis of the input and eluate fraction. Generated data were searched for cross-links as well as water-/Tris-quenched dead-links. **B)** The DSBU cross-linked peptide library was incubated with and without DSBU antibodies (+AB/-AB) in a 3:1 molar ratio overnight. On the subsequent day, samples were passed through a 30 kDa molecular weight cut off filter, washed, and peptides eluted applying acidic conditions. LC-MS/MS analysis was carried out for equal percentages of the input, flow-through, wash, and eluate fraction as well as a sample containing 4x of the eluate fraction. Generated data were searched for cross-links as well as water-/Tris-quenched dead-links.

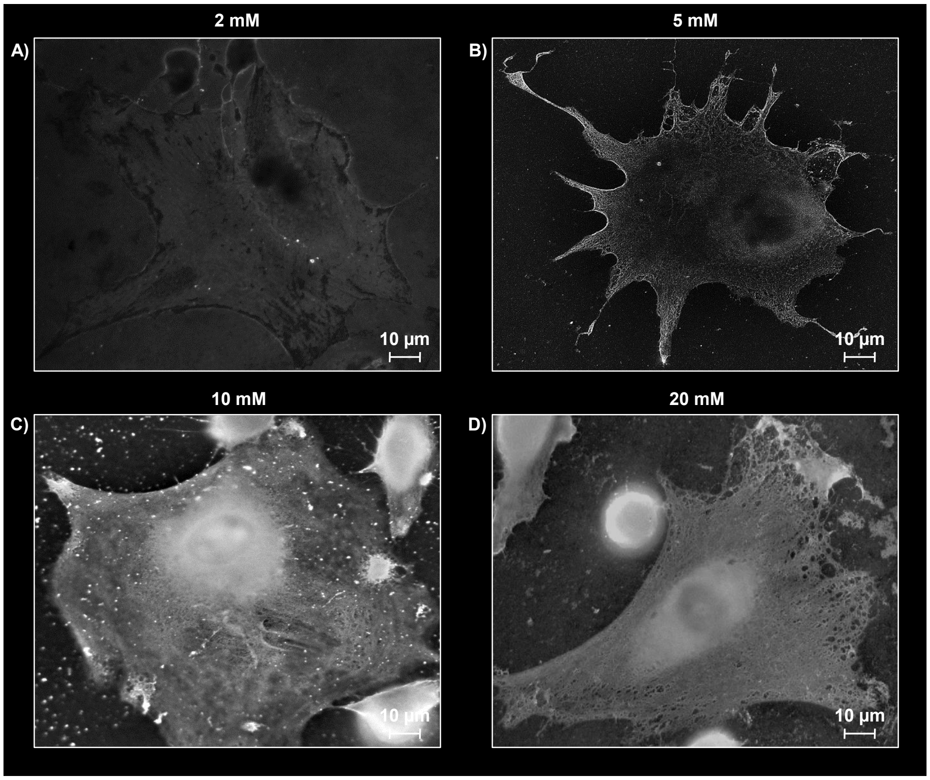


Figure S8: Immunocytochemistry of DSSO cross-linked NIH/3T3 cells at four different concentrations. **A)** 2 mM **B)** 5 mM **C)** 10 mM and **D)** 20 mM followed by staining with the purified DSSO antibody.

4.4 Conclusion

To better characterize and validate cross-link reactions, I attempted to generate antibodies against the commercially available cross-linkers DSSO and DSBU. To do so, I modified the protein bovine serum albumin (BSA) with either DSSO or DSBU and assessed the cross-linking efficiency via SDS-PAGE and LC-MS/MS **(Figure S1 A, B)**. Both methodologies showed a successful modification leading to the identification of 46/32 cross-links, 113/ 86 water-quenched dead-links, and 12/34 Tris-quenched dead-links, for DSSO and DSBU, respectively. The generated cross-link mixture was subsequently applied in an equimolar ratio to immunize rabbits, from which antisera were collected 60, 90, 120, 150, and 180-day postimmunization **(Figure 1)**.

The pre-immune- and anti-sera were further tested for their reactivity in western blot experiments against DSSO/DSBU modified BSA, chicken ovalbumin (COA), and porcine pancreas lipase (PPL) **(Figure 2A, Figure S1 C, D)**. The results showed a strong reactivity of the antisera against all three proteins, whereas the pre-immune sera resulted in no detectable signal, confirming the successful immunization. As I immunized the rabbits with a mixture of DSSO/DSBU-modified BSA, resulting in both antibodies' presence in their serum, I further developed an affinity liquid chromatography approach utilizing DSSO/DSBU modified PPL to purify the respective antibody populations. Applying 40 mL of 1:50 (v/v) diluted antisera, I successfully retrieved the respective antibodies in the eluate fractions **(Figure 2B, Figure S2 A, B)**. I obtained 11 and 12 mg of the anti-DSSO and -DSBU antibodies, respectively, indicating a comparable immune response to both cross-linkers. The possibility of cross-reactivity among both purified antibody populations was excluded by western blot experiments by detecting DSBU-modified PPL with anti-DSSO antibodies and vice versa **(Figure S2 C, D)**. The results showed that both cross-linkers were sufficiently different from being recognized exclusively by the respective antibody.

I further assessed the purity of the individual antibody populations via SDS-PAGE verifying a successful depletion of most serum proteins, as well as the presence of the characteristic bands for the heavy and light chain of the antibodies at 55 and 25 kDa **(Figure 2C)**. Additionally, I characterized the type of antibody to be immunoglobulin G (IgG), resulting from somatic hypermutation and maturation due to repeated antigen exposure **(Figure 2D)**. To further assess the capability of the antibody to react with different types of cross-linker-protein conjugates, I applied DSSO- and DSBU-modified HEK293 whole-cell lysates. The results showed a strong signal across the entire molecular weight range for both antibodies, indicating that antibody binding is independent of protein sequence and solely dependent on the cross-linker **(Figure 2E)**. Utilizing a dilution series of DSSO/DSBU modified PPL, I was further able to determine the limit of detection for the DSSO antibody in the range of 50 ng/mL and 10 ng/mL for the DSBU antibody **(Figure 2F)**.

As cross-linking reactions typically result in low yields, enrichment of cross-linked proteins or peptides presents a promising way to increase the number of cross-links identified in mass spectrometry experiments. Therefore, I tested the applicability of the antibodies for immunoprecipitation of cross-linker modified proteins applying various commonly used strategies. Those included but were not limited to strategies targeting primary amines (CnBr- and NHS-activated agarose beads) and protein A beads. However, as all conventional protocols led to the inactivation of the antibodies, I developed an alternative strategy by first biotinylating the respective antibodies, followed by their immunoprecipitation via streptavidin **(Figure 3A,**

Figure S4 A, B). Applying this approach, I was able to successfully immunoprecipitate BSA and HEK293 whole cell lysate proteins modified with DSSO and DSBU **(Figure 3B-E)**. I further used this strategy for immunoprecipitation of DSSO and DSBU modified peptides, however, the LC-MS/MS analysis results revealed that the antibody's major activity is directed toward water-quenched dead-links **(Figure S4 C, Figure S7 A, B)**. As the water-quenched dead-links made the majority of cross-linker modified peptides applied for the immunization of the rabbits **(Figure S1 B)**, a likely reason for their specificity to those is their high abundance. Hence, it was concluded that for the generation of antibodies specific for inter-links, a dead-link-free sample has to be applied for the immunization of rabbits.

Since existing protocols for assessing cross-link reactions before mass spectrometry analysis exclusively depend on changes of the molecular weight of proteins via SDS-PAGE, I utilized the antibodies to investigate the cellular localization of the cross-link reaction in live NIH/3T3 cells by immunocytochemistry experiments. For this, intact cells were modified with either DSSO or DSBU and incubated with the respective antibodies. The results showed predominant labeling of proteins localized to the plasma membrane leading to the conclusion of limited membrane permeability of both cross-linkers **(Figure 4A)**. This was since intact cellular membranes present a natural barrier for the diffusion of cross-linkers, even though the cross-linker is membrane permeable, as its reactive groups will most likely react to membrane proteins, lipids, and sugars. Therefore, in my final experiment, before cross-linking, the cells were treated with digitonin leading to reversible permeabilization of the plasma membrane. After digitonin treatment, the results revealed a striking difference showing subcellular structures that could now be visualized with the DSSO/DSBU-specific antibodies **(Figure 4B)**. Results of unmodified cells with the anti-DSSO/DSBU antibodies resulted in no signal at the same exposure time, indicating that they are specific for cross-linker-modified proteins **(Figure 4C)**.

Overall, I was able to show that rabbits immunized with cross-linked BSA (immunogenic carrier protein) can be used to generate DSSO/DSBU-specific antibodies. The affinity-purified antibodies were successfully applied in western blotting and immunocytochemistry experiments. The antibodies retained peptides modified with water-quenched dead-links and cross-linker-modified proteins in immunoprecipitation assays (most likely also through water-quenched dead-links). These findings suggest that antibodies generated against cross-linked BSA are selective for water-quenched dead-links. A new type of immunogenic carrier protein, such as one with no dead-links, would most likely have to be used for future immunization approaches to develop cross-link (intra and inter-links)-specific antibodies. Yet, the developed antibodies are beneficial for the quality control of cross-linking of intact cells or tissues, as they enable the investigation of DSSO and DSBU reactivity and (sub-)cellular distribution. Given that cross-linker-accessible proteins will always acquire a mixture of cross-linked reaction products (intra-, inter-, and dead-links), knowing which (sub)cellular sites receive water-quenched dead-links is crucial for determining reaction conditions.

5. Chapter 3: Lysosomal Interactomics and Structural Perspectives

5 This chapter has been published as a preprint: [**Singh, J.**, Elhabashy, H., Muthukottiappan, P., Stepath, M., Eisenacher, M., Kohlbacher, O., Gieselmann, V., Winter, D. (2022). Cross-linking of the Endolysosomal System Reveals Flotillin Structures and Putative Cargo. bioRxiv]

- For the online version of the paper including supplement, visit: https://doi.org/10.1101/2022.01.12.475930
- The supporting information excluding supplementary tables is available at: https://doi.org/10.1101/2022.01.12.475930
- **Remark:** The publication within the scope of chapter 3 has been done by me as the only first author and **SHOULD BE CONSIDERED** for this cumulative doctoral dissertation.

5.1 Introduction

The lysosomal proteome can be divided into luminal, membrane-associated and integral membrane proteins. While luminal proteins are primarily involved in the degradation of lysosomal substrates, the membrane acts as an interface between the lysosome and the cell, fulfilling various functions. These include, but are not limited to, lysosomal acidification, nutrient transport/sensing, lysosomal transport, and vesicle fusion [2]. While certain proteins and complexes have been investigated in great detail [101], the function, structure, and interaction partners of a significant number of lysosomal proteins have not been analyzed yet. Cross-linking mass spectrometry (XL-MS), a strategy to covalently link proteins on a large, unbiased scale, provides an ideal tool for investigating lysosomal protein interactions. In combination with state-of-the-art bioinformatic approaches, like interactive computational modeling, novel protein conformations can be further determined [71, 73].

Compared to other methods such as X-ray crystallography, XL-MS has the major advantage that transient interactions can be analyzed in addition to stable protein complexes. Moreover, the cross-link reaction is possible under physiological conditions, which favors the analysis of functionally relevant protein interactions [78]. The interest in cross-linking methodology is continuously increasing, especially due to technological advances regarding available mass spectrometers and analysis software [76, 82]. Publications demonstrate the successful use of the methodology for the structural elucidation of simple systems, such as the RNA polymerase II-TFIF complex [102]. With the development of new innovative cross-linkers and algorithms for automated data analysis, the study of protein-protein interaction in very complex samples, such as whole-cell lysates, is also possible [76]. Also, the study of protein interactions in whole organelles was performed in multiple studies [77, 87, 100]. In one of the studies, the mitochondrial membrane was successfully reconstructed using XL-MS, a large number of protein interactions were identified, and new unknown interactions were postulated [77]. The current study used XL-MS to investigate endosome/lysosome enriched fractions applying the MS-cleavable cross-linker disuccinimidyl sulfoxide (DSSO) and presents the first cross-linking dataset of human endolysosomal proteins.

5.2 Publication: Main Article

Cross-linking of the Endolysosomal System Reveals Flotillin Structures and Putative Cargo

Jasjot Singh[1,*], Hadeer Elhabashy[2,3,4], Pathma Muthukottiappan[1], Markus Stepath[5,6], Martin Eisenacher[5,6], Oliver Kohlbacher[3,4,7], Volkmar Gieselmann[1], and Dominic Winter[1,8*]

[1]Institute for Biochemistry and Molecular Biology, Medical Faculty, University of Bonn, 53115 Bonn, Germany.
[2]Department of Protein Evolution, Max-Planck-Institute for Developmental Biology, Max-Planck-Ring 5, 72076 Tübingen, Germany.
[3]Institute for Bioinformatics and Medical Informatics, University of Tübingen, Sand 14, 72076, Tübingen, Germany.
[4]Department of Computer Science, University of Tübingen, Sand 14, 72076 Tübingen, Germany.
[5]Medizinisches Proteom-Center, Medical Faculty, Ruhr-University Bochum, Bochum, Germany.
[6]Medical Proteome Analysis, Center for Protein Diagnostics, Ruhr-University Bochum, Bochum, Germany.
[7]Institute for Translational Bioinformatics, University Hospital Tübingen, 72076 Tübingen, Germany.
[8]Lead contact

***Corresponding authors**
Jasjot Singh and Dr. Dominic Winter
Institute for Biochemistry and Molecular Biology
University of Bonn
Nussallee 11, 53115 Bonn, Germany
Email: jsin@uni-bonn.de; dominic.winter@uni-bonn.de

Keywords
Lysosome, Endosome, Cross-linking, Mass spectrometry, Flotillin

SUMMARY

Lysosomes are well-established as the main cellular organelles for the degradation of macromolecules and emerging as regulatory centers of metabolism. They are of crucial importance for cellular homeostasis, which is exemplified by a plethora of disorders related to alterations in lysosomal function. In this context, protein complexes play a decisive role, regulating not only metabolic lysosomal processes, but also lysosome biogenesis, transport, and interaction with other organelles. Using cross-linking mass spectrometry, we analyzed lysosomes and early endosomes. Based on the identification of 5,376 cross-links, we investigated protein-protein interactions and structures of lysosome- and endosome-related proteins. In particular, we present evidence for a tetrameric assembly of the lysosomal hydrolase PPT1 and heterodimeric/-multimeric structures of FLOT1/FLOT2 at lysosomes and early endosomes. For FLOT1-/FLOT2-positive early endosomes, we identified >300 proteins presenting putative cargo, and confirm the latrophilin family of adhesion G protein-coupled receptors as substrates for flotillin-dependent endocytosis.

INTRODUCTION

Lysosomes, the central lytic organelles of mammalian cells, are of crucial importance for cellular homeostasis. This is underscored by the detrimental consequences resulting from impairment of lysosomal function: mutations in genes encoding lysosomal proteins are causative for a group of around 70 rare and frequently devastating diseases, so-called lysosomal storage disorders (LSDs). Moreover, lysosomal dysfunction has been demonstrated in a number of more common conditions, including neurodegenerative diseases and cancer (Fraldi et al., 2016; Platt, 2018).

In addition to the long-known role of lysosomes in the degradation of intra- and extracellular substrates, more recent findings place them at the center of metabolic signaling. The major player in this context is the mammalian target of rapamycin complex 1 (mTORC1), whose activity is regulated at the lysosomal surface. This regulation is mediated by several protein complexes located in/at the lysosomal membrane, which integrate the activity of major signaling pathways, as well as the concentration of various metabolites (Shin and Zoncu, 2020). Furthermore, protein complexes were shown to play a role in other lysosome-related processes, such as their transport, direct interaction with other cellular compartments, gene regulation, immunity, cell adhesion/migration, and plasma membrane repair (Ballabio, 2019)

For most of these functions, protein-protein interactions at the lysosomal membrane play a decisive role. Nutrient sensing and activation of mTORC1 is regulated by the interaction of at least 30 individual proteins (Liu and Sabatini, 2020), and lysosomal motility is controlled by the reversible association to microtubules through dynein and kinesin by several adaptor/scaffold complexes, such as BLOC1-related complex (BORC) (Cabukusta, 2018). The core feature of lysosomes, their acidic pH, is maintained by the 1.25 MDa vacuolar-type ATPase (V-ATPase) complex, which consists of 35 subunits (17 unique proteins), and catalyzes the transport of protons across the lysosomal membrane (Wang, 2020). Delivery of certain lysosomal proteins is achieved by members of the homotypic fusion and protein sorting (HOPS) (Garg et al., 2011), the class C core vacuole/endosome tethering (Corvet) (Balderhaar, 2013), as well as adaptor protein (AP) complexes, and the endosomal sorting complex required for transport (ESCRT) mediates repair of lysosomal membranes (Skowyra et al., 2018).

Also, for the interactions of lysosomes with other organelles, protein complexes play an essential role. This includes fusion events with cargo delivery vesicles such as endosomes, phagosomes, and autophagosomes (Cheng et al., 2010), exocytosis at the plasma membrane (Reddy et al., 2001), or direct interactions with the endoplasmic reticulum (Levin-Konigsberg et al., 2019), the Golgi apparatus (Hao et al., 2018) peroxisomes (Chu et al., 2015), RNA granules (Liao et al., 2019), and mitochondria (Wong, 2018). The latter facilitates, for example, the exchange of small molecules and was shown to regulate events such as mitochondrial fusion and fission (Ballabio, 2019).

The majority of protein complexes that facilitate these processes are poorly characterized, and novel members/interactors are continuously being identified. Given the central role of lysosomes in metabolic regulation, and the high number of cellular structures they interact with, it is highly likely that several functionally important interactors of lysosomal proteins are still unknown. Although structural data are available for a number of lysosomal luminal proteins and complexes in/at its membrane, three-dimensional information is still lacking for a significant fraction of the lysosomal proteome. The majority of existing structural data originates from crystallography experiments, heavily relying on affinity-purified proteins, or fragments thereof, from pro- or eukaryotic overexpression systems, and crystallization *in vitro*. The applicability of these structures to the *in vivo* situation remains, therefore, in some instances, questionable (Niedzialkowska et al., 2016).

A promising avenue to identify unknown interactions of lysosomal proteins, and to reveal new insights into their structure under physiological conditions, is chemical cross-linking in combination with mass spectrometry-based proteomics (XL-LC-MS/MS) (O'Reilly and Rappsilber, 2018). In cross-linking experiments, a chemical linker forms covalent bonds between certain amino acids such as lysine. In subsequent MS analyses these bonds are identified, providing direct proof for the interaction of proteins within a certain distance constraint, defined by the type of cross-linker (Yu and Huang, 2018). This allows for the identification of protein-protein interactions (PPIs), and hence localization, with high confidence. Compared to other commonly used approaches, such as immunoprecipitation (IP), proximity labeling (Kim and Roux, 2016) or lysosome-enrichment (Muthukottiappan and Winter, 2021), such data provide superior spatial evidence. Furthermore, the distance constraints of the cross-linker can serve as a basis for the molecular modeling of proteins and their complexes. This allows supplementing well-established techniques such as nuclear magnetic resonance (NMR), X-ray crystallography, or cryo-electron microscopy (cryo-EM), compensating for missing/incomplete data, and validating predicted protein structures (Barysz, 2018).

So far, XL-LC-MS/MS experiments have been performed for samples of varying complexity, ranging from individual proteins (Chen, 2010) and multi-subunit complexes (Albanese, 2020; O'Reilly et al., 2020), to whole organelles (Fasci, 2018; Schweppe et al., 2017) and cell/tissue lysates (Chavez et al., 2019; Liu et al., 2015). Dedicated analysis of lysosomal proteins by cross-linking has not been performed to date, which is certainly related to the fact that lysosomal proteins are of low abundance (estimated 0.2 % of cellular protein mass (Itzhak et al., 2017; Valm et al., 2017)).

In the current study, we present the first cross-linking dataset of the endolysosomal compartment of HEK293 cells, applying the MS-cleavable cross-linker disuccinimidyl sulfoxide (DSSO) to lysosome-/early endosome-enriched fractions. We present an interaction map of lysosomal proteins, of which we verify selected PPIs by co-IP and validate/extend existing protein structures. Based on the cross-linking data and computational modeling, we further propose higher-order structures for PPT1 and flotillin assemblies. Finally, by affinity purification and MS analysis of flotillin-positive early endosomes, we investigate the putative cargo of these vesicles.

RESULTS

Cross-Linking Mass Spectrometry Analysis of Lysosome Enriched Fractions

In mammalian cells, the majority of lysosomal proteins are of relatively low abundance, and whole-cell XL-LC-MS/MS studies typically cover only a fraction of the lysosomal proteome (Figure S1A). A way to overcome this limitation is lysosome enrichment, which we showed to increase signal intensities for certain lysosomal proteins up to 100-fold relative to whole cell lysates (Singh et al., 2020). Accordingly, we enriched lysosomes by superparamagnetic iron oxide nanoparticles (SPIONs, Figure 1A), and established cross-linking conditions for lysosome-enriched fractions utilizing the MS-cleavable cross-linker DSSO (Kao et al., 2011). Due to a limited membrane permeability of DSSO (Singh et al., 2021), we cross-linked lysosomes both in an intact (IT) and disrupted (DR) state, and determined optimal reaction conditions by silver staining and western blot (Figure S1B, C). Subsequently, we enriched lysosomes from 384 plates of HEK293 cells across three biological replicates, and assessed lysosomal intactness, recovery, and enrichment (Figure 1B, 1C). Using a non-cross-linked fraction of each sample, we acquired an LC-MS/MS reference dataset. In total, we identified 4,181 proteins, of which 474 were assigned the term “lysosome” based on GO terms and UniProt classifiers in >3 runs, indicating an excellent performance of lysosome enrichment (Akter, 2020). To assess the quantitative distribution of lysosomal proteins in our samples, we utilized these data to estimate absolute protein abundances by intensity-based absolute quantification (iBAQ) (Schwanhausser et al., 2011). This revealed a three-fold overrepresentation of lysosomal protein abundance relative to the whole dataset (Figure 1D).

We cross-linked lysosome enriched fractions in both the IT and DR state, followed by their proteolytic digestion, strong cation-exchange (SCX) peptide fractionation, and analysis by LC-MS/MS (Figure 1A). Analysis of the XL-LC-MS/MS dataset with XlinkX (Klykov et al., 2018) resulted in the assignment of 6,580 cross-link spectral matches, originating from 4,294 cross-linked peptides. Out of the 2,467 unique residue-to-residue cross-links, 524 identifications (270 intra links between different residues of the same protein and 254 inter-links between two different proteins) originated from 111 proteins assigned to the lysosomal compartment (Figure 1E, Figure S1D-H). Interestingly, only 25 % of cross-links were found both in the IT and the DR state, while the latter contributed a larger fraction to the dataset, further demonstrating the limited membrane permeability of DSSO (Figure 1F). A similar distribution was observed for the cross-links identified for non-lysosomal proteins contained in the dataset (Figure S1D). Strikingly, while cross-link spectral matches (CSMs) of cytosolic proteins were almost equally distributed between both conditions, 91 % of the dataset's CSMs assigned to lysosomal luminal proteins were identified in samples cross-linked in the DR state (Figure 1G).

As lysosomal proteins are expressed at a dynamic abundance range encompassing three orders of magnitude (Akter, 2020), we further correlated CSMs, peptide spectral matches (PSMs), and iBAQ values. Even though higher abundant proteins tended to yield more CSMs, we did not identify a strong correlation between cross-link identification and protein abundance, showing that our dataset also covered proteins of low expression levels (Figure 1H, Figure S1I). When we compared the average iBAQ abundance for proteins involved in intra- and inter-links, we observed a tendency towards the identification of more intra-links in higher abundant proteins, which was less pronounced for lysosomal proteins (Figure S1J, K).

Finally, we investigated the distribution of CSMs across a short-list of lysosomal proteins/complexes (Muthukottiappan and Winter, 2021). Most categories showed an equal distribution between DR and IT samples, with the exception of proteins involved in lysosomal substrate degradation (93 %), heat shock proteins (70 %), and annexins (75 %), for which more cross-links were annotated in the DR sample (Figure S1L).

Characterization of the Human Lysosomal Interactome

Based on all inter-links contained in the dataset, we constructed a network of 1,008 proteins engaged in 1,023 interactions, of which 254 involved lysosomal proteins (Figure 2A, Figure S2A-C). Comparison to known interactions revealed an overlap of ~30 %, confirming the validity of our dataset. While 34 % of interactions of non-lysosomal proteins were included in STRING, only 26 % of potential lysosomal PPIs have been reported previously (Figure 2B). We classified lysosomal PPIs based on the interacting subcellular compartment, revealing an overrepresentation of nuclear and cytoplasmic/cytoskeletal proteins (Figure 2C). With respect to lysosomal and lysosome-associated proteins, we identified the highest numbers of PPIs for the V-ATPase, the flotillins, mTORC1, and the syntaxins (Figure 2D, E, Figure S2D, E).

For the V-ATPase, we detected most PPIs for the D subunit of its soluble V1 part (ATP6V1D). This may be related to the capability of V1 to dissociate from the lysosomal membrane-embedded V0 part, which was shown to be involved in the regulation of V-ATPase activity (Maxson and Grinstein, 2014), exposing the D subunit to interactions. The cross-link of ATP6V1D with frizzled 9 (FZD9), a member of the WNT signaling pathway, sparked our interest. FZD9 was shown to be sorted to late endosomes/lysosomes after its internalization by endocytosis (Grainger et al., 2019), and the ATP6AP2 subunit of the V-ATPase was reported to interact with FZD8 (Cruciat et al., 2010). We therefore investigated the validity of this PPI by co-IP, confirming both the observed interaction of ATP6V1D with FZD9, as well as interaction with other subunits of the complex (Figure 2F).

We further investigated the interaction of FLOT1, FLOT2, and GNB4, for which we identified an inter-link with FLOT2. We were able to co-IP FLOT1 and FLOT2, which are known to form heterooligomers (Babuke et al., 2009), as well as GNB4 with its direct interactor FLOT2 as well as with FLOT1, indicating binding of GNB4 to FLOT1/FLOT2 heteromeric assemblies (Figure 2G).

As we observed 18 different PPIs for FLOT1 and FLOT2, we further investigated their distribution across both proteins. While FLOT1/FLOT2 inter-links were detected across most of the regions predicted to form a helical structure (amino acids 193-365 and 213-362 for FLOT1 and FLOT2, respectively) (Rivera-Milla et al., 2006), the interaction with other proteins occurred almost exclusively in confined sections of < 100 amino acids (Figure 2H, I). While the equal distribution of inter-links shows that no sequence-dependent bias towards cross-link detectability exists, localization of the majority of PPIs to a distinct part of the proteins suggests the presence of FLOT1/FLOT2 interaction hotspots.

Structural Integration of Cross-Linker Distance Constraints Suggests a Tetrameric Assembly of PPT1 *In Vivo*

Cross-links between different amino acids provide distance information (the length of DSSO links is ~35 Å) that is helpful to validate (or infer) protein structures (Kastritis et al., 2017). Initially, we used TopoLink (Ferrari et al., 2019) to match 161 unique cross-links to the resolved structures of 34 lysosomal and lysosome-associated proteins, confirming the validity of our dataset. The remaining 64 cross-links assigned to lysosomal proteins could not be integrated, as the respective regions have not been resolved yet. We matched these cross-links either to homology models based on available PDB structures from other organisms using SWISS-MODEL (Waterhouse et al., 2018), or to predicted AlphaFold models which became available during the preparation of this manuscript (Jumper et al., 2021) (Figure 3A, Figure S3).

The complex for which we identified the highest number of cross-links was the V-ATPase, a 35-mer assembly of 17 unique proteins (Wang, 2020), of which 13 were covered in our dataset. While the soluble V1 part (eight different proteins) yielded 26 cross-links, only three were identified for the membrane-embedded V0 section (seven different proteins). After confirmation of the V1 part's correct stoichiometry through the individual subunit's iBAQ values in the non-cross linked dataset ($A_3B_3E_3G_3D_1F_1C_1H_1$) (Figure 3B), we mapped the identified cross-links to a recently published structure determined by cryo-EM (Wang, 2020). Out of the 29 unique cross-links identified, 21 could readily be integrated into the published structure, while 8 originated from regions that were so-far structurally not resolved (Figure 3C). For these sections, we integrated the predicted full-length protein models by AlphaFold, aligning them with the published V-ATPase sub-unit structures in the complex, based on the identified cross-links (Figure 3D). In this combined model, we cover ~95 % of the V-ATPase sequence, and 90 % of cross-links fulfil DSSO's distance constraints. The remaining three cross-links were inter-links of the A and E subunit of the V1 part, which undergo conformational changes during ATP hydrolysis (Wang, 2020).

With respect to mTORC1, we identified most cross-links for proteins related to Ragulator, a lysosomal membrane-associated complex that is crucial for mTORC1 activity (Shin and Zoncu, 2020). For the N-terminal region of LAMTOR1, we identified three cross-links of the same lysine residue with different amino acids of LAMTOR3. Two of them violated DSSO's distance constraints relative to the crystal structure obtained in the presence of the Ragulator-associated RAG GTPases (De Araujo, 2017), indicating cross-linking of an alternative state, possibly representing Ragulator in the absence of RAG GTPases (Figure S3J). We further identified three intra-links from RRAGA that originated from the same lysine residue at its C-terminus, cross-linked to three different amino acids (Figure S3K. This could be related to a high flexibility of this region of the protein, which is in accordance with the fact that it could not be covered in a previous crystallization study (Su et al., 2017).

Of the 161 lysosomal cross-links identified, 21 exceeded DSSO's distance constraint (Figure 3E). Surprisingly, nine of them originated from PPT1, a member of the palmitoyl protein thioesterase family. The current PPT1 crystal structure (3GRO) resembles a homodimeric assembly of two identical subunits (Figure 3F). Strikingly, while all 11 PPT1 intra-links fulfilled the distance constraints of this structure, the inter-links exceeded them, indicating the possibility of an alternative oligomerization state. This possibility is in line with a previous study, detecting the maximal enzymatic activity of PPT1 at a complex size of >100 kDa (Lyly et al., 2007). We therefore used HADDOCK to perform restraint-based docking for the monomeric subunits of PPT1, extracted from 3GRO. This resulted in the prediction of a tetrameric PPT1 model, which fulfills the distance constraints for all 18 PPT1 cross-links (Figure 3G, H).

Proposal of a Heterodimeric FLOT1/FLOT2 Model Featuring Extended Alpha-Helical Domains

We identified the highest number of cross-links for the two members of the flotillin family, FLOT1 and FLOT2, which were also overrepresented in the proteomic dataset of the lysosome-enriched fraction (Figure 1H). FLOT1 and FLOT2 are lipid raft-associated proteins, which are present in

In order to elucidate if the structural assembly of FLOT1 and FLOT2 differs between early endosomes and lysosomes, we performed cross-linking of early endosome-enriched fractions. To increase the number of FLOT1-/FLOT2-positive endosomes, we co-transfected cells with FLAG-tagged FLOT1 and FLOT2, as their overexpression was shown to increase the number of FLOT-positive endosomes (Babuke et al., 2009; Frick et al., 2007). After confirmation of their correct localization (Figure S5A, B), we established SPIONs pulse-chase conditions for the enrichment of early endosomes (Figure 5C). Analysis of the early endosome-enriched fraction by western blot and LC-MS/MS verified the presence of marker proteins such as EEA1, the clathrin chains CLTA, CLTB, and CLTC, as well as the RAB-GTPases RAB5, RAB11, and RAB14. Furthermore, we detected FLOT1 and FLOT2, while markers for other organelles were depleted (Figure 5D). We then performed enrichment of early endosomes from FLOT1/-FLOT2-FLAG overexpressing HEK293 cells, established their cross-linking followed by SCX fractionation and investigated them by XL-LC-MS/MS (Figure 5E, Figure S5C-F). Importantly, western blot analysis of FLOT1/FLOT2 aggregation in response to different amounts of DSSO showed that the concentration we used for cross-linking of lysosome-enriched fractions does not result in over-cross-linking (Figure S5D), further supporting the validity of these data. In total, we identified 1,081 cross-links from 414 unique proteins (Figure 5F, Figure S5G). This early endosome dataset contains 15 unique cross-links for FLOT1 and FLOT2, which all matched to our predicted FLOT1-FLOT2 heterodimer within DSSO's distance constraints, and from which 80 % overlapped with the cross-linking dataset from lysosome-enriched fractions (Figure 5G). These data indicate that early endosome- and lysosome-localized flotillins assemble in a similar way.

It has been shown previously, that flotillins form higher-order assemblies (Solis et al., 2007; Stuermer et al., 2001). We therefore performed blue native (BN)-PAGE experiments to investigate the size distribution of FLOT1/FLOT2 structures in a native and a cross-linked state (Figure 5H, Figure S5H). These analyses revealed that a significant amount of both FLOT1 and FLOT2 migrates at a range corresponding to a tetrameric assembly, while smaller fractions migrated at sizes consistent with higher-order structures exceeding 1 MDa. The cross-linking samples, which were generated with the same reaction conditions as the early endosome and lysosome experiments, presented with the same complex sizes, indicating that the cross-link data represent the native state.

Based on the combined 32 unique cross-links from the lysosome- and endosome-cross-linking datasets, we further investigated possible structures for higher-order assemblies. All of our attempts to model the tetrameric structure did not lead to a plausible outcome. We further addressed the structures exceeding 1 MDa, which were stabilized by cross-linking with DSSO (Figure 5H). We utilized the structure of rat major vault protein (Mvp), which assembles into the 3.7 MDa rat liver vault (PDB: 4V60), as template for a higher-order hetero-oligomeric FLOT1/FLOT2 model, as Mvp shares several structural properties with flotillins. A key feature in this context are the N-terminal PHB and the 42-turn-long cap-helix domains, which are crucial for stabilizing the particle (Tanaka, 2009). We used the rat liver vault structure as a template for building a model based on FLOT1/FLOT2 hetero-oligomers, proposing a 38-mer structure, which could represent one of many possible higher-order assemblies of the FLOT1/FLOT2 heterodimer (Figure 5I). Important features of this structure are the exposed ring of palmitoylation/myristoylation sites, the central arrangement of PDZ domains, and the two rings of phosphotyrosine (pY) residues. While the pY sites known for FLOT1/FLOT2 interaction (Y160/Y163) are located at the inside of the structure, the sites possibly involved in PPI regulation (Y238/Y241) are located at its outside, making them accessible to kinases even after formation of the higher-order structure.

Analysis of Flotillin-Endosome Cargo Reveals an Overrepresentation of Membrane Proteins and Receptors

Flotillins have been proposed as defining structural components of an endocytic pathway independent of clathrin and caveolin (Frick et al., 2007; Glebov et al., 2006) and were shown to co-localize with early endosomes (Gorbea et al., 2010). In agreement with these findings, we nearly every type of vertebrate cell, and are highly conserved among organisms (Rivera-Milla et al., 2006). We confirmed their co-enrichment with lysosomes by western blotting (Figure 4A) and co-localization by immunostaining (Figure 5A), which is in agreement with previous EM studies detecting FLOT1 at the lysosomal surface (Kaushik et al., 2006; Kokubo et al., 2003). It is well-established, that FLOT1 and FLOT2 form heterodimers with a 1:1 stoichiometry (Frick et al., 2007), but only partial structural information is available from NMR analyses of the N-terminal region of mouse Flot2, as purification of the full-length proteins is problematic (Dempwolff et al., 2016).

Our dataset contains 29 unique cross-links for FLOT1 and FLOT2, including 11 intra- and 22 inter-links. In a first step, we predicted individual secondary structures for both FLOT1 and FLOT2 (Figure 4B) using PSIPRED (Jones, 1999) For the N-termini, this resulted in a cluster of beta-sheets, which is in accordance with their sequence homology to the stomatin/PHB/flotillin/HflK/C (SPFH) domains and the Flot2 NMR structure. The middle section features an extended α-helical region that is interrupted once in the case of FLOT2, while the C-terminus forms one beta-sheet and several short helices for both proteins. We further calculated coiled-coil probabilities for the helical regions with PCOILS (Zimmermann et al., 2018) with window sizes of 14, 21, and 28 amino acids (Figure 4B). Dependent on the region of both proteins, windows of 14 and 28 amino acids delivered the best results, with slightly different patterns for FLOT1 and FLOT2. Matching of our data to full-length structural models from AlphaFold showed an excellent agreement and all identified cross-links confirmed the predicted structures.

Subsequently, we built heterodimeric models using ColabFold (Mirdita et al., 2021), containing closely aligned highly similar structures for both flotillins. In particular, they feature a globular N-terminal region consisting of SPFH domains (residues 1-162 of FLOT1 and FLOT2), a central linear α-helical region (residues 163-341 of FLOT1 and 163-350 of FLOT2), and a C-terminal α-helical coiled-coil structure (residues 342-427 of FLOT1 and 351-426 of FLOT2). The SPFH domains of both FLOT1 and FLOT2 present with antiparallel β-sheets, with six repeats each, and four partially exposed α-helices, forming an ellipsoidal-like globular domain (Figure 4C, Figure S4). Based on the fact that the major helices were not interrupted, and that the C-terminus was in its most compact state, we selected model four, which was also supported by all FLOT1/FLOT2 cross-links detected (Figure 4C, Figure S4C).

We integrated several structural features which are known for both flotillins into our model (Figure 4C). The *S*-palmitoylation and *N*-myristoylation sites, which are crucial for membrane-association of both flotillins (Neumann-Giesen et al., 2004), are located at the SPFH domain's membrane interfaces. The tyrosine phosphorylation sites, which were shown to be crucial for flotillin-mediated endocytosis and FLOT1/FLOT2 interaction (Bach and Bramkamp, 2015; Riento et al., 2009), are surface-exposed and in proximity to a basic motif (HQR) on the respective other flotillin. Moreover, the PDZ3 domains of both flotillins strongly co-localize, forming a combined feature, and the EA-rich motifs, which were predicted to mediate flotillin oligomerization (Rivera-Milla et al., 2006; Solis et al., 2007), are distributed along the length of the central α-helical region (Kwiatkowska, 2020). Interestingly, the putative interaction hotspot (Figure 2 H, I) locates around the major bend observed in this structure, containing one tyrosine residue (Y238/Y241) in its center. These residues are located in highly conserved sequence motifs (A-X-A-X-L-A-pY-X-L-Q with X: [D/Q or E/Q]), possibly presenting a regulatory switch for FLOT1/FLOT2 PPIs.

Flotillins Assemble in Similar Higher Order Structures at Lysosomes and Endosomes

While FLOT1 was detected previously at the cytosolic face of lysosomes by EM (Kaushik et al., 2006; Kokubo et al., 2003) assemblies of both FLOT1 and FLOT2 were only shown for the plasma membrane or early endosomes, where they play a role in clathrin-independent endocytosis (Gorbea et al., 2010; Stuermer et al., 2001). In line with these findings, we detected FLOT1 and FLOT2 to partially co-localize with the plasma membrane and lysosomes of HeLa and HEK293 cells (Figure 5A, B). In these analyses, we also observed numerous FLOT1/FLOT2 punctae, which did not co-localize with lysosomes, presenting putative FLOT1/FLOT2-positive endosomes.

in vivo (Go et al., 2021; Liao et al., 2019) In contrast to cross-linking, proximity biotinylation only allows to determine the presence of a protein within a defined radius relative to the respective fusion construct. It cannot identify, however, if two proteins are directly interacting, or which residues/domains are close to each other. This is exemplified by a recent study utilizing LAMP1, LAMP2, and LAMP3 fused to BirA* for investigation of the lysosomal proteome (Go et al., 2021) Comparison of the interaction partners identified for all three constructs revealed that only 8-18 % were unique, while the others were enriched in at least two datasets. Therefore, the interactome presented in this study provides a level of detail that is unprecedented for the analysis of lysosomal PPIs.

Another common approach for the investigation of protein interactions is co-IP, which we used to investigate two PPIs of members of the largest interaction networks, namely the FLOT1/FLOT2 complex and the V-ATPase complex. We selected the interaction of ATP6V1D with FZD9, a G-protein coupled multi-pass transmembrane receptor for WNT2 (Karasawa et al., 2002), as another member of this family, FZD8, was previously identified to interact with ATP6AP2. Intriguingly, it was shown in this study that correct V-ATPase function, and accordingly acidification of the lysosome, is necessary for WNT signaling, and that interaction of FZD8 with the V-ATPase complex plays a decisive role (Hermle et al., 2010). The direct interaction of ATP6V1D and FZD9 identified in this study substantiates this functional connection of WNT signaling and lysosomal acidification. A possible role could be related to lysosomal acidification. V-ATPase can only acidify lysosomes when the V0 part, which is integrated in the lysosomal membrane, pairs with the cytosolic V1 part. The independently assembled V1 part (Wang, 2020) can reversibly dissociate from V0, and this appears to be a process that can be regulated through different types of stimuli (Maxson and Grinstein, 2014). Similarly, it is conceivable that frizzled proteins could control V1/V0 V-ATPase assembly, and thus possibly regulate lysosomal acidification, and hence WNT signaling.

Another major lysosomal complex covered in our dataset was mTORC1. Among others, we detected cross-links for the interaction of the same lysine in the N-terminal region of LAMTOR1 with different residues of LAMTOR3. Both proteins are members of the Ragulator complex, whose interaction with the RAG-GTPases regulates mTORC1 activity (Zhang et al., 2017). It was shown previously that the N-terminal region of LAMTOR1 could not be crystallized without the RAG GTPases, implicating that it could exist in an unordered state under these circumstances (De Araujo, 2017). Matching of the three cross-links to the Ragulator crystal structure determined in the presence of RAG GTPases fulfilled only for one of them DSSO’s distant constraints. A possible explanation for the other two cross-links is that they originate from a state where Ragulator was not interacting with the RAG GTPases, indicating an alternative structure of the LAMTOR1 N-terminal region *in situ*.

The observation of several overlength cross-links, when matched to a structure determined from overexpression or *in vitro* experiments, can indicate that it does not present the native form of a protein (or it’s complex). In our dataset, the inter-links matched to the dimeric model of PPT1 (PDB: 3GRO) were responsible for 43 % of overlength cross-links, while all intra-links matched the published structure. Importantly, our proposed tetrameric PPT1 model satisfied all distance constraints. This is in line with previous findings, detecting the majority of enzymatic activity in size exclusion chromatography fractions correlating with the molecular weight of a tetrameric assembly (Lyly et al., 2007). The observed discrepancy to the crystallography-based structure could be related to the expression system utilized (*Spodoptera frugiperda*), as it lacks the capability to glycosylate PPT1 properly. It was shown previously, that glycosylation-deficient PPT1 variants are devoid of enzymatic activity, which was attributed to improper folding (Bellizzi et al., 2000; Lyly et al., 2007).

The flotillins, which yielded the highest number of cross-links, were reported in several studies to interact with a large variety of proteins at different subcellular locations, and to be involved in a plethora of processes (Bodin et al., 2014). FLOT1 has been demonstrated to play a role in clathrin/caveolin-independent endocytosis (Glebov et al., 2006), and the cargo of such FLOT1-positive endosomes has been shown to be delivered to lysosomes (Fan et al., 2019;

detected their co-enrichment with endosomes (Figure 5C, D). To investigate the putative cargo of FLOT1-/FLOT2-positive early endosomes, we combined early endosome enrichment from HEK293 cells overexpressing FLAG-tagged FLOT1/FLOT2 and immunoprecipitation of FLAG-positive intact vesicles (Figure 6A). Subsequently, we analyzed the resulting fractions by label-free quantification (Figure S6A-D). Based on 5,089 protein groups quantified across all conditions, we were able to define protein populations that were enriched/depleted in FLOT1-/FLOT2-positive endosomes relative to the total cellular pool of early endosomes (Figure 6B). Importantly, the FLOT1-/FLOT2-depleted population of early endosomes contained the clathrin chains CLTA and CLTC, EEA1, and the endosome-related GTPases RAB5 and RAB11, confirming a separation of clathrin- and flotillin-containing early endosomes. We confirmed these findings by co-immunostaining of EEA1 with endogenous and overexpressed FLOT1/FLOT2 in HeLa and HEK293 cells, showing that FLOT1-/FLOT2-positive vesicles present a distinct population from EEA1-positive endosomes (Figure S6E).

The enriched population, which presents potential cargo of FLOT1-/FLOT2-positive early endosomes, consists of 328 proteins. GO-enrichment analysis revealed an overrepresentation of membrane proteins (Figure S6F), which is in agreement with a potential role of flotillins in the endocytosis and vesicular transport of plasma membrane proteins, demonstrated previously e.g. for NPC1L1 (Ge et al., 2011; Meister and Tikkanen, 2014). We subsequently performed STRING analyses to sub-classify the potential cargo proteins (Figure 6C). They contained both receptor tyrosine kinases and G-protein coupled receptors (50 receptors total), seven members of the solute carrier (SLC) family of transporters, and 13 members of the transmembrane protein (TMEM) family.

One group of receptors that sparked our interest was the latrophilins, as we detected all three members of this G protein-coupled receptor subfamily (LPHN1, LPHN2, and LPHN3) to be significantly enriched in FLOT1-/FLOT2-positive endosomes (Figure 7A). We confirmed their co-localization to FLOT1-/FLOT2-positive endosomes by immunofluorescence in HeLa and HEK293 cells (Figure 7 B, Figure S7). To further investigate the relation between FLOT1/FLOT2 and LPHN1/LPHN2/LPHN3, we performed IP experiments. We were able to co-IP all latrophilins with both flotillins, implicating not only the presence of latrophilins at FLOT1-/FLOT2-positive endosomes, but also a direct interaction (Figure 7C).

DISCUSSION

In the current study, we present the first XL-LC-MS/MS analysis of lysosome-enriched fractions. In comparison to whole-cell studies, we detected significantly higher numbers of cross-links for proteins reported previously to be related to the lysosome. This was especially the case for *bona fide* lysosomal proteins, which are localized in/at the organelle (Thelen et al., 2017). When considering this shortlist, we identified >100-fold more cross-links compared to previously published whole-cell studies (Chavez et al., 2019; Liu et al., 2015). A probable key factor in this context is our starting material, consisting of SPIONs-enriched lysosomes, as, relative to the total proteome, lysosomal proteins are typically of low abundance (estimated 0.2 % of cellular protein mass (Itzhak et al., 2016)). With respect to individual lysosomal proteins, we did not observe a strong correlation of their abundance and the number of cross-links identified. We identified, for example, no cross-links for the lysosome-associated membrane glycoproteins 1/2 (LAMP1 and LAMP2), which were estimated to contribute 50 % of lysosomal protein mass (Eskelinen et al., 2004) and were among the most abundant lysosomal proteins in our dataset. Also, for the lysosomal luminal hydrolase CTSD, which was detected with the third-highest iBAQ value in the whole dataset, only two unique cross-links were detected, while we found the same number for the low abundant protein CHID1 (~275-fold less abundant).

Based on the identified cross-links, we generated a protein interaction network containing 70 % potential novel PPIs. Compared to other approaches, cross-linking allows identifying interacting residues between individual proteins *in situ*. An approach able to generate data under similar near-native conditions is proximity biotinylation, utilizing e.g., BirA* or APEX2, which have been used in two different studies to investigate the lysosomal (surface) proteome and interactome *in viv*o (Go et al., 2021; Liao et al., 2019) In contrast to cross-linking, proximity biotinylation only allows to determine the presence of a protein within a defined radius relative to the respective fusion construct. It cannot identify, however, if two proteins are directly interacting, or which residues/domains are close to each other. This is exemplified by a recent study utilizing LAMP1, LAMP2, and LAMP3 fused to BirA* for investigation of the lysosomal proteome (Go et al., 2021) Comparison of the interaction partners identified for all three constructs revealed that only 8-18 % were unique, while the others were enriched in at least two datasets. Therefore, the interactome presented in this study provides a level of detail that is unprecedented for the analysis of lysosomal PPIs.

Another common approach for the investigation of protein interactions is co-IP, which we used to investigate two PPIs of members of the largest interaction networks, namely the FLOT1/FLOT2 complex and the V-ATPase complex. We selected the interaction of ATP6V1D with FZD9, a G-protein coupled multi-pass transmembrane receptor for WNT2 (Karasawa et al., 2002), as another member of this family, FZD8, was previously identified to interact with ATP6AP2. Intriguingly, it was shown in this study that correct V-ATPase function, and accordingly acidification of the lysosome, is necessary for WNT signaling, and that interaction of FZD8 with the V-ATPase complex plays a decisive role (Hermle et al., 2010). The direct interaction of ATP6V1D and FZD9 identified in this study substantiates this functional connection of WNT signaling and lysosomal acidification. A possible role could be related to lysosomal acidification. V-ATPase can only acidify lysosomes when the V0 part, which is integrated in the lysosomal membrane, pairs with the cytosolic V1 part. The independently assembled V1 part (Wang, 2020) can reversibly dissociate from V0, and this appears to be a process that can be regulated through different types of stimuli (Maxson and Grinstein, 2014). Similarly, it is conceivable that frizzled proteins could control V1/V0 V-ATPase assembly, and thus possibly regulate lysosomal acidification, and hence WNT signaling.

Another major lysosomal complex covered in our dataset was mTORC1. Among others, we detected cross-links for the interaction of the same lysine in the N-terminal region of LAMTOR1 with different residues of LAMTOR3. Both proteins are members of the Ragulator complex, whose interaction with the RAG-GTPases regulates mTORC1 activity (Zhang et al., 2017). It was shown previously that the N-terminal region of LAMTOR1 could not be crystallized without the RAG GTPases, implicating that it could exist in an unordered state under these circumstances (De Araujo, 2017). Matching of the three cross-links to the Ragulator crystal structure determined in the presence of RAG GTPases fulfilled only for one of them DSSO's distant constraints. A possible explanation for the other two cross-links is that they originate from a state where Ragulator was not interacting with the RAG GTPases, indicating an alternative structure of the LAMTOR1 N-terminal region *in situ*.

The observation of several overlength cross-links, when matched to a structure determined from overexpression or *in vitro* experiments, can indicate that it does not present the native form of a protein (or it's complex). In our dataset, the inter-links matched to the dimeric model of PPT1 (PDB: 3GRO) were responsible for 43 % of overlength cross-links, while all intra-links matched the published structure. Importantly, our proposed tetrameric PPT1 model satisfied all distance constraints. This is in line with previous findings, detecting the majority of enzymatic activity in size exclusion chromatography fractions correlating with the molecular weight of a tetrameric assembly (Lyly et al., 2007). The observed discrepancy to the crystallography-based structure could be related to the expression system utilized (*Spodoptera frugiperda*), as it lacks the capability to glycosylate PPT1 properly. It was shown previously, that glycosylation-deficient PPT1 variants are devoid of enzymatic activity, which was attributed to improper folding (Bellizzi et al., 2000; Lyly et al., 2007).

The flotillins, which yielded the highest number of cross-links, were reported in several studies to interact with a large variety of proteins at different subcellular locations, and to be involved in a plethora of processes (Bodin et al., 2014). FLOT1 has been demonstrated to play a role in clathrin/caveolin-independent endocytosis (Glebov et al., 2006), and the cargo of such FLOT1-positive endosomes has been shown to be delivered to lysosomes (Fan et al., 2019;

Stuermer et al., 2001). Furthermore, FLOT1 was detected at the lysosomes' cytosolic face (Kaushik, Kokuba). In accordance with these findings, we detected and verified the interaction of FLOT1/FLOT2 with guanine-nucleotide-binding protein subunit beta-4 (GNB4), a beta subunit of heterotrimeric G-proteins (Ruiz-Velasco et al., 2002). Variants of GNB4 were shown to cause Charcot–Marie–Tooth disease (Soong et al., 2013), and downregulation of GNB4 levels has been found in Gaucher Disease, a lysosomal storage disorder (Pawlinski et al., 2021) Association of GNB4 with lysosomes was implicated in a previous study (Schroder et al., 2007), and a yeast two-hybrid screen revealed its direct interaction with LAMP2 (Haenig et al., 2020) which co-localizes with FLOT1 and the lysosomal surface (Kaushik). Taken together, these data provide strong evidence that the lysosomal localization of GNB4 is mediated through interaction with the flotillins.

With respect to the structure of flotillins, only rudimentary information existed based on NMR analysis of the mouse Flot2 N-terminal region (PDB: 1WIN). For the remaining protein, until recently, the predicted structure featured a 180° turn in the central helix of both flotillins resulting in direct interaction of the proteins' N- and C-terminal domains (Rivera-Milla et al., 2006). The recently predicted FLOT1 and FLOT2 AlphaFold structures disagree with this model, predicting an extended helix with a bend in its middle for both of the proteins. This is supported by the cross-links detected in our dataset. We did not observe long-distance intra-links, which would be indicative for the proximity of distant regions of the proteins, and the inter-links of both proteins behave in unison, confirming intermolecular interactions along an extended region of the heterodimer.

We detected interaction with GNB4, among others, in the putative interaction hotspot, which localized in both flotillins to the major bend in the extended alpha-helical structure. Intriguingly, both the FLOT1 and the FLOT2 tyrosine residue located in this structure were reported in PhosphoSitePlus (Hornbeck et al., 2015) to be phosphorylated in >30 studies. Therefore, they could present a potential regulatory element for PPIs, e.g., to discriminate interactions regulated by tyrosine kinase- or G-protein-coupled-receptor signaling, as it was shown previously that flotillins are involved in both types of signaling pathways (Sugawara et al., 2007)

Multiple heptad coiled-coil motifs in both flotillins are known to favor higher-order protein structures (Grigoryan and Keating, 2008; Woolfson et al., 2012). In agreement with these characteristics and previously published results (Babuke et al., 2009; Solis et al., 2007), we identified different levels of higher-order flotillin structures, ranging from an abundant tetrameric assembly to assemblies exceeding 1MDa. Motivated by these findings, we generated a multi-heterodimeric model for the high molecular weight flotillin complex. Based on the close interaction of the FLOT1/FLOT2 PDZ3 binding motif, which is known to play a role in the assembly of large multiprotein complexes in stomatins (Hung and Sheng, 2002), and the reported 39-mer higher-order structure of rat liver vault protein (Tanaka, 2009), which shares characteristic features with the flotillins, we propose a model for a multimeric FLOT1/FLOT2 complex. A potential function for such an assembly could be the formation or transport of flotillin-positive endosomes (Glebov et al., 2006).

Except for a few proteins (Meister and Tikkanen, 2014), no substrates for flotillin-mediated endocytosis are known to date. Our analysis of its putative cargo identified >300 proteins, including a large number of membrane proteins and receptors. Among those, the latrophilins stood out, as all three members of this family of adhesion G protein-coupled receptors (Langenhan et al., 2016) were found to be significantly enriched. Furthermore, we observed both their co-localization and interaction with both flotillins, providing evidence for direct interaction, possibly in the course of flotillin-mediated endocytosis.

DECLARATION OF INTERESTS

The authors declare no competing interests.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Goat anti LIMP2 (1:2000)	R&D systems	# AF1966-SP
Goat anti mouse IgG HRP coupled (1:5000)	Dianova	# 115-035-044
Goat anti Rabbit IgG (H+L)-Cy3 (1:400)	Dianova	#111-165-144
Goat anti rabbit IgG HRP coupled (1:5000)	Dianova	# 111-035-003
Goat anti-Mouse IgG (H+L)-Alexa Fluor 488 (1:400)	Thermo Fisher Scientific	# A-11029
Mouse anti ACT2 (1:4000)	Sigma-Aldrich	# A5316
Mouse anti ATP6V1B2 (1:1000)	Santa Cruz	# SC166045
Mouse anti CANX (1:20,000)	Proteintech	# 66903-1-AP
Mouse anti FLOT1 (1:200)	BD Biosciences	# 610821
Mouse anti FLOT2 (1:1500)	Proteintech	# 66881-1-Ig
Mouse anti FLOT2 (1:200)	BD Biosciences	# 610383
Mouse anti FZD9 (1:1500)	Proteintech	# 67023-1-Ig
Mouse anti GAPDH (1:2500)	Cell signaling	# 5174
Mouse anti GM130 (1:1000)	BD Biosciences	# 610822
Mouse anti LAMP2 (1:1000)	Hybridoma Bank	# H4B4
Rabbit anti ATP6V1A1 (1:2000)	Thermo Fisher Scientific	# PA5-29191
Rabbit anti ATP6V1D (1:1000)	Proteintech	# 14920-1-AP
Rabbit anti CTSD (1:1000)	Proteintech	# 21327-1-AP
Rabbit anti DSSO (1:5000)	Self-made	(Singh et al., 2021)
Rabbit anti EEA1 (1:200)	Cell signaling	# 2411
Rabbit anti FLOT1 (1:2000)	Proteintech	# 15571-1-AP
Rabbit anti FLOT1 (1:200)	Cell signaling	# 18634
Rabbit anti GNB4 (1:2000)	Proteintech	# 11978-2-AP
Rabbit anti LAMP2 (1:400)	Thermo Fisher Scientific	# PA1-655
Rabbit anti LAMTOR1 (1:1000)	Sigma-Aldrich	# HPA002997
Rabbit anti LPHN1 (1:200)	Thermo Fisher Scientific	# PA5-77475
Rabbit anti LPHN2 (1:100)	Novus biologicals	#NBP2-58704
Rabbit anti LPHN3 (1:200)	Novus biologicals	#NLS1138
Rabbit anti RRAGA (1:1000)	Cell signalling	# 4357
Rabbit anti SDHA (1:800)	Proteintech	# 14865-1-AP
Rabbit anti TUBA (1:2000)	Rockland	# 600-401-880
Rat anti FLAG-HRP coupled (1:10,000)	Sigma-Aldrich	# SAB4200119
Chemicals, Peptides, and Recombinant Proteins		
2-[4-(2-Hydroxyethyl)-1-piperazinyl]-ethanesulfonic acid (HEPES)	Merck	# 110110
3xFLAG peptide	Sigma-Aldrich	# F4799
Acrylamide/bis-acrylamide	Merck	# A3669
Ammonium persulfate (APS)	Merck	# A3678
Calcium chloride ($CaCl_2$)	Carl Roth Gmbh	# 5239.1
Chloracetamide (CAA)	Thermo Fisher Scientific	# 79-07-2
cOmplete, EDTA-free protease inhibitor cocktail (PIC)	Merck	# 04693132001

Coomassie Brilliant Blue G-250	Thermo Fisher Scientific	# 800830
Dithiothreitol (DTT)	Merck	# 10197777001
Ethylenediaminetetraacetic acid (EDTA)	Merck	# E9884
Formic acid (FA)	Thermo Fisher Scientific	# 13454279
Glycerol	Merck	# 104094
L-Glutamine	Thermo Fisher Scientific	# 25030024
Magnesium acetate (MgAc)	Merck	# M5661
Magnesium chloride ($MgCl_2$)	Sigma-Aldrich	# 7786-30-3
Methanol	Carl Roth Gmbh	# 8388.6
Methanol (MeOH)	Merck	# 1060071000
PageBlue protein staining solution	Thermo Fisher Scientific	# 24620
Penicillin-Streptomycin (10,000 U/mL)	Thermo Fisher Scientific	# 15140122
Phenylmethylsulfonyl fluoride (PMSF)	Thermo Fisher Scientific	36978
Poly-L-lysine hydrobromide (PLL)	Sigma-Aldrich	# P1524
rLysC	Promega	# V1671
ROTI Mount FluorCare DAPI	Carl Roth GmbH	# HP20.1
Sodium deoxycholate (SDC)	Thermo Fisher Scientific	# 89904
Sodium Dodecyl Sulfate (SDS)	Merck	# 11667289001
Tetramethylethylenediamine (TEMED)	Carl Roth GmbH	# 2367.1
Triethylamine	AppliChem GmbH	# 471283
Triethylammonium bicarbonate (TEAB)	Sigma-Aldrich	# T7408
Trifluoroacetic acid (TFA)	Biosolve	# 202341
Tris(hydroxymethyl)aminomethane (TRIS)	Sigma-Aldrich	# 252859
Triton X-100	Sigma-Aldrich	# T8787
Trypsin	Promega	# V5111
Critical Commercial Assays		
DC-protein assay kit	BioRad	# 5000111
Dulbecco's modified Eagle medium (DMEM)	Gibco	# 11960044
Dulbecco's phosphate-buffered saline (DPBS)	Gibco	# 14190144
Enhanced chemiluminescence (ECL) kit	BioRad	# 1705060
Fetal calf serum (FCS)	Gibco	# 10270106
Fluorometric peptide assay kit	Thermo Fisher Scientific	# 23290
Silver Stain Kit	Thermo Fisher Scientific	# 24612
Deposited Data		
Cross-link MS data		http://proteomecentral.proteomexchange.org; ProteomeXchange: PXD030532 Reviewer account details: Username: reviewer_pxd030532@ebi.ac.uk Password: 26lOifDq

Structural Models (PPT1 and FLOT1-FLOT2)		Supplementary fike (ZIP) – Includes all relevant models (PDBs), docking parameters and modeling output
Unprocessed gel images presented in this manuscript can be found at		Additional proofs
Experimental Models: Cell lines		
HEK293	ATCC	# CRL-1573
HeLa	ATCC	# CCL-2
Recombinant DNA		
Plasmid: Human flotillin 1 Myc-DDK-tagged	OriGene	# RC200231
Plasmid: Human flotillin 2 Myc-DDK-tagged	OriGene	# RC220884
Softwares and algorithms		
BLAST (2.9.0)	(Camacho et al., 2009)	https://ncbiinsights.ncbi.nlm.nih.gov/2019/04/02/blast-2-9-0-now-available-with-enhanced-support-for-new-database-format-and-improved-performance/
ColabFold (1.2.0)	(Mirdita et al., 2021)	https://colabfold.mmseqs.com
Common Repository of Adventitious Proteins (cRAP database; January 2012)	GPM	https://www.thegpm.org/crap/index.htmL
Crystallography and NMR system (CNS) (1.3)	(Brunger, 2007)	http://cns-online.org/v1.3/
Cytoscape (3.8.0)	(Shannon, 2003)	https://cytoscape.org
FusionCapt advance (0.84)	Vilber Lourmat	https://www.vilber.com/fusion-fx/
Ggplot2 (3.3.2)	(Wickham, 2019)	https://www.tidyverse.org/blog/2020/03/ggplot2-3-3-0/
GraphPad Prism (6.01)	GraphPad	https://www.graphpad.com/scientific-software/prism/
HADDOCK (2.2)	(Van Zundert et al., 2016)	https://haddock.science.uu.nl/services/HADDOCK2.2/haddock.php
Mascot (2.5.1)	Matrix Science	http://www.matrixscience.com/
MaxQuant (2.0.3.0)	(Cox and Mann, 2008)	https://www.maxquant.org
Openxlsx (4.1.5)	Open source	https://www.rdocumentation.org/packages/openxlsx/versions/4.2.3
PCOILS	(Zimmermann et al., 2018)	https://toolkit.tuebingen.mpg.de/tools/pcoils
Proteome discoverer (2.4)	Thermo Fisher Scientific	https://www.thermofisher.com/order/catalog/ Thermo Fischer product/OPTON-30946#/OPTON-30946
PSIPRED	(Jones, 1999)	http://globin.bio.warwick.ac.uk/psipred/
PyMol (2.3)	Schrödinger, LLC	https://pymol.org/2/
R (4.0.2)	R Core Team, 2020	https://www.r-project.org
RStudio (1.3.1056)	R Core Team, 2020	https://www.rstudio.com/products/rstudio/download/
SCWRL (4.0)	(Krivov et al., 2009)	http://dunbrack.fccc.edu/SCWRL3.php/
Spectronaut (14.7.20)	Biognosys	https://biognosys.com/shop/spectronaut
STRING (11.0)	(Szklarczyk et al., 2019)	https://string-db.org
SWISS-MODEL (September 2021)	(Waterhouse et al., 2018)	https://swissmodel.expasy.org

Tidyverse (1.3.0)	Open source	https://www.tidyverse.org/blog/2019/11/tidyverse-1-3-0/
TopoLink (05.06.2019)	(Ferrari et al., 2019)	http://leandro.iqm.unicamp.br/topolink/lovoalign.shtml
UniProt (May 2019)	Open source	https://www.uniprot.org
Viridis (0.5.1).	Open source	http://guix.gnu.org/packages/r-viridis-0.5.1/
xiVIEW	(Graham et al., 2019)	https://www.rappsilberlab.org/software/xiview/
XlinkX node Proteome Discoverer (2.4)	(Klykov et al., 2018) Thermo Fisher Scientific	https://www.thermofisher.com/order/catalog/ Thermo Fischer product/OPTON-30946#/OPTON-30946; https://www.hecklab.com/software/xlinkx/
Zen (3.4)	Carl Zeiss microscopy GmbH	https://www.zeiss.de/mikroskopie/produkte/mikroskopsoftware/zen.html
Other		
Anti-FLAG beads	Thermo Fisher Scientific	# A36798
Anti-FLAG beads (magnetic)	Thermo Fisher Scientific	# A36797
DexoMAG40 (SPIONs)	Liquids Research LTD	# BKS25
DynaMag 2 magnet	Thermo Fisher Scientific	# 12321D
Empore C_{18} reversed-phase material	3M Bioanalytical Technologies	# 2215
LS columns	Miltenyi Biotech	# 130-042-401
Nitrocellulose membrane (0.45 µM)	Amersham	# GE10600002
Orbitrap Fusion Lumos mass spectrometer	Thermo Fisher Scientific	n.a.
P-2000 laser puller (Sutter Instruments)	Sutter Instruments	n.a.
PolySULFOETHYL A column (150x1.0 mm/5 µm particles)	PolyLC	# 101SE05
Polyvinylidene fluoride (PVDF) membrane (0.45 µM)	Amersham	# GE10600121
Protein A agarose beads	Thermo Fisher Scientific	# 15918014
QuadroMACS magnet	Miltenyi Biotech	# 130-090-976
ReproSil-Pur C_{18} AQ, 1.9 µm particles	Dr. Maisch	# r119.aq
Sep-Pak C18 cartridges	Waters	# WAT054955
TurboFect	Thermo Fisher Scientific	# R0531
UltiMate 3000 RSLCnano UHPLC system	Thermo Fisher Scientific	n.a.

METHOD DETAILS

Cell culture and enrichment of lysosomes

Tissue culture plates (10 cm) were coated with 0.5 mg/mL poly-L-lysine (PLL) in 1x phosphate-buffered saline (PBS) for 20 min at 37 °C. On each plate, 6 x 10^6 HEK293 cells were seeded in full medium supplemented with 10 % (v/v) superparamagnetic iron oxide nanoparticles (SPIONs) solution (DexoMAG40) and incubated for 24 h. Subsequently, cells were washed three times with pre-warmed PBS, fresh full medium was added, and cells were incubated for 24 h. Prior to harvesting, cells were washed twice with ice-cold PBS, scraped off the plates in 2 mL each of ice-cold isolation buffer (250 mM sucrose, 10 mM HEPES-NaOH pH 7.4, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 1.5 mM MgAc, 1x cOmplete EDTA-free protease-inhibitor cocktail), and pooled. Cell suspensions from four plates each were dounced with 25 strokes in a 15 mL dounce homogenizer, and nuclei as well as intact cells were pelleted by centrifugation for 10 min at 600 x g, 4 °C. The supernatant (post-nuclear supernatant, PNS) was transferred to a new tube, the pellet resuspended in 3 mL of isolation buffer, and dounced and centrifuged again. The supernatant from this step was combined with the first one, and the pooled PNS was used for lysosome enrichment using LS columns in combination with a QuadroMACS magnet. Columns were equilibrated with 1 mL 0.5 % (w/v) bovine serum albumin (BSA) in PBS, the combined PNS of two cell culture plates was applied to one column, and the flow through was collected. After three washing steps with 1 mL isolation buffer each, columns were removed from the magnet, and lysosomes were eluted twice in 1 mL of isolation buffer using a plunger. Individual eluate fractions were centrifuged for 30 min at 20,000 x g, 4 °C, the supernatants were discarded, the pellets were resuspended in isolation buffer, and for each biological replicate, the pellets from 64 plates were pooled. Protein concentrations were determined using the DC protein assay. The efficiency of lysosome enrichment and lysosomal integrity was assessed using the β-hexosaminidase assay (Wendeler, 2009). Fractions obtained from lysosome enrichment (25 µL each) were combined with 8 µL of 10 % Triton X-100 or 8 µL of PBS, followed by the addition of 50 µL reaction solution (100 mM sodium citrate pH 4.6, 0.2 % (w/v) BSA, 10 mM para-nitrophenyl-N-acetyl-2-B-D-glucosaminide) in a 96 well plate format. Subsequently, the plate was incubated for 15 min at 37 °C, and 200 µL of stop solution (0.4 M glycine-HCl, pH 10.4) was added to the sample. Absorbance was measured at 405 nm on a microplate reader.

Transfection of cells and enrichment of FLOT1-/FLOT2-positive early endosomes

64 tissue culture plates were coated with PLL (0.5 mg/mL), HEK293 cells were seeded at a density of 3.5 x 10^6 cells/plate, and cultivated as described above. After 24 h, cells were transfected with 6 µg of plasmid (1:1 mixture of FLOT1-FLAG and FLOT2-FLAG or GFP) using TurboFect. After 4 h, the cell culture medium was replaced, and cells were incubated for 48 h. For enrichment of endosomes, SPIONs solution was added to the cells (10 % (v/v) final concentration) for a pulse period of 5 min. Subsequently, cells were washed with pre-warmed PBS and fresh full medium was added. Following a 5 min chase, the plates were placed on ice and endosomes were enriched following the same procedure as for the enrichment of lysosomes (see above). For IP of FLOT1-FLAG/FLOT2-FLAG-positive endosomes, obtained eluate fractions (1 mL each) were subsequently incubated with magnetic anti-FLAG beads (80 µL) on an end-over-end rotator at 4 °C for 4 h. Beads were separated from samples by magnetic force (DynaMag 2 magnet), supernatants were transferred to new tubes, and the beads were washed three times with 500 µL PBS. For elution, beads were incubated with 150 µL of 150 ng/µL 3x FLAG peptide in 1x TRIS (hydroxymethyl) aminomethane buffered-saline (TBS, pH 7.6) for 30 min at 600 x g, 4 °C in a thermomixer. Subsequently, beads were separated from samples by magnetic force and eluate fractions transferred to a new tube. Protein concentrations were determined using the DC protein assay.

Cross-linking of samples

For lysosome- and early endosome-enriched fractions, a portion of the eluate containing 500 µg and 200 µg protein, respectively, was transferred to a new tube. Intact organelles were pelleted by centrifugation for 30 min at 20,000 x g, 4 °C, the supernatant discarded, and the pellets resuspended in isolation buffer at a protein concentration of 1 mg/mL. Lysosomal samples were cross-linked in two states (intact and disrupted) while endosomal samples were cross-linked only in the intact state. For disruption of lysosomes, resuspended samples were lyzed with a sonicator (Bioruptor Plus) at an amplitude of 40 with 3 cycles of 30 sec each. All samples were cross-linked at final DSSO concentrations of 5, 2, 1, 0.5, 0.25 mM for investigation of the optimal cross-linker concentration (DSSO titration). For the XL-LC-MS/MS experiments of both lysosome- as well as endosome-enriched fractions, a final concentration of 0.25 mM DSSO was applied. After addition of DSSO, the cross-link reaction was allowed to proceed for 30 min at room temperature, and quenched by the addition of TRIS-HCl pH 8.0 (20 mM final concentration). Subsequently, proteins were precipitated by addition of acetone at a ratio of 4:1 (v/v) and incubation overnight at -20 °C. The next day, samples were centrifuged for 20 min at 20,000 x g, 4 °C, the supernatant was discarded, the pellet washed twice with ice-cold acetone, air-dried, and stored at -80 °C until further use.

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

Polyacrylamide gels were prepared in-house. Both running and stacking gels were prepared with 10 % (w/v) SDS, 40 % (v/v) acrylamide, 10 % (w/v) ammonium persulfate (APS), and 1 % (v/v) tetramethylethylenediamine (TEMED), while 1.5 M TRIS-HCl pH 8.8 and 0.5 M TRIS-HCl pH 6.8 was used for running and stacking gels, respectively. Laemmli buffer (Laemmli, 1970) (4x stock, 240 mM TRIS-HCl pH 7.4, 4 % (v/v) β-mercaptoethanol, 8 % (w/v) SDS, 40 % (v/v) glycerol, 4 % (w/v) bromophenol blue) was added to samples (1x final concentration) followed by incubation for 10 min at 56 °C. Gel electrophoresis was performed at 80-140 V for up to 1.5 h. Gels were either stained overnight with Coomassie Brilliant Blue G-250 or with a Silver Stain Kit.

Immunoprecipitation of endogenous proteins

HEK293 cells were seeded at a density of 6 x 10^6 cells/10 cm plate and cultivated for 48 h in full medium. Cells were washed twice with ice-cold PBS, scraped off the plate in 1 mL of ice-cold PBS, transferred to a microtube, and centrifuged for 10 min at 600 x g, 4 °C. The supernatant was discarded and the pellet resuspended in 300 µL of RIPA lysis buffer (50 mM TRIS-HCl pH 7.4, 150 mM NaCl, 1 % Triton X-100, 0.1 % (w/v) SDS, 0.5 % (w/v) sodium deoxycholate (SDC), 1x cOmplete EDTA-free protease-inhibitor cocktail, 1 mM EDTA). Samples were incubated on ice for 30 min and passed through a 25 Gauge needle every 10 min. Subsequently, the lysate was cleared by centrifugation for 15 min at 20,000 x g, 4 °C, transferred to a new pre-cooled microtube, and protein concentrations were determined using the DC protein assay. For each sample, lysate containing 1.6 mg of protein was incubated with 3 µg of antibody overnight by end-over-end rotation at 4 °C. The next morning, 60 µL of Protein A beads were added to each sample, followed by end-over-end incubation for 1 h at 4 °C. Beads were pelleted by centrifugation for 5 min at 1000 x g, 4 °C, supernatants transferred to new tubes, and beads were washed three times with 500 µL of ice-cold PBS. Proteins were eluted from the beads by incubation in 2x Laemmli buffer for 30 min at 45 °C.

Blue native polyacrylamide gel electrophoresis (BN-PAGE)

Gradient gels (4-13 %) were prepared with a gradient mixer and overlaid with a stacking gel (Table 1). Samples were supplemented with solubilization buffer (10 mM HEPES-NaOH pH 7.4, 1 % (v/v) digitonin, 2 mM EDTA, 50 mM NaCl, 10 % (v/v) glycerol, 1 mM phenylmethylsulfonyl fluoride (PMSF)) and loading dye (10 mM Bis-TRIS-HCl pH 7.0, 0.5 % (w/v) Coomassie Brilliant Blue G-250, 50 mM ε-amino n-caproic acid), and loaded to the gel placed in a pre-cooled gel chamber. The anode buffer was 50 mM Bis-TRIS-HCL pH 7.0 and the cathode buffer 50 mM Tricine pH 7.0, 15 mM Bis-TRIS-HCl pH 7.0, 0.2 % (w/v) Coomassie Brilliant Blue G-250. The temperature of the gel chamber was maintained at 4 °C and electrophoresis was performed at 50 V for 20 h.

Table 1: BN-PAGE gel setup

	Separating gel		**Stacking gel**
Component	**4 %**	**13 %**	**3,5 %**
3x gel buffer (55 mM Imidazole, 1.5 M 6-Amino n-caproic acid) [mL]	2	2.5	1
Acrylamide/bis-acrylamide AB-3 mix (49.5 % T, 3 % C) [mL]	0.5	1.9	0.2
Glycerol [g]	-	1.5	-
10 % (w/v) APS [µL]	33.3	37.5	25
TEMED [µL]	3.3	3.7	2.5
Water [mL]	3.5	1.5	1.7
Total [mL]	6	7.5	3

Western blotting

Proteins were transferred to nitrocellulose or polyvinylidene fluoride (PVDF) membranes using a semi-dry or wet electro blotter for 1 h or 2 h at 200 mA/membrane. Membranes were blocked in 5 % non-fat dry milk in TBS containing 0.05 % (v/v) Tween 20 (TBS-T) for 1 h at RT followed by incubation with primary antibodies overnight at 4 °C (for dilutions of individual antibodies see key resource table). The next day, membranes were washed three times with TBS-T for 10 min at RT followed by incubation with secondary antibody (for dilutions of individual antibodies see key resource table) for 60 min at RT. Subsequently, membranes were washed three times for 10 min at RT with TBS-T, and the blots were developed. Protein expression signals were detected using the enhanced chemiluminescence (ECL) kit, visualized with the FUSION SOLO 4M system, and analyzed by the FusionCapt advance software.

Immunofluorescence microscopy

HeLa and HEK293 cells were seeded in 12-well plates at a density of 2 x 10^4 and 40 x 10^4 cells per well, respectively. For HEK293 cells, glass coverslips were coated using PLL (0.5 mg/mL). Cells were cultured for 36 h after seeding. Cells were transfected with 1 µg of plasmid (1:1 mixture of FLOT1-FLAG and FLOT2-FLAG) using TurboFect. After 4 h, the medium was replaced with fresh full medium and cells were incubated for 48 h. For staining, cells were washed using PBS and fixed with ice-cold methanol at -20 °C for 20 min. After fixation, cells were washed twice with PBS and blocked with 2 % BSA in PBS for 1 h at RT. Blocked cells were stained with primary (for dilutions of individual antibodies see key resource table) overnight at 4 °C in a humid chamber. Subsequently, cells were washed three times with TBS for 5 min each and incubated with secondary antibodies (for dilutions of individual antibodies see key resource table) for 1 h at RT in the dark. Coverslips were washed three times with TBS for 5 min each, rinsed once with distilled water, and mounted on specimen slides using ROTI Mount FluorCare DAPI. Images were acquired using an Axiovert 200 M microscope equipped with an AxioCam705 camera. Images were prepared using the ZEN software.

Mass spectrometry sample preparation

Precipitated proteins were resuspended in 100 µL of freshly prepared 8 M urea/100 mM triethylammonium bicarbonate (TEAB) and incubated for 45 min at 600 rpm, 37°C. Disulfide bridges were reduced with 4 mM DTT (final concentration) at 56 °C for 30 min, alkylated with 8 mM chloroacetamide (final concentration) at RT for 30 min (Müller and Winter, 2017), and the reaction was quenched by the addition of 4 mM DTT. Subsequently, samples were diluted 1:1 with 100 mM TEAB, rLysC was added at an enzyme-to-protein ratio of 1:100 (w/w), and proteolytic digestion was performed at 37 °C overnight. The following day, the urea concentration was reduced to 1.6 M by addition of 100 mM TEAB, trypsin was added at an enzyme-to-protein ratio of 1:100 (w/w), and the samples were incubated at 37 °C for 8 h. The resulting peptides were desalted using 50 mg Sep-Pak C_{18} cartridges, dried using a vacuum centrifuge, and stored at -80 °C until further use.

Strong cation-exchange (SCX) chromatography fractionation

SCX fractionation was performed with an UltiMate 3000 RSLC HPLC chromatography system in combination with a PolySULFOETHYL A column (150 mm x 1 mm, 5 µm particle size). Desalted peptides (500 µg each) were reconstituted in 20 µL of SCX solvent A (20 % acetonitrile (ACN), 10 mM KH_2PO_4 pH 2.7), loaded to the analytical column with 100 % SCX solvent A, and eluted with increasing amounts of SCX solvent B (500 mM KCl, 20 % ACN, 10 mM KH_2PO_4 pH 2.7) at a flow rate of 50 µL/min. The gradient was as follows (adapted from (Klykov et al., 2018)) 0-42 min: 0-2 % B; 42-50 min: 2-3 % B; 50-60 min: 3-8 % B; 60-70 min: 8-20 % B; 70-80 min: 20-40 % B; 80-86 min: 40-90 % B; 86-90 min: 90 % B; 90-91 min: 0 % B; 91-120 min: 0 % B. Eluting peptides were collected with a fraction collector, individual fractions dried using a vacuum centrifuge, resuspended in 20 µL of 5 % ACN, 5 % formic acid (FA), and desalted using C_{18} STAGE-tips (Rappsilber et al., 2007). STAGE tip eluates were dried using a vacuum centrifuge, resuspended in 5 % ACN, 5 % FA, and peptide concentrations were determined with a quantitative fluorometric peptide assay.

Liquid chromatography electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS)

Dried peptides were resuspended in 5 % ACN, 5 % FA and analyzed using an UltiMate 3000 RSLCnano UHPLC system coupled to an Orbitrap Fusion Lumos mass spectrometer. From each sample (25 % of the total amount for individual SCX fractions, 1 µg of non-cross-linked samples) were loaded on a 50 cm C_{18} reversed-phase analytical column at a flow rate of 600 nL/min using 100 % solvent A (0.1 % FA in water). For analytical column preparation, fused silica capillaries (360 µm outer diameter, 100 µm inner diameter) were used to generate spray tips using a P-2000 laser puller. Tips were packed with 1.9 µm ReproSil-Pur AQ C_{18} particles to a length of 50 cm. Peptide separation was performed with 120 min (SCX fractions) and 240 min (non-cross-linked samples) linear gradients from 5-35 % solvent B (90 % ACN, 0.1 % FA) at a flow rate of 300 nL/min. MS1 spectra were acquired in the Orbitrap mass analyzer from m/z 375-1,575 at a resolution of 60,000. For peptide fragmentation, charge states from 3+ to 8+ for cross-linked samples and 2+ to 5+ for non-cross-linked samples were selected, and dynamic exclusion was defined as 60 sec and 120 sec for 120 min and 240 min gradients, respectively.

Cross-linked samples were either analyzed with an MS2-MS3-MS2 strategy (Liu et al., 2017) or with a stepped collision energy approach (Stieger, 2019), where ions with the highest charge state were prioritized for fragmentation. For both methods, MS2 scans were acquired at a resolution of 30,000 in the Orbitrap analyzer with a dynamic mass window. In case of stepped collision energy, peptides were fragmented using higher collision dissociation (HCD) with 21, 26, and 31 % normalized collision energy (NCE). For the MS2-MS3-MS2 fragmentation method, sequential collision-induced dissociation (CID) and electron-transfer dissociation (ETD) spectra were acquired for each precursor. The precursor isolation width was set to m/z 1.6 with standard automatic gain control and automatic maximum injection time. The NCE for CID-MS2 scans were set to 25 % and calibrated charge-dependent ETD parameters enabled. MS3 scans were triggered by a targeted mass difference of 31.9721 detected in the MS2 scan. The MS3 scan was performed in the ion trap part of the instrument with CID at 35 % NCE with a normalized automatic gain control (AGC) target of 300 %.

Data independent acquisition (DIA) analyses were performed with the same instrumental setup as described above. For each sample, 1 µg of peptides were loaded directly on a reversed-phase analytical column packed with 3 µm ReprosilPur AQ C_{18} particles to a length of 40 cm and separated with 120 min linear gradients. After acquisition of one MS1 scan 24 static window DIA MS2 scans were performed. MS1 scans were acquired in the Orbitrap analyzer from m/z 350-1200 at a resolution of 120,000 with a maximum injection time of 20 msec and an AGC target setting of 5 $x10^5$. MS2 scans were defined to cover the MS1 scan range with 36 scan windows of

24.1 m/z each, resulting in an overlap of 0.5 m/z and a cycle time of 3.44 sec. Peptides were fragmented by HCD with an NCE of 27 %, and spectra were acquired in the Orbitrap analyzer with a resolution of 30,000, a maximum injection time of 60 msec, and an AGC target setting of $1x10^6$.

Proteome Discoverer analysis

Thermo *.raw files from cross-linked samples were analyzed using Proteome Discoverer, utilizing Mascot and XlinkX for peptide identification. The following settings were used for both algorithms: precursor ion mass tolerance: 10 ppm; Orbitrap fragment ion mass tolerance: 20 ppm; ion trap fragment ion mass tolerance: 0.5 Da; fixed modification: carbamidomethylation at cysteine; variable modification: oxidation at methionine; enzyme: trypsin; number of allowed missed cleavage sites: 2; minimum peptide length: 5 amino acids; cross-linking site: lysine (K) and N-terminus of proteins. Data were searched against UniProt *Homo sapiens* (Entries: 20,365) in combination with the common repository of adventitious proteins (cRAP) database containing common contaminants. The Proteome Discoverer workflow was split into two branches with a cross-link and standard peptide search. MS2 spectra containing DSSO reporter ions were analyzed with pre-defined “MS2-MS2-MS3” and “MS2” search options using XlinkX. Peptide identifications were accepted with a minimum XlinkX score of 40 and filtered at false discovery rates (FDRs) of 1 % and 5 % at the cross-linked peptide spectrum level. Cross-links were exported. Spectra, which did not contain reporter ions were searched using Mascot. Identified peptides were filtered at 1 % FDR on the peptide level using Percolator and proteins exported at 1 % FDR. Only high confidence peptide identifications were considered and data exported. Data from both algorithms were further analyzed applying different software packages (R, Excel, GraphPad Prism, STRING, Cytoscape, xiVIEW, TopoLink, PSIPRED, PCOILS, and PyMol)

MaxQuant analysis

Thermo *.raw files from non-cross-linked samples were analyzed using MaxQuant (Cox and Mann, 2008) for determining iBAQ values (Schwanhausser et al., 2011). The following settings were used: precursor ion mass tolerance: 4.5 ppm; Orbitrap fragment ion mass tolerance: 20 ppm; fixed modification: carbamidomethylation at cysteine; variable modifications: oxidation at methionine, acetylation at protein N-terminus, and deamidation at asparagine (N) as well as glutamine (Q); enzyme: trypsin; number of allowed missed cleavage sites: 2; minimum peptide length: 5 amino acids. Data were searched against UniProt *Homo sapiens* (Entries: 20,365) in combination with the cRAP database containing common contaminants. Data was filtered at 1 % FDR on the peptide level and protein level and exported, followed by analysis with different software packages (Excel and GraphPad Prism).

Spectronaut analysis

Thermo *.raw DIA files from FLOT1-FLAG+FLOT2-FLAG-transfected HEK293 cells were analyzed using Spectronaut. Initially, hybrid spectral libraries were generated from both DDA and DIA files with the Pulsar search engine integrated into Spectronaut, applying the following parameters: precursor ion mass tolerance: dynamic; Orbitrap fragment ion mass tolerance: dynamic; fixed modification: carbamidomethylation at cysteine; variable modifications: oxidation at methionine, acetylation at protein N-terminus and deamidation at asparagine (N) as well as glutamine (Q); enzyme: trypsin; number of allowed missed cleavage sites: 2; minimum peptide length: 5 amino acids. Data were searched against UniProt *Homo sapiens* (Entries: 20,365) in combination with the cRAP database containing common contaminants. For each peptide, the 3 - 6 most abundant b/y ions were selected for library generation, dependent on their signal intensity. Dynamic retention time alignment was performed based on the high-precision indexed retention time (iRT) concept (Bruderer, 2016). Mass tolerances (precursor and fragment ions) as well as peak extraction windows were defined automatically by Spectronaut. Normalization was disabled and data filtered at 1 % FDR on the peptide and protein level (q-value < 0.01). High confidence identifications were exported, followed by analysis with different software packages (R, Excel, STRING, Cytoscape, and GraphPad Prism).

Structural analysis

Protein sequences were obtained from UniProt and searched using BLAST against protein database (PDB) entries with an E-value of 0.0001. In case no reference structure for *Homo sapiens* was available, structures from other organisms were obtained from the SWISS-MODEL repository (Waterhouse et al., 2018), and/or predicted structures from the AlphaFold protein structure database were used (Jumper et al., 2021). Amino acid numbering was adjusted to UniProt entries, identified cross-links were mapped and topologically evaluated with TopoLink (Ferrari et al., 2019), and visualized using PyMol.

Molecular docking

Protein structure perturbation and optimization was performed with SCWRL (Krivov et al., 2009) and restraint-based docking with HADDOCK (Van Zundert et al., 2016) as well as CNS (Brunger, 2007). Distance constraints of identified cross-links (20 ± 10 Å) were used to limit the possible interaction search space, applying unambiguous restraint distances on C-beta (except for glycine, for which C-alpha was used). In line with the default HADDOCK protocol, 500 initial restraints-based complex models were generated, followed by their rigid-body energy minimization. For the best 100 models, semi-flexible refinement in torsion angle space was performed, followed by molecular dynamics refinement in explicit water. Generated models were evaluated based on the weighted sum of electrostatic and van der Waals energies, complemented by the empirical desolvation energy. Based on these parameters, the ten best-scoring models were reported. Finally, models were further clustered within a 5 Å pairwise root mean square deviation, and the lowest energy model of each cluster reported. Results were visualized using PyMol.

QUANTIFICATION AND STATISTICAL ANALYSIS

Spectronaut results were analyzed using R (R Core Team, 2020) applying the integrated development environment RStudio. For data importing, tidying, transforming and statistical modeling Tidyverse (Wickham, 2019) and R base packages were used. Results were exported using Openxlsx and visualized by Ggplot and Viridis. Protein signal intensities were initially log2-transformed for data quality assessment and visualization. Missing values were replaced by "NA", while imputation of missing values was omitted. Subsequently, data from individual replicates of the experimental conditions (GFP, SPIONs, and SPIONs+IP) were categorized into three populations as follows: background (three valid values in all three conditions), SPIONs-specific (no valid GFP value and ≥ 2 valid values in SPIONs or SPIONs+IP) and SPIONs “on/off” (≥ 2 valid values in only one SPIONs condition and no valid value in other conditions). Subsequently, non-logarithmic data of the individual data sets were normalized on the signal intensities of FLOT1 and FLOT2 in the respective datasets, while the GFP sample was not normalized, followed by log2 transformation of all datasets. Proteins with ≥ 2 valid values in each dataset were compared using a two-sided unpaired t-test. On/off proteins were defined as significant and their p-values set to zero; while p-values of proteins not matching any of the two conditions were set to one. P-value-adjustment was performed according to Benjamini-Hochberg (Benjamini and Hochberg, 1995) and proteins with a q-value ≤ 0.05 were considered differently enriched. Significantly enriched proteins were submitted to protein-protein interaction and enrichment analysis with STRING (Szklarczyk et al., 2019) while the entire list of proteins found in the experiment was defined as background gene set. Networks (type: full) were generated, in which the edges indicate both functional and physical protein associations.

FIGURE TITLES AND LEGENDS

Figure 1: Cross-linking mass spectrometry analysis of lysosome-enriched fractions.
A) Experimental workflow for the XL-LC-MS/MS analysis of lysosome-enriched fractions.
B) Normalized β-hexosaminidase activities for individual fractions from lysosome enrichment by SPIONs. Shown are average values (n=3, +STDEV).
C) Western blot analysis of lysosome-enriched fractions for contamination by other organelles. Lysosome: lysosomal proteins (CTSD, LAMP2, LIMP2, and LAMTOR1). Other: Golgi-apparatus (GM130), cytoskeleton (TUBA), cytosol (GAPDH), endoplasmic reticulum (CANX), mitochondria (SDHA).
D) Summed iBAQ abundances for proteins identified in lysosome-enriched fractions in ≥ 3 replicates.
E) Classification of unique cross-linked residue pairs.
F) Proteins detected in non-cross-linked lysosome enriched fractions (proteome) and unique lysosomal cross-linked residue pairs (interactome) for DR and IT samples.
G) Localization of CSMs for 68 lysosomal proteins cross-linked in the DR and IT state. Cytosolic: proteins located at the cytosolic face of lysosomal membrane; Lumen: lysosomal luminal proteins.
H) Correlation of cross-link identification and protein abundance for lysosomal proteins. CSMs and PSMs represent summed values of the analysis of 6 replicates (DR and IT).
Abbreviations: SPIONs: superparamagnetic iron oxide nanoparticles; DR: disrupted; IT: intact; SCX: strong cation-exchange; IN: input; FT: flow through; W: wash; EL: eluate; WCL: whole cell lysate; iBAQ: intensity based absolute quantification; XL: cross-links; CSMs: cross-link spectral matches; PSMs: peptide spectral matches.
See also: Figure S1.

Figure 2: The human lysosomal interactome.
A) PPIs identified in the XL-LC-MS/MS dataset. Lysosomal proteins (blue dots), non-lysosomal proteins (grey dots), and PPIs (grey lines) are indicated.
B) Matching of PPIs to the STRING database.
C) Numbers of proteins from distinct subcellular localizations interacting with lysosomal proteins.
D/E) Interaction networks of the V-ATPase (D) and the flotillin (E) complex.
F) Co-IP of ATP6V1D and FZD9. ATP6V1B2 and ATP6V1A1 are known members of the V-ATPase complex; LAMP2 is a lysosomal membrane protein. Control: empty beads.
G) Co-IPs of FLOT1, FLOT2, and GNB4. RRAGA is a lysosomal membrane-associated protein. Control: empty beads.
H/I) Site frequency distribution for identified FLOT1 (H) and FLOT2 (I) cross-links. Site frequency represents the percentage of cross-links detected in bins of 20 residues each. The region indicated by red dots represents the PHB domain.
Abbreviations: CY: cytoplasm; CK: cytoskeleton; NC: nucleus; ER: endoplasmic reticulum; GA: Golgi apparatus; PM: plasma membrane; LS: lysosome; OT: others; NK: not known; IN: input; SN: supernatant; W: wash; EL: eluate; X: empty lane.
See also: Figure S2.

Figure 3: Cross-link-based structural refinement of the V-ATPase and PPT1 complex.
A) Matching of cross-links to the crystal structure of lysosomal alpha-galactosidase A (GLA).
B) Average iBAQ abundances for individual V-ATPase subunits based on the analysis of lysosome-enriched fractions in a non-cross-linked state.
C) Cross-links identified for individual V-ATPase subunits. Structurally resolved regions are colored grey, unresolved regions white.
D) Refined structure of the V-ATPase complex. Identified cross-links were integrated in the V-ATPase cryo-EM structure (Wang, 2020). Missing regions were supplemented with predicted structures from AlphaFold based on the identified cross-links.

E) Distances for mapped cross-links to crystal structures and AlphaFold models for lysosomal proteins. Bin size: 2.5 Å.
F) Reported homodimeric human PPT1 structure (PDB: 3GRO).
G/H) Tetrameric PPT1 model representing the most favorable energetic state and fulfilling DSSO's distance constraints for all 18 cross-links. A, B, C, and D indicate individual subunits.
Abbreviations: DR: disrupted; IT: intact; RS: resolved structure; AF: AlphaFold.
See also: Figure S3.

Figure 4: Proposed model for a heterodimeric FLOT1-FLOT2 complex.
A) Western blot analysis of lysosome-enriched fractions. CTSD is a lysosomal luminal and LAMP2, a lysosomal transmembrane protein.
B) Identified cross-links, predicted secondary structures (PSIPRED), and coiled-coil probabilities (PCOILS) for FLOT1 and FLOT2.
C) Heterodimeric FLOT1/FLOT2 interaction model representing the lowest energy state and satisfying the distance constraints of all cross-links. The model was generated using HADDOCK based on predicted monomeric AlphaFold structures.
Abbreviations: WCL: whole cell lysate; EL: eluate; PHB: prohibitin homology domain; SPFH: stomatin/PHB/flotillin/HflK/C domain; WND: window; EA: glutamic acid/alanine; PZD3: postsynaptic density protein-95/discs large/zonula occludens-1.
See also: Figure S4.

Figure 5: Investigation of higher-order flotillin assemblies in lysosome- and early endosome-enriched fractions.
A/B) Immunostaining of HeLa and HEK293 cells for FLOT1, FLOT2, and the lysosomal marker LAMP2.
C) Analysis of FLOT1/FLOT2 in early endosomes. SPIONs pulse treatment time course followed by endosome enrichment, chase times are indicated. Pre-CTSD serves as a marker for early endosomes.
D) Western blot analysis of early endosome fractions with marker proteins for different subcellular compartments: Cytosol (ACTG2 and GAPDH), cytoskeleton (TUBA), endoplasmic reticulum (CANX), and mitochondria (SDHA).
E) Experimental workflow for early endosome enrichment and XL-LC-MS/MS analysis.
F) Overview of early endosome cross-linking dataset.
G) Overlap of unique FLOT1/FLOT2 cross-links for XL-LC-MS/MS analyses of early endosome- and lysosome-enriched fractions.
H) Western blot analysis of BN-PAGE separated FLAG-IP eluates with/without cross-linking by DSSO. IN/EL refers to fractions of FLAG-IP.
I) Nominated lowest energy model for the FLOT1/FLOT2 heteromeric 38-mer.
Abbreviations: DSSO: disuccinimidyl sulfoxide; M: marker; IN: input; WA: wash; EL: eluate; WCL: whole cell lysate; SPIONs: superparamagnetic iron oxide nanoparticles; SCX: strong cation-exchange; LS: lysosome; EE early endosome; EA: glutamic acid/alanine; PZD3: postsynaptic density protein-95/discs large/zonula occludens-1.
See also: Figure S5.

Figure 6: Identification of potential cargo of FLOT1-/FLOT2-positive early endosomes.
A) Western blot analysis of FLAG-FLOT1/FLOT2 in early endosome-enrichment (SPIONs) and intact endosome-IP fractions.
B) Data dependent acquisition (DIA)-based protein abundance fold-change ratios of SPIONs/SPIONs+IP fractions (n=3). Significantly different proteins are indicated. Cut offs: q-value: 0.05, fold change: 1.5.
C) STRING-based PPI analysis of proteins overrepresented in SPIONs+IP fraction. Node size corresponds to DIA signal intensity and line thickness to PPI confidence score.

Abbreviations: SPIONs: superparamagnetic iron oxide nanoparticles; IP: immunoprecipitation; IN: input; SN: supernatant; FT: flow through; WA: wash; EL: eluate; PPI: protein-protein interaction.
See also: Figure S6.

Figure 7: Latrophilins are a putative cargo of FLOT1-/FLOT2-positive early endosomes.
A) Average DIA abundances for LPHN1, LPHN2, LPHN3, CLTA, EEA1, and RAB5C (n=3, +STDEV). Significance based on Student's unpaired two-sided t-test: $*$ ($p < 0.05$), $**$ ($p < 0.01$), $***$ ($p < 0.001$).
B) Co-immunostaining of HeLa cells for FLOT1, FLOT2, LPHN1, LPHN2, and LPHN3.
C) Co-IP of FLOT1 and FLOT2 with LPHN1, LPHN2, and LPHN3. LDLR serves as a control for plasma membrane and clathrin-mediated endocytosis.
Abbreviations: SP (SPIONs): superparamagnetic iron oxide nanoparticles; IP: immunoprecipitation; IN: input; WA: wash; EL: eluate.
See also: Figure S7.

Figure 1

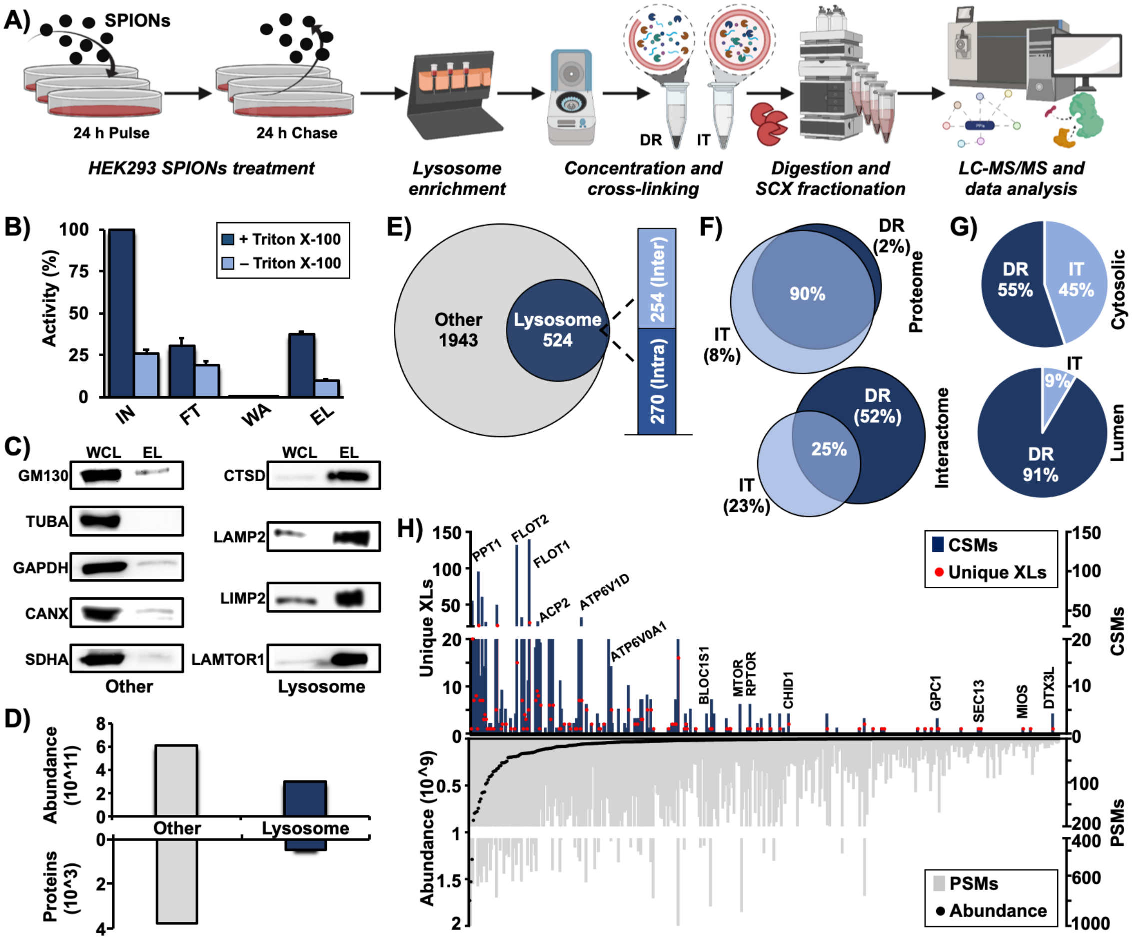

A)
SPIONs
24 h Pulse
24 h Chase
HEK293 SPIONs treatment
Lysosome enrichment
DR
IT
Concentration and cross-linking
Digestion and SCX fractionation
LC-MS/MS and data analysis
B)
Activity (%)
100
75
50
25
0
+ Triton X-100
– Triton X-100
IN
FT
WA
EL
C)
WCL
EL
GM130
TUBA
GAPDH
CANX
SDHA
Other
CTSD
LAMP2
LIMP2
LAMTOR1
Lysosome
D)
Abundance (10^11)
8
6
4
2
0
Other
Lysosome
Proteins (10^3)
2
4
E)
Other 1943
Lysosome 524
254 (Inter)
270 (Intra)
F)
DR (2%)
90%
IT (8%)
Proteome
DR (52%)
25%
IT (23%)
Interactome
G)
DR 55%
IT 45%
Cytosolic
IT
9%
DR 91%
Lumen
H)
Unique XLs
150
100
50
20
10
0
PPT1
FLOT2
FLOT1
ACP2
ATP6V1D
ATP6V0A1
BLOC1S1
MTOR
RPTOR
CHID1
GPC1
SEC13
MIOS
DTX3L
CSMs
Unique XLs
CSMs
Abundance (10^9)
0.5
1
1.5
2
PSMs
100
200
400
600
1000
PSMs
Abundance

Figure S1

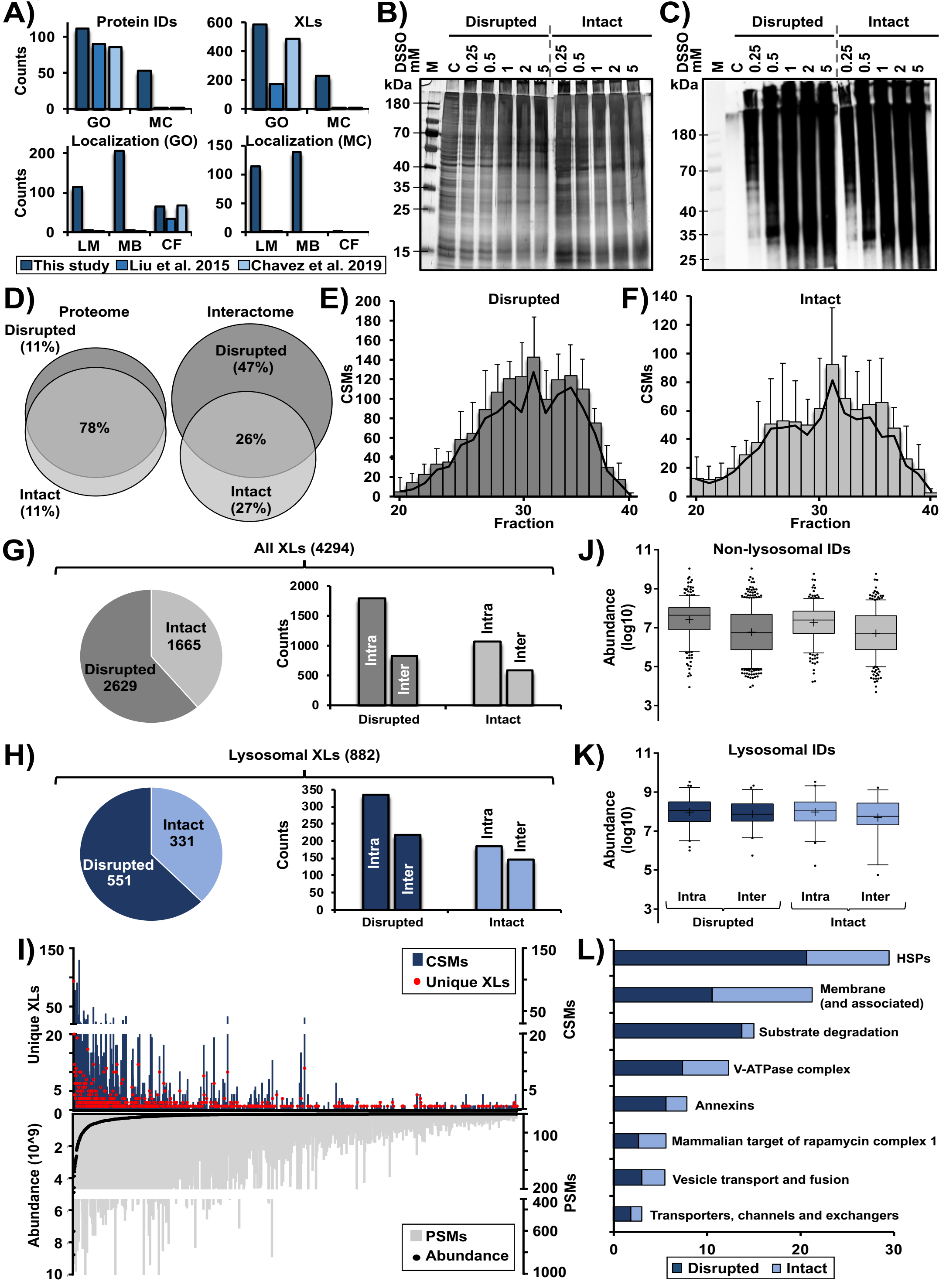

A)
Protein IDs
XLs
Counts
GO
MC
Localization (GO)
Localization (MC)
LM
MB
CF
This study
Liu et al. 2015
Chavez et al. 2019
B)
C)
DSSO mM
M
C
0.25
0.5
1
2
5
kDa
Disrupted
Intact
180
70
40
35
25
15
D)
Proteome
Interactome
Disrupted (11%)
78%
Intact (11%)
Disrupted (47%)
26%
Intact (27%)
E)
Disrupted
F)
Intact
CSMs
Fraction
G)
All XLs (4294)
Intact 1665
Disrupted 2629
Counts
Intra
Inter
Disrupted
Intact
H)
Lysosomal XLs (882)
Intact 331
Disrupted 551
J)
Non-lysosomal IDs
Abundance (log10)
K)
Lysosomal IDs
Intra
Inter
Disrupted
Intact
I)
Unique XLs
CSMs
Unique XLs
CSMs
Abundance (10^9)
PSMs
Abundance
PSMs
L)
HSPs
Membrane (and associated)
Substrate degradation
V-ATPase complex
Annexins
Mammalian target of rapamycin complex 1
Vesicle transport and fusion
Transporters, channels and exchangers
Disrupted
Intact

Figure S1: Cross-linking mass spectrometry analysis of HEK293 lysosome-enriched fractions. Related to Figure 1.

A) Identified cross-links for lysosomal and lysosome-related proteins. Proteins are categorized in two groups: GO: proteins which are related to lysosomes based on their categorization in gene ontology databases, also including proteins which are not important for lysosomal function; MC manually curated list of proteins which are localized in/at lysosomes (Thelen et al., 2017) and known to play a role in lysosomal function. Numbers from the published whole proteome cross-linking studies were extracted from the respective supplementary tables. Cross-links assigned to both GO and MC were further subcategorized based on the protein's location at the lysosome

B) Silver-stained SDS-PAGE gel of lysosome-enriched fractions cross-linked with the indicated amounts of DSSO in a disrupted and intact state at a protein concentration of 1 µg/µL. For each sample, 2.5 µg of protein were loaded to the gel. Control samples were treated with DMSO. Based on the band pattern, the optimal concentration of DSSO was determined to be 0.25 mM (final concentration), as the application of higher amounts resulted in a shift of band patterns. Especially with respect to the lower molecular weight region, less bands were observed, while the high molecular weight region (containing cross-linked proteins) increased in intensity, which is indicative for protein aggregates resulting from over-cross-linking. Shown are two separate SDS-PAGE analyses (performed under the same conditions) for samples cross-linked in the intact and the disrupted state, individual gels are indicated by a dashed grey line.

C) Western blot analysis of samples described in (B) for the visualization of cross-linker-modified proteins using an in-house produced antibody which shows immunoreactivity against water-quenched DSSO (Singh et al., 2021). For each sample, 1 µg of protein was loaded to the gel. Shown are two separate western blot analyses (performed under the same conditions) for samples cross-linked in the intact and the disrupted state, individual gels are indicated.

D) Overlap in proteins, which are not categorized as lysosomal based on their GO term, identified in mass spectrometric analyses of lysosome-enriched fractions in the disrupted and the intact state. Proteome: proteins identified in analyses of non-cross-linked samples; Interactome: unique cross-links identified in analyses of cross-linked samples.

E/F) Distribution of identified CSMs in LC-MS/MS analyses of SCX-fractionated tryptic digests of lysosome-enriched fractions cross-linked in a disrupted (E) and intact (F) state. Since later eluting SCX fractions contain the majority of higher charged cross-linked tryptic peptides, only SCX fractions 20-40 were analyzed by LC-MS/MS. Shown are average values (n=3, +STDEV).

G/H) Distribution of identified cross-links from the analysis of lysosome-enriched fractions cross-linked in the disrupted and the intact state. Shown are numbers identified in the disrupted/intact state, and how many cross-links were assigned as intra- or inter-links, for cross-links identified in the whole dataset (G) and such assigned to lysosomal proteins (H).

I) Correlation of cross-link identification and protein abundance for all non-lysosomal proteins. Proteins are sorted based on their average iBAQ abundance (Schwanhausser et al., 2011) in the LC-MS/MS analysis of non-cross-linked samples. Average iBAQ abundances as well as total numbers of unique cross-links, CSMs, and PSMs are shown. Values are based on the analysis of (n=3) each for lysosome-enriched fractions in the intact and disrupted state.

J/K) Distribution of iBAQ abundances for all (J) and lysosomal proteins (K) for which intra- or inter-links were detected. Values are based on the LC-MS/MS analysis of non-cross-linked lysosome-enriched fractions in the disrupted and the intact state. Only proteins for which cross-links were identified in the respective state were considered. Shown are combined values from three independent replicates, the median is indicated by a line and the average by "+".

L) Categorization of the 1,415 CSMs identified for 68 lysosomal proteins. Proteins were assigned to previously defined categories of lysosomal proteins (Akter, 2020).

Abbreviations: IDs: identifications; LM: lumen; MB: membrane; CF: cytosolic face; GO: gene ontology; MC: manually curated; DSSO: disuccinimidyl sulfoxide; DMSO: dimethyl sulfoxide; C: control (no DSSO); M: protein marker; iBAQ: intensity based absolute quantification; SCX: strong cation exchange; XL: cross-link; PSMs: peptide spectral matches; CSMs: cross-link spectral matches; HSPs: heat shock proteins.

Figure 2

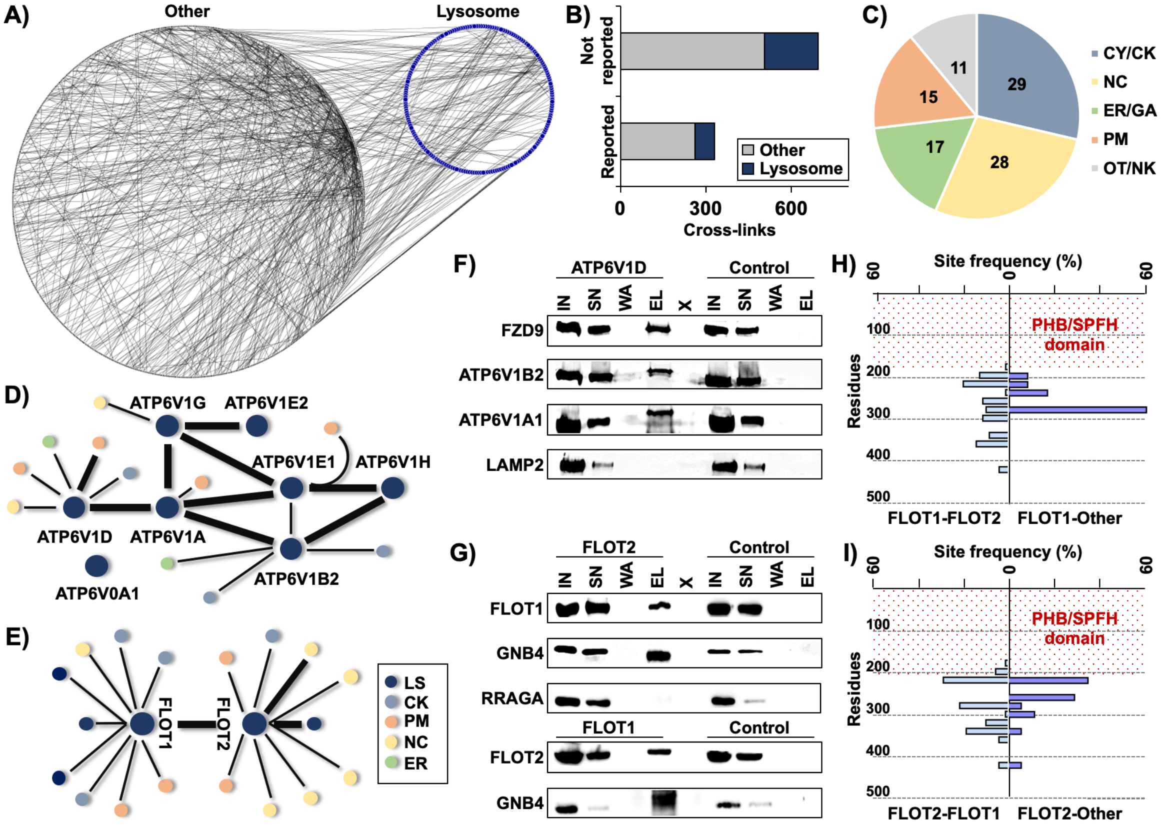

A)
Other
Lysosome
B)
Not reported
Reported
Other
Lysosome
0
300
600
Cross-links
C)
29
28
17
15
11
CY/CK
NC
ER/GA
PM
OT/NK
D)
ATP6V1G
ATP6V1E2
ATP6V1E1
ATP6V1H
ATP6V1D
ATP6V1A
ATP6V1B2
ATP6V0A1
E)
FLOT1
FLOT2
LS
CK
PM
NC
ER
F)
ATP6V1D
Control
IN
SN
WA
EL
X
IN
SN
WA
EL
FZD9
ATP6V1B2
ATP6V1A1
LAMP2
G)
FLOT2
Control
IN
SN
WA
EL
X
IN
SN
WA
EL
FLOT1
GNB4
RRAGA
FLOT1
Control
FLOT2
GNB4
H)
Site frequency (%)
60
0
60
Residues
100
200
300
400
500
PHB/SPFH domain
FLOT1-FLOT2
FLOT1-Other
I)
Site frequency (%)
60
0
60
Residues
100
200
300
400
500
PHB/SPFH domain
FLOT2-FLOT1
FLOT2-Other

Figure S2

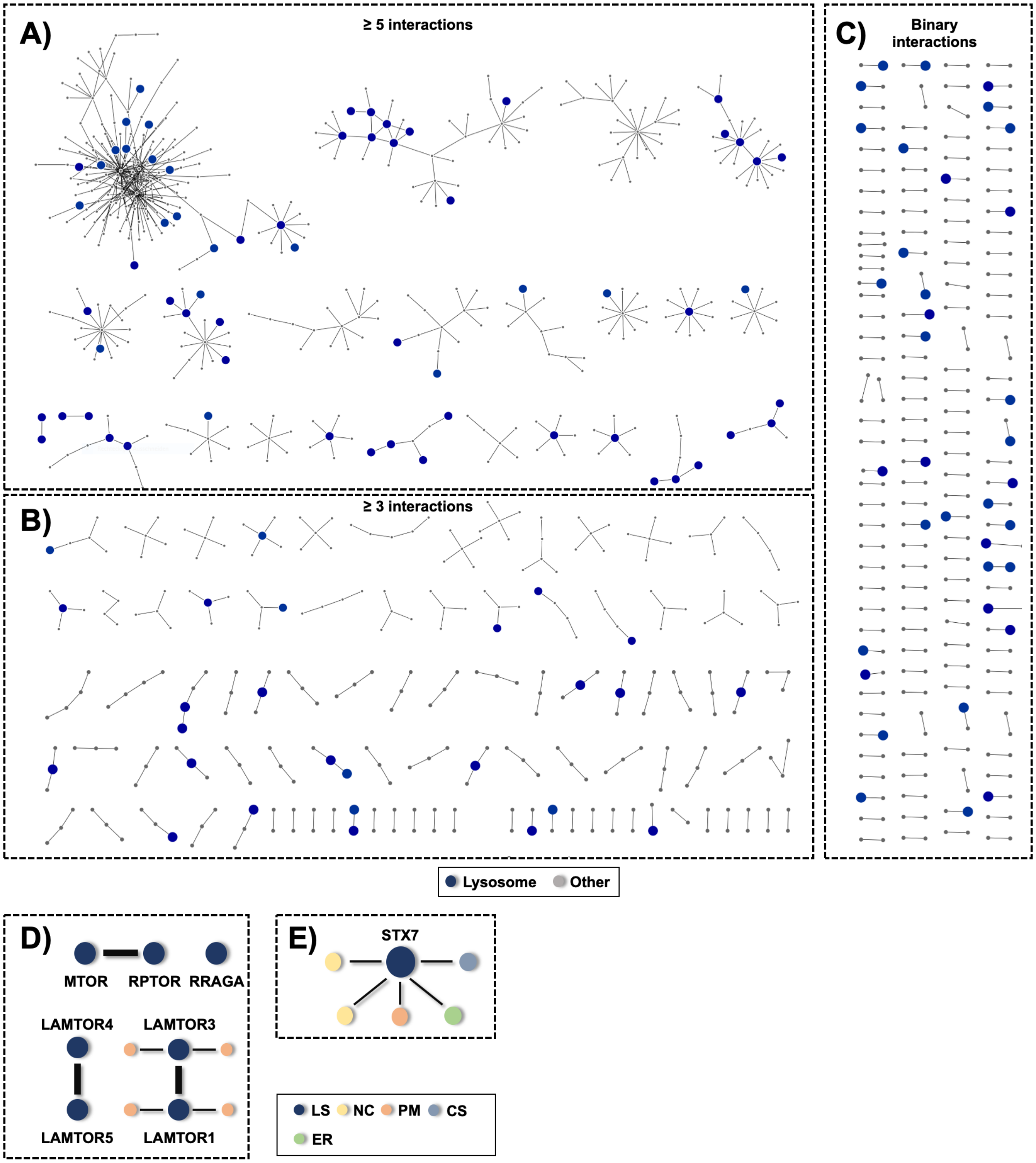

A)
≥ 5 interactions
B)
≥ 3 interactions
C)
Binary interactions
● Lysosome
● Other
D)
MTOR
RPTOR
RRAGA
LAMTOR4
LAMTOR3
LAMTOR5
LAMTOR1
E)
STX7
● LS
● NC
● PM
● CS
● ER

Figure S2: Analysis of protein-protein interactions in lysosome-enriched fractions. Related to Figure 2.

A-C) Protein-protein interaction networks based on cross-links identified between two different proteins (inter-links, 1,023 in total). Lysosomal proteins are highlighted as large blue filled circles while non-lysosomal proteins are depicted as small grey dots. Interactions were extracted from the entire XL-LC-MS/MS dataset from disrupted and intact lysosomes. Based on the number of interactors involved in a particular network, data were grouped into different classes: A) ≥ 5 interactors, B) ≥3 interactors, C) binary interactions. The networks were generated using Cytoscape.

D) Protein-protein interaction networks for proteins related to mTORC1; subcellular localization of interactors is indicated by color. The networks were generated using Cytoscape.

E) Protein-protein interactions identified for Syntaxin 7 (STX7), subcellular localizations of interactors are indicated by color. The network was generated using Cytoscape.

Abbreviations: LS: lysosome; NC: nucleus; PM: plasma membrane; CS: cytoskeleton; ER: endoplasmic reticulum.

Figure 3

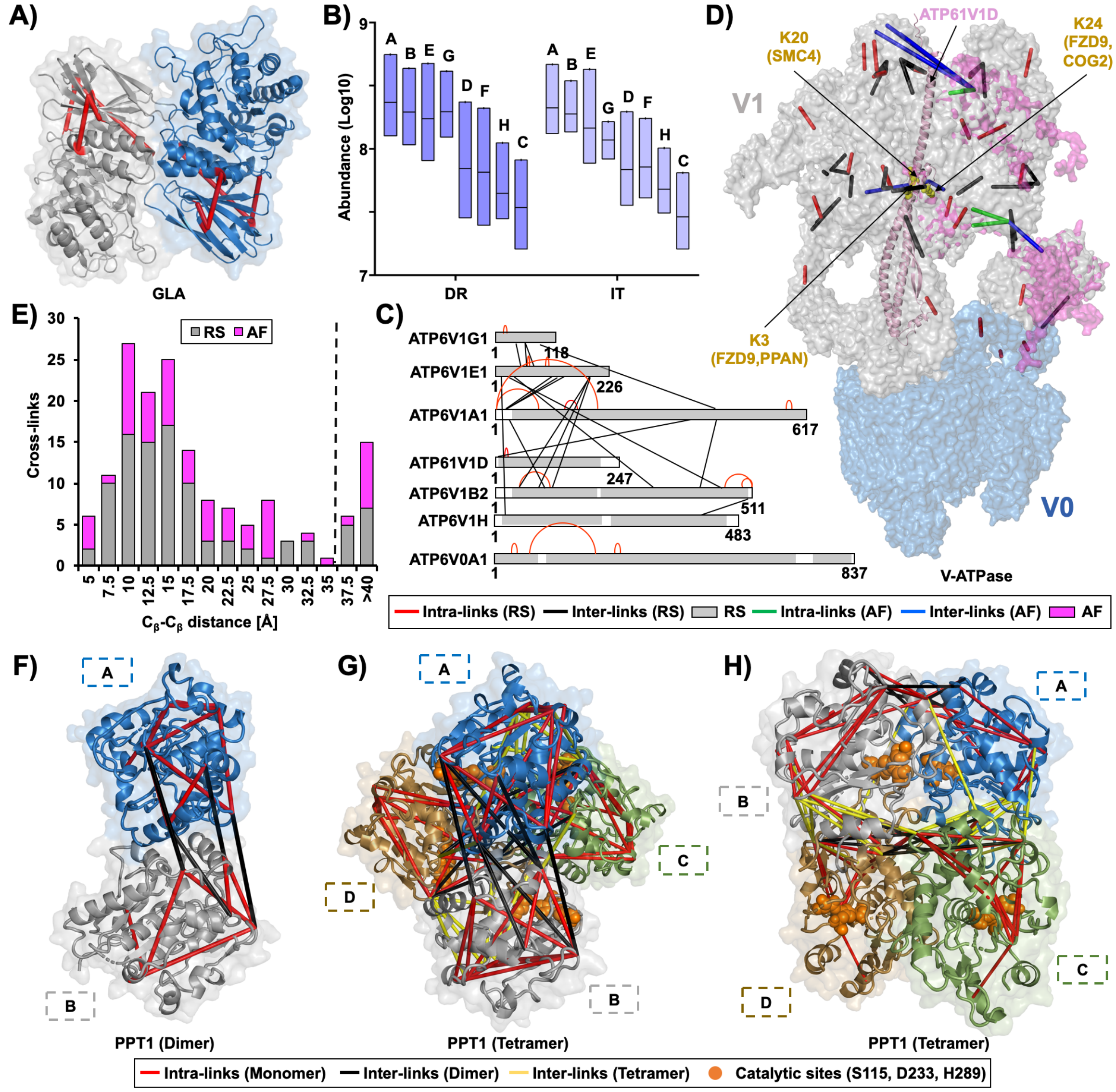

A)
GLA
B)
Abundance (Log10)
9
8
7
A B E G D F H C
A B E G D F H C
DR
IT
D)
ATP61V1D
K20
(SMC4)
K24
(FZD9,
COG2)
V1
K3
(FZD9,PPAN)
V0
V-ATPase
E)
Cross-links
30
25
20
15
10
5
0
RS
AF
5
7.5
10
12.5
15
17.5
20
22.5
25
27.5
30
32.5
35
37.5
>40
C_β-C_β distance [Å]
C)
ATP6V1G1
1
118
ATP6V1E1
1
226
ATP6V1A1
1
617
ATP61V1D
1
247
ATP6V1B2
1
511
ATP6V1H
1
483
ATP6V0A1
1
837
Intra-links (RS)
Inter-links (RS)
RS
Intra-links (AF)
Inter-links (AF)
AF
F)
A
B
PPT1 (Dimer)
G)
A
B
C
D
PPT1 (Tetramer)
H)
A
B
C
D
PPT1 (Tetramer)
Intra-links (Monomer)
Inter-links (Dimer)
Inter-links (Tetramer)
Catalytic sites (S115, D233, H289)

Figure S3

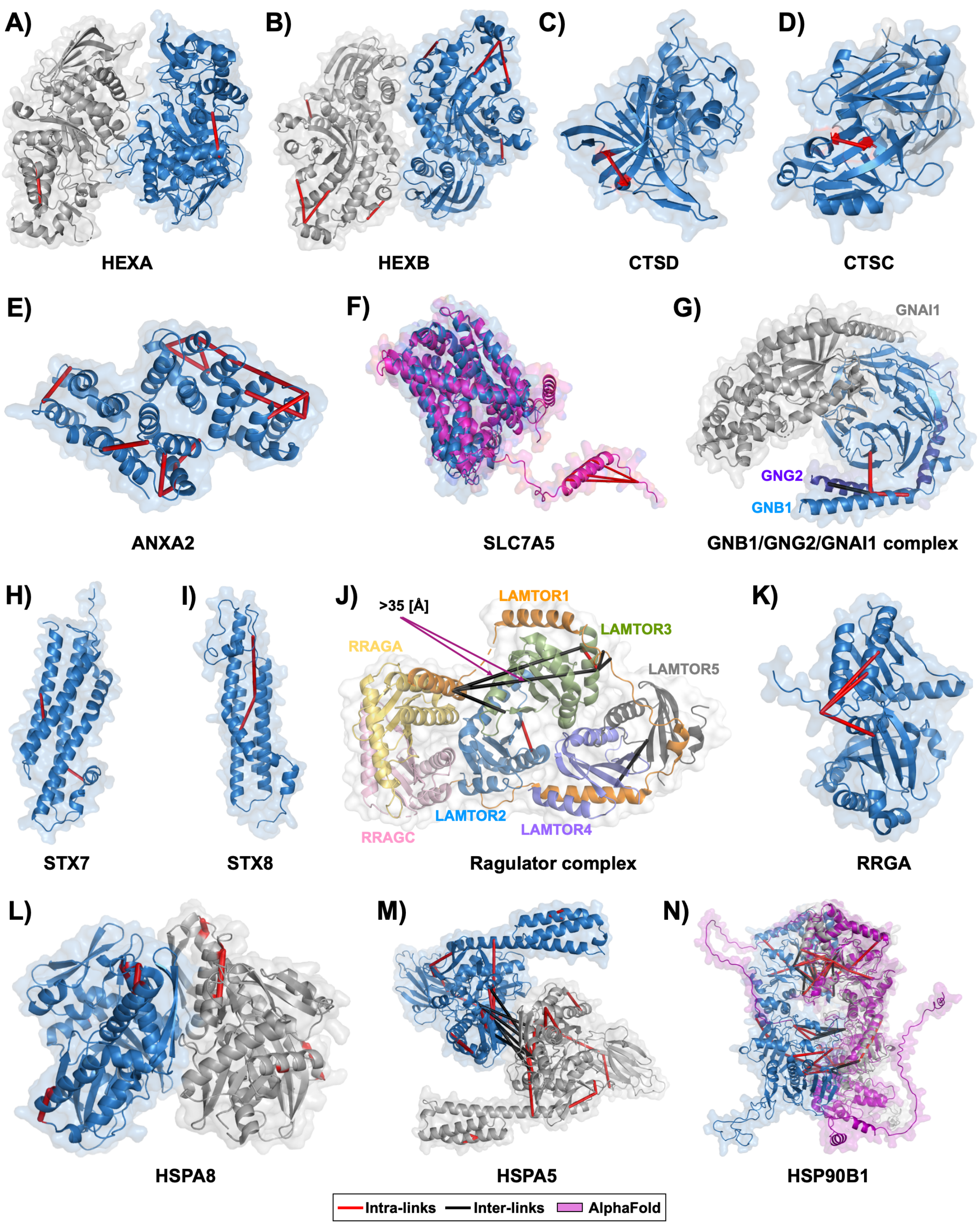

A)
HEXA
B)
HEXB
C)
CTSD
D)
CTSC
E)
ANXA2
F)
SLC7A5
G)
GNAI1
GNG2
GNB1
GNB1/GNG2/GNAI1 complex
H)
STX7
I)
STX8
J)
>35 [Å]
LAMTOR1
LAMTOR3
RRAGA
LAMTOR5
LAMTOR2
LAMTOR4
RRAGC
Ragulator complex
K)
RRGA
L)
HSPA8
M)
HSPA5
N)
HSP90B1
Intra-links
Inter-links
AlphaFold

Figure S3: Integration of cross-links into resolved structures and AlphaFold models of lysosomal and lysosome-associated proteins. Related to Figure 3.
A) Integration of cross-links into the RS of beta-hexosaminidase subunit alpha (HEXA).
B) Integration of cross-links into the RS of Beta-hexosaminidase subunit beta (HEXB).
C) Integration of cross-links into the RS of cathepsin D (CTSD).
D) Integration of cross-links into the RS of cathepsin C (CTSC).
E) Integration of cross-links into the RS of annexin A2 (ANXA2).
F) Integration of cross-links into the AF (Jumper et al., 2021) model of the large neutral amino acids transporter small subunit 1 (SLC7A5).
G) Integration of cross-links into the RS of the complex of GNB1, GNG2, and GNAI1.
H) Integration of cross-links into the homology model of syntaxin 7 (STX7).
I) Integration of cross-links into the homology model of syntaxin 8 (STX8).
J) Integration of cross-links into the RS of the Ragulator complex (LAMTOR1-5) in the presence of RRAGA and RRAGC.
K) Integration of cross-links into the AF model of Ras-related GTP-binding protein A (RRGA).
L) Integration of cross-links into the RS of heatshock protein 8 (HSPA8).
M) Integration of cross-links into the RS of heatshock protein 5 (HSPA5).
N) Integration of cross-links into the mixed model based on the RS and the AF model for heatshock protein 90B1 (HSP90B1). Magenta structures were extracted from AlphaFold while grey/blue structures are based on crystal or cryo-EM structures retrieved from PDB.
Legend: Inter-links (cross-links between two different proteins or two subunits of the same protein) are shown in black, and intra-links (cross-links within the same protein) are shown in red. Overlength cross-links (>35 Å) are indicated.

Figure 4

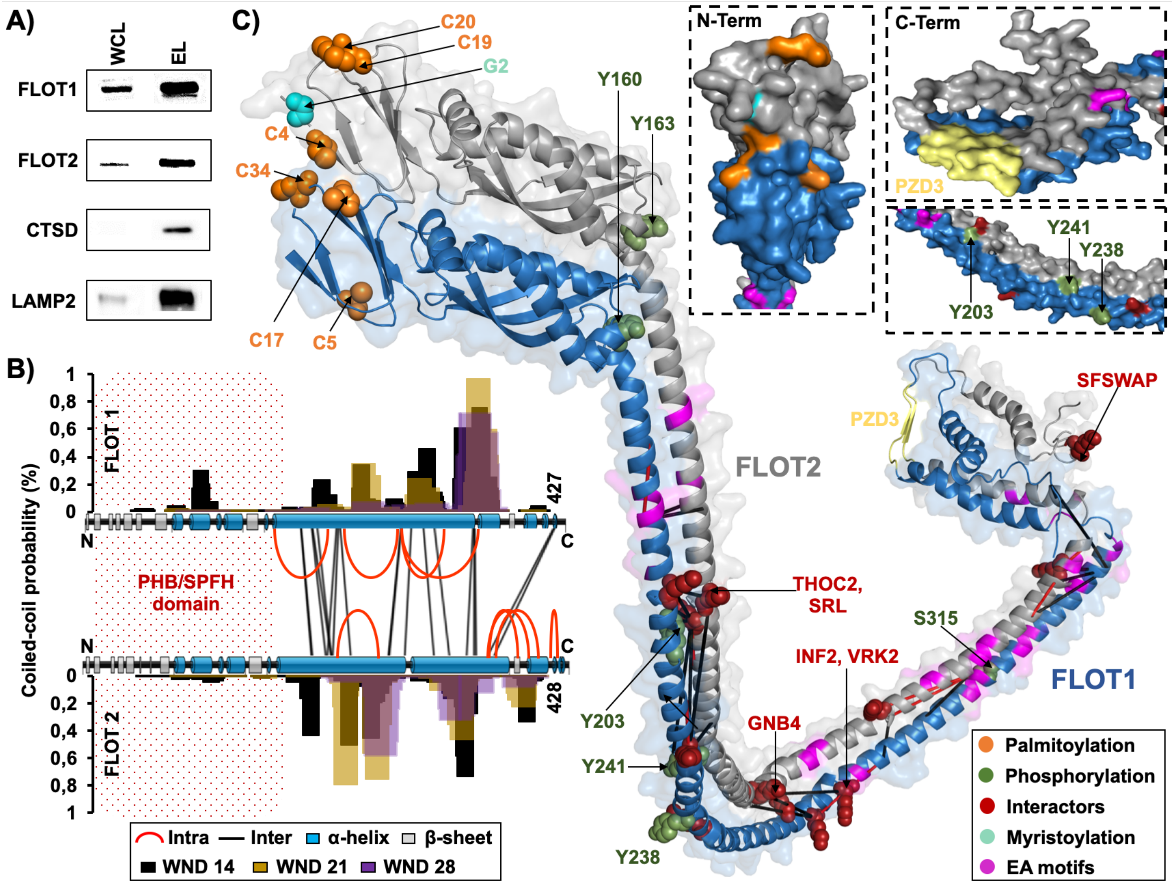

A)
WCL
EL
FLOT1
FLOT2
CTSD
LAMP2
B)
Coiled-coil probability (%)
1
0,8
0,6
0,4
0,2
0
FLOT 1
427
N
C
PHB/SPFH domain
N
C
FLOT 2
428
0
0,2
0,4
0,6
0,8
1
Intra
Inter
α-helix
β-sheet
WND 14
WND 21
WND 28
C)
C20
C19
G2
C4
C34
C17
C5
Y160
Y163
N-Term
C-Term
PZD3
Y241
Y238
Y203
FLOT2
PZD3
SFSWAP
THOC2, SRL
S315
INF2, VRK2
GNB4
FLOT1
Y203
Y241
Y238
Palmitoylation
Phosphorylation
Interactors
Myristoylation
EA motifs

Figure S4

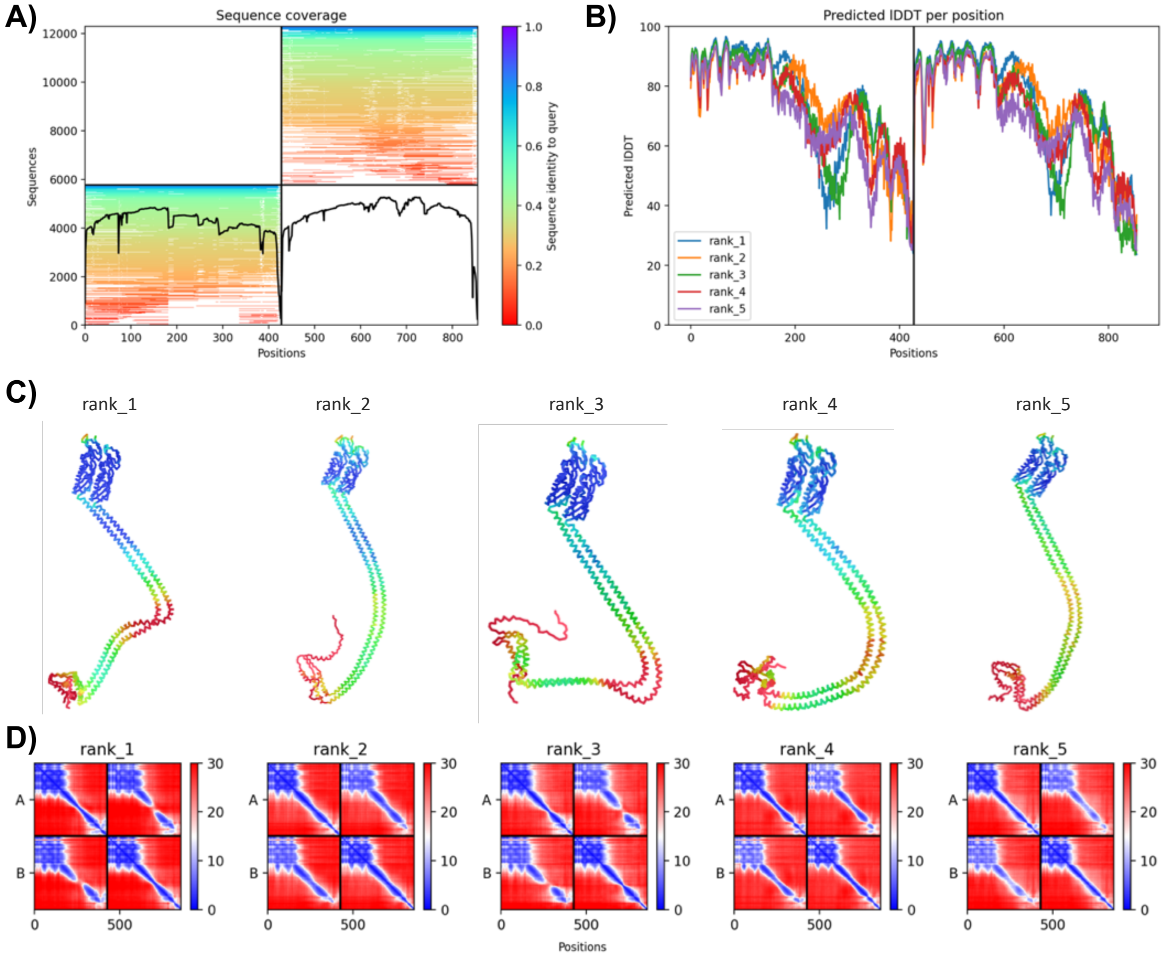

A)
Sequence coverage
Sequences
Positions
Sequence identity to query
B)
Predicted lDDT per position
Predicted lDDT
Positions
rank_1
rank_2
rank_3
rank_4
rank_5
C)
rank_1
rank_2
rank_3
rank_4
rank_5
D)
rank_1
rank_2
rank_3
rank_4
rank_5
A
B
Positions

Figure S4: Quality assessment of different AlphaFold-based ColabFold models of the FLOT1/FLOT2 heterodimeric complex. Related to Figure 4.
A) Visualization of depth and diversity of the multiple sequence alignment (MSA).
B) Predicted local distance difference test (plDDT) assessing local structural confidence of the different models. LDDT is a metric that evaluates local distance differences between all heavy atoms in a model, including validation of stereochemical plausibility. It ranges from 0 to 100, where 100 is the most confident.
C) Five highest scoring ColabFold models for the FLOT1-FLOT2 heterodimer. Color coding of individual structural features is based on their plDDTs per residue.
D) Inter-chain predicted aligned error (PAE), which aims to evaluate the position error at residue x, if the predicted and the true structures were aligned on residue y. It indicates the pairwise confidence of the respective prediction and ranks the models.
Abbreviations: MSA: multiple sequence alignment; plDDT: predicted local distance difference test; PAE: predicted aligned error

Figure 5

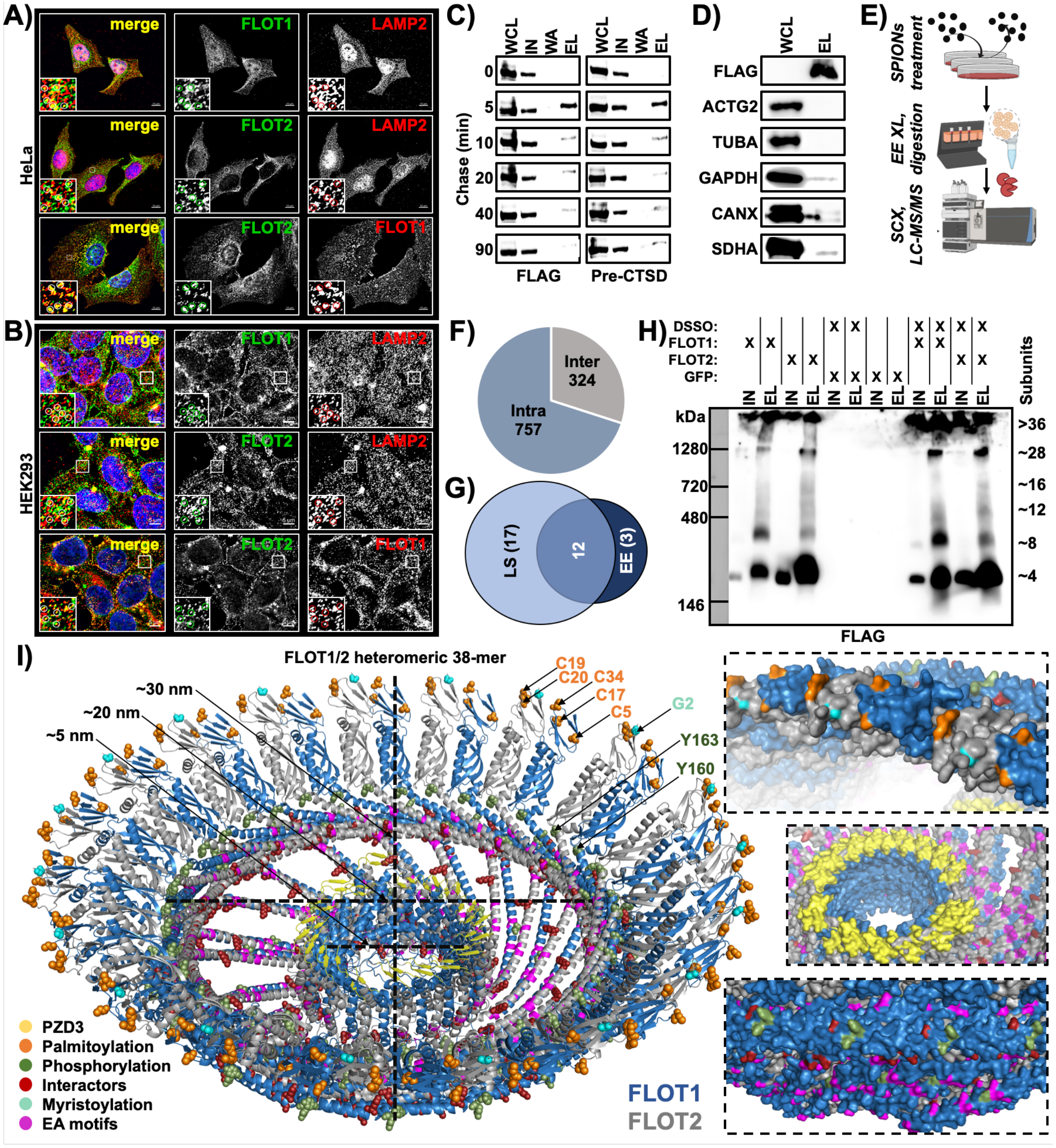

A)
merge
FLOT1
LAMP2
merge
FLOT2
LAMP2
merge
FLOT2
FLOT1
HeLa
B)
merge
FLOT1
LAMP2
merge
FLOT2
LAMP2
merge
FLOT2
FLOT1
HEK293
C)
WCL
IN
WA
EL
WCL
IN
WA
EL
Chase (min)
0
5
10
20
40
90
FLAG
Pre-CTSD
D)
WCL
EL
FLAG
ACTG2
TUBA
GAPDH
CANX
SDHA
E)
SPIONs treatment
EE XL, digestion
SCX, LC-MS/MS
F)
Inter 324
Intra 757
G)
LS (17)
12
EE (3)
H)
DSSO:
FLOT1:
FLOT2:
GFP:
IN
EL
kDa
1280
720
480
146
FLAG
Subunits
>36
~28
~16
~12
~8
~4
I)
FLOT1/2 heteromeric 38-mer
~30 nm
~20 nm
~5 nm
C19
C20
C34
C17
C5
G2
Y163
Y160
PZD3
Palmitoylation
Phosphorylation
Interactors
Myristoylation
EA motifs
FLOT1
FLOT2

Figure S5

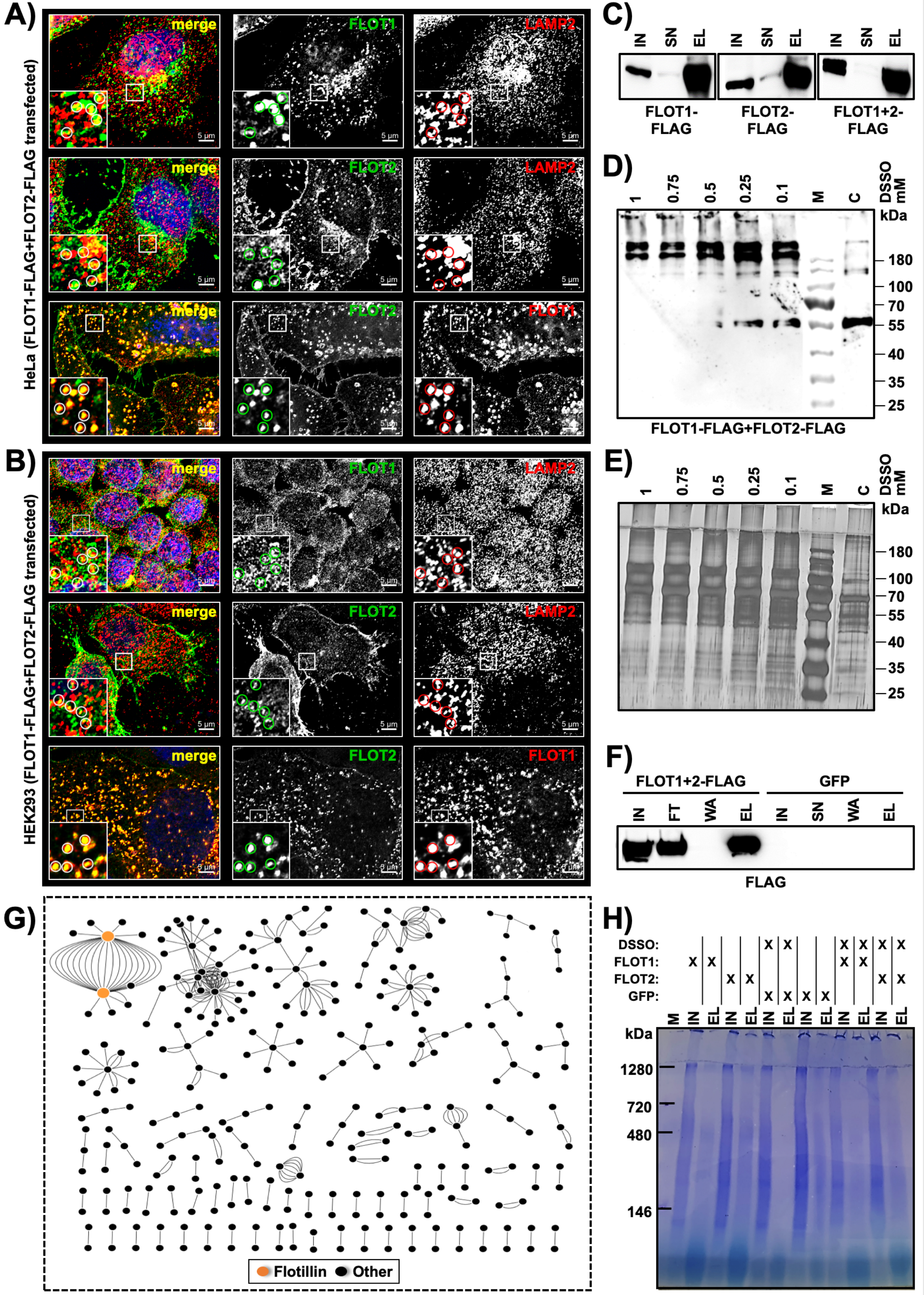

A)
HeLa (FLOT1-FLAG+FLOT2-FLAG transfected)
merge
FLOT1
LAMP2
merge
FLOT2
LAMP2
merge
FLOT2
FLOT1
5 µm
B)
HEK293 (FLOT1-FLAG+FLOT2-FLAG transfected)
merge
FLOT1
LAMP2
merge
FLOT2
LAMP2
merge
FLOT2
FLOT1
5 µm
C)
IN SN EL IN SN EL IN SN EL
FLOT1-FLAG
FLOT2-FLAG
FLOT1+2-FLAG
D)
1 0.75 0.5 0.25 0.1 M C
DSSO mM
kDa
180
100
70
55
40
35
25
FLOT1-FLAG+FLOT2-FLAG
E)
1 0.75 0.5 0.25 0.1 M C
DSSO mM
kDa
180
100
70
55
40
35
25
F)
FLOT1+2-FLAG
GFP
IN FT WA EL IN SN WA EL
FLAG
G)
Flotillin
Other
H)
DSSO:
FLOT1:
FLOT2:
GFP:
M IN EL IN EL IN EL IN EL IN EL IN EL IN EL
kDa
1280
720
480
146

Figure S5: Investigation of higher order FLOT1/FLOT2 assemblies and XL-LC-MS/MS analysis of early endosomes. Related to Figure 5.

A/B) Immunostaining analysis of HeLa and HEK293 cells transfected with FLOT1-FLAG+FLOT2-FLAG. Analyses were performed with anti-FLOT1 and anti-FLOT2 antibodies, co-localization with lysosomes was assessed through the lysosomal marker protein LAMP2.

C) Western blot analysis for verification of FLOT1-FLAG, FLOT2-FLAG, and FLOT1-FLAG+FLOT2-FLAG expression and enrichment by FLAG-IP.

D) Western blot analysis for determination of the optimal DSSO concentration for the cross-linking of SPIONs-enriched early endosomes from FLOT1-FLAG+FLOT2-FLAG transfected HEK293 cells. Early endosome-enriched fractions were cross-linked with the indicated amounts of DSSO at a protein concentration of 1 µg/µL followed by SDS-PAGE and anti-FLAG western blot. For each sample, 5 µg of protein were loaded to the gel. The non-cross-linked control sample (C) was treated with DMSO. Efficient cross-linking of FLOT1/FLOT2 complexes was judged by detection of the anti-FLAG band pattern. These analyses revealed that application of > 0.25 mM DSSO resulted in a near-quantitative loss of the FLOT1/FLOT2 monomers (~55 kDa) with almost exclusive detection of higher molecular weight assemblies (>180 kDa) and increasing amounts of protein aggregates which did not enter the separation gel.

E) Silver staining-based determination of the optimal DSSO concentration for the cross-linking of SPIONs-enriched early endosomes from FLOT1-FLAG+FLOT2-FLAG transfected HEK293 cells. Samples were generated and cross-linked as described in (D) followed by SDS-PAGE and silver staining of the gel. For each sample, 2.5 µg were loaded. The optimal DSSO concentration was determined to be 0.25 mM (final concentration), as the application of higher DSSO amounts resulted in reduced intensities in the low molecular weight region, indicating possible over cross-linking which could result in protein aggregates. Therefore, in accordance to the western blot results (see D), 0.25 mM DSSO was chosen.

F) Verification of presence of FLOT1 and FLOT2 in SPIONs-enriched early endosomes from FLOT1-FLAG+FLOT2-FLAG overexpressing HEK293 cells prior to the XL-LC-MS/MS experiment. Loading amount: 10 % of each fraction.

G) Interaction networks based on cross-links identified in the XL-LC-MS/MS dataset for DSSO-treated early endosome-enriched fractions from FLOT1-FLAG+FLOT2-FLAG-overexpressing HEK293 cells. In total, 324 inter-links were identified. FLOT1 and FLOT2 are highlighted in orange. The network was generated using Cytoscape.

H) Coomassie staining of a BN-PAGE analysis of anti-FLAG-IP eluate fractions from HEK293 cells transfected with FLOT1-FLAG or FLOT2-FLAG. Input and eluate fractions were analyzed in an untreated and a DSSO-cross-linked state. Control cells were transfected with GFP. Loading amounts: IN: 50 µg, EL: 25 % of eluate fraction.

Abbreviations: DSSO: disuccinimidyl sulfoxide; DMSO: dimethyl sulfoxide; PNS: post nuclear supernatant; M: marker; IN: input; SN: supernatant; FT: flow through; WA: wash; EL: eluate; C: control (no DSSO); IP: immunoprecipitation; BN-PAGE: blue native polyacrylamide gel electrophoresis; GFP: green fluorescent protein.

Figure 6

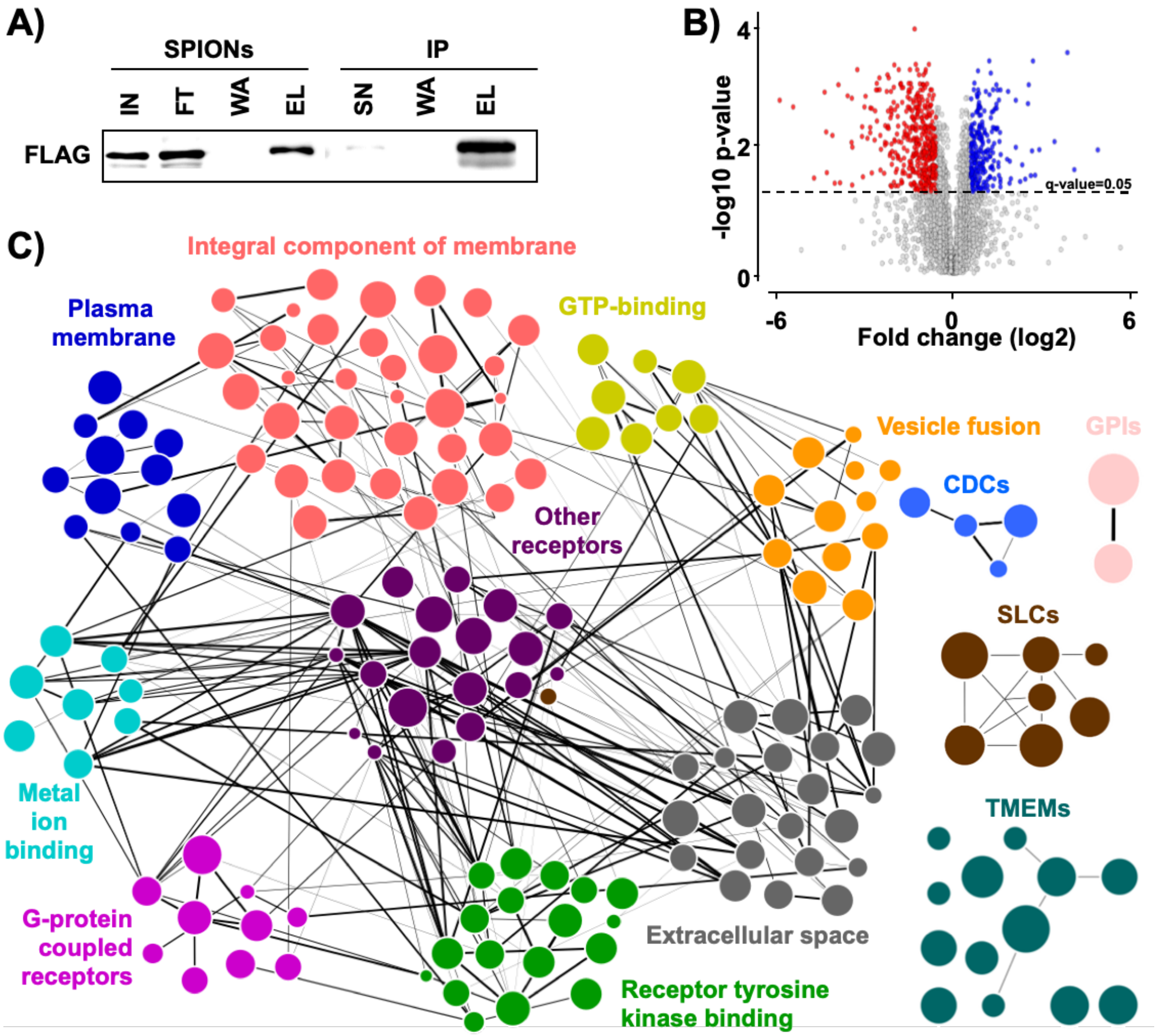

A)
SPIONs
IP
IN
FT
WA
EL
SN
WA
EL
FLAG
B)
-log10 p-value
4
2
0
q-value=0.05
-6
0
6
Fold change (log2)
C)
Integral component of membrane
Plasma membrane
GTP-binding
Vesicle fusion
GPIs
CDCs
Other receptors
SLCs
Metal ion binding
G-protein coupled receptors
Extracellular space
TMEMs
Receptor tyrosine kinase binding

Figure S6

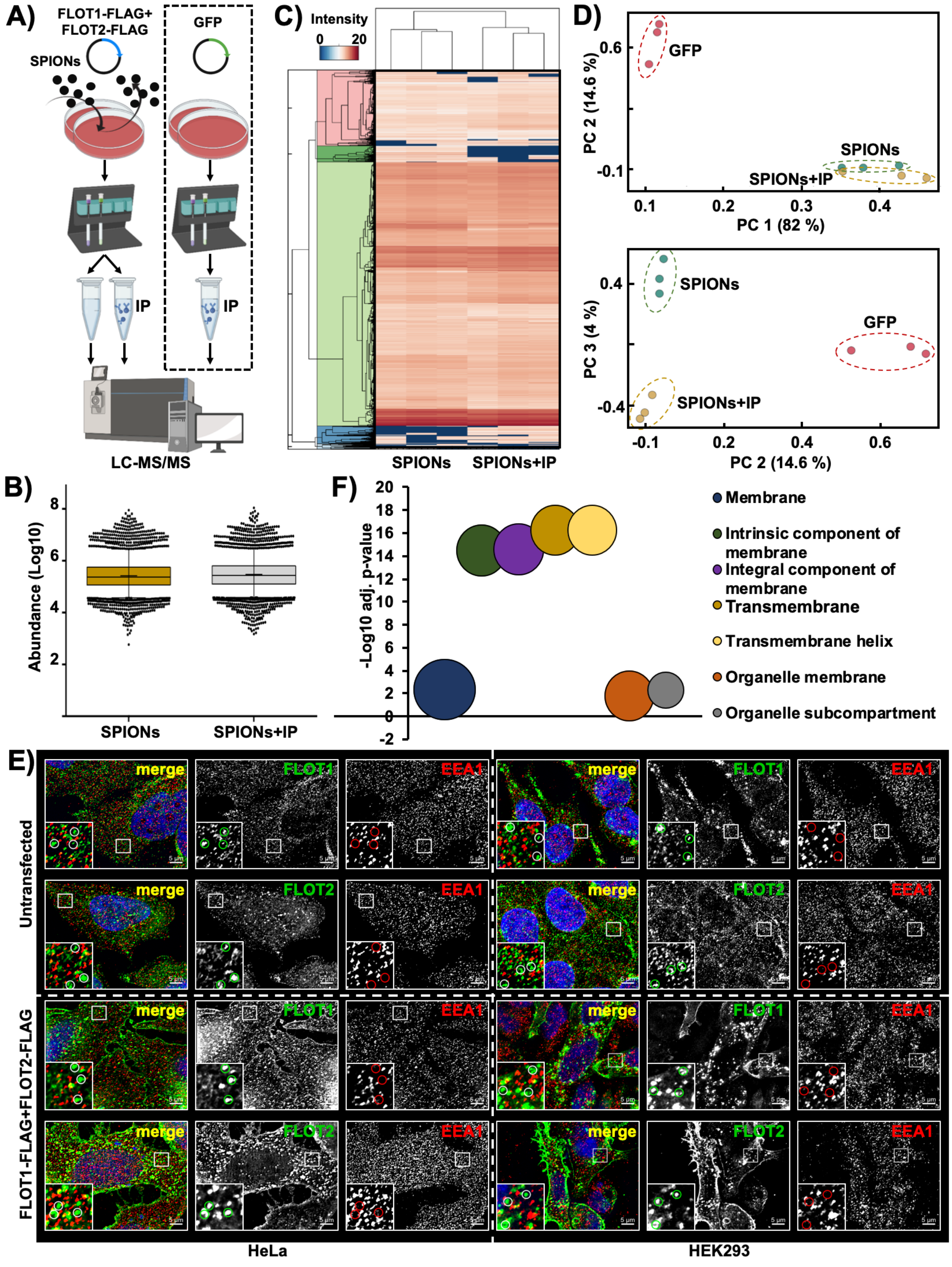

A)
FLOT1-FLAG+
FLOT2-FLAG
SPIONs
GFP
IP
IP
LC-MS/MS
B)
Abundance (Log10)
8
6
4
2
SPIONs
SPIONs+IP
C)
Intensity
0
20
SPIONs
SPIONs+IP
D)
PC 2 (14.6 %)
0.6
-0.1
GFP
SPIONs
SPIONs+IP
0.1
0.2
0.3
0.4
PC 1 (82 %)
PC 3 (4 %)
0.4
-0.4
SPIONs
GFP
SPIONs+IP
-0.1
0.2
0.4
0.6
PC 2 (14.6 %)
F)
-Log10 adj. p-value
20
18
16
14
12
10
8
6
4
2
0
-2
● Membrane
● Intrinsic component of membrane
● Integral component of membrane
● Transmembrane
● Transmembrane helix
● Organelle membrane
● Organelle subcompartment
E)
Untransfected
FLOT1-FLAG+FLOT2-FLAG
merge
FLOT1
EEA1
FLOT2
5 µm
HeLa
HEK293

Figure S6: Determination of flotillin vesicular cargo. Related to Figure 6.
A) Workflow for the enrichment of early endosomes by SPIONs and the subsequent enrichment of FLOT1-/FLOT2-positive early endosomal subpopulations by FLAG-IP. HEK293 cells were double-transfected with FLOT1-FLAG+FLOT2-FLAG, control cells with GFP. All samples were analyzed by LC-MS/MS using DIA.
B) Analysis of protein abundances in DIA datasets of SPIONs-enriched early endosomes and FLAG-IP-enriched FLOT1-/FLOT2-positive early endosomes. Shown are combined values from three independent biological replicates, the median is indicated by a line, the average is marked with a “+”.
C) Unsupervised clustering of protein abundances from three independent biological replicates for SPIONs-enriched early endosomes and FLAG-IP-enriched FLOT1-/FLOT2-positive early endosomes. Color coding correlates with the intensity of individual proteins.
D) PCA for all SPIONs-enriched early endosomes, FLAG-IP-enriched FLOT1-/FLOT2-positive early endosomes, and GFP-transfected control cells. Four individual principal components were used (PC1 and PC2, as well as PC3 and PC4).
E) Immunostaining analysis of HeLa and HEK293 cells. FLOT1, FLOT2, and EEA1 was analyzed both in untransfected cells as well as such transfected with FLOT1-FLAG+FLOT2-FLAG. A potential co-localization of FLOT1 and FLOT2 with the early endosome marker EEA1 was investigated.
F) GO enrichment analysis applying STRING for proteins which are overrepresented in FLOT1-/FLOT2-positive early endosomes. Results from the GO-category cellular component, as well as the UniProt keywords transmembrane and transmembrane helix are shown. Bubble size correlates with the number of proteins assigned to an individual category. Shown are p-values corrected for multiple testing within each category using the Benjamini–Hochberg procedure (cut off < 0.05) representing the significance of enrichment (n=3).
Abbreviations: SPIONs: superparamagnetic iron oxide nanoparticles; DIA: data independent acquisition; IP: immunoprecipitation; PCA: principal component analysis; adj.: adjusted; GO: gene ontology.

Figure 7

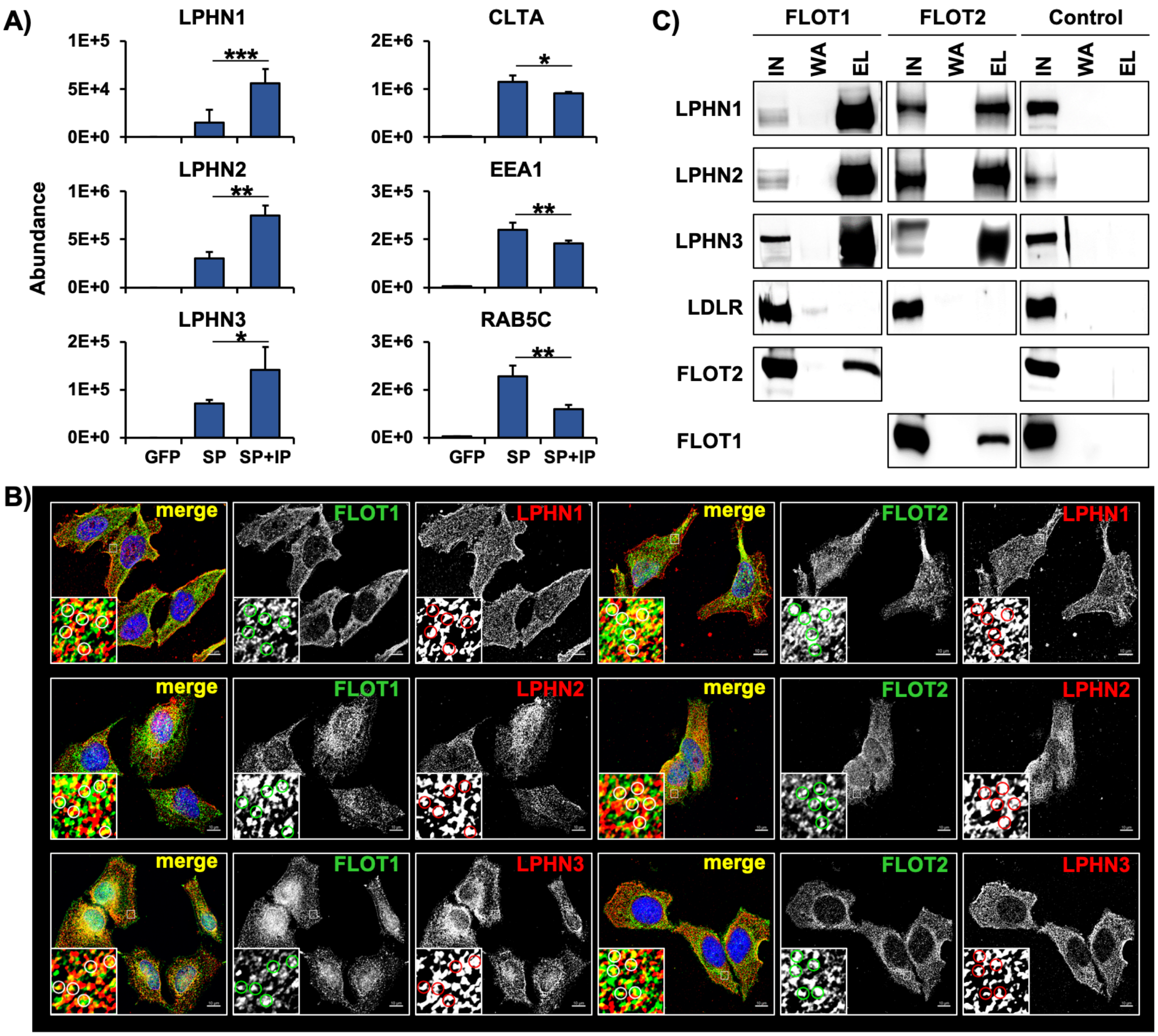

A)
Abundance
LPHN1
1E+5
5E+4
0E+0
***
LPHN2
1E+6
5E+5
0E+0
**
LPHN3
2E+5
1E+5
0E+0
*
GFP
SP
SP+IP
CLTA
2E+6
1E+6
0E+0
*
EEA1
3E+5
2E+5
0E+0
**
RAB5C
3E+6
2E+6
0E+0
**
GFP
SP
SP+IP
C)
FLOT1
FLOT2
Control
IN
WA
EL
IN
WA
EL
IN
WA
EL
LPHN1
LPHN2
LPHN3
LDLR
FLOT2
FLOT1
B)
merge
FLOT1
LPHN1
merge
FLOT2
LPHN1
merge
FLOT1
LPHN2
merge
FLOT2
LPHN2
merge
FLOT1
LPHN3
merge
FLOT2
LPHN3

Figure S7

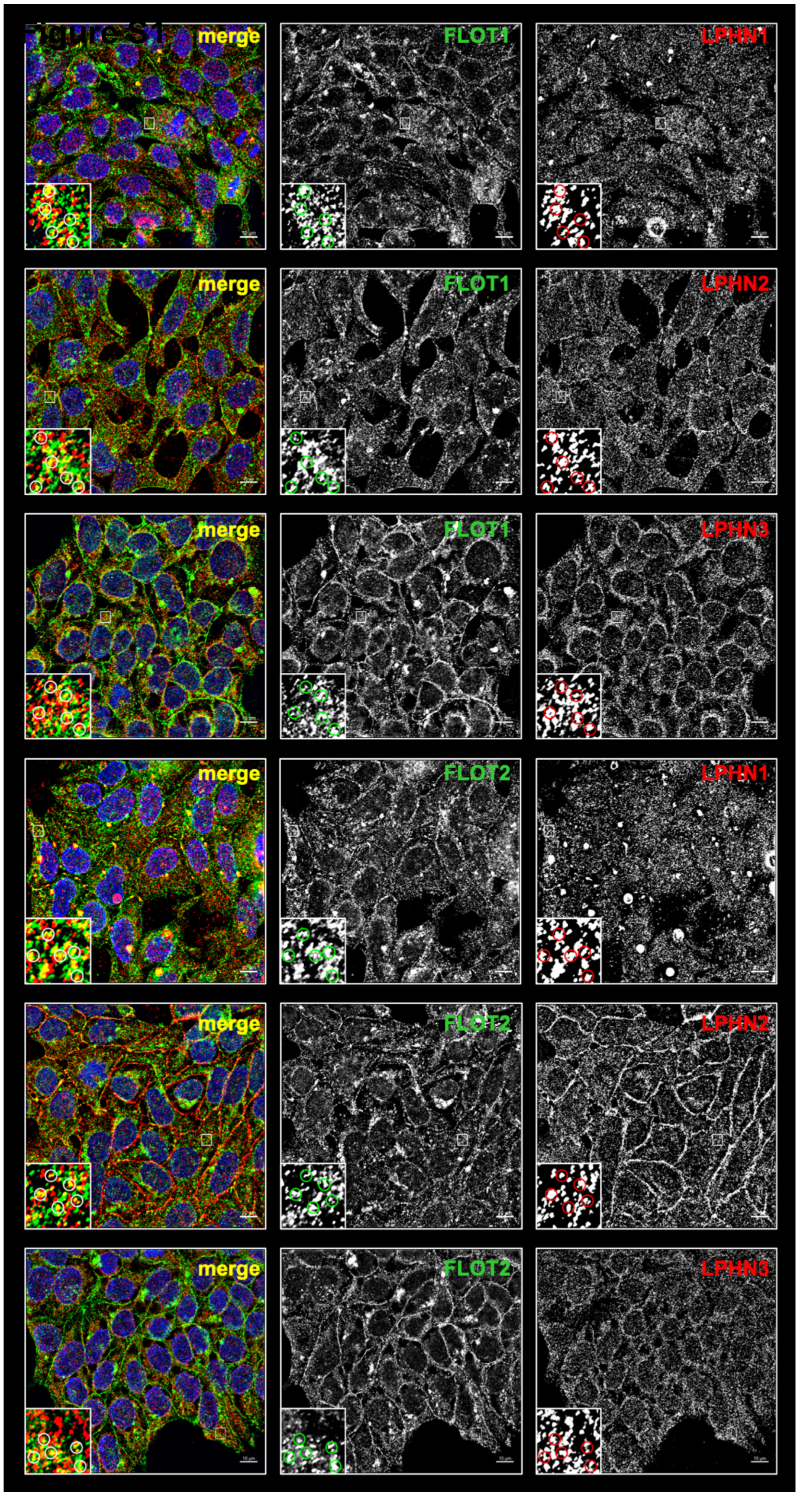


Figure S1
merge
FLOT1
LPHN1
merge
FLOT1
LPHN2
merge
FLOT1
LPHN3
merge
FLOT2
LPHN1
merge
FLOT2
LPHN2
merge
FLOT2
LPHN3

Figure S7: Latrophillins co-localize with FLOT1 and FLOT2 in punctate stainings. Related to Figure 7.
Immunostaining analysis of HEK293 cells for FLOT1 and FLOT2 in combination with LPHN1, LPHN2, or LPHN3. Dependent on the combination of antibodies, different degrees of co-localization can be observed.

REFERENCES

Akter, F., Ponnaiyan, S., Kögler-Mohrbacher, B., Bleibaum, F., Damme, M., Renard, B.Y., Winter, D. (2020). Multi cell line analysis of lysosomal proteomes reveals unique features and novel lysosomal proteins. bioRxiv.

Albanese, P., Tamara, S., Saracco, G., Scheltema, R.A., Pagliano, C. (2020). How paired PSII-LHCII supercomplexes mediate the stacking of plant thylakoid membranes unveiled by structural mass-spectrometry. Nat. Commun. *11*, 1-14.

Babuke, T., Ruonala, M., Meister, M., Amaddii, M., Genzler, C., Esposito, A., and Tikkanen, R. (2009). Hetero-oligomerization of reggie-1/flotillin-2 and reggie-2/flotillin-1 is required for their endocytosis. Cell Signal. *21*, 1287-1297.

Bach, J.N., and Bramkamp, M. (2015). Dissecting the Molecular Properties of Prokaryotic Flotillins. PLoS One *10*, e0116750.

Balderhaar, H.J., Ungermann, C. (2013). CORVET and HOPS tethering complexes - coordinators of endosome and lysosome fusion. J. Cell Sci. *126*, 1307-1316.

Ballabio, A., Bonifacino, J.S. (2019). Lysosomes as dynamic regulators of cell and organismal homeostasis. Nat. Rev. Mol. Cell Biol. *21*, 101-118.

Barysz, H.M., Malmström, J. (2018). Development of Large-scale Cross-linking Mass Spectrometry. Mol. Cell. Proteomics *17*, 1055–1066.

Bellizzi, J.J., Widom, J., Kemp, C., Lu, J.Y., Das, A.K., Hofmann, S.L., and Clardy, J. (2000). The crystal structure of palmitoyl protein thioesterase 1 and the molecular basis of infantile neuronal ceroid lipofuscinosis. Proc. Natl. Acad. Sci. U.S.A. *97*, 4573-4578.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. *57*, 289-300.

Bodin, S., Planchon, D., Rios Morris, E., Comunale, F., and Gauthier-Rouvière, C. (2014). Flotillins in intercellular adhesion - from cellular physiology to human diseases. J. Cell Sci. *127*, 5139–5147.

Bruderer, R., Bernhardt, O. M., Gandhi, T., Reiter, L. (2016). High-precision iRT prediction in the targeted analysis of data-independent acquisition and its impact on identification and quantitation. Proteomics *16*, 2246-2256.

Brunger, A.T. (2007). Version 1.2 of the Crystallography and NMR system. Nat. Protoc. *2*, 2728–2733.

Cabukusta, B., Neefjes, J. (2018). Mechanisms of lysosomal positioning and movement. Traffic *19*, 761-769.

Chavez, J.D., Keller, A., Zhou, B., Tian, R., and Bruce, J.E. (2019). Cellular Interactome Dynamics during Paclitaxel Treatment. Cell Rep. *29*, 2371-2383.

Chen, Z.A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski-Wills, J.C., Nilges, M., Cramer, P., Rappsilber, J. (2010). Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. EMBO J. *29*, 717–726.

Cheng, X., Shen, D., Samie, M., and Xu, H. (2010). Mucolipins: Intracellular TRPML1-3 channels. FEBS Lett. *584*, 2013-2021.

Chu, B.B., Liao, Y.C., Qi, W., Xie, C., Du, X., Wang, J., Yang, H., Miao, H.H., Li, B.L., and Song, B.L. (2015). Cholesterol transport through lysosome-peroxisome membrane contacts. Cell *161*, 291-306.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367-1372.

Cruciat, C.-M., Ohkawara, B., Acebron, S.P., Karaulanov, E., Reinhard, C., Ingelfinger, D., Boutros, M., and Niehrs, C. (2010). Requirement of Prorenin Receptor and Vacuolar H+-ATPase-Mediated Acidification for Wnt Signaling. Science *327*, 459–463.

De Araujo, M.E., Naschberger, A., Fürnrohr, B.G., Stasyk, T., Dunzendorfer-Matt, T., Lechner, S., Welti, S., Kremser, L., Shivalingaiah, G., Offterdinger, M., Lindner, H.H., Huber, L.A., Scheffzek, K. (2017). Crystal structure of the human lysosomal mTORC1 scaffold complex and its impract on signaling. Science *358*, 377-381.

Dempwolff, F., Schmidt, F.K., Hervas, A.B., Stroh, A., Rosch, T.C., Riese, C.N., Dersch, S., Heimerl, T., Lucena, D., Hulsbusch, N., et al. (2016). Super Resolution Fluorescence Microscopy and Tracking of Bacterial Flotillin (Reggie) Paralogs Provide Evidence for Defined-Sized Protein Microdomains within the Bacterial Membrane but Absence of Clusters Containing Detergent-Resistant Proteins. PLoS Genet. *12*, e1006116.

Eskelinen, E.L., Schmidt, C.K., Neu, S., Willenborg, M., Fuertes, G., Salvador, N., Tanaka, Y., Lullmann-Rauch, R., Hartmann, D., and Heeren, J. (2004). Disturbed cholesterol traffic but normal proteolytic function in LAMP-1/LAMP-2 double-deficient fibroblasts. Mol. Biol. Cell *15*, 3132-3145.

Fan, W.L., Guo, J., Gao, B.C., Zhang, W.B., Ling, L.C., Xu, T., Pan, C.J., Li, L., Chen, S., Wang, H., et al. (2019). Flotillin-mediated endocytosis and ALIX-syntenin-1-mediated exocytosis protect the cell membrane from damage caused by necroptosis. Sci. Signal. *12*, 10.

Fasci, D., van Ingen, H., Scheltema, R.A., Heck, A.J. (2018). Histone interaction landscapes visualized by crosslinking mass spectrometry in intact cell nuclei. Mol. Cell. Proteomics *17*.

Ferrari, A.J.R., Clasen, M.A., Kurt, L., Carvalho, P.C., Gozzo, F.C., Martínez, L., and Valencia, A. (2019). TopoLink: Evaluation of structural models using chemical crosslinking distance constraints. Bioinformatics *35*, 3169–3170.

Fraldi, A., Klein, A.D., Medina, D.L., and Settembre, C. (2016). Brain Disorders Due to Lysosomal Dysfunction. Annu. Rev. Neurosci. *39*, 277-295.

Frick, M., Bright, N.A., Riento, K., Bray, A., Merrified, C., and Nichols, B.J. (2007). Coassembly of Flotillins Induces Formation of Membrane Microdomains, Membrane Curvature, and Vesicle Budding. Curr. Biol. *17*, 1151-1156.

Garg, S., Sharma, M., Ung, C., Tuli, A., Barral, D.C., Hava, D.L., Veerapen, N., Besra, G.S., Hacohen, N., and Brenner, M.B. (2011). Lysosomal trafficking, antigen presentation, and microbial killing are controlled by the Arf-like GTPase Arl8b. Immunity *35*, 182-193.

Ge, L., Qi, W., Wang, L.J., Miao, H.H., Qu, Y.X., Li, B.L., and Song, B.L. (2011). Flotillins play an essential role in Niemann-Pick C1-like 1-mediated cholesterol uptake. Proc. Natl. Acad. Sci. U.S.A *108*, 551-556.

Glebov, O.O., Bright, N.A., and Nichols, B.J. (2006). Flotillin-1 defines a clathrin-independent endocytic pathway in mammalian cells. Nat. Cell Biol. *8*, 46-54.

Go, C.D., Knight, J.D.R., Rajasekharan, A., Rathod, B., Hesketh, G.G., Abe, K.T., Youn, J.Y., Samavarchi-Tehrani, P., Zhang, H., Zhu, L.Y., et al. (2021). A proximity-dependent biotinylation map of a human cell. Nature *595*, 120-124.

Gorbea, C., Pratt, G., Ustrell, V., Bell, R., Sahasrabudhe, S., Hughes, R.E., and Rechsteiner, M. (2010). A protein interaction network for Ecm29 links the 26 S proteasome to molecular motors and endosomal components. J. Biol. Chem. *285*, 31616-31633.

Grainger, S., Nguyen, N., Richter, J., Setayesh, J., Lonquich, B., Oon, C.H., Wozniak, J.M., Barahona, R., Kamei, C.N., Houston, J., et al. (2019). EGFR is required for Wnt9a-Fzd9b signalling specificity in haematopoietic stem cells. Nat. Cell Biol. *21*, 721–730.

Grigoryan, G., and Keating, A.E. (2008). Structural specificity in coiled-coil interactions. Curr. Opin. Struct. Biol. *18*, 477-483.

Haenig, C., Atias, N., Taylor, A.K., Mazza, A., Schaefer, M.H., Russ, J., Riechers, S.P., Jain, S., Coughlin, M., Fontaine, J.F., et al. (2020). Interactome Mapping Provides a Network of Neurodegenerative Disease Proteins and Uncovers Widespread Protein Aggregation in Affected Brains. Cell Rep. *32*, 108050.

Hao, F., Kondo, K., Itoh, T., Ikari, S., Nada, S., Okada, M., and Noda, T. (2018). Rheb localized on the Golgi membrane activates lysosome-localized mTORC1 at the Golgi-lysosome contact site. J. Cell Sci. *131*.

Hermle, T., Saltukoglu, D., Grünewald, J., Walz, G., and Simons, M. (2010). Regulation of Frizzled-Dependent Planar Polarity Signaling by a V-ATPase Subunit. Curr. Biol. *20*, 1269–1276.

Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V., and Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. *43*, D512-D520.

Hung, A.Y., and Sheng, M. (2002). PDZ domains: structural modules for protein complex assembly. J. Biol. Chem. *277*, 5699-5702.

Itzhak, D.N., Davies, C., Tyanova, S., Mishra, A., Williamson, J., Antrobus, R., Cox, J., Weekes, M.P., and Borner, G.H.H. (2017). A Mass Spectrometry-Based Approach for Mapping Protein Subcellular Localization Reveals the Spatial Proteome of Mouse Primary Neurons. Cell Rep. *20*, 2706-2718.

Itzhak, D.N., Tyanova, S., Cox, J., and Borner, G.H. (2016). Global, quantitative and dynamic mapping of protein subcellular localization. eLife *5*, e16950.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. *292*, 195–202.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583-589.

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

Karasawa, T., Yokokura, H., Kitajewski, J., and Lombroso, P.J. (2002). Frizzled-9 is activated by Wnt-2 and functions in Wnt/beta -catenin signaling. J. Biol. Chem. *277*, 37479-37486.

Kastritis, P.L., O'Reilly, F.J., Bock, T., Li, Y., Rogon, M.Z., Buczak, K., Romanov, N., Betts, M.J., Bui, K.H., Hagen, W.J., et al. (2017). Capturing protein communities by structural proteomics in a thermophilic eukaryote. Mol. Syst. Biol. *13*, 936.

Kaushik, S., Massey, A.C., and Cuervo, A.M. (2006). Lysosome membrane lipid microdomains: novel regulators of chaperone-mediated autophagy. EMBO J. *25*, 3921-3933.

Kim, D.I., and Roux, K.J. (2016). Filling the Void: Proximity-Based Labeling of Proteins in Living Cells. Trends Cell Biol. *26*, 804–817.

Klykov, O., Steigenberger, B., Pektas, S., Fasci, D., Heck, A.J.R., and Scheltema, R.A. (2018). Efficient and robust proteome-wide approaches for cross-linking mass spectrometry. Nat. Protoc. *13*, 2964–2990.

Kokubo, H., Helms, J.B., Ohno-Iwashita, Y., Shimada, Y., Horikoshi, Y., and Yamaguchi, H. (2003). Ultrastructural localization of flotillin-1 to cholesterol-rich membrane microdomains, rafts, in rat brain tissue. Brain Res. *965*, 83-90.

Krivov, G.G., Shapovalov, M.V., and Dunbrack, R.L. (2009). Improved prediction of protein side-chain conformations with SCWRL4. Proteins: Struct., Funct., Bioinf. *77*, 778–795.

Kwiatkowska, K., Matveichuk, O. V., Fronk, J., Ciesielska, A. (2020). Flotillins: At the Intersection of Protein S-Palmitoylation and Lipid-Mediated Signaling. Int. J. Mol. Sci. *21*, 2283.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Langenhan, T., Piao, X., and Monk, K.R. (2016). Adhesion G protein-coupled receptors in nervous system development and disease. Nat. Rev. Neurosci. *17*, 550-561.

Levin-Konigsberg, R., Montaño-Rendón, F., Keren-Kaplan, T., Li, R., Ego, B., Mylvaganam, S., DiCiccio, J.E., Trimble, W.S., Bassik, M.C., Bonifacino, J.S., et al. (2019). Phagolysosome resolution requires contacts with the endoplasmic reticulum and phosphatidylinositol-4-phosphate signalling. Nat. Cell Biol. *21*, 1234-1247.

Liao, Y.C., Fernandopulle, M.S., Wang, G., Choi, H., Hao, L., Drerup, C.M., Patel, R., Qamar, S., Nixon-Abell, J., Shen, Y., et al. (2019). RNA Granules Hitchhike on Lysosomes for Long-Distance Transport, Using Annexin A11 as a Molecular Tether. Cell *179*, 147-164.

Liu, F., Lossl, P., Scheltema, R., Viner, R., and Heck, A.J.R. (2017). Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. Nat. Commun. *8*, 1-8.

Liu, F., Rijkers, D.T.S., Post, H., and Heck, A.J.R. (2015). Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. Nat. Methods *12*, 1179–1184.

Liu, G.Y., and Sabatini, D.M. (2020). mTOR at the nexus of nutrition, growth, ageing and disease. Nat. Rev. Mol. Cell Biol. *21*, 183-203.

Lyly, A., Schantz, C.v., Salonen, T., Kopra, O., Saarela, J., Jauhiainen, M., Kyttälä, A., and Jalanko, A. (2007). Glycosylation, transport, and complex formation of palmitoyl protein thioesterase 1 (PPT1)--distinct characteristics in neurons. BMC cell biology *8*, 22.

Maxson, M.E., and Grinstein, S. (2014). The vacuolar-type H^+-ATPase at a glance - more than a proton pump. Journal of cell science *127*, 4987–4993.

Meister, M., and Tikkanen, R. (2014). Endocytic trafficking of membrane-bound cargo: a flotillin point of view. Membranes *4*, 356-371.

Mirdita, M., Ovchinnikov, S., and Steinegger, M. (2021). ColabFold - Making protein folding accessible to all. bioRxiv.

Müller, T., and Winter, D. (2017). Systematic Evaluation of Protein Reduction and Alkylation Reveals Massive Unspecific Side Effects by Iodine-containing Reagents. Mol. Cell. Proteomics *16*, 1173-1187.

Muthukottiappan, P., and Winter, D. (2021). A proteomic view on lysosomes. Mol Omics.

Neumann-Giesen, C., Falkenbach, B., Beicht, P., Claasen, S., Lüers, G., Stuermer, C.A., Herzog, V., and Tikkanen, R. (2004). Membrane and raft association of reggie-1/flotillin-2: role of myristoylation, palmitoylation and oligomerization and induction of filopodia by overexpression. Biochem. J. *378*, 509-518.

Niedzialkowska, E., Gasiorowska, O., Handing, K.B., Majorek, K.A., Porebski, P.J., Shabalin, I.G., Zasadzinska, E., Cymborowski, M., and Minor, W. (2016). Protein purification and crystallization artifacts: The tale usually not told. Protein Sci. *25*, 720-733.

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

O'Reilly, F.J., Xue, L., Graziadei, A., Sinn, L., Lenz, S., Tegunov, D., Blötz, C., Singh, N., Hagen, W.J.H., Cramer, P., et al. (2020). In-cell architecture of an actively transcribing-translating expressome. Science *369*, 554–557.

Pawlinski, L., Polus, A., Kałużna, M., Sordyl, M., Tobór-Swietek, E., Krawczyk, M., Bednarek, M., Solnica, B., Ruchała, M., and Kieć-Wilk, B. (2021). Gene expression with corresponding pathways analysis in Gaucher disease. Exp. Mol. Pathol. *123*, 104679.

Platt, F.M., d'Azzo, Al., Davidson, B.L., Neufeld, E.F., Tifft, C.J. (2018). Lysosomal storage diseases. Nat. Rev. Dis. Primers *4*, 1-25.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. *2*, 1896–1906.

Reddy, A., Caler, E.V., and Andrews, N.W. (2001). Plasma membrane repair is mediated by Ca(2+)-regulated exocytosis of lysosomes. Cell *106*, 157-169.

Riento, K., Frick, M., Schafer, I., and Nichols, B.J. (2009). Endocytosis of flotillin-1 and flotillin-2 is regulated by Fyn kinase. J .Cell Sci. *122*, 912-918.

Rivera-Milla, E., Stuermer, C.A.O., and Málaga-Trillo, E. (2006). Ancient origin of reggie (flotillin), reggie-like, and other lipid-raft proteins: Convergent evolution of the SPFH domain. Cell. Mol. Life Sci. *63*, 343–357.

Ruiz-Velasco, V., Ikeda, S.R., and Puhl, H.L. (2002). Cloning, tissue distribution, and functional expression of the human G protein β4-subunit. Physiol. Genomics *8*, 41-50.

Schroder, B., Wrocklage, C., Pan, C., Jager, R., Kosters, B., Schafer, H., Elsasser, H.P., Mann, M., and Hasilik, A. (2007). Integral and associated lysosomal membrane proteins. Traffic *8*, 1676-1686.

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337-342.

Schweppe, D.K., Chavez, J.D., Lee, C.F., Caudal, A., Kruse, S.E., Stuppard, R., Marcinek, D.J., Shadel, G.S., Tian, R., and Bruce, J.E. (2017). Mitochondrial protein interactome elucidated by chemical cross-linking mass spectrometry. Proc. Natl. Acad. Sci. U.S.A. *114*, 1732–1737.

Shin, H.R., and Zoncu, R. (2020). The Lysosome at the Intersection of Cellular Growth and Destruction. Dev. Cell *54*, 226-238.

Singh, J., Kaade, E., Muntel, J., Bruderer, R., Reiter, L., Thelen, M., and Winter, D. (2020). Systematic Comparison of Strategies for the Enrichment of Lysosomes by Data Independent Acquisition. J. Proteome Res. *19*, 371-381.

Singh, J., Ponnaiyan, S., Gieselmann, V., and Winter, D. (2021). Generation of Antibodies Targeting Cleavable Cross-Linkers. Anal. Chem. *93*, 3762-3769.

Skowyra, M.L., Schlesinger, P.H., Naismith, T.V., and Hanson, P.I. (2018). Triggered recruitment of ESCRT machinery promotes endolysosomal repair. Science *360*.

Solis, G.P., Hoegg, M., Munderloh, C., Schrock, Y., Malaga-Trillo, E., Rivera-Milla, E., and Stuermer, C.A.O. (2007). Reggie/flotillin proteins are organized into stable tetramers in membrane microdomains. Biochem. J. *403*, 313–322.

Soong, B.W., Huang, Y.H., Tsai, P.C., Huang, C.C., Pan, H.C., Lu, Y.C., Chien, H.J., Liu, T.T., Chang, M.H., Lin, K.P., et al. (2013). Exome Sequencing Identifies GNB4 Mutations as a Cause of Dominant Intermediate Charcot-Marie-Tooth Disease. Am. J. Hum. Genet. *92*, 422-430.

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

Stuermer, C.A.O., Lang, D.M., Kirsch, F., Wiechers, M., Deininger, S.O., Plattner, H., and Krieger, M. (2001). Glycosylphosphatidyl Inositol-anchored Proteins and fyn Kinase Assemble in Noncaveolar Plasma Membrane Microdomains Defined by Reggie-1 and -2. Mol. Biol. Cell *12*, 3031–3045.

Su, M.Y., Morris, K.L., Kim, D.J., Fu, Y., Lawrence, R., Stjepanovic, G., Zoncu, R., and Hurley, J.H. (2017). Hybrid Structure of the RagA/C-Ragulator mTORC1 Activation Complex. Mol. Cell *68*, 835-846.e833.

Sugawara, Y., Nishii, H., Takahashi, T., Yamauchi, J., Mizuno, N., Tago, K., and Itoh, H. (2007). The lipid raft proteins flotillins/reggies interact with Galphaq and are involved in Gq-mediated p38 mitogen-activated protein kinase activation through tyrosine kinase. Cell. Signalling *19*, 1301–1308.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., and Bork, P. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47*, D607-D613.

Tanaka, H., Kato, K., Yamashita, E., Sumizawa, T., Zhou, Y., Yao, M., Ivasaki, K., Yoshimura, M., Tsukihara, T (2009). The structure of rat liver vault at 3.5 angstrom resolution. Science *323*, 384-388.

Thelen, M., Winter, D., Braulke, T., and Gieselmann, V. (2017). SILAC-Based Comparative Proteomic Analysis of Lysosomes from Mammalian Cells Using LC-MS/MS. Methods Mol. Biol. *1594*, 1–18.

Valm, A.M., Cohen, S., Legant, W.R., Melunis, J., Hershberg, U., Wait, E., Cohen, A.R., Davidson, M.W., Betzig, E., and Lippincott-Schwartz, J. (2017). Applying systems-level spectral imaging and analysis to reveal the organelle interactome. Nature *546*, 162-167.

Van Zundert, G.C.P., Rodrigues, J.P.G.L.M., Trellet, M., Schmitz, C., Kastritis, P.L., Karaca, E., Melquiond, A.S.J., van Dijk, M., Vries, S.J.d., and Bonvin, A.M.J.J. (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J. Mol. Biol. *428*, 720–725.

Wang, L., Wu, D., Robinson, C.V., Wu, H., Fu, T.M. (2020). Structures of a Complete Human V-ATPase Reveal Mechanisms of Its Assembly. Mol. Cell *80*, 501-511.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., De Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Res. *46*, W296–W303.

Wendeler, M., Sandhoff, K. (2009). Hexosaminidase assays. Glycoconjugate J. *26*, 945–952.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D.A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Miller, E., Bache, S., Ooms, J., Robinson, D., Seidel, D. Spinu, V., Takahashi, K., Wilke, C., Woo, K.,Yutani, H. (2019). Welcome to the Tidyverse. J. Open Source Softw. *4*, 1686.

Wong, Y.C., Ysselstein, D., Krainc, D. (2018). Mitochondria-lysosome contacts regulate mitochondrial fission via RAB7 GTP hydrolysis. Nature *554*, 382-386.

Woolfson, D.N., Bartlett, G.J., Bruning, M., and Thomson, A.R. (2012). New currency for old rope: from coiled-coil assemblies to α-helical barrels. Curr. Opin. Struct. Biol. *22*, 432-441.

Yu, C., and Huang, L. (2018). Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. Anal. Chem. *90*, 144-165.

Zhang, T., Wang, R., Wang, Z., Wang, X., Wang, F., and Ding, J. (2017). Structural basis for Ragulator functioning as a scaffold in membrane-anchoring of Rag GTPases and mTORC1. Nat. Commun. *8*, 1394.

Zimmermann, L., Stephens, A., Nam, S.Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A.N., and Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J. Mol. Biol. *430*, 2237–2243.

5.3 Publication: Supporting Information

Supplementary tables for chapter 3 were not published online and will be only published after the peer review process. Supplementary tables can be accessed from the provided CD.

- Table S1: List of lysosomal proteins
- Table S2: Protein IDs & iBAQ values (LEF)
- Table S3: Cross-links (LEF)
- Table S4: Distance constraints cross-links
- Table S5: Protein IDs & iBAQ values (EEF)
- Table S6: Cross-links (EEF)
- Table S7: Flotillin positive endosome cargo

5.4 Conclusion

In the presented study, I analysed the lysosomal interactome on a large, unbiased scale and characterized lysosome-associated protein structures, as currently only for 69 % of the lysosomal human Gene Ontology (GO) annotated proteins structural information is accessible. Available protein structures elucidated by applying crystallography or cryo-electron microscopy experiments mostly depend on overexpression constructs from pro- or eukaryotic organisms. Therefore, the applicability of such structures resembling *in vivo* conditions frequently lacking post translational modifications (PTMs) remains to a certain extent questionable. Therefore, to provide a more comprehensive picture on lysosomal interactions and structures resembling near native physiological cellular conditions, I applied the MS-cleavable disuccinimidyl sulfoxide (DSSO) cross-linker to lysosome-enriched fractions of HEK293 cells, as existing studies failed to cover the majority of low abundant lysosomal proteins **(Figure 1A, Figure S1 A)**.

In a first step, I verified the integrity of lysosomes through a β-hexosaminidase assay and their enrichment by western blot **(Figure 1B, C)**. I further applied LC-MS/MS to assess the overall coverage of the lysosomal proteome resulting in the identification of 4,181 proteins, of which 474 were lysosomal, indicating an efficient coverage. The assessment of protein abundance of samples applying the label-free intensity-based absolute quantification (iBAQ) approach resulted in a ~3-fold overrepresentation for the average abundance of lysosomal proteins compared to the whole dataset, further verifying the successful enrichment of the lysosomal proteome **(Figure 1D, Table S1, S2)**. Subsequently, lysosome enriched fractions were divided into two fractions and cross-linking conditions optimized applying DSSO **(Figure S1 B, C)**. One fraction was mechanically disrupted, representing the disrupted (DR) fraction while the other represented the intact (IT) status of the lysosomes. Followed by their cross-linking, samples were proteolytically digested and cross-linked peptides enriched via strong cation exchange chromatography (SCX) and samples measured via LC-MS/MS **(Figure 1A, Figure S1 D-F)**. The analysis resulted in the assignment of 6,580 cross-link spectral matches, originating from 4,294 cross-linked peptides covering 524 cross-links (intra and inter-links) involving lysosomal proteins located to 111 lysosomal proteins **(Figure 1E, Figure S1 G, H, Table S3)**. Furthermore, the LC-MS/MS results verified the limited membrane permeability of DSSO as only 26 % of cross-links were found both in the IT and the DR state, while the latter contributed a bigger fraction to the dataset **(Figure 1F)**. Along this line, further analysis revealed that 91 % of the dataset's CSMs assigned to lysosomal luminal proteins were identified in samples cross-linked in the DR state, while the remaining CSMs were equally distributed among the IT and DR datasets **(Figure 1G)**. Additionally, I showed that despite the identification of on average more CSMs for higher abundant proteins, those did not show a strong correlation with protein abundance, ensuring that the dataset also covered proteins of low abundance **(Figure 1H, Figure S1 I)**. Additional analysis revealed a trend towards the identification of intra-links for higher abundant proteins compared to inter-links, whereas for the lysosomal proteome the effect was not that pronounced **(Figure S1 J, K)**. The cross-linking dataset primarily covered proteins located at the cytosolic face of the lysosomal membrane, further demonstrating the inaccessibility of the lysosomal lumen by the cross-linker **(Figure S1 L)**.

Overall the data covered 1,008 proteins engaged in 1,023 interactions, from which 254 (inter-links) were lysosomal **(Figure 2A, B, Figure S2 A-C, Table S3)**. Based on this I further classified lysosomal interactors in five distinct subcellular categories, of which the majority were assigned to the cytoplasm/cytoskeleton and nucleus **(Figure 2C)**. Most interactions identified among the lysosomal proteins were associated with the v-ATPase complex, the two flotillins (FLOT1 and FLOT2), Ragulator complex, and syntaxins **(Figure 2D, E, Figure S2 E, F)**. To verify the identified interactions biologically, I performed co-immunoprecipitation experiments for two unknown novel interactions. For the identified interaction of ATP6V1D with FZD9, a member of the WNT signaling pathway, I successfully co-immunoprecipitated FZD9 and ATP6V1D **(Figure 2F)**. Additionally, I successfully co-immunoprecipitated GNB4 with FLOT1 and FLOT2, confirming the validity of the cross-linking dataset **(Figure 2G)**. It should be emphasized that both co-immunoprecipitations were performed with cells not overexpressing the interaction partners. Moreover, analysis of the ~20 different PPIs identified for both flotillin proteins revealed an interaction hotspot between the amino acids 200-300 of the sequence **(Figure 2H)**.

The distance constraints of DSSO (up to ~35 Å) allow for the investigation of protein structures by, on the one hand, validating structurally defined regions and, on the other hand facilitating molecular modeling of such originating from regions that were not resolved so far. In total, I mapped 161 unique cross-links to high-resolution crystal structures of 34 lysosomal proteins **(Figure 3A, Figure S3 A-N, Table S4)**. As 64 lysosomal cross-links could not be mapped due to unresolved, flexible regions of the crystal/cryo-EM structures, I validated them based on homology and AlphaFold models. Applying the iBAQ approach, I first verified the stoichiometry of the 35meric v-ATPase complex **(Figure 3B)**. Subsequently, I supplemented missing amino acid regions of the resolved cryo-EM structure with the predicted topology by AlphaFold and validated the structure covering ~95 % of the v-ATPase sequence **(Figure 3C, D, Table S4)**.

The Ragulator complex consists of five subunit proteins (LAMTOR1-5) acting as a guanine nucleotide exchange factor (GEF) for the Rag GTPase heterodimer (RagA/B-RagC/D) complex. Hereby, Ragulator complex converts Rags to their active state, in which RagA/B are loaded with guanosine triphosphate (GTP) and RagC/D are bound to guanosine diphosphate (GDP) and thereafter Rags complex localizes mTORC1 at the lysosomal surface. The identified cross-links for the complex for instance suggested that in the absence of the RAG GTPases, the N-terminal region of LAMTOR1 exists in an unordered state, indicating a certain flexibility of this region *in vivo* **(Figure S3 J)**.

Out of the 161 mapped cross-links, 21 cross-links exceeded the distance constraint of DSSO, from which strikingly 9 originated from the same protein palmitoyl protein thioesterase 1 (PPT1) **(Figure 3E)**. These high number of overlength cross-links suggested a higher oligomerization state, contrary to the proposed homodimeric PPT1 crystal structure **(Figure 3F)**. Hence, I performed restraint-based docking analysis, which resulted in the prediction of a novel tetrameric PPT1 model fulfilling the distance constraints for all 18 PPT1 cross-links identified **(Figure 3G, H)**. The tetrameric model was further supported by a previous study demonstrating maximal enzymatic activity of PPT1 at a complex size >100 kDa, in line with the molecular weight of the proposed tetrameric model.

Since structural information was limited for the flotillin proteins, I applied various algorithms to predict their secondary structure **(Figure 4B)**. The results showed a cluster of beta-sheets and alpha helices at the proteins' N and C-termini, which were in line with the recently predicted AlphaFold models. As flotillins frequently assemble as hetero-oligomers with a postulated 1:1 stoichiometry, I applied the identified 29 unique cross-links among those proteins to propose the first heterodimeric flotillin model **(Figure 4C)**. The N-terminal region of the model consists of antiparallel β-sheets, with six repeats each, and four partially exposed α-helices, forming an ellipsoidal-like globular domain. PTMs such as the S-palmitoylation and N-myristoylation sites crucial for the membrane association of flotillin were located towards the membrane interface, while phosphorylation sites were surface exposed.

As flotillins were also detected in the plasma membrane and early endosomes **(Figure 5A, B)**, I further investigated whether and to which extent flotillin on lysosomes and early endosomes exhibit the same structure. In order to do so, I optimized the enrichment of early endosome fractions containing flotillin **(Figure 5C)**, as well as other early endosomal marker proteins (EEA1, CLTA, CLTB, CLTC) via LC-MS/MS **(Table S5)** and verified depletion of proteins from other cellular compartments **(Figure 5D)**. Along this line, appropriate transfection as well as cross-linking conditions were established, showing correct localization and efficient cross-linking of flotillin **(Figure S5 A-D)**. Applying the established methods (**Figure 5E, Figure S5E)**, XL-LC-MS/MS analysis revealed that flotillins localizing to the lysosome and early endosome have the same structural orientation **(Figure 5F, G, Figure S5 G, Table S6)**. As further distinctive structural characteristics such as the presence of the PDZ3 domains, EA-rich motifs, and highest number of identified cross-links suggested a higher order flotillin structure, I performed a blue native electrophoresis experiment determining its overall size **(Figure 5H, Figure S5 H)**. The results showed that most of the complex resides as a tetramer, while a substantial amount of the protein was engaged in higher-order assemblies of > 1 MDa, which was in line with published size estimations via EM. Additionally, based on a published template of a protein structure with a similar N-terminal conserved domain to flotillins, a heterooligomeric flotillin model was proposed consisting of 38 subunits **(Figure 5I).**

To further investigate a possible function of the 38mer and its role in endocytosis, I enriched for flotillin vesicles by first enriching an early endosome fraction (SPIONs), followed by immunoprecipitation (IP) of C-terminal tagged flotillin particles **(Figure S6 A)**. I verified the double enrichment of flotillin particles via western blot **(Figure 6A)** and label-free quantified differences among the early endosome enriched fraction (SPIONs) and double enriched flotillin particles (SPIONs+IP) applying DIA-LC-MS/MS **(Figure S6 B-D)**. I identified 5,089 proteins quantified across all datasets, from which I was able to define protein populations that were depleted or enriched in FLOT1/FLOT2-positive endosomes relative to the whole pool of early endosomes **(Figure 6B, Table S7)**. The results showed a depletion of early endosome marker in the SPION+IP sample, further validating the depletion of early endosomal proteins in this fraction. In total, I was able to identify 328 proteins that were either exclusively present in the SPIONs+IP fraction, or significantly upregulated compared to the SPIONs-enriched early endosomes, presenting the potential cargo of flotillin particles. The potential cargo consisted of various protein categories such as 50 receptors, including receptor tyrosine kinases and G-protein coupled receptors **(Figure 6C)**. Strikingly, for instance, all three members of latrophilins a conserved subfamily of G-protein-coupled receptors were identified to be significantly overrepresented in the SPIONs+IP dataset **(Figure 7A)**. To further investigate whether all three members of the latrophilin family are endocytosed in a flotillin dependent manner, I performed immunofluorescence microscopy (HEK293 and HeLa cells) verifying co-localization, as well co-IPs verifying the interactions **(Figure 7B, C, Figure S7)**.

6. Chapter 4: Turnover of Lysosomal and Lysosome Associated Proteins

6 The presented results in chapter 4 have not been published yet, as the manuscript is under preparation. Please find below all the authors involved in this project:

- **Singh, J.**, Stepath, M., Dang, H.M., Möller, C., Muchamedin, A., Hardt, R., Bernhardt, O., Gandhi, T., Bruderer, R., Reiter, L., Eisenacher, M., Vakhrushev, S., Gieselmann, V., Winter, D.
- **Remark:** The presented results within the scope of chapter 4 has been done by me as the only first author and **SHOULD BE CONSIDERED** for this cumulative doctoral dissertation.
- Supplementary tables for chapter 4 can be accessed from the provided CD.
  - Table S1: List of lysosomal proteins
  - Table S2: Protein half-lives
  - Table S3: Phosphoproteomics dataset

6.1 Introduction

Cellular homeostasis depends on a balance of macromolecule synthesis and degradation, commonly referred to as *“turnover”* [103]. The turnover of a protein is defined as the time required for both degradation and resynthesis of half of the proteins at the given cellular state, also defined in terms of half-life ($T_{1/2}$) [95]. The metabolic state of a cell has to be separated between a “steady-state” and those perturbated undergoing a dynamic change within experiments. The general assumption at a steady state is that 1) all proteins have a net change in protein level equaling to zero, which implies 2) that the number of molecules produced equals the number of proteins degraded. Hence, under a steady state the constant “degradation rate” simply equals to the time it takes to remove the existing protein pool by half. However, when a new steady state is reached after perturbation, the proteins might be expressed at very different levels than before. Yet, its half-life will only be different if the constant degradation rate changes. In other words, this means that changes in protein synthesis alone will not change/affect proteins degradation rate, but rather only its abundance [95]. Interdependencies between synthesis, degradation, and overall proteome level are shown and further discussed [95] **(Figure 8)**.

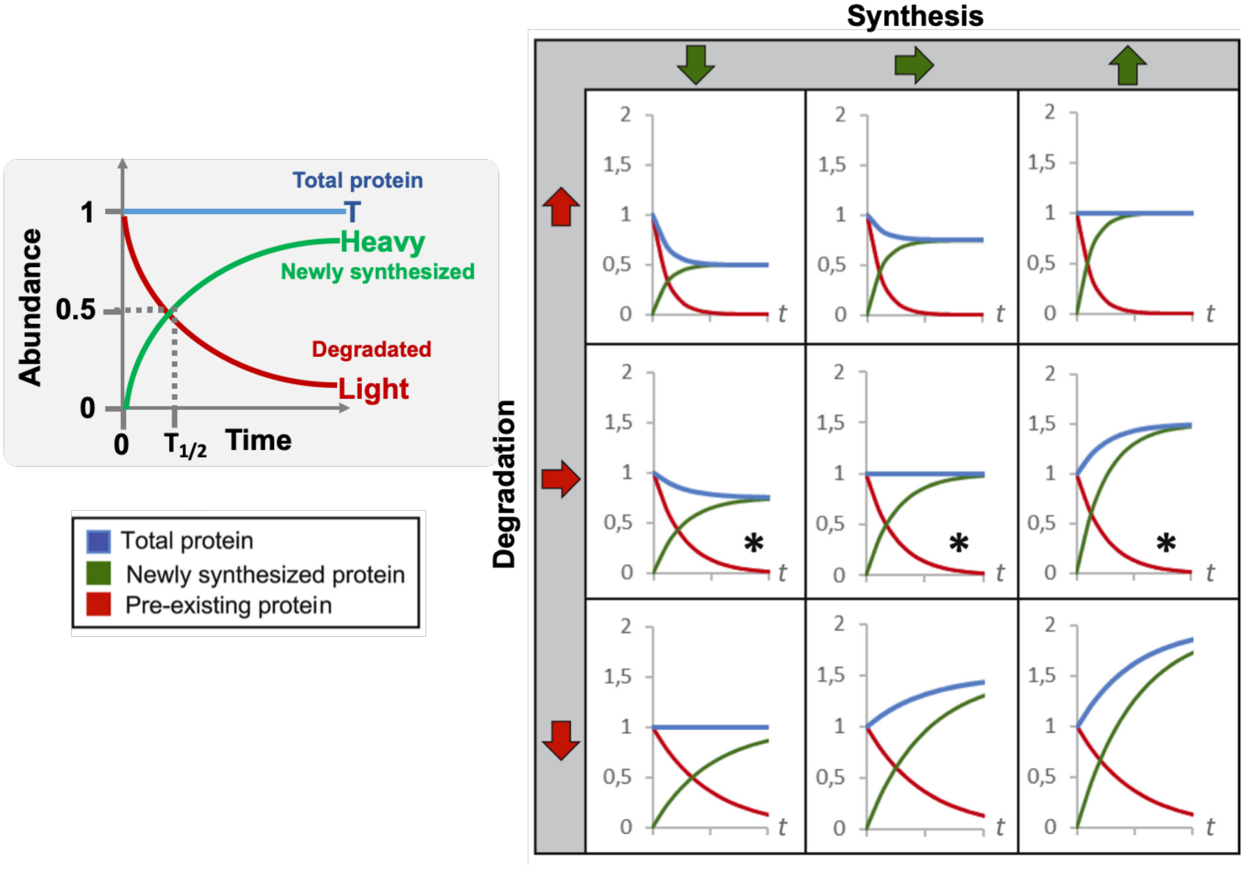


Figure 8: Schematic illustration of interdependencies between parameters of turnover. At a steady-state of a cell, the assumption is that the amount of protein synthesized equals the amount of protein degraded, leading to the constant *“total protein”* (T) in blue. Usually, in such experiments, cells are cultured in SILAC (stable isotope labeling in cell culture) medium containing amino acids labelled with light (L) isotopes and switched to SILAC medium containing amino acids labelled with heavy (H) isotopes for different time periods. From this, either linear regression or ratio-based models can be generated by building ratios between heavy (green) / light (red) (*“synthesis rate”* / *“degradation rate”)* labeled protein intensities representing the *“half-live ($T_{1/2}$)”*. The asterisks highlight the constant *“degradation rate”* at the same level (resulting to the same half-live), however, absolute protein amount could change as the constant synthesis rate changes. Copied and modified from [95].

The half-live is not only an intrinsic feature of a particular protein but is among others also dependent on the cell type, post-translational modifications, organelle affiliation, cellular environment, and can range between few seconds to several weeks [104, 105]. The turnover is constant under a steady-state; however, synthesis and degradation rates can be adjusted to external or internal stimuli based on dynamic responses. Such proteostatic mechanisms belong to one of the crucial cellular processes, ensuring the maintenance of functional proteins at their respective locations, as well as proper concentrations by removing aged and misfolded proteins from the cellular protein pool [95]. In eukaryotic cells, degradation of proteins is either carried out via the ubiquitin proteasome degradation pathway or via lysosomal proteolysis. Lysosomes receive their cargo via vesicles such as fusion of endocytic vesicles and autophagosomes with lysosomes [101].

Proteomic studies in the majority of experiments aim to screen for differentially regulated proteins across various cellular states. However, the analysis of synthesis and degradation rates over time would provide another layer of crucial information to quantitatively proteomic approaches, especially as the regulation of the proteome to various stimuli is a dynamic process [104]. For instance, with the application of “pulse-chase experiments” with isotope-labeled amino acids, protein turnover can be investigated. Classically, such experiments are performed with radioactive amino acids and immunoprecipitation of the protein of interest, restricting individual analyses to the investigation of single proteins. On the other hand, pulse stable isotope labeling by amino acids in cell culture (pSILAC), combined with mass spectrometry (MS) based proteomics, enables the study of protein half-lives on a global scale [104, 106]. In an ideal case, the experimental setup would allow determining quantitative proteome differences, the determination of half-lives, and tracking synthesis and degradation rates separately.

One of the commonly applied approaches for the determination of half-lives includes the generation of protein abundance ratios between the SILAC heavy and light channels (heavy/light). In order to additionally determine synthesis and degradation rates, the experimental setup has to be extended with the inclusion of a SILAC medium channel [104], or by combining pSILAC with isobaric labeling further discussed in the review [95]. In conventional pSILAC approaches, the biggest challenge is related to the experimental variability arising from sample preparation and MS measurements. To circumvent the issue, pSILAC can be combined with data independent acquisition (DIA) MS/MS, allowing to reproducibly label-free quantify the individual channels over time. The combination of pSILAC with DIA MS/MS, hence allows to determine half-lives of proteins as well as the constants synthesis and degradation rates separately, an approach which has been yet recently only proposed, although not applied [107].

6.1.1 Aims of the Study

Following goals were covered within this project:

- Investigation of protein half-lives of HEK293 WT cells with a focus on lysosomal proteins.
- Establishment of conditions impairing lysosome and proteasome function.
- Extending the investigation of protein half-lives under five different perturbational states.
- Generation of a phosphoproteomic resource dataset of lysosome enriched fractions for future protein half-life correlation analysis.

6.2 Materials and Methods

For a more detailed list please refer to the key resource table and method section of chapter 3.

6.2.1 Consumables and Equipment

- Analytical balance, Sartorius (Göttingen, Germany)
- Balance scale, PC4400 digital top load precision, Mettler Toledo (USA)
- Bis-Tris (4-12 %) gradient gel, Thermo Fisher Scientific (Bremen, Germany)
- Cation-SR Empore^{TM} Extraction Disks, Supelco Merck KGaA (Darmstadt, Germany)
- C18 octadecyl Empore^{TM} Extraction Disks, Supelco Merck KGaA (Darmstadt, Germany)
- Cell counter, automatic cell counter Nano EnTek (St. Ingbert, Germany)
- Cell culture dishes (100 cm), Sarstedt (Nümbrecht, Germany)
- Cell Scraper (25 cm), Sarstedt (Nümbrecht, Germany)
- Centrifuge, Eppendorf 5424 R (Hamburg, Germany)
- Centrifuge, Heraeus Labofuge 400 (Hanau, Germany)
- Clean bench, CA/R6 Clean Air (Minneapolis, USA)
- Counting chamber, Fuchs-Rosenthal, marienfeld superior (Laida-Koenigshoven, Germany)
- Cryovials, Nalgene Nunc (Wiesbaden, Germany)
- Glassware (beakers, bottles, measurement cylinders) Carl Roth GmbH and Co. KG (Karlsruhe, Germany)
- Ice-machine, Ziegra (Isernhagen, Germany)
- Incubator, Heracel 240i (Bremen, Germany)
- Incubator, VWR collection INCU-line digital mini incubator (Leuven, Belgien)
- Orbitrap Fusion Lumos mass spectrometer, Thermo Fisher Scientific (Bremen, Germany).
- Microcentrifuge, MiniStar silverline, VWR (Leuven, Belgien)
- Micropipette, Eppendorf (Wesseling-Berzdorf, Germany)
- Microplate 96-well, BD Falcon (Heidelberg, Germany)
- Microplate reader, Genios Tecan (Männedorf, Switzerland)
- Microscope Zeiss Axio Lab A1, Carl Zeiss (Oberkochen, Germany)
- Miltenyi LS column, Miltenyi Biotec (Bergisch Gladbach, Germany)
- Parafilm, Pechiney Plastic packaging (Brampton, Canada)
- Pasteur pipettes (230 mm), BRAND (Wertheim, Germany)
- pH-Meter, Calimatic 761, Knick (Berlin, Germany)
- Pipette tips, maximum recovery, Axygen (Union City, USA)
- Plastic conical tubes (15 ml / 50 ml), Sarstedt (Nümbrecht, Germany)
- Plastic pipette (5 ml / 10 ml / 25 ml), Sarstedt (Nümbrecht, Germany)
- Plates (96-well), BD falcon (Heidelberg, Germany)
- PowerPac basic power supply, Biorad Laboratories GmbH (Munich, Germany)
- Quadro MACS miltenyi magnet, Miltenyi Biotec (Bergisch Gladbach, Germany)
- ReproSil-Pur 120 C18-AQ, 1.9 µm, Dr. Maisch GmbH, Ammerbuch-Entringen, Germany
- Shaking water baths, Thermolab (Burgwedel, Gernmany)
- Sonicator, BRANSON 2510, Branson (Danbury, USA)
- Thermomixer Comfort, Eppendorf (Wesseling-Berzdorf, Germany)
- Vortex, Ika vortex 3, Merck KGaA (Darmstadt, Germany)

6.2.2 Chemicals and Solutions

- 1,4-Dithiothritol (DTT), Sigma (Seelze, Germany)
- 3-(N-morpholino)propanesulfonic acid (MOPS), Merck KGaA (Darmstadt, Germany)

- 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, AppliChem (Darmstadt, Germany)
- Acetic acid, Biosolve (Dieuze, France)
- Acrylamide, Merck KGaA (Darmstadt, Germany)
- Ammonium acetate (AA), Merck KGaA (Darmstadt, Germany)
- Ammonium persulfate (APS), Carl Roth GmbH and Co. KG (Karlsruhe, Germany)
- Bio-Rad DC^{TM} protein assay reagent kit, BioRad (Hercules, USA)
- Bovine serum albumin (BSA), Sigma (Seelze, Germany)
- Calcium chloride ($CaCl_2$), Merck KGaA (Darmstadt, Germany)
- Chloroacetamide, Merck KGaA (Darmstadt, Germany)
- Dimethyl sulfoxide (DMSO), Sigma-Aldrich (Munich, Germany)
- Dulbecco’s modified Eagle’s media (DMEM), without L-glutamine and sodium pyruvate Invitrogen Gibco (Paisley, UK)
- Dulbecco’s phosphate-buffered saline (DPBS), without magnesium (Mg) and calcium (Ca), Gibco (Paisley, UK)
- EDTA-free Protease Inhibitor Cocktail (Mini), Merck KGaA (Darmstadt, Germany)
- Ethanol, Hoffman Lamson (Bentleyville, USA)
- Ethylenediaminetetraacetic acid (EDTA), Sigma (Seelze, Germany)
- Fetal calf serum (FCS), Invitrogen, Gibco (Paisley, UK)
- Formic acid (FA), Biosolve (Dieuze, France)
- Glycine, Carl Roth GmbH and Co. KG (Karlsruhe, Germany)
- Instant Blue colloidal Coomassie stain (Thermo Fisher Scientific (Rockford, USA)
- L-glutamin, Sigma (Seelze, Germany)
- Magnesium acetate (MgAc), Merck KGaA (Darmstadt, Germany)
- Magnesium chloride ($MgCl_{2)},$ Merck KGaA (Darmstadt, Germany)
- Magnetite dextran (magnetic / pulse media) coated 10 nm particle, Liquids Research Ltd (Bangor, UK)
- Methanol, Merck KGaA (Darmstadt, Germany)
- N,N,N’,N’-Tetramethylethylenediamine (TEMED), Carl Roth GmbH and Co. KG (Karlsruhe, Germany)
- PageRule prestained protein ladder (Marker - 10 - 180 kDa), Thermo Fisher Scientific (Rockford, USA)
- Para-nitrophenyl-N-acetyl-2-B-D-glucosaminide (β-hex substrate), Merck KGaA (Darmstadt, Germany)
- Pepstatin A, PanReac AppliChem (Germany)
- Potassium chloride (KCl), Merck KGaA (Darmstadt, Germany)
- Potassium phosphate (K_2HPO_4), Merck KGaA (Darmstadt, Germany)
- Sodium chloride (NaCl), Merck KGaA (Darmstadt, Germany)
- β-mercaptoethanol, Merck KGaA (Darmstadt, Germany)
- Sucrose, Carl Roth GmbH and Co. KG (Karlsruhe, Germany)
- Trifluoric acid (TFA), Biosolve (Dieuze, France)
- Triton-X-100, Carl Roth GmbH and Co. KG (Karlsruhe, Germany)
- Trypsin, Invitrogen, Gibco (Paisley, UK)

6.2.3 Software

- FusionCapt Advance Solo 4 (Vilbert Lourmat (Collégien, France))
- GraphPad Prism (San Diego, CA, USA)
- MaxQuant (Max Planck Institute of Biochemistry (Planegg, Germany))
- Proteome Discoverer (Thermo Fisher Scientific (Waltham, USA))
- R studio (R Core Team, 2020)
- Spectronaut (Biognosys (Schlieren, Switzerland))

6.2.4 Methods

6.2.4.1 Pulse Stable Isotope Labelling by Amino Acids in Cell Culture (pSILAC)

Poly L-lysine (PLL) (0.5 mg/mL in phosphate buffered saline (PBS)) was used to coat 42 tissue culture plates / condition (10 cm dishes) for 20 min at 37 °C. Subsequently, plates were washed three times with pre-warmed PBS and human embryonic kidney cells 293 (HEK293) ($6x10^6$ cells / 10 cm dish) seeded on PLL coated plates and cultured for 48 h at 37 °C and 5 % CO_2 in stable isotope labelling by amino acids (SILAC) Dulbecco's modified essential media (DMEM) supplemented with dialyzed fetal calf serum (FCS-10 % v/v) + glutamine (2 mM) + penicillin (100 IU/mL) + streptomycin (100 µg/mL), light or $^{13}C_6{}^{15}N_2$ labeled lysine, and 87.8 mg/mL light or $^{13}C_6{}^{15}N_4$ labeled arginine. Per experimental condition in total 42x 10 cm dishes with $6x10^6$ HEK293 cells were seeded, corresponding to 7 time points with three biological replicates / time point (2 plates / time point). Cells were seeded in SILAC media containing amino acids labeled with light isotopes supplemented with 10 % (v/v) magnetite dextran solution (superparamagnetic iron oxide nanoparticles (SPIONs)) and incubated for 24 h at 37 °C and 5 % CO_2. To remove the magnetite dextran solution, cells were washed three times with pre-warmed PBS before being re-supplemented with SILAC DMEM and incubated for 24 h. Cells were washed three times with pre-warmed PBS and the SILAC media containing amino acids labeled with light isotopes was replaced with SILAC medium containing amino acids labelled with heavy isotopes for 1, 2, 4, 8, 16, and 32 h, whereas the time point 0 h served as a control (no SILAC medium containing amino acids labelled with heavy isotopes).

6.2.4.2 Enrichment of Lysosomes

Please refer to the method section of chapter 3.

6.2.4.3 β-hexosaminidase Activity Assay

Please refer to the method section of chapter 3.

6.2.4.4 Pepstatin A Treatment and Cathepsin D Activity Assay

HEK293 cells ($3.5x10^6$ / 10 cm dish) were treated and incubated with 40 µM Pepstatin A (PSTA - dissolved in DMSO) for 48 h at 37 °C and 5 % CO_2, while the control cells were not treated with PSTA (supplemented with the same amount of DMSO). Subsequently, cells were harvested and lysed in ice-cold lysis buffer (500 mM NaCl, 50 mM Tris – HCl, 2 % Triton X-100, 0.4 % sodium dodecyl sulfate (SDS), pH 7.4) for 30 min on ice, centrifuged (20,000 x g), followed by protein measurement of the supernatant (DC-assay). 10 µL of samples with equal protein amounts (10 µg) were loaded onto a black 96-well plate, followed by the addition of 90 µL of acetate buffer (500 mM sodium acetate, 1 mM EDTA, 0.05 % Brij35, pH 4.0) and 11.11 µL of the substrate MCA-GKPILFFRLK-D-NH2 (final concentration: 20 µM) to each sample. The plate was incubated for 6 h at 37 °C. The fluorescence was measured at an excitation wavelength of 330 nm and an emission wavelength of 405 nm at 37 °C on a Tecan pro microplate reader (number of reads 3). Activity assays were performed in three biological replicates. Data analysis was performed in GraphPad Prism and Excel, where the readings were normalized against the control sample. Performed by H.D.

6.2.4.5 Bortezomib Treatment and Proteasome Activity Assay

The experimental setup was based on the "Proteasome 20S Activity Assay Kit" (Sigma-Aldrich, Munich, Germany). HEK293 cells (80,000 / well) were seeded in a 96-well plate. After 24 h, cells were treated with 100 nM Bortezomib (BTZ), while the control cells were not treated with BTZ. After the treatment, 100 µL of the proteasome assay loading solution was added to each well. The plate was incubated overnight at 37 °C and protected from light. The fluorescence was measured at an excitation wavelength of 490 nm and an emission wavelength of 525 nm at 37 °C on a Tecan pro microplate reader (number of reads 3). Activity assays were performed in three biological replicates. Data analysis was performed in GraphPad Prism and Excel, where the readings were normalized against the control sample. Performed by H.D.

6.2.4.6 Cell Viability Assay

The experimental setup was based on the “MTT cell viability assay” (Thermo Fisher, Rockford, USA). HEK293 cells (40,000 / well) were seeded in a 96-well plate. After 24 h, cells were treated with 100 nM BTZ and incubated for different time lengths (32, 16, 8, 4, 2, and 1 h) at 37 °C and 5 % CO_2. Control cells were not treated with BTZ. After the treatment, DMEM was discarded, and 50 µL FCS-free DMEM and 50 µL MTT Reagent were added to each well. A blank consisting of FCS-free DMEM and MTT reagent was prepared. The plate was incubated at 37 °C for 3 h. After the incubation, 150 µL of MTT solvent was added to each well and the plate was shaken on an orbital shaker for 15 min. The absorbance was measured at 590 nm on a Tecan pro microplate reader (number of reads 3). Data analysis was performed in GraphPad Prism and Excel, where the readings were normalized against the control sample. Performed by H.D.

6.2.4.7 Generation of a CRISPR-Cas9 Niemann-Pick type C1 Gene Knock out Cell Line

HEK293 cells (150,000 / well) were seeded in a 24-well plate in DMEM supplemented with fetal calf serum (FCS-10 % v/v) + glutamine (2 mM) + penicillin (100 IU/mL) + streptomycin (100 µg/mL) and grown for 24 h at 37 °C and 5 % CO_2. For transfection, clustered regularly interspaced short palindromic repeats ribonucleic acid (crRNA - target sequence: ATGGTGCAGT TCGTGTTATACGG; protospacer adjacent motif (PAM): CGG, Exon: 9), tracerRNA and Cas9 plasmids (all three from Dharmacon (Horizon, Cambridge, UK)) were dissolved in 10 mM Tris-HCl (pH 7.5) buffer to a final concentration of 10 µM, 10 µM and 100 ng/µL, respectively. All three reagents were mixed together and incubated for 30 min at room temperature. Prior to transfection, the mixture of crRNA, tracerRNA, Cas9 plasmid, and DharmaFECT transfection solution (Dharmacon) were added to 500 µL DMEM media, resulting in final concentrations of 50 nM, 50 nM, 2 ng/µL and 6 ng/µL. Subsequently, cells were transfected and incubated with the mixture for 48 h at 37 °C and 5 % CO_2, while for control purposes (negative control) one well was only supplemented with the Cas9 plasmid. Following transfection, the DMEM medium was replaced against fresh DMEM media supplemented with 2 µg/mL puromycin (except the negative control) and the cells again incubated for 48 h at 37 °C and 5 % CO_2. After 48 h, cells were diluted (1:5) in DMEM media supplemented with puromycin (2 µg/mL) and further cultivated for 24 h at 37 °C and 5 % CO_2. Followed by the limiting dilution, singly selected cells were cultured to 90 % confluency and split to three 96 well plates, while two were frozen at -80°C in DMEM media (20 % FCS, 10 % DMSO) and the other was lysed in 50 µL lysis buffer (1 mM $CaCl_2$, 3 mM $MgCl_2$, 1 mM EDTA, 0.2 mg/mL Proteinase K, 1 % Triton X-100 and 10 mM Tris-HCl) to extract genomic DNA. Polymerase chain reaction (PCR) was performed followed by sequencing and western blot analysis to verify the clones with a knock out of the NPC1 gene. Performed in parts with the assistance of A.M.

6.2.4.8 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Please refer to the method section of chapter 3.

6.2.4.9 Western-blotting

Please refer to the method section of chapter 3.

6.2.4.10 Proteolytic Digestion

Please refer to the method section of chapter 3.

6.2.4.11 Peptide Assay

The experimental setup was based on the protocol “Pierce Quantitative Fluorometric Peptide Assay” by Thermo Fisher. In brief, a peptide assay standard was diluted freshly in the following concentrations: 1000; 500; 250; 125; 62.5; 31.3; 15.6 and 7.8 µg/mL. The peptides, dissolved in 5 % ACN, were diluted 1:10. 10 µL of each sample and standard were pipetted into a black 96-well microtiter plate. To each well 70 µL of fluorometric peptide, assay buffer were added, and afterwards, 20 µL of fluorometric peptide assay reagent were added to each well. The microtiter plate was incubated for 5 min at RT. The fluorescence was measured using a Tecan Pro microplate reader applying an excitation wavelength of 390 nm, and an emission wavelength of 475 nm. Data analysis was performed in Excel.

6.2.4.12 Phosphopeptide Enrichment

Phosphopeptides were enriched from 1.8 mg of urea in solution digested, Oasis cartridges desalted lysosome enriched protein digest by the SMOAC (Sequential enrichment using Metal Oxide Affinity Chromatography) approach. The procedure utilizes both the High-SelectTM TiO_2 Phosphopeptide Enrichment kit (Thermo Scientific, Darmstadt) and the High-SelectTM Fe-NTA Phosphopeptide Enrichment kit (Thermo Scientific, Darmstadt). Briefly, for TiO_2 enrichment the sample was reconstituted in 150 µL of the provided binding/equilibration buffer and then loaded on an equilibrated TiO_2 spin tip and passed through two times for 5 min by centrifugation at 1000xg. Then the spin tip was washed with 20 µL binding/equilibration buffer followed by 20 µL wash buffer at 3,000xg for 2 min each. The wash procedure was repeated once. Finally, the spin tip was washed once with 20 µL water and then eluted 2x with 50 µL phosphopeptide elution buffer. The eluate was immediately dried in a vacuum centrifuge to avoid phosphopeptide degradation. For the subsequent Fe-NTA enrichment, the combined flow-through and wash fractions from the TiO_2 enrichment were dried in a vacuum centrifuge, dissolved in 200 µL of binding/wash buffer and applied to an equilibrated IMAC spin column. The sample was mixed with the IMAC resin by column tapping and incubated for 30 min to allow phosphopeptide binding. Sample mixing was repeated every 10 min. After centrifugation at 1,000xg for 30 sec, the IMAC column was washed three times with 200 µL of binding/wash buffer, followed by 200 µL water. Finally, phosphopeptides were eluted with 2x100 µL elution buffer and afterwards immediately dried in a vacuum centrifuge. Samples were subsequently resuspended in 20 µL of 5 % ACN / 5 % FA from which 25 % were measured on the Orbitrap Fusion Lumos. Data analysis was performed via Proteome Discoverer and MaxQuant. Performed in parts with the assistance of R.H.

6.2.4.13 Liquid Chromatography Electrospray Ionization Mass Spectrometry

Please refer to the method section of chapter 3.

6.2.4.14 Spectronaut Analysis

Please refer to the method section of chapter 3. Modifications: For the generation of a spectral library, in the labeling setting the “Labelling Applied” option was enabled, as well as the SILAC labels (“Arg10” and “Lys8”) were specified in the second channel. To ensure a complete heavy/light labeling of the whole library the “In-Silico Generate Missing Channels” option in the workflow settings was enabled. Data were exported from Spectronaut in a csv. format and further processed in software including Excel, Prism, R-studio.

6.2.4.15 Normalization and Post Data analysis

Data from Spectronaut was imported and analyzed using R Version 4.0.2 (R Core Team, 2020) with the integrated development environment RStudio (Version 1.3.1056). For data importing, tidying, transforming and statistical modeling tidyverse [108], and R base packages were used. Computed linear and nonlinear models were tidied with “Broom” (R-package). Advanced normalization steps (normalization between arrays) were performed using “limma” [109] with BiocManager to manage Bioconductor packages. For data visualization “ggplot2” (R-package) was utilized. First, protein groups and the corresponding non-normalized signal intensities of the individual MS2 channels (MS2 channel 1 = light, and MS2 channel 2 = heavy) were imported. Next, explicit and implicit missing values were replaced by zero value. Implicit missing values were made transparent by pivoting steps between long and wide data format. Based on these results the sum of both individual channels was computed and defined as total signal for each protein group and condition. Here, a condition consists of a treatment and a time point. All non-zero intensities were log2 transformed and otherwise left unchanged. After inspection of the resulting intensity distributions, all processed intensities, which had values lower than one, were considered as signal noise and replaced by “NA”. Data was filtered based on the total intensities and only entries with at least two valid values within a condition were kept. Subsequent data normalization was conducted based on the total intensities. Normalization methods (e.g.: non-normalized (log2 transformed), quantile, median scaling, Loess) were separately fit for each treatment. Resulting normalization factors were applied to transform (normalize) the intensities of light and heavy channel individually. All normalization approaches assume that the majority of total expressions (i.e. summed protein amount of light (degradation) and heavy (synthesis)) is stable across the analyzed time points. Based on the visual inspection of a variety of descriptive statistical measures (e.g.: corresponding MA plots, PCA plots, boxplots, histograms) the normalization results were compared to one another and to the non-normalized log2 transformed data. Based on these findings, quantile normalization showed the best and most reasonable performance in removing technical biases. Following analysis was conducted based on quantile normalized data. Half-lives were estimated by fitting linear models to the data. Here, the half-life is defined as time at which the amount of light and heavy labeled protein per protein group are equal. In this case, the ratio of the intensities between the non-logarithmic intensities of the two channels is one. In brief, the ratios of the non-logarithmic intensities (heavy/light) were computed and represented the dependent response variable contrary to the time in hours as independent explanatory variable. To obtain linearity, the ratios were transposed with the natural logarithm + 1. Next, linear regression was applied to the prepared data and based on the slope estimates and considering the transformed ratios the half-lives were computed. In addition, R^2 values were computed to evaluate the quality of the models. Only for visualization purposes and better comparability, the data was scaled to the intensity of the time point zero of the light channels for each protein group and treatment. Performed in parts with the assistance of M.S.

6.3 Results and Discussion

6.3.1 Investigation of Protein Half-Lives of HEK293 WT Cells

As lysosomes are aside from the ubiquitin proteasome degradation pathway, the major degradative organelle of the cells, they play a crucial role in maintaining cellular homeostasis by degrading proteins [2]. Hence, there is a great interest in a better understanding of their role in protein degradation and investigation of their protein half-lives. Numerous large-scale turnover studies performed on the whole-cell level did not allow for reproducible half-life determination of the low abundant lysosomal proteins [103, 104, 106, 107, 110]. Therefore, to study the role of lysosomes in protein degradation as well as half-lives of lysosomal proteins on a large scale, the pSILAC DIA MS/MS [107] approach was applied. Due to the overall low abundance of lysosomal proteins, lysosomes were enriched by superparamagnetic iron oxide nanoparticles (SPIONs) from HEK293 cells, allowing for a comprehensive coverage of the lysosomal proteome [34, 44]. Preliminary experiments (results not shown) led to the development of a 48 h workflow from the day of cell seeding in SILAC cell-culture growth medium containing amino acids labelled with light isotopes (Lys0, Arg0) to the day of lysosome enrichment. The overall workflow was based on seven different incubation lengths (0, 1, 2, 4, 8, 16, 32 h) of SILAC medium containing amino acids labelled with heavy isotopes (Lys8, Arg10) with three biological replicates per time point **(Figure 9)**.

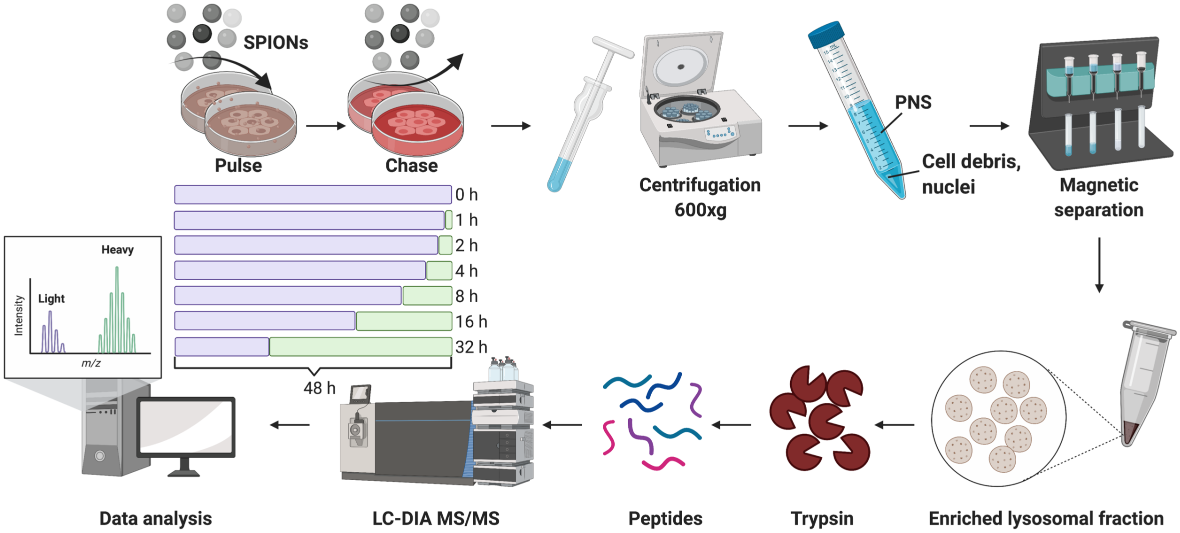


Figure 9: Scheme of pSILAC DIA MS/MS workflow. Lysosome enrichments were performed from HEK293 cells. Cells were incubated with superparamagnetic iron oxide nanoparticles (SPIONs) for 24 h (pulse) followed by a 24 h chase for delivery to the lysosomes. The bar graph in purple shows the incubation time with medium containing amino acids labelled with light isotopes and in green medium containing amino acids labelled with heavy isotopes. On the day of enrichment, cells were harvested, homogenized and the post-nuclear supernatant (PNS) passed through columns retaining lysosomes in a magnetic field. Lysosomal fractions were proteolytically digested with trypsin, samples desalted, and measured on an Orbitrap Fusion Lumos, followed by data analysis. **Abbreviations:** SPIONs: superparamagnetic iron oxide nanoparticles, LC-DIA MS/MS: liquid chromatography data independent acquisition tandem mass spectrometry, PNS: post nuclear supernatant.

MS analysis of enriched lysosomal fractions from HEK293 WT cells (21 DIA and 7 DDA runs) resulted in the identification of about 510 lysosomal and 5,600 proteins associated to other cellular organelles **(Figure 10A, B)**. To calculate half-lives based on the pSILAC DIA MS/MS approach, the assumption is the following: The synthesis rate of a protein is tracked based on the increase of SILAC heavy labeled peptides abundances, while the rate of degradation corresponds to the decrease of SILAC light labeled peptides abundances. Two methods were proposed for the determination of protein half-lives: 1) In the first method (linear regression model), the ratio of SILAC labeled peptide abundances for the heavy and light SILAC channels for the respective protein are calculated and plotted against time. The time where the protein abundance ratio (heavy/light) equals 1 is defined as the half-life of the protein [111] **(Figure 10C)**. In the second method individual SILAC labeled peptide abundances for the heavy and light SILAC channels for the respective protein were plotted against the time. The intersection point of the two profiles corresponds to the time where 50 % of the protein is turned over (=half-life) [103, 106] **(Figure 10D)**. Applying the first method, half-lives for 475 lysosomal and 4,173 proteins associated to other cellular organelles were determined resulting in a median half-life of 17 and 16 h for the lysosomal and rest of the proteome respectively **(Figure 10E, F)**. Further, the top five proteins with the longest and shortest half-lives for non-lysosomal and lysosomal proteins were shown **(Table 1)**.

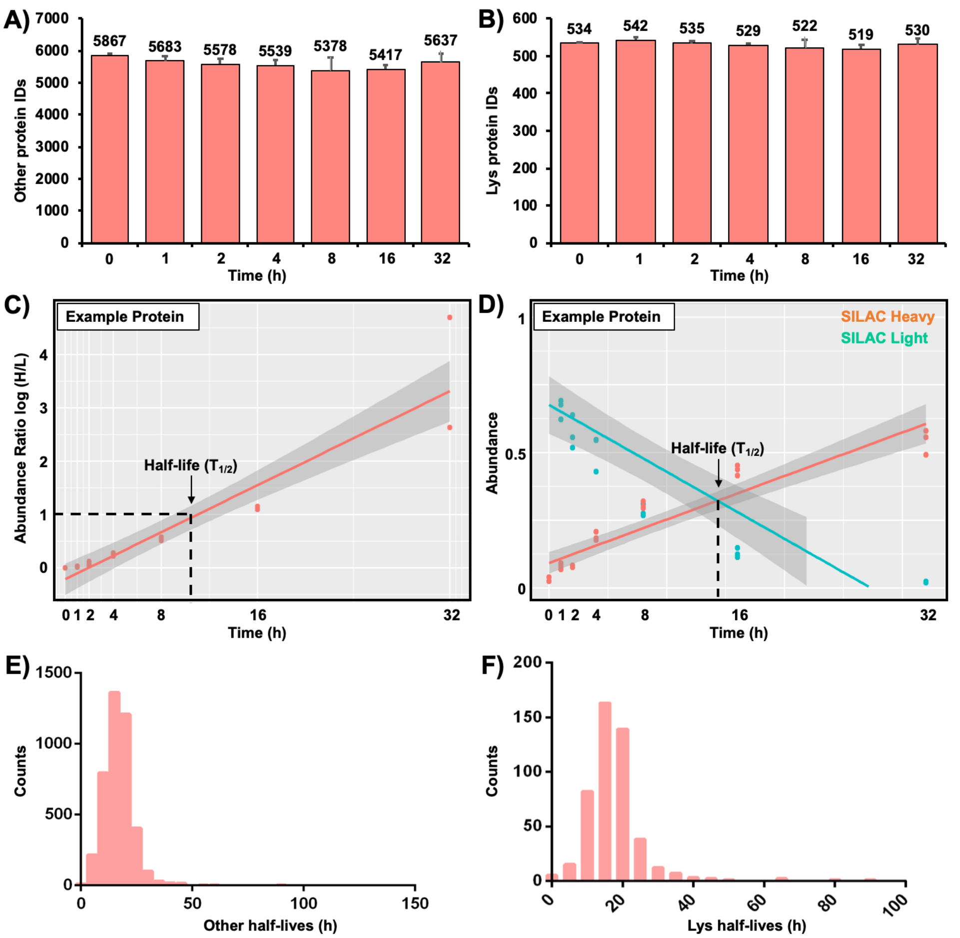


Figure 10: Determination of protein half-lives from lysosome enriched fractions (HEK293-WT cells). A/B) Bar graphs showing proteomic data inspection (protein IDs) of lysosome enriched fractions from HEK293-WT cells. Lysosomes were enriched from a total of 43 plates of 10 cm HEK293 cells using SPIONs cultured in medium containing amino acids labelled with light (L) isotopes and switched to SILAC medium containing amino acids labelled with heavy (H) isotopes for different time periods for seven different time points including 0, 1, 2, 4, 8, 16, and 32 h pulse with three replicates each. Data represents 21 pSILAC DIA LC-MS/MS runs. Data was analysed in the Spectronaut software. **C)** Determination of protein half-lives applying the linear regression model, where the ratio of SILAC labeled peptide abundances for the heavy and light SILAC channels for the respective protein are calculated and plotted against time. The time point where the protein abundance ratio of the SILAC heavy/light channels equals 1 is defined as the half-life of the protein. **D)** Determination of protein half-lives by plotting abundances of SILAC labeled protein abundances for the heavy and light SILAC channels against the time. The intersection point of the two profiles corresponds to the half-life of the protein. **E/F)** Binned frequency distribution analysis of protein half-lives for identified non-lysosomal (other) and lysosomal proteins (Lys), applying a bin range of 5 h on the x-axis, plotted against the number of identified proteins (counts). **Abbreviations**: IDs: identifications, H: heavy, L: light.

Table 1: Example of proteins (top 5) with the longest and shortest half-lives for non-lysosomal and lysosomal proteins.

Other half-lives (h) - Longest		Lys half-lives (h) -Longest	
Other	**Time**	**Genes**	**Time**
RPL23	106	CMTM6	81
H3C1	90	ITFG2	66
DPYSL5	88	SOD1	64
SLC18B1	81	CTSL	50
CKB	75	SGSH	44
Other half-lives (h) - Shortest		**Lys half-lives (h) - Shortest**	
MRFAP1	1	ATP6AP2	4
NEO1	1	CTSV	5
RNF149	2	SPPL2B	6
MORF4L2	2	GJA1	6
TGOLN2	2	SLC39A14	6

6.3.2 Establishment of Conditions Impairing Lysosome and Proteasomal Function

As the majority of performed large-scale turnover studies to date were performed on cells in a steady-state, the turnover study was further extended to conditions impairing lysosome and proteasome function – Both organelles associated to the two main systems for degradation of proteins associated to the autophagy-lysosome pathway and ubiquitin-proteasome system, respectively. This includes the impairment of 1) autophagosome-lysosome fusion (autophagy-related gene 5 – knock out (ATG5-KO) cells), 2) lysosomal degradation (pepstatin A (PSTA) treatment), 3) cholesterol homeostasis (Niemann-Pick cholesterol transporter 1 – knock out (NPC1-KO) cells), 4) protein glycosylation (multiple glycosylation machinery related genes knock out (MGM-KO) cells), and 5) proteasomal degradation (bortezomib (BTZ) treatment).

6.3.2.1 Verification of ATG5-KO Impairing Autophagosome-Lysosome Fusion

Substrates targeted for degradation are delivered to the lysosomes by vesicle transport, such as through endosomes (for extracellular material), autophagosomes (for intracellular material), or alternatively by chaperone mediated autophagy or endosomal microautophagy [2]. To identify proteins delivered by the autophagy-lysosome pathway, an ATG5-KO (autophagy-related gene 5 – knock out) cell line impairing autophagosome-lysosome fusion was used [112]. Autophagosomes are formed by the development and growth of a double-layered isolation membrane in the cytoplasm, which is enabled by specific autophagy proteins such as the microtubule-associated protein light chain 3 (LC3), engulfing cargo to be degraded [113]. Followed by the closure of the autophagosomal membrane, the lysosomes and autophagosome fuse via the action of different proteins like soluble N-ethylmaleimide-sensitive-factor attachment receptors (SNARES) [2]. In this process ATG5 interacts with ATG12 and ATG16 to convert LC3-I into a membrane bound form called LC3-II. The conversion of LC3-I into LC3-II is accompanied by a change in molecular weight, which is frequently used as a marker for autophagy [113]. Thus, knock out of ATG5 blocks autophagy and should lead to a loss of LC3-II [112]. This was verified via western blot, while CTSD (lysosomal protein) served as positive control showing that equal amounts of samples were loaded on the gel **(Figure 11)**.

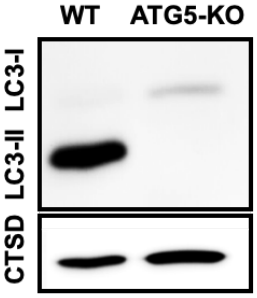


Figure 11: Verification of ATG5-KO. A) Western blot analysis of ATG5-KO cells by comparing LC3-II signal for 12 µg of wild type (WT) and knock out (KO) cells. LC3B antibody (detecting LC3-I and LC3-II) dilution 1:1000 (v/v) in 5 % BSA in TBST overnight. As a loading control, the membrane was stripped and incubated with the cathepsin D (CTSD) antibody (1:2000 (v/v) ratio in 5 % BSA in TBST for 2 h). **Abbreviations:** WT: wild type, ATG5-KO: autophagy related gene knock out 5, CTSD: cathepsin D, BSA: bovine serum albumin, TBST: Tris-buffered saline Tween 20.

6.3.2.2 Verification of CTSD Inhibition Impairing Lysosomal Degradation

Cathepsin D (CTSD) belongs to one of the most abundant lysosomal luminal hydrolases with the main function to degrade proteins [44, 114]. In order to determine the substrates degraded within the lysosomes, lysosomal degradation was impaired applying the inhibitor pepstatin A (PSTA) [115]. PSTA inhibits acid proteases such as CTSD and CTSE in a pseudo-irreversible manner (i.e. it does not bind covalently to the enzymes), significantly impairing lysosomal protein degradation. In order to validate the extent of inhibition, a cathepsin D activity assay was performed in lysates of HEK293 cells, which were treated with 40 µM pepstatin A for 6 h (results leading to the development of the workflow are not shown). This resulted in a 80 % inhibition of the CTSD activity compared to the non-treated control sample (DMSO) **(Figure 12)**.

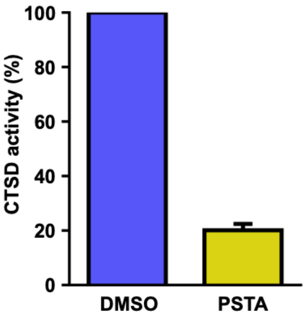


Figure 12: Verification of CTSD inhibition. Bar graph with standard deviation showing average readings from three biological replicates for HEK293 cell lysates either treated with 40 µM pepstatin A or DMSO (control) for 48 h. Assay readings have been normalized against the control. **Abbreviations:** CTSD: cathepsin D, DMSO: Dimethyl sulfoxide, PSTA: Pepstatin A.

6.3.2.3 Verification of NPC1-KO Impairing Cholesterol Homeostasis

Niemann-Pick cholesterol transporter 1 (NPC1) is a lysosomal transmembrane protein involved in trafficking of low-density lipoprotein (LDL)-derived cholesterol from lysosomes to the endoplasmic reticulum (ER) [31, 32]. A mutation or complete absence of this protein leads to the failure of cholesterol transport and its lysosomal accumulation, leading to the Niemann-Pick type C1 (NPC1) disease [116]. In order to investigate the underlying disease causing proteomic changes upon cholesterol accumulation, a NPC1-KO cell line was generated using the CRISPR-Cas9 system [117]. The knock out was verified via western blot by detecting NPC1, while TUBA1 (cytosolic protein) served as a positive control showing that equal amounts of samples were loaded on the gel **(Figure 13)**.

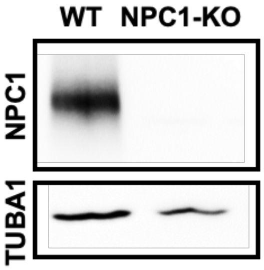


Figure 13: Verification of NPC1-KO. Western blot analysis of NPC1-KO cells by comparing NPC1 signal for 10 µg of wild-type (WT) and knock out (KO) cells. NPC1 antibody dilution 1:2000 (v/v) in 5 % BSA in TBST overnight. As a loading control, the membrane was stripped and developed with the alpha-tubulin antibody (1:5000 (v/v) ratio in 5 % BSA in TBST for 2 h). **Abbreviations:** WT: wild-type, NPC1-KO: Niemann-pick type C1 knock out, TUBA1: alpha-tubulin.

6.3.2.4 MGM-KO Cell Line Impairing Protein Glycosylation

The HEK293 MGM-KO (multiple glycosylation machinery related genes knock out) cell line of [C1GALT1C1 (C1GALT1 specific chaperone 1), GNPTAB (N-acetylglucosamine-1-phosphotransferase subunits alpha/beta), MGAT1 (alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase), and ALG3 (Dol-P-Man:Man(5)GlcNAc(2)-PP-Dol alpha-1,3-mannosyltransferase)] was kindly provided by our collaboration partners from the University of Copenhagen. The applied cell line (termed MGM-KO) in this study has not been published yet and was generated specifically for this project. The cell line was used to study whether and to which extent the impairment of glycosylation might influence the half-life of especially lysosomal proteins. The cell line is limited to produce rather homogenous truncated O-glycosylation (GalNAc-alpha1-Ser/Thr) through the knock out of C1GALT1 specific chaperone 1 specific to the C1GALT1 gene (A glycosyltransferase primarily responsible to build the core 1 O-glycan Gal-beta1-3GalNAc-alpha1-Ser/Thr, which serves as precursor for several extended O-glycans). The cell line is primarily supposed to express homogeneous mannose 3-5 N-glycan structures without mannose-6-phosphate.

6.3.2.5 Verification of Proteasome Inhibition Impairing Proteasomal Degradation

Proteins targeted for degradation by the ubiquitin-proteasome system reach the proteasome by ubiquitination - A tag of 76 amino acids covalently formed between the C-terminal glycine (G76) of ubiquitin and the sidechain of a lysine of its substrate proteins [118]. To determine which lysosomal associated proteins are degraded by the proteasome, a workflow for the inhibition of proteasomal activity using bortezomib (BTZ) [119] was established. BTZ is a dipeptide with leucine containing boronic acid applied for the inhibition of the proteasome, in which the boron atom binds chymotrypsin-like sites of the 26S proteasome. In order to validate the extent of proteasome inhibition, a proteasome activity assay was performed in lysates of HEK293 cells, which were treated with 100 nM for 16 h, as incubations longer than 16 h led to the death of the cells. This resulted to a 78 % inhibition of proteasomal activity compared to the non-treated control sample (DMSO) **(Figure 14)**.

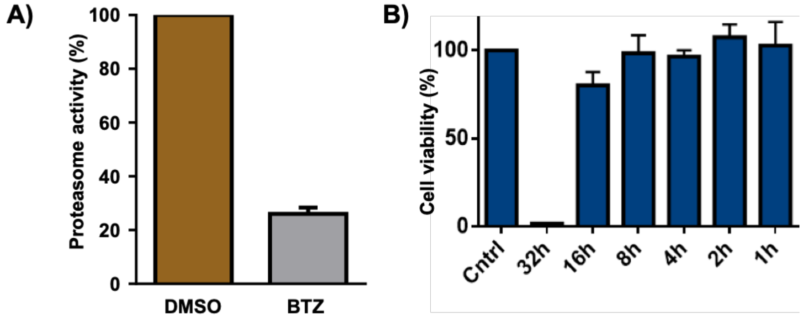


Figure 14: Verification of proteasome inhibition. A) Bar graph with standard deviation showing average readings from three biological replicates for HEK293 cells either treated with 100 nM BTZ or DMSO (control) for 16 h. **B)** Bar graph with standard deviation showing average readings from the cell viability assay for HEK293 cells incubated with 100 nM BTZ for different time lengths. Assay readings have been normalized against the control (DMSO). **Abbreviations**: DMSO: dimethyl sulfoxide, BTZ: bortezomib.

6.3.3 Quantification of Lysosome Enrichment and Lysosome Integrity

The efficiency of lysosome enrichments and their integrity was assessed using β-hexosaminidase assays, an enzyme, which resides inside the lumen of intact lysosomes [45]. The artificial substrate of this enzyme “para-nitrophenyl-N- acetyl-2-ß-D-glucosaminide” cannot pass the lysosomal membrane. Thus, when lysosomes are incubated under isotonic conditions the substrate is inaccessible for the enzyme β-hexosaminidase. Through the treatment of lysosomes with detergents such as Triton X-100, the lysosomal membrane can be lysed, leading to the release of the enzyme. Thus, activity determination in the presence of Triton X-100 treatment gives the total activity present in the sample, whereas activity under isotonic conditions only represents broken lysosomes. The activity difference (Triton X-100-Isotonic) thus reflects the intact lysosomes. β-hexosaminidase activity assays were performed for all six lysosomal conditions with three replicates/time points after lysosome enrichment. All 21 enzymatic assays/condition were normalized and summarized for each of the seven time points **(Figure S1-6).** The results show normalized enzymatic activity readings for the "input" fraction representing the post nuclear supernatant and the eluate fraction representing the lysosome enriched fraction. From the comparison of enzymatic activity readings among the different conditions from the input fraction one could conclude that overall, the lysosomal yield in the eluate fractions among all conditions was quite similar (20-40 %). However, the analysis showed that lysosomal integrity varied widely among the different conditions. Especially, for the MGM and NPC1-KO conditions, more lysosomes were broken in the input and the eluate indicating more fragile lysosomes **(Figure 15)**.

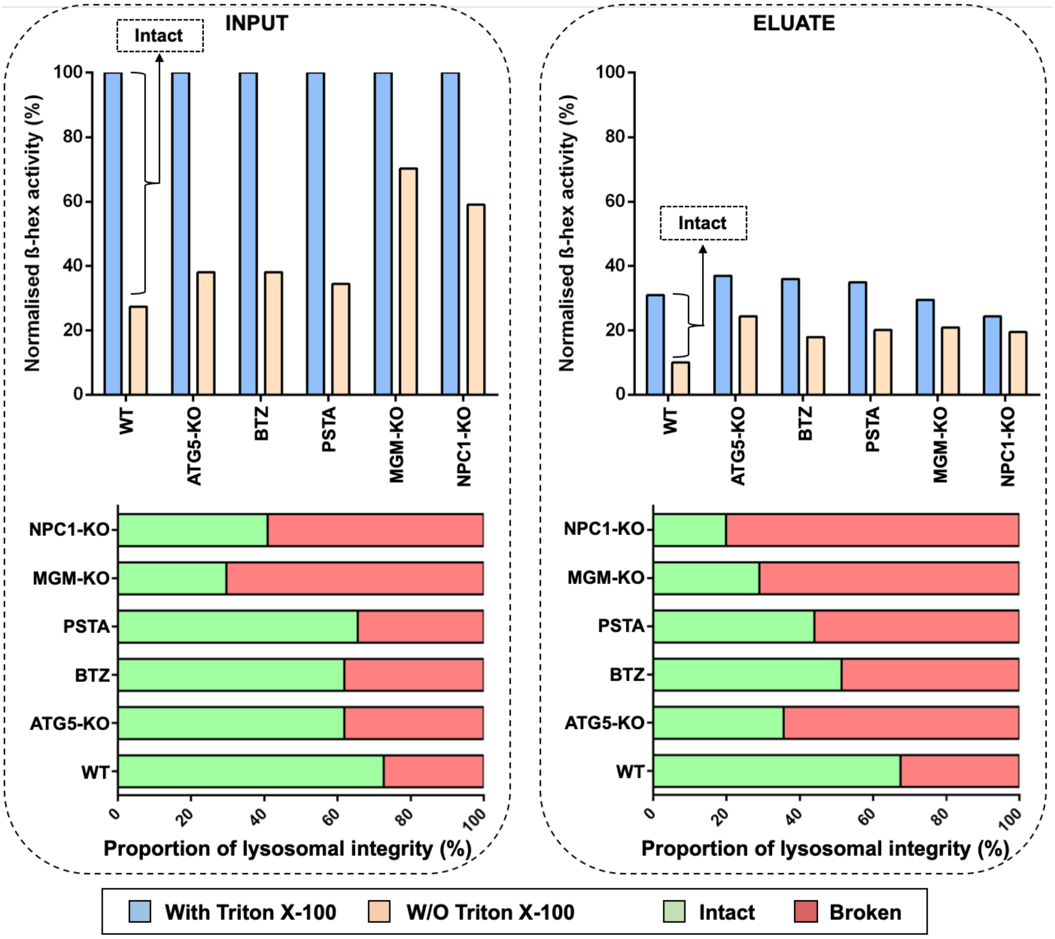


Figure 15: Summary of normalized β-hexosaminidase activity assay readings to compare lysosome enrichments and integrity. Shown are mean value for samples incubated with SILAC medium containing amino acids labelled with heavy (H) isotopes for different time periods for seven different time points including 0, 1, 2, 4, 8, 16, and 32 h pulse with three replicates each for each of the six conditions. For each of the individual time points and replicates the data was combined. Samples treated with Triton X-100 represent broken lysosomes for the input (=post nuclear supernatant) and eluate fractions. Difference in activity readings between the intact and broken status of lysosomes represents the lysosomal integrity. **Abbreviations:** WT: wild-type, ATG5-KO: autophagy related gene 5 knock out, BTZ: bortezomib, PSTA: pepstatin A, MGM-KO: multiple glycosylation machinery related genes knock out, NPC1-KO: Niemann-Pick type C1 knock out. Also see Figure S1-6.

6.3.4 Investigation of Protein Half-Lives from Lysosome Enriched Fractions

Determining protein half-lives representing "true" changes under cellular perturbational conditions requires more mathematical adjustments than those at a steady state. This open problem still needs to be addressed in the field [120]. Dynamic changes have only been estimated approximately applying assumptions based on linear rate changes, which most probably do not represent the "true" change/response of the dynamically adjusting proteome. Yet, the steady-state assumptions were successfully applied to effectively compare relative end-point synthesis and degradation rates between conditions [103, 105, 121]. One must add, that the critical constant degradation rate for determining half-life, corresponding to the time for a protein to be reduced from the given existing pool by half, is certainly true for non-dividing cells. However, for dividing cells such as HEK293 cells used in this study, this is certainly not true, as, with every cell division, the protein pool will be diluted. Hence, in dividing cells, the reduction of protein amount is based on a combination of actual degradation and dilution due to cell division [122]. Therefore, to calculate the "true" loss of a protein, a dilution factor (=doubling time of cells) was considered to determine the degradation rate. Further, categorically, synthesis, and degradation rates differ in their mathematical properties. Synthesis is a zero-order process where the change is measured in units (protein amount over time). On the other hand, degradation is a first-order process, as its rate corresponds to the removal of proteins from an existing pool following an exponential decay function [95].

6.3.4.1 Proteomic Overview and Data Inspection

Acquired pSILAC DIA MS/MS (138 raw files), and DDA MS/MS runs (48 raw files) were used to generate a hybrid peptide spectral library using Pulsar integrated within the Spectronaut software. The spectral library contains 236,234 precursors, covering 153,115 unique peptides annotated to 8,208 proteins. Followed by the generation of a spectral library, all 138 DIA files were analysed in Spectronaut [69]. The final output contained individual quantificational values (=protein abundances) for the light and heavy channels for each of the proteins. The data were further exported and analysed using in-house-built scripts in R-studio (methods 6.2.4.15). The data analysis led to the identification of about 5,500 proteins each modified with SILAC light and heavy isotopes showing an efficient labeling across all conditions and time points **(Figure 16A)**. The data covered at least 550 lysosomal proteins for all of the lysosome enriched fractions, compared to 400 IDs for the whole cell lysate illustrating an efficient coverage of the lysosomal proteome, which is line with published studies [34, 44] **(Figure 16B)**. Further, a heatmap was applied to visualize clusters of all identified protein IDs in 0, 1, 2, and 3 replicates across all conditions and time points indicating a good reproducibility among the respective replicates **(Figure 16C)**. The non-normalized protein abundances showed, as expected, a reduction of the proteins labeled with SILAC light isotopes (green) (correlation of intensity (log2) on the x-axis) and an increase of proteins labeled with the SILAC heavy isotopes over time **(Figure 16D)**. Quantificational analyses revealed two major insights: 1) significant differences in median protein abundance among the different conditions, 2) outliers, where for example SILAC heavy intensities for 16 h are higher than compared to the 32 h for the ATG5-KO, and vice versa for the SILAC light intensities **(Figure 16E)**. These discrepancies in quantificational data were not surprising due to experimental variations arising from sample handling, but mostly associated to factors related to the MS measurements (chromatographic separation, ionization, differences in sensitivity over time). With this, the data remains biased, especially if not measured in the same measurement slot, which is however hardly avoidable due to restrictions in measurement times and loss of sensitivity over a period of measurement. The initial analysis of the whole dataset led to conclusion that although the entirety of the whole dataset shows a good coverage of the whole as well lysosomal proteome with efficient labeling and reproducibility on the level of protein identification, quantificational biases remain requiring an appropriate normalization strategy.

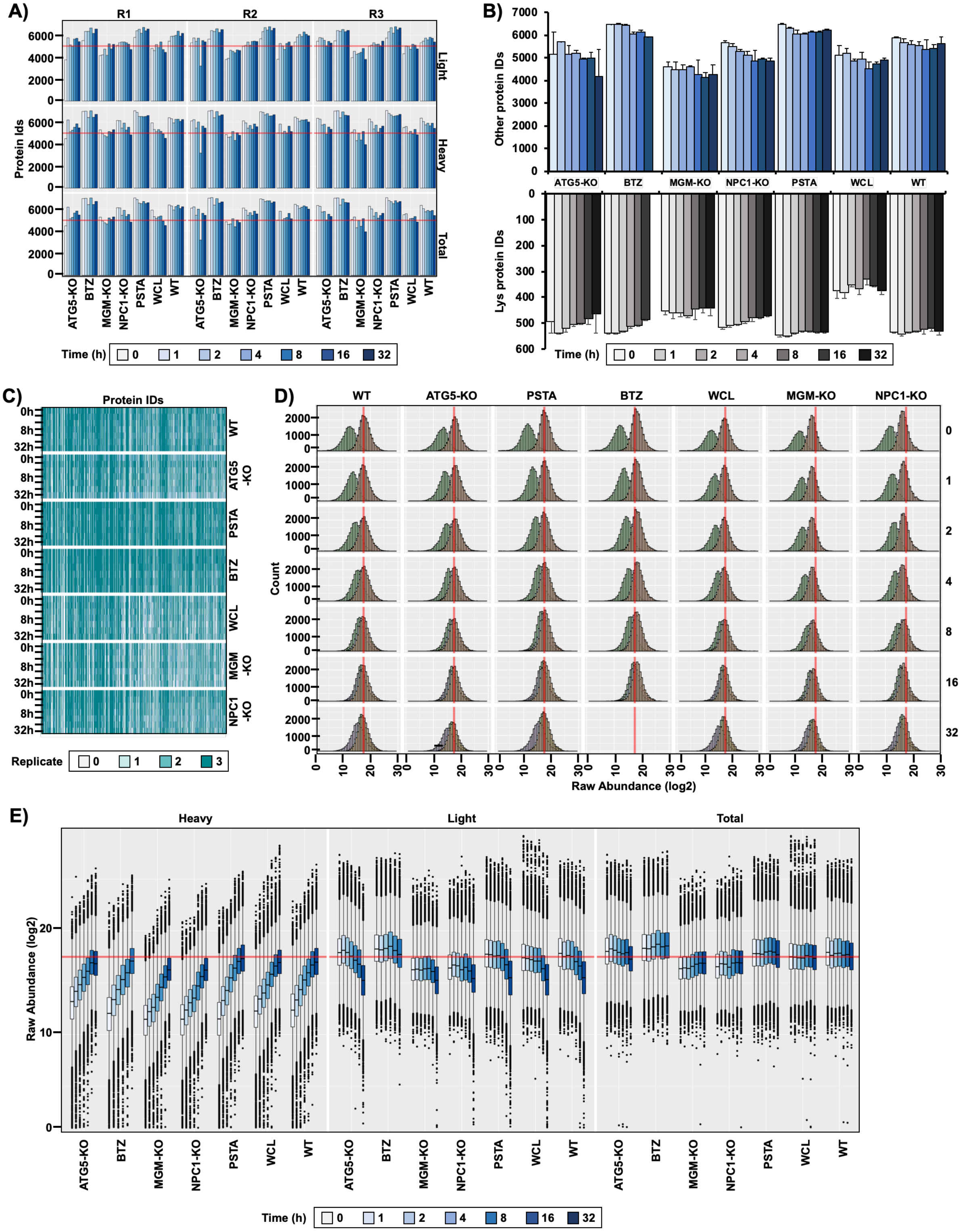


Figure 16: Proteomic overview for the pSILAC DIA MS/MS turnover study. A) Bar graph showing proteomic data inspection (protein IDs) of whole-cell lysate (WCL) from HEK293 cells serving as a control, as well lysosome enriched fractions isolated under six different lysosomal perturbation conditions including: the impairment of autophagosome-lysosome fusion (ATG5-KO cells), proteasome activity (BTZ treatment), protein glycosylation (MGM-KO), cholesterol homeostasis (NPC1-KO), cathepsin activity (PSTA treatment), and wild type (WT). Lysosomes were enriched from a total of 294 plates of 10 cm HEK293 cells using SPIONs cultured in medium containing amino acids labelled with light (L) isotopes and switched to SILAC medium containing amino acids labelled with heavy (H) isotopes for different time periods for seven different time points including 0, 1, 2, 4, 8, 16, and 32 h pulse with three replicates each. Data represents 186 LC-MS/MS runs (applying the data-independent acquisition (DIA) and data-dependent acquisition (DDA)). Data was analysed in the Spectronaut software and post analysis performed in house-built R-scripts. **B)** Bar graph showing a summary of bar graph A displaying average proteins IDs for all identified proteins, in which proteins have been shown for a list of lysosomal gene ontology (GO) annotated proteins (LYS (GO)), and proteins associated to other cellular compartments (other). The list of lysosomal proteins associated with the gene ontology list and the manually curated list can be viewed under (Table S1). **C)** Visualization of missing/valid values for all identified proteins/condition/time points, where the number of identifications per replicate were color-coded from 0-3. **D)** Histograms showing the normal distribution of data separated on the x-axis based on bins of protein abundance and plotted against the number of counts. Each bin visualizes the distribution for the quantified values for the SILAC light (green) and heavy channel (orange) over time for each of the conditions and time points (replicates for each of the time conditions have been combined). The vertical red line indicates a fix value on the x-axis for a better comparison among each of the graphs. **E)** Box-whisker plots (95 % quantile box/2.5 % +/- whisker, median=line) showing raw quantitative values (=protein abundance) for each of the seven-time points for each of the conditions, where the replicates have been clustered and visualized for the SILAC light and heavy channel. A red line was inserted for a better comparability of protein abundance among the conditions. **Abbreviations:** R: replicate, IDs: identifications (number of identified proteins), Lys: lysosomal, ATG5-KO: autophagy related gene 5 knock out, BTZ: bortezomib, MGM-KO: multiple glycosylation machinery related genes knock out, NPC1-KO: Niemann-Pick type C1 knock out, PSTA: pepstatin A, WCL: whole cell lysate, WT: wild-type.

6.3.4.2 Normalization of Proteomic Data Applying a “MasterMix” Approach

A normalization approach to account for technical biases was developed based on the generation and measurement of a "MasterMix". The rationale behind the approach was to pool all 21 samples for a given condition covering all seven-time points (including its three replicates). Subsequently, samples from all seven conditions were measured in triplicates in a randomized order in the same measurement slot. This strategy proved to be highly successful, as the measured "MasterMix" data showed no measurement artefacts (“batch measurement effects”) among the conditions **(Figure 17A, B)**. Hence, normalization factors were calculated from these results and applied to the acquired initial dataset showing a normal distributed narrow range of factors **(Figure 17C)**. After bioinformatic assessment and validation of different normalization approaches, the dataset was normalized (see section 6.2.4.15) showing a steady decrease of the SILAC light channel intensities and an increase for the SILAC heavy channel intensities over time **(Figure 17D)**. One can further observe that the total intensities after normalization were adjusted, as the sum of SILAC light and heavy channel should be the same within a condition over the seven-time points. The data after normalization allowed us to compare the conditions among each other and revealed: 1) The overall proteome abundance of lysosome enriched fractions and whole cell lysate is similar, representing the steady-state and hence excluding artifacts (with a high probability) as a result of lysosome enrichment via SPIONs, 2) the highest median proteome abundance for the NPC1 condition (total), indicating a major impairment of degradational processes (highest light channel), while the synthesis rate (comparable heavy channel) seems not to be affected, 3) pepstatin A treated cells with an impaired lysosomal degradation revealed a significant proportion of proteins not being able to be degraded (an increase of SILAC light channel), visible as outliers in the box whisker plots **(Figure 17D)**. In the last step, quality control (QC) criteria were developed and assessed to determine half-lives of identified proteins. The QC criteria included the identification of the protein in at least two replicates over at least a series of five-time points with a coefficient of variance (CV) threshold of 50 % **(Figure 17E)**.

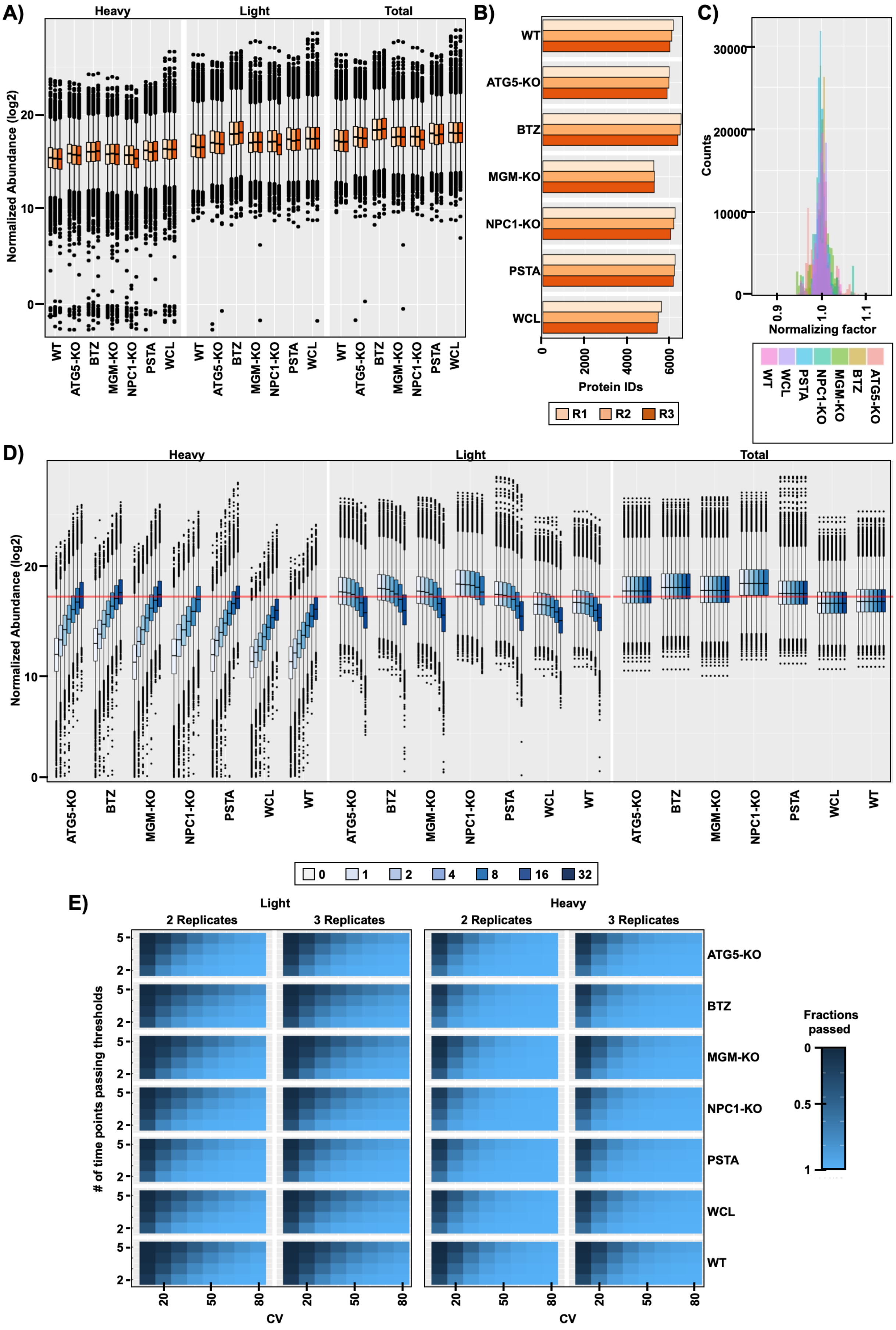

A)
Heavy
Light
Total
Normalized Abundance (log2)
20
10
0
WT
ATG5-KO
BTZ
MGM-KO
NPC1-KO
PSTA
WCL
B)
WT
ATG5-KO
BTZ
MGM-KO
NPC1-KO
PSTA
WCL
0
2000
4000
6000
Protein IDs
R1
R2
R3
C)
30000
20000
10000
0
Counts
0.9
1.0
1.1
Normalizing factor
WT
WCL
PSTA
NPC1-KO
MGM-KO
BTZ
ATG5-KO
D)
Heavy
Light
Total
Normalized Abundance (log2)
20
10
0
ATG5-KO
BTZ
MGM-KO
NPC1-KO
PSTA
WCL
WT
0
1
2
4
8
16
32
E)
Light
Heavy
2 Replicates
3 Replicates
# of time points passing thresholds
5
2
ATG5-KO
BTZ
MGM-KO
NPC1-KO
PSTA
WCL
WT
20
50
80
CV
Fractions passed
0
0.5
1

Figure 17: Normalization strategy for the pSILAC DIA MS/MS turnover study. A) Box-whisker plots (95% quantile box / 2.5 % +/- whisker, median=line) showing quantitative values (=protein abundance) for each of the conditions, where three replicates have been clustered and visualized for the SILAC light and heavy channel. For these results, 21 additional LC-MS/MS runs were acquired in which for each of the conditions all of the seven time points across all three replicates have been pooled and measured in the same measurement slot to minimize technical variances and apply these for normalization purposes. **B)** For the same dataset, protein IDs across all the seven different conditions have been plotted for each of the three replicates. **C)** Histogram visualizing the range of calculated normalization factors (in bins) plotted on the x-axis and matched against the number of valid values on which it has been applied for the valid values, including the SILAC light, and heavy channel for the seven conditions. **D)** Box-whisker plots (95% quantile box / 2.5 % +/- whisker, median=line) showing quantitative values (=protein abundance) for each of the 7 time points for each of the conditions, where the replicates were clustered and visualized for the SILAC light and heavy channel. A red line was inserted for a better comparability of protein abundance among the conditions. **E)** Color coded matrix visualizing the % of fractions passing the coefficient of variation (CV) thresholds for the SILAC light and heavy channels. **Abbreviations:** R: replicate, IDs: identifications (number of identified proteins), ATG5-KO: autophagy related gene 5 knock out, BTZ: bortezomib, MGM-KO: multiple glycosylation machinery related genes knock out, NPC1-KO: Niemann-Pick type C1 knock out, PSTA: pepstatin A, WCL: whole cell lysate, WT: wild-type.

6.3.4.3 Determination of Protein Half-lives

Filtered, normalized values passing the established quality control criteria were used to determine protein half-lives for 6,915 proteins across all conditions applying ratio-based linear regression models [111] **(Figure 18A)**. The dataset covered 539 lysosomal proteins based on the lysosomal gene ontology list **(Table S1)**. In total, across all conditions, out of 48,405 (6,915 x seven conditions) possible determinable half-lives, the study covered 33,968 including 2,947 associated to the lysosomal proteome **(Figure 18B, Table S2)**.

The data showed a wide range of protein half-lives, where for a small number of proteins negative half-lives were determined, which could not be true and represent an artifact of missing values for the respective SILAC channel. Hence, to compensate for such artifacts, the r-square value as a further quality control criterion was introduced. Based on this, the half-lives for the lysosomal and rest of the proteome were categorized, eliminating artifacts to a minimum for the categories "High" and "Medium" (Med) **(Figure 18B-D)**. One of the most important observations was that the half-lives from proteins of the whole cell lysate and lysosome enriched fractions in wild type show a similar range, excluding artifacts with a high probability resulting from lysosome enrichment via SPIONs. Looking into a narrower range of protein half-lives ranging from a scale of 0-50 h, representing the majority of our data, one could observe major differences in protein half-lives among the different conditions **(Figure 18E, F)**.

PSTA treatment for example resulted to the lowest protein half-lives compared to other conditions **(Figure 18E, F)**. Interestingly, at the same time a number of proteins appear to be more stable upon PSTA treatment, visible as outliers in the box whisker plots **(Figure 17D)**. Surprisingly, the ATG5-KO causing a substantial decrease of autophagy did not lead to an overall increase in protein half-lives. In contrast, impaired lysosomal cholesterol egress, proteasome inhibition and impaired protein glycosylation led to an overall increase in protein half-lives **(Figure 18E, F)**.

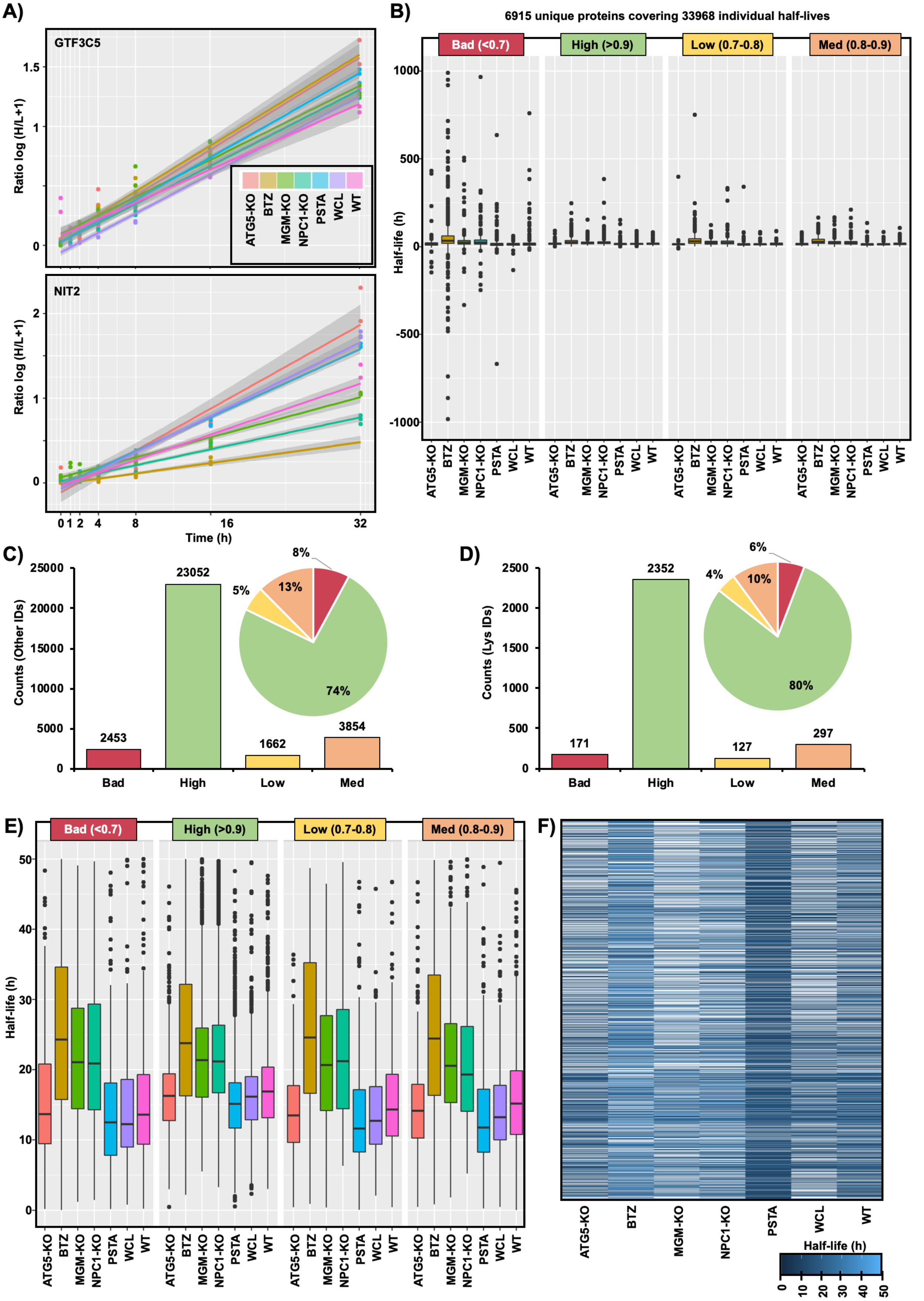

A)
GTF3C5
NIT2
Ratio log (H/L+1)
Time (h)
0 1 2 4 8 16 32
ATG5-KO
BTZ
MGM-KO
NPC1-KO
PSTA
WCL
WT
B)
6915 unique proteins covering 33968 individual half-lives
Bad (<0.7)
High (>0.9)
Low (0.7-0.8)
Med (0.8-0.9)
Half-life (h)
1000
500
0
-500
-1000
C)
Counts (Other IDs)
2453
23052
1662
3854
Bad
High
Low
Med
8%
5%
13%
74%
D)
Counts (Lys IDs)
171
2352
127
297
6%
4%
10%
80%
E)
Half-life (h)
F)
Half-life (h)
0 10 20 30 40 50

Figure 18: Determination of protein half-lives. A) Line dot graphs show two examples for protein half-life determination using linear regression models. The x-axis shows incubation time with media containing amino acids labelled with heavy isotopes, while the y-axis shows the normalized, quality controlled (passing CV thresholds) abundance ratios log (heavy/light+1). Graphs show: Whole cell lysate (WCL) from HEK293 cells serving as a control, and lysosome enriched fractions isolated under six different lysosomal perturbation conditions including the impairment of autophagosome-lysosome fusion (ATG5-KO), proteasome activity (BTZ treatment), protein glycosylation (MGM-KO), cholesterol homeostasis (NPC1-KO), cathepsin activity (PSTA treatment), and wild type (WT). **B)** Box-whisker plots (95 % quantile box/2.5 % +/- whisker, median=line) show half-lives for all seven conditions. Half-lives were categorized into four groups based on the r-square values (R^2), where 1 represents the highest possible positive linear correlation. **C-D)** Classification of individual half-lives related to the rest and the lysosomal proteome. **E)** Box-whisker plots (95% quantile box / 2.5 % +/- whisker, median=line) show half-lives for all seven conditions with a zoomed scale between 0-50 h. **F)** Color-coded visualization of proteins with a half-life between 0-50 h. **Abbreviations:** Lys: lysosomal, IDs: identifications, GO: gene ontology, ATG5-KO: autophagy related gene 5 knock out, BTZ: bortezomib, MGM-KO: multiple glycosylation machinery related genes knock out, NPC1-KO: Niemann-Pick type C1, PSTA: pepstatin A, WCL: whole cell lysate, WT: wild-type.

6.3.4.4 Investigation of Half-lives of Lysosomal and Lysosomal Associated Proteins

Overall, I was able to determine in average about 510 (in total unique 539) protein half-lives for lysosomal and lysosomal associated proteins compared to 277 for the whole cell lysate sample. These results were not surprising, as I showed the benefit of enrichment of lysosomes prior to DIA analysis significantly improving reproducibility [44], which was crucial for the determination of protein half-lives. Binned frequency distribution analysis of protein half-lives associated to 539 lysosomal proteins showed average half-lives for the WT: 18 h, ATG5-KO: 16 h, PSTA: 17 h, BTZ: 70 h, MGM-KO: 23 h, and NPC1-KO: 30 h **(Figure 19A-G, Table S2).**

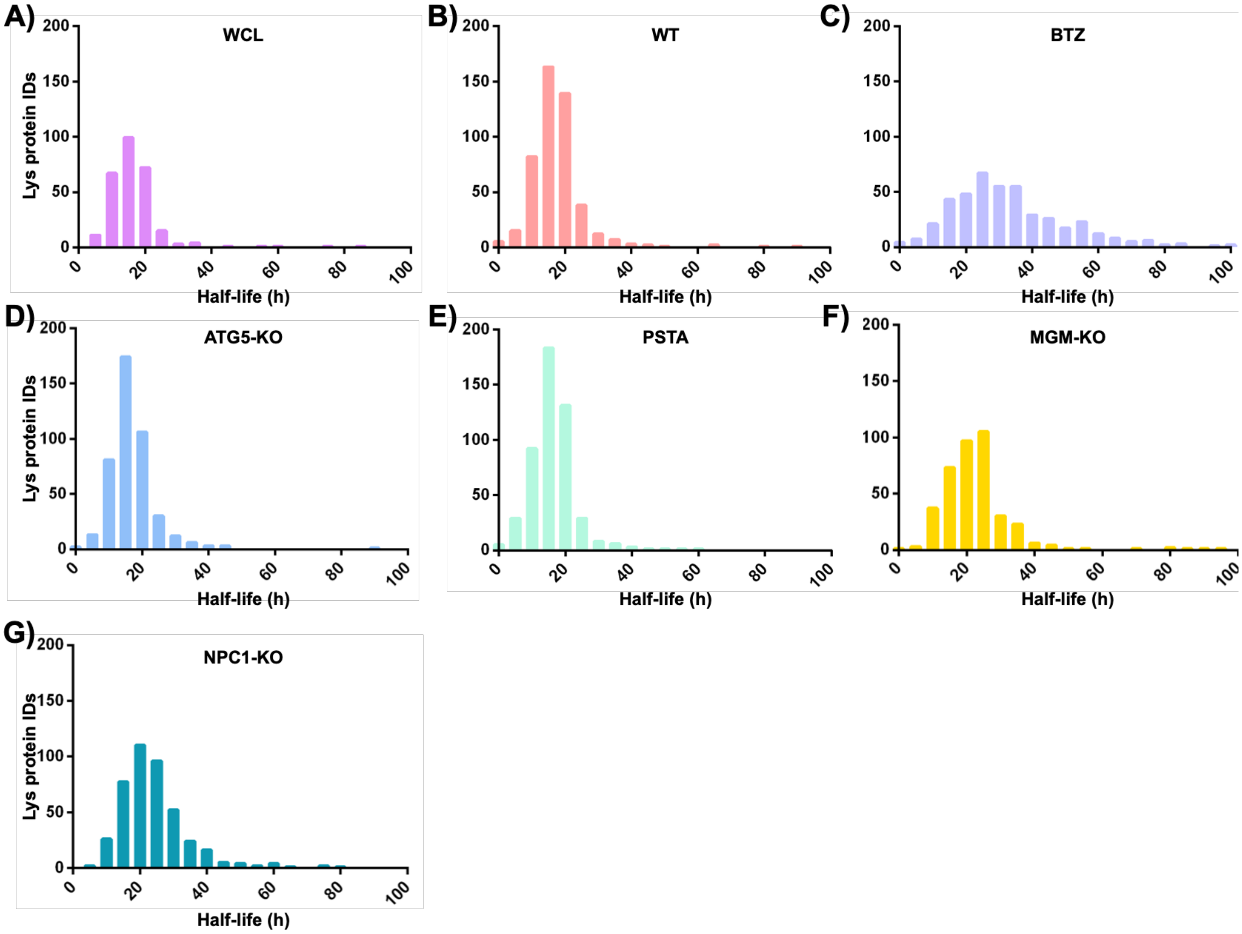


Figure 19: Determination of half-lives of lysosomal proteins. A-G) Binned frequency distribution analysis for identified lysosomal proteins, applying a bin range of 5 h on the x-axis, plotted against the number of identified lysosomal proteins for each of the conditions on the y-axis. Classification of half-lives of lysosomal and lysosomal associated proteins based on protein class categories (Table S1). **Abbreviations:** ATG5-KO: autophagy related gene 5 knock out, BTZ: bortezomib, MGM-KO: multiple glycosylation machinery related genes knock out, NPC1-KO: Niemann-Pick type C1 knock out, PSTA: pepstatin A, WCL: whole cell lysate, WT: wild-type.

BTZ treated samples were particularly interesting, as although no significant differences among the median abundance of the individual light and heavy channels could be observed for those **(Figure 17D)**, a dramatic increase in half-lives were observed **(Figure 18E, F, Figure 19C)**. These observations were in line with a recently published study applying pSILAC DIA MS/MS to assess half-lives of proteins after BTZ treatment of primary human foreskin fibroblasts (HFF-1) [123]. These results are not surprising, because if the proteasome is inhibited, target proteins known to be degraded by the proteasome like HNRNPK, EIF3A, and EIF4A accumulate and cannot be degraded [123] **(Figure 20A-C).** This would imply that upon inhibition of the proteasome, no redirection of these proteins for degradation to the lysosomes is possible. Vice versa, I also investigated whether proteasomal inhibition affects the turnover of lysosomal proteins. The three lysosomal proteins with the most extended half-life upon proteasome inhibition are the alpha-subunit of beta-hexosaminidase (HEXA), alpha-glucosidase (GAA) and beta-galaktosidase. Common property of these enzymes is that all of them cleave glycosidic bonds **(Figure 20D-F)**.

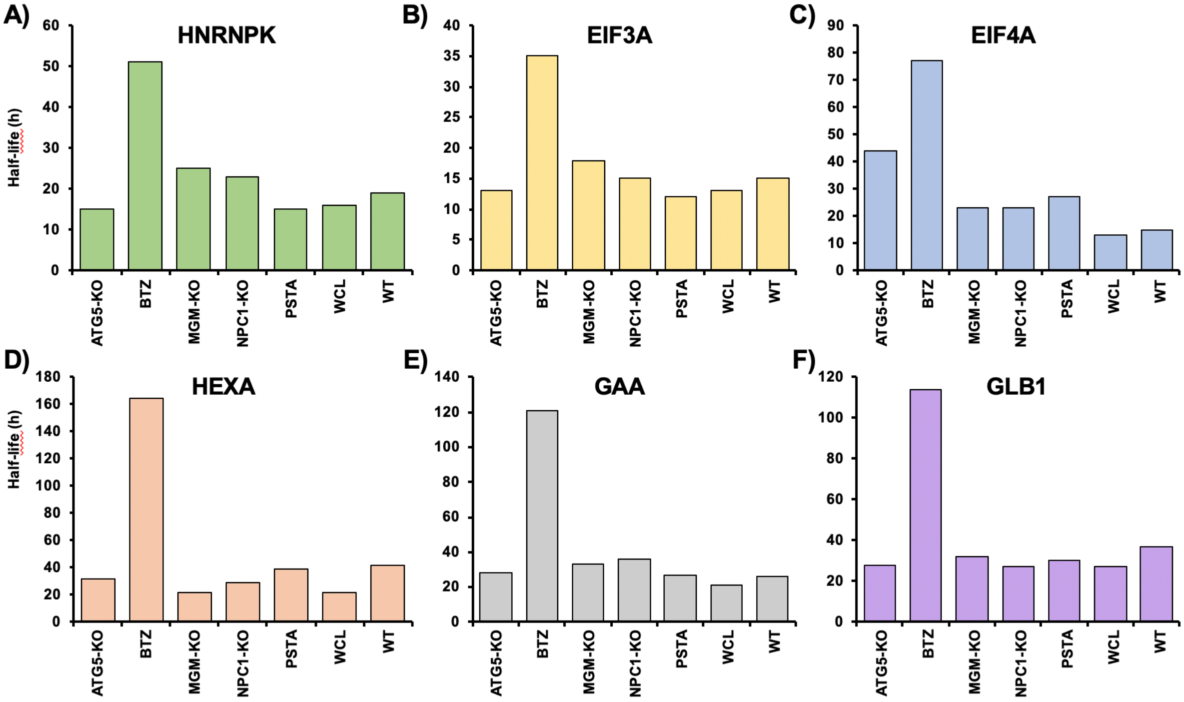


Figure 20: Example of half-life of proteins targeted to be degraded by the proteasome. A-C) Known proteasomal targets**. D-F)** Potential proteasomal targets, among specifically lysosomal proteins. **Abbreviations:** ATG5-KO: autophagy related gene 5 knock out, BTZ: bortezomib, MGM-KO: multiple glycosylation machinery related genes knock out, NPC1-KO: Niemann-Pick type C1 knock out, PSTA: pepstatin A, WCL: whole cell lysate, WT: wild-type.

The half-life of a protein is influenced by factors such as intrinsic (primary amino acid sequence features, secondary structure features), functional (relative protein abundance, complex association and assembly type, molecular pathway relationships), and contextual parameters (subcellular localization, cell/tissue type, and microenvironment) [95]. Hence, a similar yet more simplified analysis of average half-lives focusing on a subcategory of manually classified lysosomal protein classes was performed. For this I extracted half-lives of lysosomal proteins associated to three different “protein classes” including 10 members of cathepsins (proteases), 13 members of the v-ATPase complex, and 4 membrane proteins. Looking into the three individual classes of lysosomal proteins, major differences among the half-lives of proteins belonging to the same protein class could be observed **(Figure 21)**. The differences in half-lives for individual proteins were even more pronounced when compared among the different cellular conditions.

Among all the different enzymes within the lysosomes, cathepsins represent a family of lysosomal proteases involved in protein degradation and antigen processing among others (reviewed in [114, 124]). All cathepsins identified in my data belong to one of the three family of proteases dependent on the type amino acid at their active site, namely: serine protease (CTSA), aspartic protease (CTSD), while all other eight members (CTSV, CTSF, CTSO, CTSZ, CTSH, CTSC, CTSB, and CTSL) belong to the cysteine proteases. CTSL for example showed to have the longest, whereas CTSV showed to have the shortest half-life compared to the other cathepsins among almost all conditions (excluding PSTA) **(Figure 21)**. In the ATG5-KO condition the half-life of CTSL increased ~40h compared to the WT condition, while CTSV had the same half-life. Overall one could conclude that the conditions seem to have specific effects on the half-life of individual cathepsins, rather than affecting all members in a same way.

One of the most reproducible findings in turnover studies are that proteins engaged in the same complex do not entirely exhibit the same turnover. They are rather typically subclustered based on the architecture of the multimeric complex, where studies reported that dynamic subunits revealed higher half-lives than “core” subunits stabilizing the complex, such as seen for the proteasome [105, 106, 125]. To further follow up on these findings, half-lives for the 13 different subunits of the 32 multimeric lysosomal v-ATPase complex were compared [86]. The v-ATPase consists of a membrane integrated V0 part and a V1 part facing the cytosol [86]. Analysis revealed overall similar half-lives of the cytosolic facing subunits (ATP6V1A1 and ATP6V1B2 - both building the head of the complex with three subunits each) to be the ones with the highest with ~20h **(Figure 21)**. Whereas for example all three subunits of the membrane integrated core V0 subunit (ATP6V0A2, ATP6V0A2, ATP6V0D1) had similar half-lives of ~16h each. These results were in line with the literature, where differences in half-lives for subunits, especially for members of large multimembered complexes, are based on their locations. It was reported that for certain complexes, specific subunits might be translated in excess, which are then degraded to stoichiometric equivalencies [110, 126]. Studies suggest that this might imply that the excess of subunits may favor complex formation in one location distinct from the one of its function [106]. Especially in the view that v-ATPase as a complex is associated with a wide range of functions and different subcellular localizations by the dissociation of the V1 domain (under different stimuli reviewed in [127]) apart from the lysosomal membrane, further investigations are needed for better mechanistic understandings.

For the third and last class including four lysosomal associated membrane proteins, SCARB2 had the longest half-life among all conditions **(Figure 21)**. Interestingly, the BTZ treatment showed to selectively stabilize SCARB1 and 2, indicating that both of the lysosome associated membrane proteins might be degraded by the proteasome.

Overall it can be concluded that the generated large-scale data serves as a valuable resource. The data for instance can hint towards different mechanistic processes such as by identifying distinctive class of proteins exhibiting different sensitivities to lysosomal proteases or endocytic transportation routes.

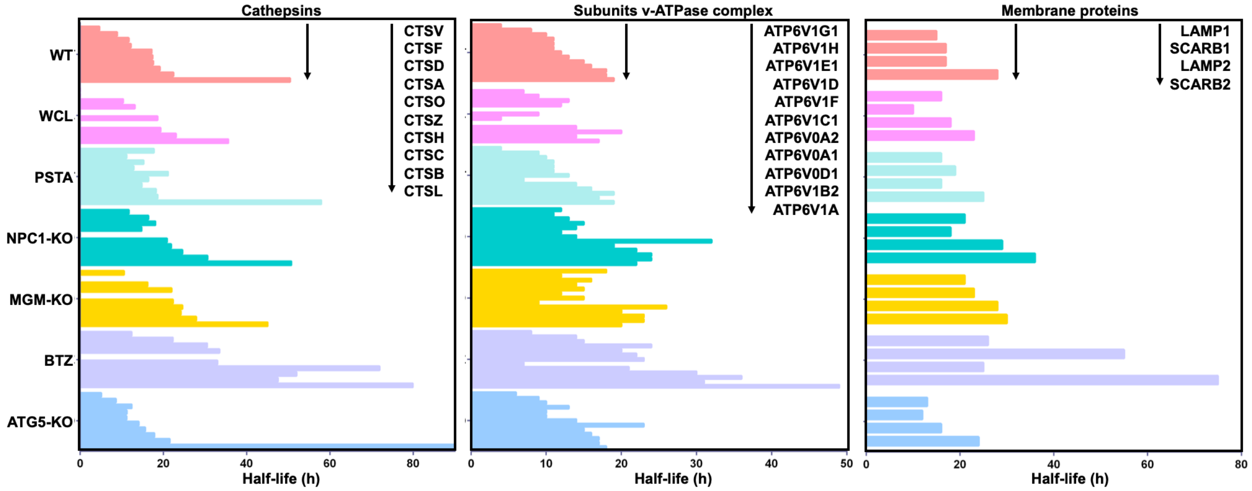


Figure 21: Investigation of half-lives of lysosomal proteins among different protein class subcategories. Binned half-life frequency distribution analysis for lysosomal proteins belonging to each of the three classes (cathepsins, v-ATPase, and membrane proteins), applying a bin range of 5 h on the x-axis. **Abbreviations:** ATG5-KO: autophagy related gene 5 knock out, BTZ: bortezomib, MGM-KO: multiple glycosylation machinery related genes knock out, NPC1-KO: Niemann-Pick type C1 knock out, PSTA: pepstatin A, WCL: whole cell lysate, WT: wild-type.

6.3.5 Comparison of Protein Half-lives Among Our and Published Studies

A number of large-scale turnover studies have been published [103, 104, 111, 120]. However, due to experimental variations / differences across these studies a direct comparison to the data presented here is hardly possible. Only the study by Larance et al. [104] is sufficiently similar for a direct comparison of data. Since Larance et al. determined half-lives of proteins in Hela cells, a comparison of his and our data set allows to reveal cell type specific differences in protein turnover. To my knowledge, the presented study comprises the yet most extensive turnover dataset with the determination of 33,968 individual half-lives covering 6,915 unique proteins under five different perturbation conditions applying the pSILAC DIA MS/MS approach **(Figure 22A)**. Whereas all the other studies covered one to two conditions determining ~6000-12,000 half-lives.

The comparison of half-lives of 3,125 proteins identified in their HeLa dataset and our HEK293 dataset showed a median half-life of 20 and 18 h, with the majority of them having a half-life between 20-25 and 15-20 h, respectively **(Figure 22B, C)**. Although the overall half-life differences might be marginal between both studies, the hierarchical clustering analysis revealed a discrepancy of >15 h for specific clusters of proteins and up to 46 h for individual proteins **(Figure 22D)**. Proteins with increased half-life differences in HeLa (>25 h compared to HEK293) included the proteins RAPGEF6, MET, POP7, TYSND1, and RRP9. Proteins with increased half-life differences in HEK293 (>40 h compared to HeLa) included the proteins CIAPIN1, CKB, SEPHS1, SDCBP, and SSU72.

The results are not surprising as previously reported, both cell lines despite having a similar proteome, exhibit different expression levels of proteins [128]. Therefore, comparisons and conclusions should be taken carefully and assessed critically, especially considering the methodological/experimental setup. For instance, the importance of the studied biological system for comparison purposes is exemplified by the fact that half-lives for non-dividing primary cells such as B-cells exhibit half-lives of >1000 h [105]. Overall it can be concluded that the presented data can be further used as a resource to reveal cell type specific differences in protein turnover among different cell lines.

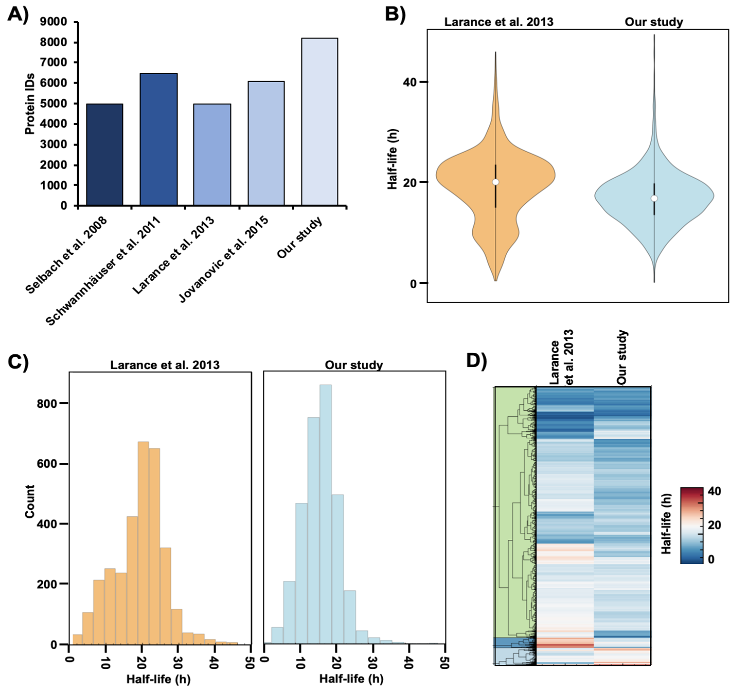


Figure 22: Comparison of determined protein half-lives across published studies. A) Comparison of our turnover study to published datasets including the study from Sehlbach et al. 2008 (HeLa cells) [111], Schwannhäuser et al. 2011 (NIH-3T3 cells) [103], Larance et al. 2013 (HeLa cells) [104], Jovanovic et al. 2015 (bone marrow derived dendritic cells) [120] and our study (HEK293 cells) **B-D)** For comparison purposes 3,125 proteins covered by the turnover study conducted by Larance et al. 2013 [104] in HeLa cells and our study in HEK293 cells were compared and visualized **B)** Violin plots displaying half-life distributions and the median as white point as well as the interquartile range as a black line **C)** Binned frequency distribution of half-lives visualized in histograms with their respective counts (=protein identifications) **D)** Unsupervised hierarchical clustering of half-lives visualized as a heatmap showing lower half-lives in blue and longer half-lives in red.

6.3.6 Benefits and Limitation of pSILAC DIA MS/MS for Half-life Determination

The pSILAC DIA MS/MS approach has the advantage that the experiment could be combined with the enrichment of lysosomes using superparamagnetic iron oxide nanoparticles (SPIONs). This was a crucial factor in allowing comprehensive coverage of the lysosomal proteome [34, 44]. DIA MS/MS, in particular, has the benefit that synthesis and degradation kinetics can be determined separately on a large-scale, allowing one to study dynamic proteome changes over time [107]. In addition, the method allows a comprehensive analysis of the proteome, where the applied tracers behave identically to natural amino acids, hence, eliminating the need to modify proteins such as in other approaches like fluorescence- or tag-based detections. However, as in any other method, limitations do exist: 1) Change of media required for metabolic uptake might induce cellular stress, 2) target systems with a slow metabolism might display a limited sensitivity for short time points due to low uptake, 3) the assumption, that due to recycling of amino acids, heavy signal incorporation may be lower than the actual synthesis rate [74].

6.3.7 Phosphoproteomics of Lysosome Enriched Fractions

Post translational modifications (PTMs) such as phosphorylation, glycosylation, acetylation, and methylation can influence the structure, stability, and function of proteins in a cell. A recently published study for example revealed that certain phosphorylated peptides have a longer half-live compared to their non-phosphorylated counterparts [107]. In the following section results of a phosphoproteomics dataset are presented, which will be used in future analysis to further correlate PTMs with protein half-lives.

Phosphorylation belongs to one of the most dominant PTMs, with an estimated 1/3 of the proteome being phosphorylated. Thus, phosphorylation (=attachment of phosphoryl group) plays a vital role in numerous cellular processes, including cell proliferation, signaling events, and protein-protein interactions [129]. To study the phosphoproteome on a large, unbiased scale, mass-spectrometry-based proteomics has proven its strength. The methodology allows to identify phosphorylated proteins, pinpoint the phosphorylation sites, and provide quantitative data

[130]. Given the multiple functions of lysosomes and their interaction with other cellular compartments, phosphoproteomics of lysosomal and lysosome associated proteins may yield new insights into sites of dynamic regulation controlling lysosomal function.

Due to the low abundance of lysosomal proteins and phosphorylated peptides within a tryptic digested sample, lysosome enrichment followed by phosphopeptide enrichment was performed **(Figure 23A)**. Acquired LC-MS/MS data were analysed using two different proteomic analysis software: Proteome Discoverer (PD) and MaxQuant (MQ). Analysis with PD resulted in the identification of a total of 3,660 proteins. Based on a list of lysosomal (LYS) gene ontology (GO) annotated proteins and a more stringent manually curated (MC) list of lysosomal proteins, (part of the LYS (GO)) the data covered 279 lysosomal proteins **(Figure 23B, Table S1, S3)**. Peptide groups and peptide spectral matches (PSMs) showed a similar percentage distribution as among the identified proteins for the category other cellular compartments and lysosomes **(Figure 23C, D)**. From all of identified proteins, 2,683 proteins were found to be phosphorylated **(Figure 23E)**. Further assessment of the data revealed that with 86 %, the majority of phosphorylation were assigned to serine, followed by threonine and tyrosine with 13 and 1 % each **(Figure 23F)**. In its entirety, the data analysis with PD led to the identification of 6,786 unique phosphorylation sites. In contrast, the analysis with MaxQuant resulted in the identification of a higher number of phosphorylated proteins and outnumbered PD with the identification of additional 1,879 phosphorylation sites **(Figure 23G-I)**. The identified phosphorylation sites were in addition annotated with a site probability score (>75%), serving as a quality control criterion for a highly reliable site specification **(Figure 23I, Table S3)**. Further, all identified phosphorylation sites were compared to the PhosphoSitePlus database - A source for phosphorylation sites reported in the literature. The analysis resulted in an overlap of 93 % with previously reported phosphorylation sites verifying our dataset, while 7 % present potential novel sites **(Figure 23J)**.

The 86 phosphorylated lysosomal proteins listed **(Table S3)** cover not only lysosomal proteins but also proteins involved in transmission of signals from and to the lysosomes and various vesicular fusion and sorting processes involved in lysosomal movement and biogenesis [2]. As far as proper, well known lysosomal enzymes [10] are concerned, phosphorylation can be seen for a few soluble intra-lysosomal enzymes (DNAse2, GAA, LIPA, PRCP, CTSC, PLD2). These data are difficult to explain, since lysosomal enzymes should not be exposed to kinases during their transport from the endoplasmic reticulum (ER). However, the presence of kinases active within the ER has been reported [131-133], so that phosphorylation of soluble lysosomal enzymes appears possible, yet its functional significance remains to be elucidated. Phosphorylation was also found on a number of lysosomal transporters/exchangers. These include the sialic acid transporter (Sialin - SLC17A5), SIDT2 which has recently been shown to transport RNA/DNA across the lysosomal membrane directly into the lysosome [134, 135], and the cobalamin transporter (TCN1). All phosphorylation sites are located on the cytosolic domains of these transporters making them accessible for cytoplasmic kinases.

It is certainly interesting to speculate that phosphorylation may control transport from the lysosome to the cytosol and vice versa. Since some of these transporters such as Sialin are neither strictly specific for compounds nor for subcellular localization, phosphorylation may control transport substrate specificity or localization. For example, Sialin transports sialic acid but also glucuronic acid out of the lysosome [136] but it can also locate to the membrane of synaptic vesicles where it pumps aspartate and glutamate into the vesicles [137]. Similarly, yet not investigated roles of phosphorylation may apply to lysosomal H+/Cl-exchangers CLCN3 and

CLCN7/OSTM1. The data set also reveals the extensive phosphorylation of proteins involved in lysosomal mediated mTORC1 activity control, which include a wide variety of proteins (LAMTOR1, LAMTOR5, DEPDC5, MTOR, TSC2, WDR59, RPTOR, DPTOR, MLST8, and RRAG) [138]. Finally, proteins regulating lysosomal positioning (e.g. BORCS6, SPAG) suggest that phosphorylation impacts lysosomal positioning. STARD3, STARDNI3 and VAPA are three proteins which interact to tether lysosomes to the ER and mediate cholesterol transport from the lysosome to the ER [139]. Interestingly all three proteins were found to be phosphorylated suggesting that lysosome-ER tethering and cholesterol transport are tightly regulated by kinases. Overall the phosphoproteomic data generated will serve as a resource to investigate the role of dynamic phosphorylation in the control of various aspects of lysosomal function.

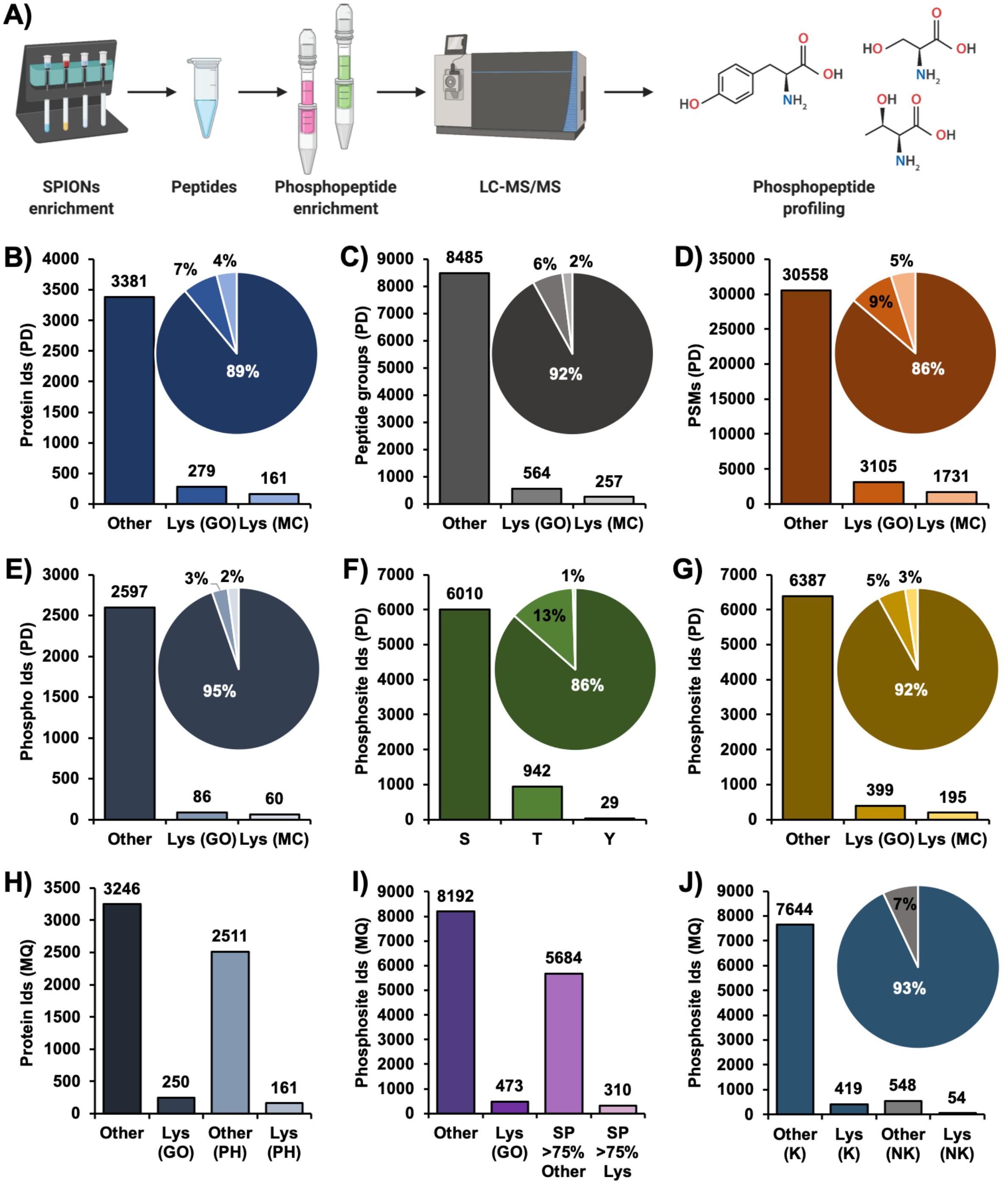


Figure 23: Workflow and overview of the lysosomal phosphoproteomic dataset. A) Schematic workflow for the enrichment of phosphopeptides from lysosome enriched fractions. Lysosomes were enriched from 50 plates of 10 cm HEK203 cells using SPIONs. 1.8 mg of protein from the eluate fraction were proteolytically digested, desalted and applied for phosphopeptide enrichment applying the SMOAC (Sequential enrichment using Metal Oxide Affinity Chromatography) approach. After enrichment and resuspension, 25 % of the sample have been measured on the Orbitrap Fusion Lumos, followed by data analysis on two different software. **B-G)** Data analysis with Proteome Discoverer (PD) and **H-J)** MaxQuant (MQ). Bar graphs and pie charts showing number and percentage distribution of: **B)** number of all identified proteins **C)** peptide groups **D)** peptide spectral matches (PSMs) and **E)** phosphorylated proteins for a list of lysosomal gene ontology (GO) annotated proteins (LYS (GO)) and a more stringent manually curated (MC) list of lysosomal proteins (LYS (MC)), as well as proteins associated to other cellular compartments (other). Bar graph and pie chart showing the total number and percentage distribution of identified phosphosites distributed among **F)** serine, threonine, and tyrosine (S, T, Y) and **G)** proteins for a list of lysosomal (GO, MC) and other proteins. Bar graphs showing **H)** number of all identified and phosphorylated proteins. **I)** Bar graph showing the total of identified phosphosites distributed among the list of lysosomal proteins and other as well as those with a site probability (SP) of >75% (representing a high confidentiality control factor for correct phosphosite localization). **J)** All identified phosphosites have been further compared to the PhosphoSitePlus to compare their status (known (K) or not known (NK). **Abbreviations:** SPIONs: superparamagnetic iron oxide nanoparticles, LC-MS/MS: liquid chromatography mass spectrometry, Ids: identifications, PSMs: peptide spectral matches, PD: Proteome Discoverer, Lys: lysosomal, GO: gene ontology, MC: manually curated, MQ: MaxQuant, SP: site probability, PH: phosphorylated, K: known, NK: not known.

6.4 Conclusion and Future Perspectives

Results showed the successful application of the pSILAC DIA MS/MS approach to determine 33,968 half-lives for 6,915 proteins. The study presents the so-far most extensive turnover study covering a steady-state and different perturbational conditions, including lysosomal and proteasomal impairment. The analysis, for instance, revealed significant differences in half-lives across conditions, such as a drastic increase of half-lives upon proteasomal inhibition or decrease upon cathepsin D inhibition. Analyses indicate a complex interplay of synthesis and degradation on protein half-life, where perturbations seem to primarily influence degradation rather than synthesis rates. Further, protein half-lives differed significantly dependent among others on their protein class subcategory and functions. Investigation of half-lives of subunits of the multimeric lysosomal v-ATPase complex for instance revealed a correlation between A and B's cytosolic facing V1 subunits (~20 h) and V0 membrane anchored subunits (~16 h) with significantly lower half-lives. Such results might imply yet unknown regulatory mechanisms where excess translation might be controlled by increased degradation to stoichiometric equivalencies. The presented data overall is in line with other large-scale published studies, in which the majority of proteins exhibit to have half-lives ranging between 15-25 h.

The data presented so far are merely descriptive and only scratch the surface of what remains to be truly explored in this study. For one, the proteomic comparison for each condition compared to the wild-type representing the steady-state needs to be studied, followed by a separate examination of synthesis and degradation kinetics. With the implementation of an approach to determine synthesis and degradation kinetics separately, the final data analysis is under proceeding. Based on this, the final dataset is supposed to include, additionally to the presented 33,968 half-lives, similar numbers for synthesis and degradation rates covering about 6,000 proteins. The dataset will be, among others, subsequently mined for associations to gene ontology classes (biological function, subcellular localization, molecular pathway). In an ideal case, such an endeavor would elucidate regulatory trends for known protein networks. Conversely, identifying similarly functioning proteins with uncorrelated half-lives could open pathways in discovering their specific roles. Above all, however, the dataset will aim to investigate which proteins are primarily degraded by the proteasome versus the lysosome and allow us further to specify the lysosomal cargo and its routes of transportation.

7. Literature

1. Muthukottiappan, P. and D. Winter, A proteomic view on lysosomes. Mol Omics, 2021. **17**(6): p. 842-859.

2. Ballabio, A., Bonifacino, J.S., Lysosomes as dynamic regulators of cell and organismal homeostasis. Nat. Rev. Mol. Cell Biol., 2019. **21**(2): p. 101-118.

3. Settembre, C., Fraldi, A., Medina, D. L., Ballabio, A., Signals from the lysosome: a control centre for cellular clearance and energy metabolism. Nat. Rev. Mol. Cell Biol., 2013. **14**(5): p. 283-296.

4. Mindell, J.A., Lysosomal acidification mechanisms. Annu. Rev .Physiol., 2012. **74**: p. 69-86.

5. De Duve, C., Pressman, B. C., Gianetto, R., Wattiaux, R., & Appelmans, F., Tissue fractionation studies. 6. Intracellular distribution patterns of enzymes in rat-liver tissue. Biochem. J., 1955. **60**(4): p. 604.

6. Bautista, S.J., Boras, I., Vissa, A., Mecica, N., Yip, C. M., Kim, P. K., Antonescu, C. N., mTOR complex 1 controls the nuclear localization and function of glycogen synthase kinase 3beta. J. Biol. Chem., 2018. **293**(38): p. 14723-14739.

7. Chantranupong, L., Wolfson, R. L., Orozco, J. M., Saxton, R. A., Scaria, S. M., Bar-Peled, L., Spooner, E., Isasa, M., Gygi, S. P., Sabatini, D. M., The Sestrins interact with GATOR2 to negatively regulate the amino-acid-sensing pathway upstream of mTORC1. Cell Rep., 2014. **9**(1): p. 1-8.

8. Sardiello M, P.M., di Ronza A, Medina DL, Valenza M, Gennarino VA, Di Malta C, Donaudy F, Embrione V, Polishchuk RS, Banfi S, Parenti G, Cattaneo E, Ballabio A., A gene network regulating lysosomal biogenesis and function. Science, 2009. **325**(5939): p. 473-477.

9. Settembre, C., Zoncu, R., Medina, D. L., Vetrini, F., Erdin, S., Huynh, T., Ferron, M., Karsenty, G., Vellard, M. C., Facchinetti, V., Sabatini, D. M., Ballabio, A., A lysosome-to-nucleus signalling mechanism senses and regulates the lysosome via mTOR and TFEB. EMBO J., 2012. **31**(5): p. 1095-1108.

10. Thelen, M., Winter, D., Braulke, T., Gieselmann, V., SILAC-Based Comparative Proteomic Analysis of Lysosomes from Mammalian Cells Using LC-MS/MS. Methods Mol. Biol., 2017. **1594**: p. 1–18.

11. Efeyan, A., Zoncu, R., Sabatini, D.M., Amino acids and mTORC1: From lysosomes to disease. Trends Mol. Med., 2012. **18**(9): p. 524–533.

12. Abbas, Y.M., Di Wu Bueler, S.A., Robinson, C.V., Rubinstein, J.L., Structure of V-ATPase from the mammalian brain. Science, 2020. **367**(6483): p. 1240–1246.

13. De Araujo, M.E., Naschberger, A., Fürnrohr, B.G., Stasyk, T., Dunzendorfer-Matt, T., Lechner, S., Welti, S., Kremser, L., Shivalingaiah, G., Offterdinger, M., Lindner, H.H., Huber, L.A., Scheffzek, K., Crystal structure of the human lysosomal mTORC1 scaffold complex and its impract on signaling. Science, 2017. **358**(6361): p. 377-381.

14. Cabukusta, B., Neefjes, J., Mechanisms of lysosomal positioning and movement. Traffic, 2018. **19**(10): p. 761-769.

15. Garg, S., Sharma, M., Ung, C., Tuli, A., Barral, D. C., Hava, D. L., Veerapen, N., Besra, G. S., Hacohen, N., Brenner, M. B., Lysosomal trafficking, antigen presentation, and microbial killing are controlled by the Arf-like GTPase Arl8b. Immunity, 2011. **35**(2): p. 182-193.

16. Balderhaar, H.J., Ungermann, C., CORVET and HOPS tethering complexes - coordinators of endosome and lysosome fusion. J. Cell Sci., 2013. **126**(6): p. 1307-1316.

17. Skowyra, M.L., P.H. Schlesinger, T.V. Naismith, and P.I. Hanson, Triggered recruitment of ESCRT machinery promotes endolysosomal repair. Science, 2018. **360**(6384).

18. Friedman, J.R., J.R. Dibenedetto, M. West, A.A. Rowland, and G.K. Voeltz, Endoplasmic reticulum-endosome contact increases as endosomes traffic and mature. Mol. Biol. Cell, 2013. **24**(7): p. 1030-1040.

19. Hao, F., K. Kondo, T. Itoh, S. Ikari, S. Nada, M. Okada, and T. Noda, Rheb localized on the Golgi membrane activates lysosome-localized mTORC1 at the Golgi-lysosome contact site. J. Cell Sci., 2018. **131**(3).

20. Chu, B.B., Y.C. Liao, W. Qi, C. Xie, X. Du, J. Wang, H. Yang, H.H. Miao, B.L. Li, and B.L. Song, Cholesterol transport through lysosome-peroxisome membrane contacts. Cell, 2015. **161**(2): p. 291-306.

21. Wong, Y.C., Ysselstein, D., Krainc, D., Mitochondria-lysosome contacts regulate mitochondrial fission via RAB7 GTP hydrolysis. Nature, 2018. **554**(7692): p. 382-386.

22. Platt, F.M., d'Azzo, Al., Davidson, B.L., Neufeld, E.F., Tifft, C.J., Lysosomal storage diseases. Nat. Rev. Dis. Primers, 2018. **4**(1): p. 1-25.

23. Dorfman, A., Matalon, R., The mucopolysaccharidoses (A Review). Proc. Natl. Acad. Sci. U.S.A., 1975. **73**(2): p. 630-637.

24. Mistry, P.K., G. Lopez, R. Schiffmann, N.W. Barton, N.J. Weinreb, and E. Sidransky, Gaucher disease: Progress and ongoing challenges. Mol. Genet. Metab., 2017. **120**(1-2): p. 8-21.

25. Masson, C., Cissé, I., Simon, V., Insalaco, P., Audran, M., Fabry disease: a review. Jt. Bone Spine, 2004. **71**(5): p. 381-383.

26. Biffi, A., Lucchini, G., Rovelli, A., Sessa, M., Metachromatic leukodystrophy: an overview of current and prospective treatments. Bone Marrow Transplant., 2008. **42**(2): p. S2-S6.

27. Zimran, A., Elstein, D., Management of Gaucher disease: enzyme replacement therapy. Pediatr. Endocrinol. Rev., 2014. **12** p. 82-87.

28. Hironaka, K., Yamazaki, Y., Hirai, Y., Yamamoto, M., Miyake, N., Miyake, K., Okada, T., Morita, A., Shimada, T.,, Enzyme replacement in the CSF to treat metachromatic leukodystrophy in mouse model using single intracerebroventricular injection of self-complementary AAV1 vector. Sci. Rep., 2015. **5**(1): p. 1-12.

29. Burlina, A., Niemann-Pick disease type C: introduction and main clinical features. J. Neurol., 2014. **261**(2): p. 525-527.

30. Vanier, M.T., Niemann-Pick disease type C. ORJD, 2010. **5**(1): p. 1-8.

31. Mellon, S.H., W. Gong, and M.D. Schonemann, Endogenous and synthetic neurosteroids in treatment of Niemann-Pick Type C disease. Brain Res. Rev., 2008. **57**(2): p. 410-420.

32. Goldman, S.D. and J.P. Krise, Niemann-Pick C1 functions independently of Niemann-Pick C2 in the initial stage of retrograde transport of membrane-impermeable lysosomal cargo. J. Biol. Chem., 2010. **285**(7): p. 4983-4994.

33. Chandramouli, K., Qian, P. Y., Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. Hum. Genomics Proteomics, 2009.

34. Mosen, P., A. Sanner, J. Singh, and D. Winter, Targeted Quantification of the Lysosomal Proteome in Complex Samples. Proteomes, 2021. **9**(1).

35. Itzhak, D.N., S. Tyanova, J. Cox, and G.H. Borner, Global, quantitative and dynamic mapping of protein subcellular localization. eLife, 2016. **5**: p. e16950.

36. Itzhak, D.N., C. Davies, S. Tyanova, A. Mishra, J. Williamson, R. Antrobus, J. Cox, M.P. Weekes, and G.H.H. Borner, A Mass Spectrometry-Based Approach for Mapping Protein Subcellular Localization Reveals the Spatial Proteome of Mouse Primary Neurons. Cell Rep., 2017. **20**(11): p. 2706-2718.

37. Valm, A.M., S. Cohen, W.R. Legant, J. Melunis, U. Hershberg, E. Wait, A.R. Cohen, M.W. Davidson, E. Betzig, and J. Lippincott-Schwartz, Applying systems-level spectral imaging and analysis to reveal the organelle interactome. Nature, 2017. **546**(7656): p. 162-167.

38. Arai, K., Kanaseki, T., Ohkuma, S., Isolation of highly purified lysosomes from rat liver: identification of electron carrier components on lysosomal membranes. J. Biochem., 1991. **110**(4): p. 541-547.

39. Diettrich, O., Mills, K., Johnson, A. W., Hasilik, A., Winchester, B. G., Application of magnetic chromatography to the isolation of lysosomes from fibroblasts of patients with lysosomal storage disorders. FEBS Lett., 1998. **441**(3): p. 369-372.

40. Walker, M.W. and E. Lloyd-Evans, A rapid method for the preparation of ultrapure, functional lysosomes using functionalized superparamagnetic iron oxide nanoparticles. Methods Cell Biol, 2015. **126**: p. 21-43.

41. Ponsford, A.H., T.A. Ryan, A. Raimondi, E. Cocucci, S.A. Wycislo, F. Frohlich, L.E. Swan, and M. Stagi, Live imaging of intra-lysosome pH in cell lines and primary neuronal culture using a novel genetically encoded biosensor. Autophagy, 2021. **17**(6): p. 1500-1518.

42. Zoncu, R., Bar-Peled, L., Efeyan, A., Wang, S., Sancak, Y., Sabatini, D. M., mTORC1 senses lysosomal amino acids through an inside-out mechanism that requires the vacuolar H(+)-ATPase. Science, 2011. **334**(6056): p. 678-683.

43. Abu-Remaileh, M., Wyant, G.A., Kim, C., Laqtom, N.N., Abbasi, M., Chan, S.H.,Freinkman, E., Sabatini, D.M., Lysosomal metabolomics reveals V-ATPase- and mTOR-dependent regulation of amino acid efflux from lysosomes. Science, 2017. **358**(6364): p. 807–813.

44. Singh, J., E. Kaade, J. Muntel, R. Bruderer, L. Reiter, M. Thelen, and D. Winter, Systematic Comparison of Strategies for the Enrichment of Lysosomes by Data Independent Acquisition. J. Proteome Res., 2020. **19**(1): p. 371-381.

45. Wendeler, M., Sandhoff, K., Hexosaminidase assays. Glycoconjugate J., 2009. **26**(8): p. 945–952.

46. Chapel, A., Kieffer-Jaquinod, S., Sagné, C., Verdon, Q, Ivaldi, C., Mellal, M., Thirion, J., Jadot, M., Bruley, C., Garin, J., Gasnier, B., Journet, A, An Extended Proteome Map of the Lysosomal Membrane Reveals Novel Potential Transporters. Mol. Cell. Proteomics, 2013. **12**(6): p. 1572–1588.

47. Sleat, D.E., M. Jadot, and P. Lobel, Lysosomal proteomics and disease. Proteomics Clin. Appl., 2007. **1**(9): p. 1134–1146.

48. Aebersold, R., Mann, M., Mass spectrometry-based proteomics. Nature, 2003. **422**(6928): p. 198-207.

49. Fenn, J.B., M. Mann, C.K. Meng, S.F. Wong, and C.M. Whitehouse, Electrospray ionization for mass spectrometry of large biomolecules. Science, 1989. **246**(4926): p. 64-71.

50. Karas, M., & Hillenkamp, F. , Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal. Bioanal. Chem., 1988. **60**(20): p. 2299-2301.

51. Steen, H., Mann, M., The ABC's (and XYZ's) of peptide sequencing. Nat. Rev. Mol. Cell Biol., 2004. **5**(9): p. 699-711.

52. Schubert, O.T., H.L. Rost, B.C. Collins, G. Rosenberger, and R. Aebersold, Quantitative proteomics: challenges and opportunities in basic and applied research. Nat. Protoc., 2017. **12**(7): p. 1289-1294.

53. Lossl, P., M. van de Waterbeemd, and A.J. Heck, The diverse and expanding role of mass spectrometry in structural and molecular biology. EMBO J., 2016. **35**(24): p. 2634-2657.

54. Pitt, J.J., Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. Clin. Biochem. Rev., 2009. **30**(1): p. 19.

55. Demarque, D.P., A.E. Crotti, R. Vessecchi, J.L. Lopes, and N.P. Lopes, Fragmentation reactions using electrospray ionization mass spectrometry: an important tool for the structural elucidation and characterization of synthetic and natural products. Nat. Prod. Rep., 2016. **33**(3): p. 432-455.

56. Han, H., Xia, Y., & McLuckey, S. A. , Ion trap collisional activation of c and z• ions formed via gas-phase ion/ion electron-transfer dissociation. J. Proteome Res., 2007. **6**(8): p. 3062-3069.

57. Perkins, D.N., Pappin, D. J., Creasy, D. M., Cottrell, J. S. , Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis, 1999. **20**(18): p. 3551-3567.

58. Bantscheff, M., Lemeer, S., Savitski, M. M., Kuster, B., Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Anal. Bioanal. Chem., 2012. **404**(4): p. 939-965.

59. Ong, S.E., B. Blagoev, I. Kratchmarova, D.B. Kristensen, H. Steen, A. Pandey, and M. Mann, 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. 376-386.

60. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. Anal. Bioanal. Chem., 2007. **389**(4): p. 1017-1031.

61. Zanivan, S., Krueger, M., & Mann, M., In vivo quantitative proteomics: the SILAC mouse. . Methods Mol. Biol. , 2012. **757**: p. 435-450.

62. Dayon, L., Hainard, A., Licker, V., Turck, N., Kuhn, K., Hochstrasser, D. F., Burkhard, P. R., Sanchez, J. C., Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. Anal. Chem., 2008. **80**(8): p. 2921-2931.

63. Ross, P.L., Y.N. Huang, J.N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, and D.J. Pappin, Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics, 2004. **3**(12): p. 1154-1169.

64. Gerber, S.A., J. Rush, O. Stemman, M.W. Kirschner, and S.P. Gygi, Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc. Natl. Acad. Sci. U. S. A., 2003. **100**(12): p. 6940-6945.

65. Zhu, W., Smith, J. W., Huang, C. M., Mass spectrometry-based label-free quantitative proteomics. J. Biomed. Biotechnol., 2010.

66. Picotti, P. and R. Aebersold, Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat. Methods, 2012. **9**(6): p. 555-66.

67. Chapman, J.D., Goodlett, D. R., Masselon, C. D., Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. Mass. Spectrom. Rev., 2014. **33**(6): p. 452-470.

68. Gillet, L.C., P. Navarro, S. Tate, H. Rost, N. Selevsek, L. Reiter, R. Bonner, and R. Aebersold, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol. Cell. Proteomics, 2012. **11**(6).

69. Bruderer, R., Bernhardt, O. M., Gandhi, T., Reiter, L., High-precision iRT prediction in the targeted analysis of data-independent acquisition and its impact on identification and quantitation. Proteomics, 2016. **16**(15-16): p. 2246-2256.

70. Sinitcyn, P., H. Hamzeiy, F. Salinas Soto, D. Itzhak, F. McCarthy, C. Wichmann, M. Steger, U. Ohmayer, U. Distler, S. Kaspar-Schoenefeld, N. Prianichnikov, S. Yilmaz, J.D. Rudolph, S. Tenzer, Y. Perez-Riverol, N. Nagaraj, S.J. Humphrey, and J. Cox, MaxDIA enables library-based and library-free data-independent acquisition proteomics. Nat. Biotechnol., 2021. **39**(12): p. 1563-1573.

71. O'Reilly, F.J. and J. Rappsilber, Cross-linking mass spectrometry: Methods and applications in structural, molecular and systems biology. Nat. Struct. Mol. Biol., 2018. **25**(11): p. 1000–1008.

72. Götze, M., C. Iacobucci, C.H. Ihling, and A. Sinz, A Simple Cross-Linking/Mass Spectrometry Workflow for Studying System-wide Protein Interactions. Anal. Chem., 2019. **91**(15): p. 10236–10244.

73. Klykov, O., B. Steigenberger, S. Pektas, D. Fasci, A.J.R. Heck, and R.A. Scheltema, Efficient and robust proteome-wide approaches for cross-linking mass spectrometry. Nat. Protoc., 2018. **13**(12): p. 2964–2990.

74. Barysz, H.M., Malmström, J., Development of Large-scale Cross-linking Mass Spectrometry. Mol. Cell. Proteomics, 2018. **17**(6): p. 1055–1066.

75. Chavez, J.D., Bruce, J.E., Chemical cross-linking with mass spectrometry: A tool for systems structural biology. Curr. Opin. Chem. Biol. , 2019. **48**: p. 8–18.

76. Liu, F., D.T.S. Rijkers, H. Post, and A.J.R. Heck, Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. Nat. Methods, 2015. **12**(12): p. 1179–1184.

77. Schweppe, D.K., J.D. Chavez, C.F. Lee, A. Caudal, S.E. Kruse, R. Stuppard, D.J. Marcinek, G.S. Shadel, R. Tian, and J.E. Bruce, Mitochondrial protein interactome elucidated by chemical cross-linking mass spectrometry. Proc. Natl. Acad. Sci. U.S.A., 2017. **114**(7): p. 1732–1737.

78. Leitner, A., R. Reischl, T. Walzthoeni, F. Herzog, S. Bohn, F. Forster, and R. Aebersold, Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. Mol. Cell. Proteomics, 2012. **11**(3).

79. Kao, A., C.L. Chiu, D. Vellucci, Y. Yang, V.R. Patel, S. Guan, A. Randall, P. Baldi, S.D. Rychnovsky, and L. Huang, Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. Mol. Cell. Proteomics, 2011. **10**(1).

80. Müller, M.Q., F. Dreiocker, C.H. Ihling, M. Schäfer, and A. Sinz, Cleavable cross-linker for protein structure analysis: Reliable identification of cross-linking products by tandem MS. Anal. Chem., 2010. **82**(16): p. 6958–6968.

81. Liu, F., P. Lossl, R. Scheltema, R. Viner, and A.J.R. Heck, Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. Nat. Commun., 2017. **8**(1): p. 1-8.

82. Gotze, M., J. Pettelkau, R. Fritzsche, C.H. Ihling, M. Schafer, and A. Sinz, Automated assignment of MS/MS cleavable cross-links in protein 3D-structure analysis. J. Am. Soc. Mass Spectrom., 2015. **26**(1): p. 83-97.

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

84. Yates, J.R., Gilchrist, A., Howell, K. E., Bergeron, J. J., Proteomics of organelles and large cellular structures. Nat. Rev. Mol. Cell. Biol., 2005. **6**(9): p. 702-714.

85. Ponnaiyan, S., F. Akter, J. Singh, and D. Winter, Comprehensive draft of the mouse embryonic fibroblast lysosomal proteome by mass spectrometry based proteomics. Sci. Data, 2020. **7**(1): p. 68.

86. Wang, L., Wu, D., Robinson, C.V., Wu, H., Fu, T.M., Structures of a Complete Human V-ATPase Reveal Mechanisms of Its Assembly. Mol. Cell, 2020. **80**(3): p. 501-511.

87. Fasci, D., van Ingen, H., Scheltema, R.A., Heck, A.J., Histone interaction landscapes visualized by crosslinking mass spectrometry in intact cell nuclei. Mol. Cell. Proteomics, 2018. **17**(10).

88. Albanese, P., Tamara, S., Saracco, G., Scheltema, R.A., Pagliano, C., How paired PSII-LHCII supercomplexes mediate the stacking of plant thylakoid membranes unveiled by structural mass-spectrometry. Nature communications, 2020. **11**(1): p. 1-14.

89. Klykov, O., C. van der Zwaan, A.J.R. Heck, A.B. Meijer, and R.A. Scheltema, Missing regions within the molecular architecture of human fibrin clots structurally resolved by XL-MS and integrative structural modeling. Proc. Natl. Acad. Sci. U.S.A., 2020. **117**(4): p. 1976–1987.

90. O'Reilly, F.J., L. Xue, A. Graziadei, L. Sinn, S. Lenz, D. Tegunov, C. Blötz, N. Singh, W.J.H. Hagen, P. Cramer, J. Stülke, J. Mahamid, and J. Rappsilber, In-cell architecture of an actively transcribing-translating expressome. Science, 2020. **369**(6503): p. 554–557.

91. Ryl, P.S.J., M. Bohlke-Schneider, S. Lenz, L. Fischer, L. Budzinski, M. Stuiver, M.M.L. Mendes, L. Sinn, F.J. O'Reilly, and J. Rappsilber, In Situ Structural Restraints from Cross-Linking Mass Spectrometry in Human Mitochondria. J. Proteome Res., 2019. **19**(1): p. 327–336.

92. Chavez, J.D., Lee, C. F., Caudal, A., Keller, A., Tian, R., Bruce, J. E., Chemical Crosslinking Mass Spectrometry Analysis of Protein Conformations and Supercomplexes in Heart Tissue. Cell Syst., 2018. **6**(1): p. 136-141.

93. Singh, J., S. Ponnaiyan, V. Gieselmann, and D. Winter, Generation of Antibodies Targeting Cleavable Cross-Linkers. Anal. Chem., 2021. **93**(8): p. 3762-3769.

94. Singh, J., H. Elhabashy, P. Muthukottiappan, M. Stepath, M. Eisenacher, O. Kohlbacher, V. Gieselmann, and D. Winter, Cross-linking of the Endolysosomal System Reveals Flotillin Structures and Putative Cargo. bioRxiv, 2022.

95. Ross, A.B., J.D. Langer, and M. Jovanovic, Proteome Turnover in the Spotlight: Approaches, Applications, and Perspectives. Mol. Cell. Proteomics, 2020. **20**.

96. Millar, A.H. and N.L. Taylor, Subcellular proteomics-where cell biology meets protein chemistry. Front. Plant Sci., 2014. **5**: p. 55.

97. Eubel, H., C.P. Lee, J. Kuo, E.H. Meyer, N.L. Taylor, and A.H. Millar, Free-flow electrophoresis for purification of plant mitochondria by surface charge. Plant J., 2007. **52**(3): p. 583-594.

98. Liu, D., P. Lössl, B.M. Rabbitts, R.S. Balaban, and A.J.R. Heck, The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes. Mol. Cell. Proteomics, 2018. **17**(2): p. 216–232.

99. Smith, D.L., M. Götze, T.K. Bartolec, G. Hart-Smith, and M.R. Wilkins, Characterization of the Interaction between Arginine Methyltransferase Hmt1 and Its Substrate Npl3: Use of Multiple Cross-Linkers, Mass Spectrometric Approaches, and Software Platforms. Anal. Chem., 2018. **90**(15): p. 9101–9108.

100. Gonzalez-Lozano, M.A., F. Koopmans, P.F. Sullivan, J. Protze, G. Krause, M. Verhage, K.W. Li, F. Liu, and A.B. Smit, Stitching the synapse: Cross-linking mass spectrometry into resolving synaptic protein interactions. Sci. Adv., 2020. **6**(8).

101. Huber, L.A. and D. Teis, Lysosomal signaling in control of degradation pathways. Curr. Opin. Cell Biol., 2016. **39**: p. 8-14.

102. Chen, Z.A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski-Wills, J.C., Nilges, M., Cramer, P., Rappsilber, J., Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. EMBO J., 2010. **29**(4): p. 717–726.

103. Schwanhausser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach, Global quantification of mammalian gene expression control. Nature, 2011. **473**(7347): p. 337-342.

104. Larance, M., Y. Ahmad, K.J. Kirkwood, T. Ly, and A.I. Lamond, Global subcellular characterization of protein degradation using quantitative proteomics. Mol. Cell. Proteomics, 2013. **12**(3): p. 638-50.

105. Mathieson, T., H. Franken, J. Kosinski, N. Kurzawa, N. Zinn, G. Sweetman, D. Poeckel, V.S. Ratnu, M. Schramm, I. Becher, M. Steidel, K.M. Noh, G. Bergamini, M. Beck, M. Bantscheff, and M.M. Savitski, Systematic analysis of protein turnover in primary cells. Nat. Commun., 2018. **9**(1): p. 689.

106. Boisvert, F.M., Ahmad, Y., Gierlinski, M., Charriere, F., Lamont, D., Scott, M., Barton, G., Lamond, A. I., A quantitative spatial proteomics analysis of proteome turnover in human cells. Mol. Cell. Proteomics, 2012. **11**(3).

107. Wu, C., Ba, Q., Li, W., Salovska, B., Hou, P., Mueller, T., Rosenberger, G.,, Gao, E., Di, Y., Liu, Y., Global and Site-Specific Effect of Phosphorylation on Protein Turnover. . Dev. Cell 2021. **56**(1): p. 111-124.

108. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D.A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Miller, E., Bache, S., Ooms, J., Robinson, D., Seidel, D. Spinu, V., Takahashi, K., Wilke, C., Woo, K.,Yutani, H., Welcome to the Tidyverse. J. Open Source Softw., 2019. **4**(43): p. 1686.

109. Ritchie, M.E., B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, and G.K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res., 2015. **43**(7): p. e47.

110. McShane, E., Sin, C., Zauber, H., Wells, J. N., Donnelly, N., Wang, X., Hou, J., Chen, W., Storchova, Z., Marsh, J. A., Valleriani, A., Selbach, M., Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation. Cell, 2016. **167**(3): p. 803-815.

111. Selbach, M., B. Schwanhausser, N. Thierfelder, Z. Fang, R. Khanin, and N. Rajewsky, Widespread changes in protein synthesis induced by microRNAs. Nature, 2008. **455**(7209): p. 58-63.

112. Goebel, T., S. Mausbach, A. Tuermer, H. Eltahir, D. Winter, V. Gieselmann, and M. Thelen, Proteaphagy in Mammalian Cells Can Function Independent of ATG5/ATG7. Mol. Cell. Proteomics, 2020. **19**(7): p. 1120-1131.

113. Runwal, G., E. Stamatakou, F.H. Siddiqi, C. Puri, Y. Zhu, and D.C. Rubinsztein, LC3-positive structures are prominent in autophagy-deficient cells. Sci. Rep., 2019. **9**(1): p. 10147.

114. Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B., Turk, D., Cysteine cathepsins: from structure, function and regulation to new frontiers. Biochim. Biophys. Acta., 2012. **1824**(1): p. 68-88.

115. Knight, C.G., & Barrett, A. J. , Interaction of human cathepsin D with the inhibitor pepstatin. Biochem. J., 1976. **155**(1): p. 117-125.

116. Brauer, A.U., Kuhla, A., Holzmann, C., Wree, A., Witt, M., Current Challenges in Understanding the Cellular and Molecular Mechanisms in Niemann-Pick Disease Type C1. Int. J. Mol. Sci., 2019. **20**(18): p. 4392.

117. Adli, M., The CRISPR tool kit for genome editing and beyond. Nat. Commun., 2018. **9**(1): p. 1-13.

118. Pickart, C.M., Back to the future with ubiquitin. Cell, 2004. **116**(2): p. 181-90.

119. Paramore, A. and S. Frantz, Bortezomib. Nat. Rev. Drug Discov., 2003. **2**(8): p. 611-613.

120. Jovanovic, M., Rooney, S., Mertins, P., Przybylski, D., Chevrier, N., Satija, R., Rodriguez, E., Fields, A., Schwartz, S., Raychowdhury, R., Mumbach, M., Eisenhaure, T., Rabani, M., Gennert, D., Lu, D., Delorey, T., Weissman, J., Carr, S., Hacohen, N., Regev, A., Dynamic profiling of the protein life cycle in response to pathogens. Science, 2015. **347**(6222).

121. Savitski, M.M., N. Zinn, M. Faelth-Savitski, D. Poeckel, S. Gade, I. Becher, M. Muelbaier, A.J. Wagner, K. Strohmer, T. Werner, S. Melchert, M. Petretich, A. Rutkowska, J. Vappiani, H. Franken, M. Steidel, G.M. Sweetman, O. Gilan, E.Y.N. Lam, M.A. Dawson, R.K. Prinjha, P. Grandi, G. Bergamini, and M. Bantscheff, Multiplexed Proteome Dynamics Profiling Reveals Mechanisms Controlling Protein Homeostasis. Cell, 2018. **173**(1): p. 260-274.

122. Baum, K., Schuchhardt, J., Wolf, J., Busse, D., Of Gene Expression and Cell Division Time: A Mathematical Framework for Advanced Differential Gene Expression and Data Analysis. Cell Syst., 2019. **9**(6): p. 569-579.

123. Pino, L.K., J. Baeza, R. Lauman, B. Schilling, and B.A. Garcia, Improved SILAC Quantification with Data-Independent Acquisition to Investigate Bortezomib-Induced Protein Degradation. J. Proteome Res., 2021. **20**(4): p. 1918-1927.

124. Vasiljeva, O., Reinheckel, T., Peters, C., Turk, D., Turk, V., Turk, B., Emerging roles of cysteine cathepsins in disease and their potential as drug targets. Curr. Pharm. Des., 2007. **13**(4): p. 387-403.

125. Martin-Perez, M. and J. Villen, Determinants and Regulation of Protein Turnover in Yeast. Cell Syst., 2017. **5**(3): p. 283-294.

126. Eisenberg, A.R., Higdon, A., Keskin, A., Hodapp, S., Jovanovic, M., Brar, G. A., Precise Post-translational Tuning Occurs for Most Protein Complex Components during Meiosis. Cell Rep., 2018. **25**(13): p. 3603-3617.

127. Maxson, M.E. and S. Grinstein, The vacuolar-type H^+-ATPase at a glance - more than a proton pump. J. of Cell Scl., 2014. **127**(Pt 23): p. 4987–4993.

128. Geiger, T., Wehner, A., Schaab, C., Cox, J., Mann, M., Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. Mol. Cell. Proteomics, 2012. **11**(3).

129. Khoury, G.A., R.C. Baliban, and C.A. Floudas, Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. Sci. Rep., 2011. **1**(1): p. 1-5.

130. Bekker-Jensen, D.B., Bernhardt, O. M., Hogrebe, A., Martinez-Val, A., Verbeke, L., Gandhi, T., Kelstrup, C. D., Reiter, L., Olsen, J. V., Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. Nat. Commun., 2020. **11**(1): p. 1-12.

131. Sroubek, J., McDonald, T. V., Protein kinase A activity at the endoplasmic reticulum surface is responsible for augmentation of human ether-a-go-go-related gene product (HERG). J. Biol. Chem., 2011. **286**(24): p. 21927-21936.

132. Liu, Z., Y. Lv, N. Zhao, G. Guan, and J. Wang, Protein kinase R-like ER kinase and its role in endoplasmic reticulum stress-decided cell fate. Cell Death Dis., 2015. **6**(7): p. e1822.

133. Farhan, H., Tyrosine kinase signaling in and on the endoplasmic reticulum. Biochem. Soc. Trans., 2020. **48**(1): p. 199-205.

134. Aizawa, S., Y. Fujiwara, V.R. Contu, K. Hase, M. Takahashi, H. Kikuchi, C. Kabuta, K. Wada, and T. Kabuta, Lysosomal putative RNA transporter SIDT2 mediates direct uptake of RNA by lysosomes. Autophagy, 2016. **12**(3): p. 565-578.

135. Aizawa, S., V.R. Contu, Y. Fujiwara, K. Hase, H. Kikuchi, C. Kabuta, K. Wada, and T. Kabuta, Lysosomal membrane protein SIDT2 mediates the direct uptake of DNA by lysosomes. Autophagy, 2017. **13**(1): p. 218-222.

136. Courville, P., M. Quick, and R.J. Reimer, Structure-function studies of the SLC17 transporter sialin identify crucial residues and substrate-induced conformational changes. J. Biol. Chem., 2010. **285**(25): p. 19316-19323.

137. Miyaji, T., N. Echigo, M. Hiasa, S. Senoh, H. Omote, and Y. Moriyama, Identification of a vesicular aspartate transporter. Proc. Natl .Acad. Sci .U.S.A, 2008. **105**(33): p. 11720-11724.

138. Shin, H.R. and R. Zoncu, The Lysosome at the Intersection of Cellular Growth and Destruction. Dev. Cell, 2020. **54**(2): p. 226-238.

139. Wilhelm, L.P., C. Wendling, B. Védie, T. Kobayashi, M.P. Chenard, C. Tomasetto, G. Drin, and F. Alpy, STARD3 mediates endoplasmic reticulum-to-endosome cholesterol transport at membrane contact sites. EMBO J., 2017. **36**(10): p. 1412-1433.

8. Supplement

List of Contents:

List of Tables:

1. Supplementary tables for chapter 1 (published) can be accessed online.
2. Supplementary tables for chapter 2 (published) can be accessed online.
3. Supplementary tables for chapter 3 (published as a preprint) can be accessed from the provided CD.
    - Table S1: List of lysosomal proteins
    - Table S2: Protein IDs & iBAQ values (LEF)
    - Table S3: Cross-links (LEF)
    - Table S4: Distance constraints cross-links
    - Table S5: Protein IDs & iBAQ values (EEF)
    - Table S6: Cross-links (EEF)
    - Table S7: Flotillin positive endosome cargo
4. Supplementary tables for chapter 4 (not published) can be accessed from the provided CD.
    - Table S1: List of lysosomal proteins
    - Table S2: Protein half-lives
    - Table S3: Phosphoproteomics dataset

8.1 Co-author Publication I:

2 Part two of chapter 1 has been published: [Ponnaiyan, S., Akter, F., **Singh, J.**, Winter, D. (2020). Comprehensive draft of the mouse embryonic fibroblast lysosomal proteome by mass spectrometry-based proteomics. Scientific data, 7(1), 1-13].

- For the online version of the paper and supporting information, visit: https://doi.org/10.1038/s41597-020-0399-5
- **Remark:** The second publication within the scope of chapter 1 **SHOULD NOT BE CONSIDERED** for this cumulative doctoral dissertation, as the majority of the work has been done by the first two co-authors.

www.nature.com/scientificdata

SCIENTIFIC DATA

Check for updates

OPEN

DATA DESCRIPTOR

Comprehensive draft of the mouse embryonic fibroblast lysosomal proteome by mass spectrometry based proteomics

Srigayatri Ponnaiyan[1,3], Fatema Akter[1,2,3], Jasjot Singh[1] & Dominic Winter[1]

Lysosomes are the main degradative organelles of cells and involved in a variety of processes including the recycling of macromolecules, storage of compounds, and metabolic signaling. Despite an increasing interest in the proteomic analysis of lysosomes, no systematic study of sample preparation protocols for lysosome enriched fractions has been performed to date. In the current study, we used samples enriched for lysosomes by paramagnetic nanoparticles and systematically evaluated experimental parameters for the analysis of the lysosomal proteome. This includes different approaches for the concentration of lysosome-containing fractions; desalting of samples by solid phase extraction; fractionation of peptide samples; and different gradient lengths for LC-MS/MS analyses of unfractionated samples by data dependent and data independent acquisition. Furthermore, we evaluated four different digestion methods including filter aided sample preparation (FASP), in-gel digestion, and in-solution digestion using either RapiGest or urea. Using the combined data, we generated a benchmark lysosomal proteome data set for mouse embryonic fibroblasts as well as a spectral library for the analysis of lysosomes by data independent acquisition.

Background & Summary

Lysosomes are the main degradative compartments of mammalian cells and contain a variety of hydrolases which catalyze the breakdown of virtually any cellular macromolecule. Malfunctions of hydrolases leads to the accumulation of their respective substrate, resulting in so-called lysosomal storage disorders (LSDs), a group of about 50 genetically different but phenotypically connected severe diseases[1]. Due to the direct relation between lysosomal hydrolase malfunction and disease phenotype, this group of enzymes has been thoroughly investigated, and the mechanisms of cellular macromolecule degradation in lysosomes are relatively well understood[2]. While no mechanisms for the regulation of lysosomal hydrolases are known, it is becoming more and more apparent that lysosomes play an important role in the distribution and regulation of cellular metabolites, and that they are significantly involved in cellular signaling, which is regulated e.g. by phoshorylation[3]. Furthermore, it is by now well-established that the impairment of lysosomal function plays often a crucial role in more common diseases as e.g. neurodegenerative disorders[4] and cancer[5]. Therefore, there is an increasing interest in the analysis of lysosomes.

The method of choice for the unbiased analysis of organelle-specific proteomes is mass spectrometry-based proteomics. Due to the low abundance of lysosomal proteins in mammalian cells, their enrichment is a prerequisite for proteomic analysis. Generally, it can be differentiated if a method aims at analyzing the whole proteome, which is divided in subcellular fractions, and lysosomes are one of them[6], or if lysosomes are the main target of the analysis[7].

One of the most commonly used methods for subcellular fractionation is density gradient centrifugation[6,8]. Several discontinuous approaches have been developed for the generation of lysosome enriched fractions using Mitrizamide[9], Nycodenz[10], and Percoll[11,12]. To further increase specificity, a change in lysosome density can

[1]Institute for Biochemistry and Molecular Biology, Medical Faculty, University of Bonn, Bonn, 53115, Germany. [2]Department of Pharmacology, Faculty of Veterinary Science, Bangladesh Agricultural University, Mymensingh, 2202, Bangladesh. [3]These authors contributed equally: Srigayatri Ponnaiyan, Fatema Akter. ✉e-mail: dominic.winter@uni-bonn.de

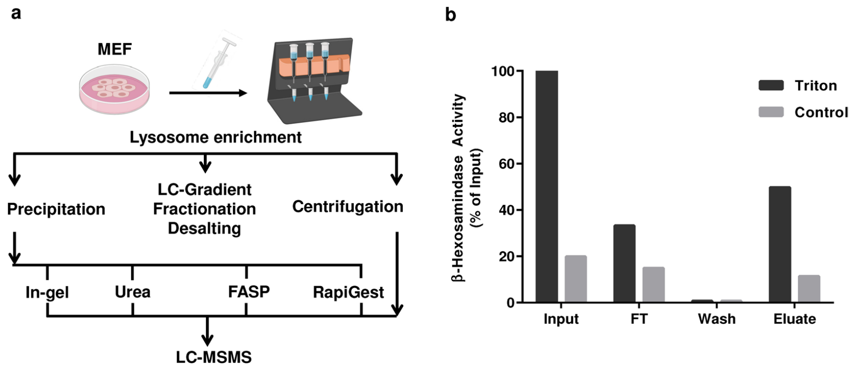


Fig. 1 Sample processing workflow and assessment of lysosomal yield. **(a)** Workflow for the sample preparation of MEF lysosome enriched fractions for LC-MS/MS analysis. Individual parts of the protocol which were evaluated are indicated. **(b)** Assessment of the efficiency of lysosome enrichment and lysosomal integrity by enzymatic β-hexosaminidase assays. Differences in enzymatic activity between control (intact lysosomes) and triton treated (disrupted lysosomes) samples allow to determine the amount of intact lysosomes in the individual fractions. MEFs: mouse embryonic fibroblasts; FASP: filter aided sample preparation.

be induced e.g. by injection of Triton-WR1339 in animals, which leads to a liver-specific change of lysosomal density[13].

Another commonly used method utilizes the specific targeting of magnetic nanoparticles (iron dextran particles (FeDex[14])/superparamagnetic iron oxide nanoparticles (SPIONs[15])) to the lysosomal compartment by delivery through unspecific fluid-phase endocytosis. The particle-containing lysosomes can then be isolated through a magnetic field. This approach has been utilized in several studies for comparative proteomics experiments as well as lipidomics studies[7,16]. Recently, another approach has been introduced for the enrichment of lysosomes, the immunoprecipitation via tagged lysosomal membrane proteins. This approach was initially established using a RFP-Flag tagged version of the lysosomal membrane protein Lamp1[17], and later extended to a HA-tagged version of TMEM192, which was utilized for the metabolomic and proteomic analysis of isolated lysosomes[18,19].

Besides the enrichment of lysosomes as an intact organelle, a unique feature of lysosomal matrix proteins has been used extensively for their affinity purification: their posttranslational modification with mannose 6 phosphate (M6P), which acts as a lysosomal targeting signal[20] and is removed by the acid phosphatases ACP2 and ACP5[21] in the lysosomal lumen. Lysosomal proteins which are still carrying the M6P residue can be enriched by immobilized domains of the M6P receptors MPR46/MPR300, or resins used for the enrichment of phosphopeptides, such as IMAC[22]. This approach has been applied to mouse embryonic fibroblasts deficient for both MPRs[23], to human brain and plasma samples[24,25], and to 17 individual rat tissues[26], revealing novel proteins of potential lysosomal origin. Additionally, changes in the lysosomal proteome in a mouse model of Niemann Pick Disease Type C[27] and in patients affected by LSDs of unknown etiology[28] were investigated. In order to increase the amount of M6P modified proteins, mice and cells deficient for ACP2 and ACP5 were used in several studies[22,29,30].

Despite many studies dealing with the isolation and mass spectrometric investigation of lysosomes, the protocols for their proteomic analysis have not been optimized to date, and sample preparation varies strongly between different datasets. In the current study, we used lysosomes isolated from mouse embryonic fibroblasts (MEFs) to systematically evaluate major steps of sample preparation and mass spectrometric analysis for lysosome-enriched fractions. We analyzed LC gradient lengths, solid phase extraction resins, peptide fractionation, and concentration approaches for lysosomes and lysosomal proteins in combination with protocols for proteolytic digestion (Fig. 1a, Table 1). Using these datasets, we generated a high confidence draft of the proteome of MEF lysosome enriched fractions, and a spectral library for their analysis by data independent acquisition (DIA).

Methods

Cell culture and isolation of lysosomes. Mouse embryonic fibroblasts (MEFs) were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal calf serum (FCS), 100 IU/mL penicillin, 100 µg/mL streptomycin, and 2 mM L-glutamine at 37 °C and 5% CO$_2$. For lysosome isolation, 3×10^6 cells were seeded per 10 cm plate and cultured in DMEM with 2.5% FCS for 72 h[7]. To each plate, 1 mL magnetite solution (EndoMAG40, Liquids Research, North Wales, UK) was added followed by 24 h incubation. Subsequently, the medium was exchanged, the cells were washed twice with 1x phosphate buffered saline (PBS), and a 24h chase was performed in DMEM with 10% FCS. For harvesting, the cells were washed twice with ice-cold 1x PBS and scraped off the plate in 2 mL isolation buffer (250 mM sucrose, 10 mM HEPES/NaOH pH 7.4, 1 mM CaCl$_2$, 15 mM KCl, 1 mM MgCl$_2$, 1.5 mM MgAc, 1 mM dithiothreitol (DTT), 1x cOmplete EDTA-free protease inhibitor cocktail (Roche Diagnostics, Mannheim, Germany)) per plate. The cell suspension was homogenized with a 15 mL douncer, nuclei and intact cells were pelleted by centrifugation at 4 °C, 600 g for 10 min, and the post-nuclear supernatant was transferred to a new tube. This process was repeated and the post-nuclear supernatants

Sample	Source	Enriched Organelle	Method	Acquisition Strategy	Replicates
1 h	MEFs	Lysosome	Gradient	DDA	3
2 h	MEFs	Lysosome	Gradient	DDA	3
3 h	MEFs	Lysosome	Gradient	DDA	3
4 h	MEFs	Lysosome	Gradient	DDA	3
6 fractions	MEFs	Lysosome	Fractionation	DDA	3
3 fractions	MEFs	Lysosome	Fractionation	DDA	3
Unfractionated	MEFs	Lysosome	Fractionation	DDA	3
Stage Tip	MEFs	Lysosome	Desalting	DDA	3
Oasis	MEFs	Lysosome	Desalting	DDA	3
Sep-Pak	MEFs	Lysosome	Desalting	DDA	3
In-gel CN	MEFs	Lysosome	Digestion	DDA	3
In-gel PN	MEFs	Lysosome	Digestion	DDA	3
Urea CN	MEFs	Lysosome	Digestion	DDA	3
Urea PN	MEFs	Lysosome	Digestion	DDA	3
FASP CN	MEFs	Lysosome	Digestion	DDA	3
FASP PN	MEFs	Lysosome	Digestion	DDA	3
RapiGest CN	MEFs	Lysosome	Digestion	DDA	3
RapiGest PN	MEFs	Lysosome	Digestion	DDA	3
0.5 h	MEFs	Lysosome	Gradient	DIA	3
1 h	MEFs	Lysosome	Gradient	DIA	3
2 h	MEFs	Lysosome	Gradient	DIA	3
3 h	MEFs	Lysosome	Gradient	DIA	3
4 h	MEFs	Lysosome	Gradient	DIA	3

Table 1. Samples analyzed by mass spectrometry. For all analyses, lysosomes isolated from MEFs were used. Different methods were applied for sample digestion, desalting, fractionation, liquid chromatography, and mass spectrometric data acquisition. Which method parameter was compared with the individual *.raw files is indicated in the column "Method". MEFs: mouse embryonic fibroblasts; CN: concentration by centrifugation; PN: concentration by precipitation; DDA: data dependent acquisition; DIA: data independent acquisition.

were pooled. The resulting samples were then passed by gravity through a LS column located in a MidiMACS Separator (both Miltenyi Biotech, Auburn, CA). Subsequently, the column was washed with 5 mL isolation buffer, removed from the stand, and the lysosome fraction was eluted by a plunger in 2 × 1 mL isolation buffer. The efficiency of lysosome enrichment and lysosomal integrity was assessed using the β-hexosaminidase assay[7]: lysosome samples were incubated with the substrate 4-Nitrophenyl N-acetyl-β-D-glucosaminide in 0.1 M sodium citrate buffer (pH 4.6) containing 0.2% BSA with/without addition of 0.8% Triton X-100 (final concentration) at 37 °C for 15 min. The reaction was stopped by addition of 0.4 M glycine (pH 10.4) and the absorbance was measured at 405 nm. The protein concentration was determined using the DC protein assay (BioRad, Hercules, CA)[31]. For pelleting of intact lysosomes, the elution fraction was centrifuged at 4 °C, 20,000 g for 30 min, and the supernatant discarded. For protein precipitation, lysosomes were lyzed by addition of 1% Triton X-100 (final concentration), ice-cold chloroform/methanol (2:1, v/v) was added at a ratio of 5 to 1 (v/v), and the mixture was incubated on ice for 1 h. Following centrifugation at 4 °C, 20,000 g for 30 min, the liquid phases were discarded, the protein layer was washed with 1 mL ice cold methanol, and centrifuged at 4 °C, 14,000 g for 20 min and the supernatant discarded.

In-gel digestion. Samples were denatured and reduced by addition of 1x modified Laemmli buffer[32] (62.5 mM Tris-HCl, pH 6.8, 2% SDS, 10% glycerol, 5 mM DTT, 0.001% bromophenol blue) and incubation at 40 °C for 10 min. Proteins were alkylated with 20 mM acrylamide at room temperature (RT) for 30 min in the dark, loaded onto a 10% SDS gel, and electrophoresis was performed until the sample had migrated for ~1 cm into the separation gel. The gel was stained with Coomassie brilliant blue and the whole section of the gel containing the sample cut into ~1 mm^3 cubes. In-gel digestion was performed as described elsewhere[33]. Briefly, the gel pieces were destained by 30% ACN/0.07 M NH_4HCO_3, dehydrated by 100% ACN, dried in a vacuum centrifuge, and digested with 1 µg trypsin (Promega, Madison, WI) in 0.1 M NH_4HCO_3 at 37 °C overnight. For the recovery of peptides, the supernatant was transferred to a new tube and the gel pieces were incubated consecutively with 0.1% TFA/50% ACN, 0.1 M NH_4HCO_3, and 100% ACN. The supernatants of the individual steps were pooled and dried using a vacuum centrifuge.

Urea in-solution digestion. Samples were resuspended in 8 M urea/0.1 M TEAB[34], and incubated at RT, 800 rpm for 45 min. Proteins were reduced with 5 mM DTT (final concentration) at 56 °C, 800 rpm for 25 min, alkylated with 20 mM acrylamide at RT for 30 min in the dark, and the reaction was quenched by addition of 5 mM DTT. Subsequently, the concentration of urea was reduced to 4 M, rLys-C (Promega) added at an enzyme to protein ratio of 1 to 100, and the sample incubated at 37 °C overnight. Subsequently, the samples were diluted to a final concentration of 1.6 M urea with 0.1 M TEAB, trypsin was added at an enzyme to protein ratio of 1 to 100, and the sample was incubated at 37 °C for 10 h. Finally, the samples were acidified using acetic acid (AcOH, 0.1% final concentration).

Filter aided sample preparation (FASP). Samples were solubilized in 20 µL 4% SDS/0.1 M Tris-HCl, pH 7.6 at 40 °C for 5 min, and reduced with 0.1 M DTT (final concentration) at 56 °C for 5 min. Subsequently, FASP digestion was performed as described elsewhere[35] with slight modifications. Briefly, samples were mixed with 200 µL of UA (8 M urea/0.1 M TEAB) and added to a filter unit (Microcons, 30 kDa cut off, Merck Millipore, Darmstadt, Germany). Subsequently, buffers were exchanged with UA solution by two consecutive centrifugation steps at RT, 14,000 g for 15 min, and proteins were alkylated by addition of 100 µL AA solution (0.05 M acrylamide in UA)[36] at RT for 20 min. The filter units were then washed twice with 100 µl of 0.05 M NH_4HCO_3 by centrifugation at 14,000 g for 10 min. Subsequently, 60 µl of 0.05 M NH_4HCO_3 and 10 µl of trypsin solution (0.1 µg/µl) were added and the sample was incubated in a wet chamber at 37 °C overnight. The digested peptides were recovered from the filter units by centrifugation at 14,000 g for 10 min and subsequent elution with 50 µl 0.5 M NaCl followed by centrifugation. Eluted peptides were acidified with AcOH (0.1% final concentration).

RapiGest in-solution digestion. Samples were solubilized in 1% RapiGest (Waters, Milford, MA)/0.1 M NH_4HCO_3, pH 7.8 at 37 °C for 45 min and diluted 1:1 with 0.1 M NH_4HCO_3. Proteins were reduced with DTT (5 mM final concentration) at 56 °C for 25 min, alkylated with acrylamide (20 mM final concentration) at RT for 30 min, and the reaction was quenched by addition of 5 mM DTT. Samples were further diluted to a final concentration of 0.1% RapiGest with 0.1 M NH_4HCO_3 (protein concentration: 1 µg/µL). Proteins were digested with trypsin (enzyme to protein ratio of 1 to 100) at 37 °C overnight. The next day, RapiGest was hydrolyzed by the addition of 1% TFA (final concentration) and incubation at 37 °C, 800 rpm for 30 min, followed by its precipitation at RT, 20,000 g for 10 min. The supernatants were transferred to new tubes.

Desalting of peptides. Peptides were desalted either using 10 mg Oasis HLB cartridges (Waters), 10 mg Sep-Pak C_{18} 1 cc Vac cartridges (Waters), or Stage Tips[37] assembled with 3 M™ Empore™ C_{18} reversed phase material (3 M, St. Paul, MN). Oasis and Sep-Pak cartridges were equilibrated with MeOH, 70% ACN/0.5% AcOH, and 0.5% AcOH, the samples loaded, and the cartridges washed three times with 1 mL 0.5% AcOH. Peptides were eluted consecutively with 500 µL 30% ACN/0.5% AcOH, 300 µL 50% ACN/0.5% AcOH and 300 µL 70% ACN/0.5% AcOH and the eluate fractions were combined. In case of C_{18} Stage Tips, tip columns were prepared with 3 layers of C_{18} disks in a 200 µL pipette tip. Tip columns were equilibrated sequentially with MeOH, 80% ACN/0.5% AcOH, and 0.5% AcOH by centrifugation at 600 g for 1 min. Acidified samples were loaded onto the column, washed with 100 µL 0.5% AcOH, and peptides were eluted twice with 20 µL of 80% ACN/0.5% AcOH by centrifugation at 600 g for 1min. Eluate fractions were dried using a vacuum centrifuge.

Pipette tip based strong anion exchange (SAX) fractionation of peptides. SAX fractionation was performed as described elsewhere[38]. Briefly, a pipet tip SAX column was assembled using 12 disks of Empore Anion-SR material (3 M) and C_{18} Stage Tips were generated using 3 disks of Empore C_{18} material. SAX buffers were composed of 20 mM AcOH, 20 mM phosphoric acid, and 20 mM boric acid. The pH of the individual solutions was adjusted to pH 11, 8, 6, 5, 4, and 3 by addition of NaOH. Subsequently, NaCl was added to the final elution buffer (pH 3) at a concentration of 0.25 M. SAX columns were equilibrated by sequential addition of 100 µL of MeOH, 1 M NaOH, and loading buffer (pH 11), each in combination with subsequent centrifugation at 7,000 g for 3 min. Stage Tips were equilibrated with 100 µL of MeOH, 80% ACN/0.5% AcOH, and water. The dried peptide samples were resuspended in 200 µL pH 11 buffer, loaded on the SAX column, and fractionation was performed by centrifugation at 7,000 g for 3 min for each step. The flow-through and the individual elution fractions were captured on the C_{18} Stage Tips, which were further washed with 100 µL of 0.5% AcOH, and eluted by 80% ACN/0.5% AcOH. The desalted peptides were dried using a vacuum centrifuge.

UHPLC-MS/MS data acquisition. Analyses were performed using a Dionex Ultimate 3000 system coupled to an Orbitrap Fusion Lumos mass spectrometer (both Thermo Scientific, Bremen, Germany). Columns were produced in-house as follows: 50 cm spray tips were generated from 360 µm outer diameter/100 µm inner diameter fused silica capillaries using a P-2000 laser puller (Sutter Instruments, Novato, CA) and packed with 1.9 µm Reprosil AQ C_{18} particles (Dr. Maisch, Ammerbuch-Entringen, Germany). Peptides were resuspended in 5% ACN/5% FA and loaded on the analytical column at a flow rate of 600 nL/min, 100% solvent A (0.1% FA in water). Subsequently, the separation was performed at a flow rate of 300 nL/min with 60, 120, 180, and 240 min linear gradients from 5–35% solvent B (95% ACN/0.1% FA). Survey spectra were acquired in the Orbitrap mass analyzer with a mass range of m/z 375–1,575 at a resolution of 60,000. MS/MS fragmentation was performed in the data dependent acquisition mode for charge states between 2–4 by HCD and data were acquired in the Orbitrap at a resolution of 30,000. The cycle time was set to 5 s and the precursor isolation width to 1.6 m/z using the quadrupole. For MS1 and MS2 scans, the automatic gain control (AGC) was set to 4×10^5 and 5×10^5, respectively. Fragmented ions were excluded from further fragmentation for 30 s, 60 s, 90 s, and 120 s, respectively, for the four different gradient lengths. For data-independent acquisition (DIA) analysis of the samples, the following method was applied: One MS1 scan with a resolution of 120,000, an AGC target setting of 5×10^5, and a maximum injection time of 20 ms covering a mass range of 350 to 1,200 m/z was performed followed by 18 to 58 static DIA scans depending on the gradient length (0.5 h: 18 scans, 1 h: 24 scans, 2 h: 36 scans, 3 h: 47 scans, 4 h: 58 scans). The isolation window widths were adjusted for each gradient length to cover the same mass range as the MS1 scan including a 0.5 m/z overlap (0.5 h: 47.7 m/z, 1 h: 35.9 m/z, 2 h: 24.1 m/z, 3 h: 18.6 m/z, 4 h: 15.2 m/z). The DIA scans were acquired with a resolution of 30,000, an AGC target setting of 1×10^6, and a maximum injection time of 60 ms. The HCD collision energy was set to 27% and the resulting cycle times based on the window designs were as follows for the individual methods: 0.5 h: 1.89 s, 1 h: 2.34 s, 2 h: 3.44 s, 3 h: 4.46 s and 4 h: 5.45 s.

Data analysis – data dependent acquisition (DDA). Thermo *.raw data were analyzed with Proteome Discoverer 2.2 (Thermo Fisher Scientific, Bremen, Germany) in combination with Mascot (www.matrixscience.com). For database searching, Uniprot *Mus musculus* (release 2019_04, 54,425 entries) in combination with the cRAP database (ftp://ftp.thegpm.org/fasta/cRAP/crap.fasta) including common contaminants was used with the following parameters: variable modifications: oxidation of methionine, acetylation of protein N-termini; fixed modification: propionamide at cysteine; mass tolerance: 10 ppm for precursor ions, 50 mmu for fragment ions; enzyme: trypsin except proline was the next amino acid; missed cleavage sites: 2. Data were filtered with a false discovery rate (FDR) of 1% at the peptide level using Percolator and proteins were exported with a FDR of 1%. Label free quantification was performed using the Minora feature detector node in Proteome Discoverer.

Data processing – data dependent acquisition (DDA). Only high confidence identifications were exported to MS Excel for further analyses. Numbers of lysosomal proteins were determined from protein files by comparison to a list of confirmed lysosomal proteins (figshare deposit[39]: Table 9_Lysosomal Protein List) generated by merging of a manually curated bona fide list[6–8,10,19,40–42] and a publicly available gene ontology database (www.pantherdb.org). Peptide spectral match (PSM) and peptide numbers were determined from the PSM files. For label free quantification, proteins with an average intensity ratio of $\log2 > 1$ or $\log2 < 0.5$ and a $p\text{-value} < 0.05$ were considered to be significantly over-/underrepresented. Missed cleavage rates for the individual digestion methods were determined from the PSM files by calculating the number of peptides with one or more missed cleavage sites and normalization on the total number of identified peptides. For identification of semi-tryptic peptides, database searches were repeated with enzyme specificity set to semi-trypsin, followed by normalization of identified semi-tryptic peptides on the number of total peptides identified.

Data analysis – data independent acquisition (DIA). DIA data were analyzed using the Pulsar[43] algorithm available in Spectronaut (Version: 13.2.19, Biognosys, Schlieren, Switzerland). A spectral library was generated based on the same parameters as defined for the analysis of the DDA data with Proteome Discoverer 2.2 except the mass tolerances, which were assigned dynamically by the Pulsar algorithm. To build the library, 3 to 6 fragment ions per peptide were selected based on their intensity. All DIA data were analyzed using this library in combination with the default settings of Spectronaut. For retention time alignment, the high precision iRT concept was applied[43]. Peak extraction windows, as well as the mass tolerances for the matching of precursor and fragment ions, were determined automatically by Spectronaut. For peak detection, a minimum requirement of 3 fragment ions was defined, whereby precursor information was only used to enhance peak detection. Data normalization was performed using local regression localization with enabled interference correction. Data were filtered at 1% FDR on the peptide precursor and protein level applying a Q-value cut-off of <0.01[44]. The generated Spectronaut project file can be viewed using the freely available Spectronaut viewer.

Data Records

The mass spectrometry data and analysis files have been deposited to the ProteomeXchange Consortium (http://www.proteomexchange.org) via the PRIDE partner repository[38]. The DDA dataset includes 75 *.raw files representing all experimental conditions (Gradient tests: 4 conditions; Desalting tests: 3 conditions; Fractionation tests: 3 conditions; Digestion tests: 8 conditions) from three experimental replicates each. The fractionation dataset includes *.raw files for each individual fraction. The DIA dataset includes 15 *.raw files comprising 0.5, 1, 2, 3 and, 4 h gradient length tests with three replicates each. Furthermore, the dataset includes the result files originating from Proteome Discoverer (7x .pdResult files, 7x pepXML search result files, 7x .pdStudy files and 15x MSF files) and one result file from Spectronaut. In addition, the protein list data from the .pdResult files are available as excel tables for each experiment. These individual analyses, as well as the list of confirmed lysosomal proteins, can be accessed through a figshare deposit[39].

Technical Validation

In order to provide a reproducible starting material for all analyses, we generated a large batch of lysosome enriched fractions from forty-eight 10 cm dishes of mouse embryonic fibroblasts (MEFs) employing superparamagnetic iron oxide nanoparticles (SPIONs)[7,15]. To assess the purity and the amount of intact lysosomes, we performed enzyme activity assays for β-hexosaminidase, a hydrolase residing in the lysosomal lumen. We were able to recover ~80% of the intact lysosomes contained in the starting material and the enrichment efficiency of the magnetic column was 62% (Fig. 1b). In the eluate fraction, 77% of lysosomes were intact (determined by the difference in enzyme activity with/without Triton X-100, Fig. 1b). When enriched by SPIONs, lysosomes are eluted from a magnetic column in a rather big volume and therefore the sample needs to be concentrated. For this purpose, and the removal of the isolation buffer which may interfere with tryptic digestion, we employed two strategies: 1) the precipitation of all proteins by chloroform/methanol (precipitation samples, PN); and 2) the concentration of intact lysosomes by centrifugation (centrifugation samples, CN). For each approach, we prepared 24 identical aliquots, which were stored at −80 °C until further use and determined the protein concentration for one representative aliquot.

Impact of LC gradient length on protein identification. Initially, we determined the impact of the LC gradient on the identification rates of peptides and proteins by analyzing 1 µg of urea digested PN sample in triplicates with four different gradient lengths (1 h, 2 h, 3 h, and 4 h). In comparison to 1 h gradients (2367 protein groups on average, 1962 identified in all 3 replicates), an increase in analysis time resulted in an average gain of 24%, 69%, and 84% proteins groups for the 2 h, 3 h, and 4 h gradients, respectively (Fig. 2a, and

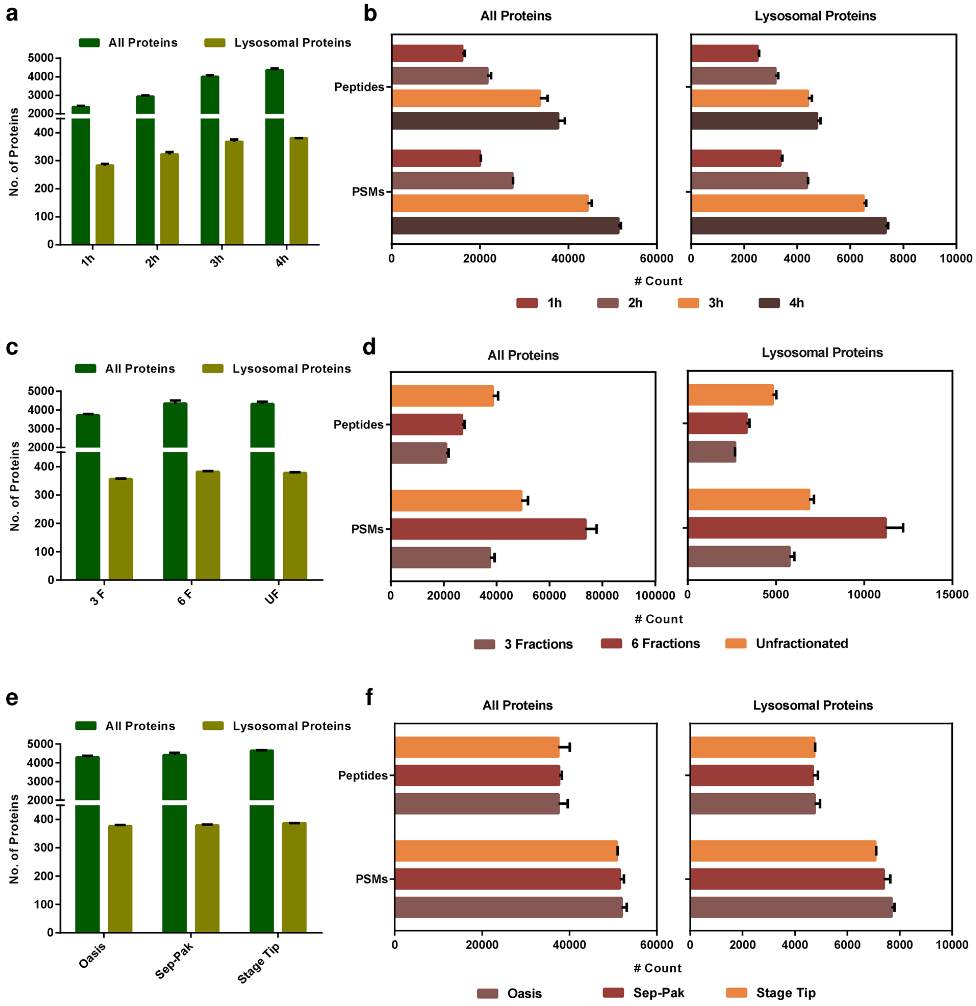


Fig. 2 Evaluation of the impact of gradient length, sample fraction, and desalting on protein identification. **(a)** Analysis of unfractionated tryptic digests of lysosome enriched fractions by LC-MS/MS using different gradient lengths (1 h, 2 h, 3 h, and 4 h). Shown are the average numbers of identified total and lysosomal protein groups. **(b)** Average numbers of PSMs and total peptides identified with different gradient lengths for total and lysosomal protein groups. **(c)** Analysis of the impact of peptide fractionation on the number of identified total and lysosomal protein groups. **(d)** Average numbers of PSMs and total peptides identified for the different fractionation methods. **(e)** Evaluation of the impact of different desalting approaches on the numbers of identified proteins groups using Stage Tip- or cartridge-based formats. **(f)** Average numbers of PSMs and total peptides for the different desalting methods. Shown are average values of 3 independent replicates + standard deviation. 3 F: 3 fractions; 6 F: 6 fractions; UF: unfractionated; PSMs: peptide spectral matches.

in figshare deposit[39]: Table 1_Proteins_Gradient Length). We further focused on a subset of lysosomal and lysosome-associated proteins (figshare deposit[39]: Table 9_Lysosomal Protein List) in the dataset. The effects observed for this group of proteins were less pronounced compared to the whole protein population with a maximal increase of 34% (Fig. 2a). When we also took the reproducibility of identification into account, however, the difference between the 1 h and 4 h gradient increased to 47% (only proteins identified in all 3 replicates). Furthermore, the number of peptide spectral matches and unique peptides assigned to lysosomal proteins raised by 2.2 fold and 1.9 fold, respectively (Fig. 2b). Also for the whole dataset, we observed similar trends but with slightly higher fold-change values (Fig. 2a,b). Compared to the total population of proteins identified in the dataset, the reproducibility of identification for lysosomal proteins was higher for all gradient lengths tested, reaching to a reproducibility of 94% of proteins detected in all three replicates for the analysis with 4 h gradients.

Impact of peptide fractionation and desalting on protein and peptide identification. With ~4,350 protein groups on average for the 4 h measurements, the comparison of LC gradients revealed a higher complexity of the lysosome-enriched samples than anticipated. We therefore evaluated, if further fractionation could improve identification rates as it allows for higher amounts of starting material and the individual fractions are of lower complexity. We employed SAX-tip based fractionation[35] generating 3 or 6 fractions in three independent replicates and analyzed them with linear gradients of 60 min, as well as an unfractionated sample with a 4 h gradient (figshare deposit[39]: Table 2_Proteins_Fractionation Methods). With regard to the total number of identified protein groups, the sample divided into 6 fractions resulted in virtually similar numbers compared to the unfractionated sample (4,349 and 4,323 protein groups on average, respectively) while the sample divided into 3 fractions yielded on average only 3,719 protein groups (Fig. 2c). When considering just the proteins identified in all three replicates, the unfractionated sample outperformed both fractionation methods. For our subset of bona fide lysosomal proteins, we observed a similar trend with the unfractionated sample yielding the highest number of identified lysosomal proteins and the most reproducible results (Fig. 2c). These results indicate that the 4 h gradient is sufficient for the complexity of the lysosomal fractions and no under-sampling occurs. With respect to the number of identified peptide spectral matches (PSMs) for total and lysosomal proteins, the sample divided into 6 fractions yielded the best results, and for numbers of unique peptides identified, the unfractionated sample performed best (Fig. 2d).

In order to evaluate if desalting influences the identification of peptides and proteins, we performed solid phase extraction with three different resins and compared the results. This included a tip-based format using C_{18} Stage Tips[37] as well as two solid phase extraction cartridges containing different stationary phases: Oasis HLB (Hydrophilic-Lipophilic Balance) cartridges, and Sep-Pak C_{18} cartridges. Using the urea digested PN sample, we desalted 10 µg of peptides with Stage Tips, and with both cartridge types 40 µg of peptides in triplicates followed by analysis of 1 µg each with 4 h gradients (figshare deposit[39]: Table 3_Proteins_Desalting Methods). Desalting with Stage Tips resulted in the highest number of lysosomal and total protein groups identified in all 3 replicates, followed by Sep-Pak and Oasis, which both delivered similar results (Fig. 2e). For both the whole dataset and the lysosomal proteins, however, the Oasis cartridges slightly outperformed both other approaches concerning the average number of PSMs and unique peptides (Fig. 2f). Concerning reproducibility, Stage Tip-based desalting outperformed both other methods.

Investigation of sample concentration and digestion procedures. Eluate fractions obtained from SPIONs enrichment are often highly diluted and it is necessary to further concentrate the contained lysosomes (or lysosomal proteins). Dependent on the subsequent experiments, it may be required to preserve the organelles' integrity (for example for enzymatic assays) excluding the application of denaturing protein precipitation (PN) approaches. Furthermore, precipitation could result in protein aggregates which may not be fully re-solubilized during sample preparation for proteomic analysis. Therefore, the pelleting of intact lysosomes by centrifugation (CN) is an attractive alternative. This approach should not result in any solubilization issues and may, as a positive side effect, lead to the depletion of unspecifically enriched soluble proteins, which will not be pelleted. However, CN may not succeed in the recovery of all lysosomes (e.g. such damaged during isolation) and proteins interacting weakly with the lysosomal surface may be lost. Furthermore, due to the lack of a denaturation step, lysosomal proteases may retain residual activity during proteolytic digestion possibly influencing the results obtained from these samples.

To compare these two individual concentration approaches, we pelleted intact lysosomes by centrifugation (CN) or chemically precipitated proteins (PN) contained in the lysosome enriched fractions. We combined both approaches with four commonly used methods for proteolytic digestion including in-gel digestion, filter aided sample preparation (FASP), and in-solution digestion using either RapiGest or urea, resulting in 8 different combinations in total (Fig. 1a). For each combination of sample concentration and digestion, we prepared three independent replicates. 10 µg of peptides were desalted by Stage Tips, and 1 µg each was analyzed with a 4 h gradient (figshare deposit[39]: Table 4_Proteins_Digestion Methods).

For the in-solution digestion with urea and RapiGest, we observed virtually no differences between both the digestion and the concentration strategies concerning the number of identified lysosomal and total protein groups (Fig. 3a,b). For sample preparation by FASP, the CN sample resulted in a markedly reduced number of total as well as lysosomal proteins, with high variability in total protein numbers between the individual replicates. In-gel digested samples yielded slightly better results for the CN samples for both lysosomal and total proteins. To further investigate differences for the individual approaches, we performed label free quantification using a combined database search (figshare deposit[39]: Table 5_LFQ_Digestion Methods). We filtered for protein groups identified with all eight workflows and performed binary comparisons for the individual digestion approaches within the same sample concentration setup (PN or CN). We then determined for each combination the number of proteins which were overrepresented in a given sample with a p-value < 0.05 and fold-change of ≥ 2 (Fig. 3c, and in figshare deposit[39]: Table 5_LFQ_Digestion Methods). For each individual approach, a specific subset of proteins was overrepresented suggesting that the choice of sample preparation should be adapted if specific proteins are of special interest. Furthermore, these data suggest that results from published studies employing different digestion strategies can be compared in a qualitative but not a quantitative way. To further investigate the regulated protein populations, we performed GO analyses for proteins which were significantly up- or downregulated, the results can be found in our figshare deposit[39]: Table 5_LFQ_Digestion Methods.

As protein precipitation may result in aggregates which could influence the efficiency of proteolytic digestion, we further investigated the percentage of missed cleavage sites (Fig. 3d). While in-gel digestion resulted in

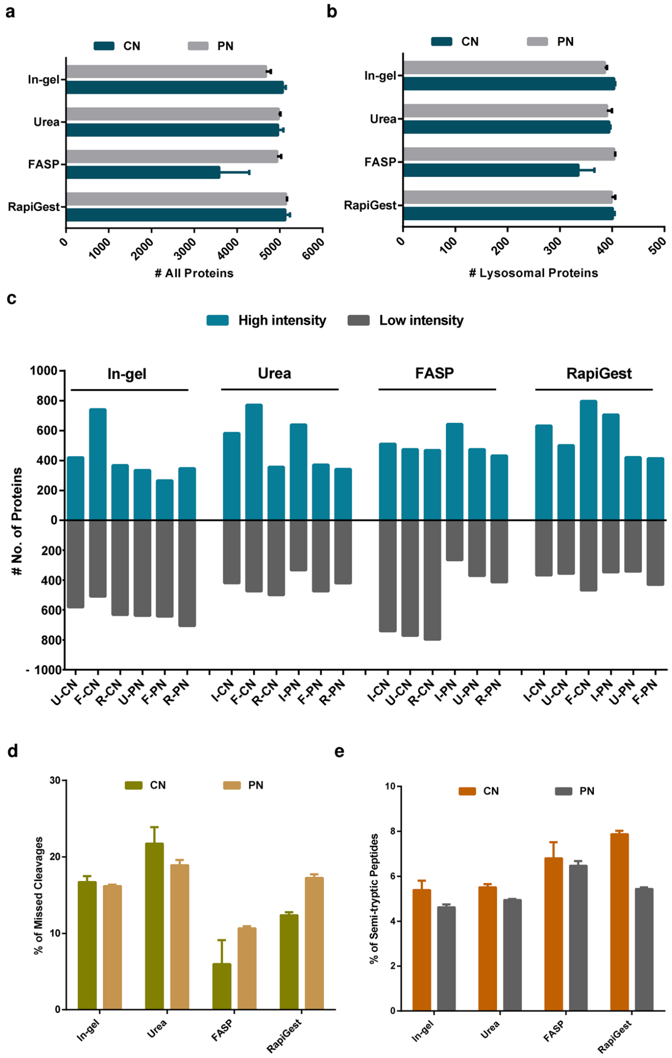


Fig. 3 Evaluation of concentration and digestion methods. (**a**) Average numbers of identified protein groups for the individual combinations of concentration and digestion methods. (**b**) Average numbers of identified lysosomal protein groups for the individual combinations of concentration and digestion methods. (**c**) Number of proteins significantly over-/underrepresented for binary comparisons of the individual datasets (fold-change up or down regulation of ≥2 and p-value < 0.05). (**d**) Average numbers of missed cleavage sites for the individual combinations of concentration and digestion methods. (**e**) Average percentage of semi-tryptic peptides for the individual combinations of concentration and digestion methods. Shown are average values of 3 independent replicates + standard deviation (**a**,**b**,**d**,**e**). PN: Precipitation; CN: centrifugation; U-CN: Urea CN; F-CN: FASP CN; R-CN: RapiGest CN; U-PN: Urea PN; F-PN: FASP PN; R-PN: RapiGest PN; I-CN: In-gel CN; I-PN: In-gel PN.

similar rates for both the PN and the CN sample, we observed for FASP and RapiGest higher numbers of missed cleavages for PN. For the urea digested sample, the CN sample resulted in a slightly higher percentage of missed cleavage sites than PN and in general in a less efficient cleavage than for the other approaches.

Lysosomes contain >20 different proteases of which ~50% belong to the family of the cathepsins, catalyzing the degradation of a wide variety of proteins[45]. Dependent on their concentration, both urea and RapiGest retain the activity of the proteases Lys-C and trypsin which are used for mass spectrometry sample processing. It was

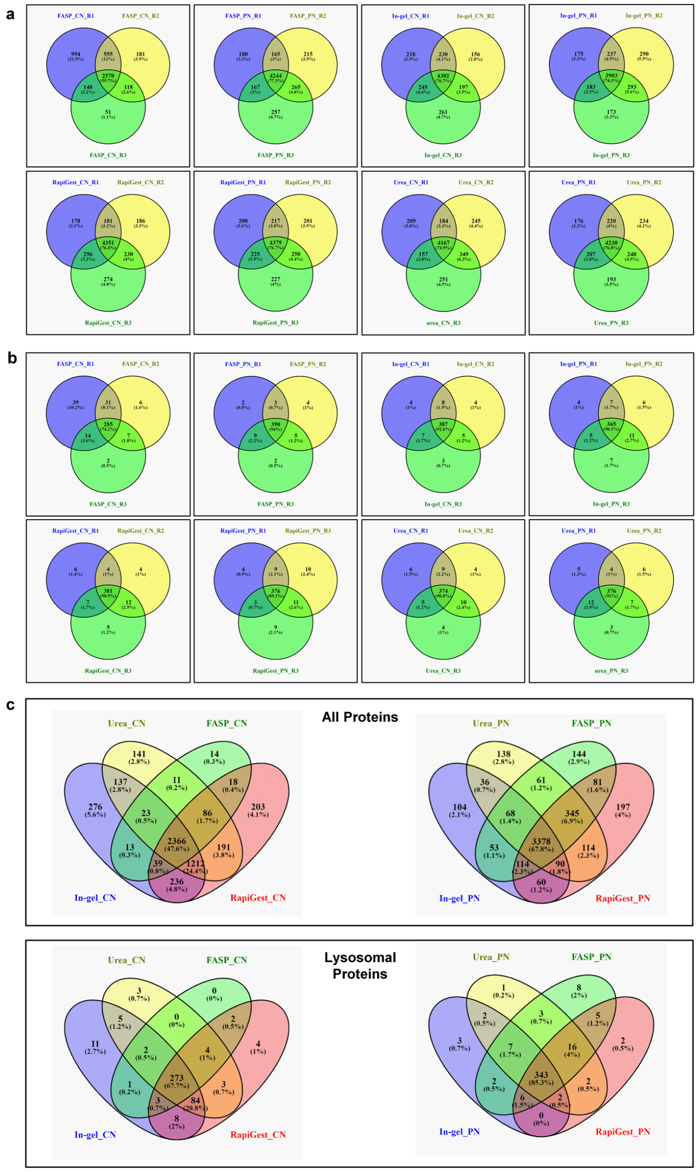


Fig. 4 Reproducibility of protein identification for the individual sample preparation methods. **(a)** Common and unique total proteins among individual replicates for each combination of sample concentration approach and digestion method. **(b)** Common and unique lysosomal proteins among individual replicates for each combination of sample concentration approach and digestion method. **(c)** Overlap of identified total or lysosomal proteins for the combination of a specific type of sample concentration and digestion method. PN: concentration by precipitation; CN: concentration by centrifugation.

shown before, that also in pH values higher than those usually present in the lysosomal lumen, cathepsins can be enzymatically active[46]. We, therefore, investigated if the CN samples still contained active lysosomal proteases, since no protein precipitation step was part of this protocol. As active cathepsins should result in peptides cleaved

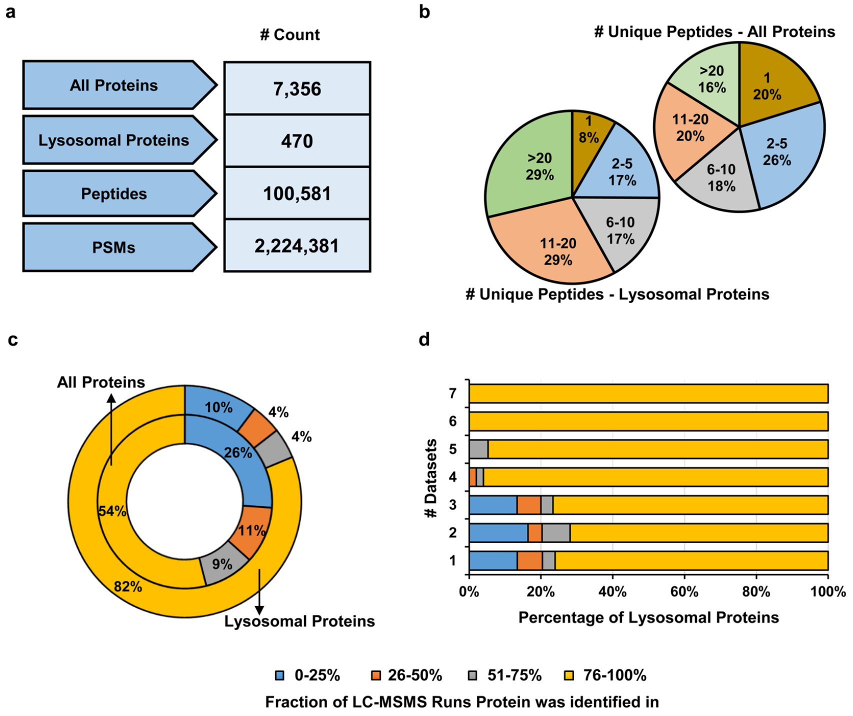


Fig. 5 Draft map of the MEF lysosomal proteome. **(a)** Number of identified proteins, lysosomal proteins, peptides, and PSMs from the combined database search of 39 × 4 h gradient measurements. **(b)** Distribution of total and lysosomal proteins identified with a specific number of unique peptides. **(c)** Reproducibility of proteins and lysosomal proteins identified across the individual analyses. Shown is the percentage of proteins identified within a certain fraction of the 39 LC-MS/MS runs analyzed. **(d)** Correlation of the identification rate of lysosomal proteins and their occurrence in published datasets.

at other sites than arginine or lysine (for digests with trypsin and Lys-C), we performed database searches for semi-tryptic peptides (figshare deposit[39]: Table 6_Semi-tryptic peptides_Digestion Methods). For all digestion methods, we observed a slight increase in semi-tryptic peptides for those concentrated by CN (Fig. 3e). This was especially pronounced for samples digested in RapiGest for which ~1,600 additional semi-tryptic peptides were identified in the CN relative to the PN sample (increase of 67%). The markedly higher difference for digests carried out in RapiGest suggests that certain lysosomal proteases may still be enzymatically active in these samples.

Concerning reproducibility among individual replicates, all approaches (with the exception of the FASP CN samples) performed similar resulting in an approximate overlap of 75% for total proteins and ~90% for lysosomal proteins (Fig. 4a,b). For the individual digestion strategies within each concentration method, the PN samples showed with 68%/85% a better performance than the CN samples (48%/68%) for the whole population of proteins and such located at the lysosome, respectively (Fig. 4c).

Draft map of the mouse embryonic fibroblast (MEF) lysosomal proteome. Utilizing measurements from all conditions, for which we analyzed the samples with 4 h gradients (39 LC-MS/MS runs in total), we performed a combined database search in order to assemble a draft map of the MEF lysosomal proteome (figshare deposit[39]: Table 7_Combined_Database Search_4h Gradient Length). In total, we identified 7,356 proteins from 100,581 peptides and 2,224,381 high confidence PSMs (Fig. 5a). For the unique peptides identified (83,619 in total), we observed a trend towards higher numbers for known lysosomal proteins (75% identified with >5 unique peptides) compared to the whole dataset (54% identified with >5 unique peptides (Fig. 5b). On the protein level, we were able to identify 470 out of 740 proteins of known lysosomal origin in total. Of these proteins, 82% were detected in >75% of LC-MS/MS runs while for the whole dataset only 54% were identified at the same rate (Fig. 5c). We further matched the detection rate of lysosomal proteins with their occurrence in published datasets (Fig. 5d, figshare deposit[39]: Table 9_Lysosomal Protein List). We observed a correlation of the number of datasets which list a given proteins as lysosomal and the identification rate in our data: the likelihood to be reproducibly detected in our analyses increases with the number of published datasets including the protein.

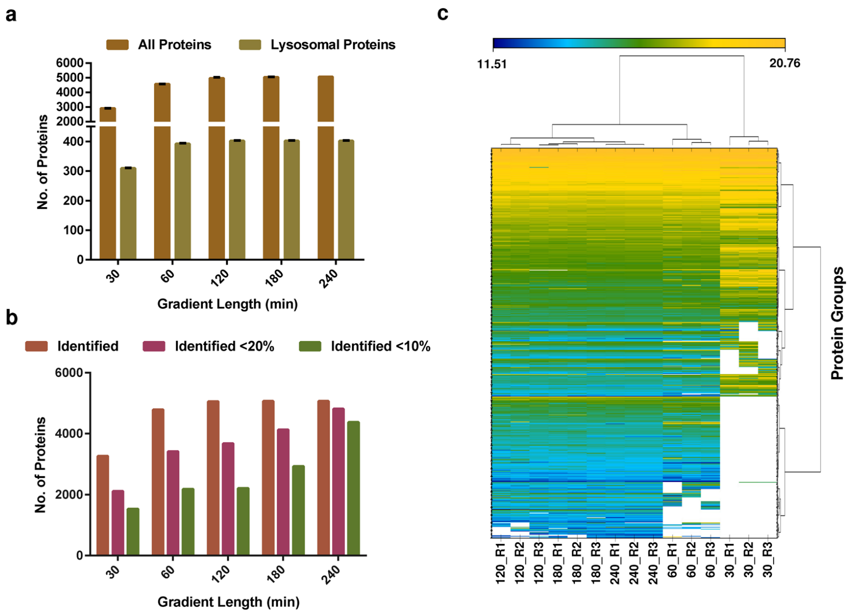


Fig. 6 Analysis of the MEF lysosomal proteome by data independent acquisition. (**a**) Average numbers of total and lysosomal proteins identified for different gradient lengths. (**b**) Numbers of total and lysosomal proteins determined with values below a certain coefficient of variation (CV) among the individual replicates. (**c**) Unsupervised clustering of protein intensities determined for the different individual gradient lengths. Each column represents the normalized protein intensities for an individual replicate. In (**a**), average values of 3 independent replicates + standard deviation are shown. CV: coefficient of variation.

Taken together, this dataset presents to our knowledge the so-far most extensive analysis of lysosomes from a single cell type identifying a highly reproducible core proteome for lysosome enriched fractions from MEFs.

Analysis of the MEF lysosomal proteome by data independent acquisition (DIA). In order to facilitate an efficient quantification of lysosome enriched fractions from MEFs in future studies by DIA, we generated a spectral library based on our dataset obtained from the combined searches by Proteome Discoverer. We imported the PD result file into the PulsarX algorithm integrated into the Spectronaut software and generated a spectral library covering 7421 proteins, 98,371 peptides, and 118,269 precursors[38]. We then analyzed 1 µg of urea in-solution digested lysosome enriched samples in 3 replicates with five different gradient lengths by DIA (30, 60, 120, 180 and 240 min, (figshare deposit[39]: Table 8_Proteins_Gradient length_DIA). While, not surprisingly, the 240 min gradient resulted in the highest number of protein identifications and the 30 min gradient in the lowest (Fig. 6a), the differences were much less pronounced as for the DDA measurements (Fig. 2a). Furthermore, when comparing gradient lengths ≥120 min, we virtually did not observe any differences in numbers of identified proteins. When we assessed the reproducibility of signal intensities for label free quantification, however, we found a continuous increase of proteins with low coefficients of variation (<10% CV) with the 240 min gradients delivering superior results (Fig. 6b). Finally, we visualized differences in protein identification and abundance for the individual DIA analyses clustering the data in a row and column-wise manner for all data points with high confidence (Fig. 6c). We observed a highly reproducible clustering of intensities among the independent biological replicates and highly similar profiles among the gradients with ≥120 min.

Usage Notes

For all analyses, the identified protein groups including the most important information can be found in the respective table in the figshare collection[39]. Furthermore, for the analysis of missed cleavage sites, the individual peptide sequences are provided in our figshare deposit[39] Table 6_Semi-tryptic Peptides_Digestion methods. If more details, like the exact peptide sequences assigned to a given protein in a specific analysis, are desired, the Proteome Discoverer (PD) result files can be accessed through the public repository[38]. For accession of these files, the Proteome Discoverer Software can be obtained from www.thermofisher.com. Furthermore, the PD study files are included which can be used to re-analyze the *.raw files with different parameters. If an analysis by a different algorithm is desired, the provided *.raw files can be analyzed with any other mainstream proteomic data analysis software. For manual analysis of the *.raw files, tools like Xcalibur or FreeStyle from Thermo Fisher Scientific can be used or freeware such as MSFileReader or the ProteoWizard toolkit.

A unique value of the dataset presented in this study is related to the planning of experiments for the analysis of lysosomes from MEFs by mass spectrometry. If the analysis of a given lysosomal or lysosome-associated protein in these cells is desired, it is possible to determine from the individual datasets which lysosome concentration method, proteolytic digestion strategy, desalting approach, fractionation method, and gradient length should be used to obtain an optimal result. For the development of targeted mass spectrometry assays for mouse samples, high confidence peptides and their fragment ions can be extracted from our combined dataset. This allows assessing how reproducible the identification of these peptides is across different experimental conditions, if the signal intensity is sufficient, and which fragment ions can be utilized for SRM/MRM assay design. Based on the information in figshare deposit[39]: Table 5_LFQ_Digestion Methods, it can furthermore be assessed if a certain digestion method results in higher intensities for the protein(s) of interest. Finally, the spectral library which was generated for the DIA analyses can be readily used for the analysis of lysosome-enriched MEF samples with the algorithm Spectronaut (www.biognosys.com). If analysis with other algorithms is desired, the data can be re-exported from the available PD study in the desired format. The definition of a high confidence lysosomal proteome by combination of 39 individual LC-MS/MS analyses presents, to our knowledge, the largest analysis of isolated MEF lysosomes so-far. This resource is valuable for the identification of proteins which are of potential lysosomal origin in MEF cells covering such which have been proposed to be located at the lysosome and such which have not been assigned to the lysosome yet.

Taken together, this dataset presents a toolbox for the conceptualization of experiments for the analysis of lysosome enriched samples from MEFs, and a valuable resource for the targeted analysis of lysosomal proteins in mouse samples.

Code availability

DDA data were analyzed with Proteome Discoverer 2.2 and DIA data were analyzed with Spectronaut (Version: 13.2.19). Data comparison was performed with Microsoft Excel 2016, GraphPad Prism 6.07, and Venny (https://bioinfogp.cnb.csic.es › tools › venny).

Received: 19 November 2019; Accepted: 23 January 2020;

References

1. Ballabio, A. The awesome lysosome. *EMBO Mol. Med.* **8**, 73–76 (2016).
2. Ballabio, A. & Gieselmann, V. Lysosomal disorders: from storage to cellular damage. *Biochim. Biophys. Acta Mol. Cell Res.* **1793**, 684–696 (2009).
3. Lim, C.-Y. & Zoncu, R. The lysosome as a command-and-control center for cellular metabolism. *J. Cell Biol.* **214**, 653–664 (2016).
4. Fraldi, A., Klein, A. D., Medina, D. L. & Settembre, C. Brain disorders due to lysosomal dysfunction. *Annu. Rev. Neurosci.* **39**, 277–295 (2016).
5. Davidson, S. M. & Vander Heiden, M. G. Critical functions of the lysosome in cancer biology. *Annu. Rev. Pharmacol. Toxicol.* **57**, 481–507 (2017).
6. Geladaki, A. *et al*. Combining LOPIT with differential ultracentrifugation for high-resolution spatial proteomics. *Nat. Comm.* **10**, 15 (2019).
7. Thelen, M., Winter, D., Braulke, T. & Gieselmann, V. in *Methods in Molecular Biology* Vol. 1594 (eds. Öllinger, K. & Appelqvist, H.) 1–18 (Humana Press, 2017).
8. Itzhak, D. N., Tyanova, S., Cox, J. & Borner, G. H. H. Global, quantitative and dynamic mapping of protein subcellular localization. *Elife* **5**, 36 (2016).
9. Wattiaux, R., Wattiauxdeconinck, S., Ronveauxdupal, M. F. & Dubois, F. Isolation of rat-liver lysosomes by isopycnic centrifugation in a metrizamide gradient. *J. Cell Biol.* **78**, 349–368 (1978).
10. Chapel, A. *et al*. An Extended Proteome Map of the Lysosomal Membrane Reveals Novel Potential Transporters. *Mol. Cell. Proteom.* **12**, 1572–1588 (2013).
11. Schroeder, B. *et al*. Integral and associated lysosomal membrane proteins. *Traffic* **8**, 1676–1686 (2007).
12. Zhang, H., Fan, X., Bagshaw, R., Mahuran, D. J. & Callahan, J. W. in *Methods in Molecular Biology* Vol. 432 (eds. Pflieger, D. & Rossier, J.) 229–241 (Humana Press, 2008).
13. Leighton, F. *et al*. Large-scale separation of peroxisomes mitochondria and lysosomes from livers of rats injected with Triton WR-1339- improved isolation procedures automated analysis biochemical and morphological properties of fractions. *J. Cell Biol.* **37**, 482–513 (1968).
14. Diettrich, O., Mills, K., Johnson, A. W., Hasilik, A. & Winchester, B. G. Application of magnetic chromatography to the isolation of lysosomes from fibroblasts of patients with lysosomal storage disorders. *FEBS Lett.* **441**, 369–372 (1998).
15. Walker, M. W. & Lloyd–Evans, E. in *Methods in Cell Biology*. Vol. 126 (eds. Platt, F. & Platt, N.) 21–43 (Academic Press, 2015).
16. Tharkeshwar, A. K. *et al*. A novel approach to analyze lysosomal dysfunctions through subcellular proteomics and lipidomics: the case of NPC1 deficiency. *Sci. Rep.* **7**, 41408 (2017).
17. Zoncu, R. *et al*. mTORC1 Senses Lysosomal Amino Acids Through an Inside-Out Mechanism That Requires the Vacuolar H+-ATPase. *Science* **334**, 678–683 (2011).
18. Abu-Remaileh, M. *et al*. Lysosomal metabolomics reveals V-ATPase- and mTOR-dependent regulation of amino acid efflux from lysosomes. *Science* **358**, 807–813 (2017).
19. Wyant, G. A. *et al*. NUFIP1 is a ribosome receptor for starvation-induced ribophagy. *Science* **360**, 751–758 (2018).
20. Vonfigura, K. & Hasilik, A. Lysosomal-enzymes and their receptors. *Annu. Rev. Biochem.* **55**, 167–193 (1986).
21. Makrypidi, G. *et al*. Mannose 6 Dephosphorylation of Lysosomal Proteins Mediated by Acid Phosphatases Acp2 and Acp5. *Mol. Cell. Biol.* **32**, 774–782 (2012).
22. Caval, T. *et al*. Targeted Analysis of Lysosomal Directed Proteins and Their Sites of Mannose-6-phosphate Modification. *Mol. Cell. Proteom.* **18**, 16–27 (2019).
23. Kollmann, K. *et al*. Identification of novel lysosomal matrix proteins by proteome analysis. *Proteomics* **5**, 3966–3978 (2005).
24. Sleat, D. E. *et al*. The human brain mannose 6-phosphate glycoproteome: A complex mixture composed of multiple isoforms of many soluble lysosomal proteins. *Proteomics* **5**, 1520–1532 (2005).
25. Sleat, D. E. *et al*. Identification and validation of mannose 6-phosphate glycoproteins in human plasma reveal a wide range of lysosomal and non-lysosomal proteins. *Mol. Cell. Proteom.* **5**, 1942–1956 (2006).

26. Sleat, D. E., Della Valle, M. C., Zheng, H., Moore, D. F. & Lobel, P. The mannose 6-phosphate glycoprotein proteome. *J. Proteome Res.* **7**, 3010–3021 (2008).
27. Sleat, D. E. *et al*. Proteomic analysis of mouse models of Niemann-Pick C disease reveals alterations in the steady-state levels of lysosomal proteins within the brain. *Proteomics* **12**, 3499–3509 (2012).
28. Sleat, D. E. *et al*. Mass Spectrometry-based Protein Profiling to Determine the Cause of Lysosomal Storage Diseases of Unknown Etiology. *Mol. Cell. Proteom.* **8**, 1708–1718 (2009).
29. Qian, M. Q., Sleat, D. E., Zheng, H. Y., Moore, D. & Lobel, P. Proteomics analysis of serum from mutant mice reveals lysosomal proteins selectively transported by each of the two mannose 6-phosphate receptors. *Mol. Cell. Proteom.* **7**, 58–70 (2008).
30. Sleat, D. E. *et al*. Extending the Mannose 6-Phosphate Glycoproteome by High Resolution/Accuracy Mass Spectrometry Analysis of Control and Acid Phosphatase 5-Deficient Mice. *Mol. Cell. Proteom.* **12**, 1806–1817 (2013).
31. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* **193**, 265–275 (1951).
32. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
33. Winter, D. & Steen, H. Optimization of cell lysis and protein digestion protocols for the analysis of HeLa S3 cells by LC-MS/MS. *Proteomics* **11**, 4726–4730 (2011).
34. Kollipara, L. & Zahedi, R. P. Protein carbamylation: *in vivo* modification or *in vitro* artefact? *Proteomics* **13**, 941–944 (2013).
35. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359 (2009).
36. Muller, T. & Winter, D. Systematic Evaluation of Protein Reduction and Alkylation Reveals Massive Unspecific Side Effects by Iodine-containing Reagents. *Mol. Cell. Proteom.* **16**, 1173–1187 (2017).
37. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
38. Comprehensive proteomic analysis of mouse embryonic fibroblast lysosomes by mass spectrometry. *PRIDE*. https://identifiers.org/pride.project:PXD014576 (2019).
39. Ponnaiyan, S., Akter, F., Singh, J. & Winter, D. Comprehensive draft of the mouse embryonic fibroblast lysosomal proteome by mass spectrometry based proteomics. *figshare*. https://doi.org/10.6084/m9.figshare.c.4743716.v1 (2019).
40. Foster, L. J. *et al*. A mammalian organelle map by protein correlation profiling. *Cell* **125**, 187–199 (2006).
41. Lübke, T., Lobel, P. & Sleat, D. E. Proteomics of the lysosome. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1793**, 625–635 (2009).
42. Mulvey, C. M. *et al*. Using hyperLOPIT to perform high-resolution mapping of the spatial proteome. *Nature protocols* **12**, 1110 (2017).
43. Bruderer, R., Bernhardt, O. M., Gandhi, T. & Reiter, L. High-precision iRT prediction in the targeted analysis of data-independent acquisition and its impact on identification and quantitation. *Proteomics* **16**, 2246–2256 (2016).
44. Rosenberger, G. *et al*. Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses. *Nat. Methods* **14**, 921 (2017).
45. Chen, S., Dong, H., Yang, S. & Guo, H. Cathepsins in digestive cancers. *Oncotarget* **8**, 41690–41700 (2017).
46. Biniossek, M. L., Nagler, D. K., Becker-Pauly, C. & Schilling, O. Proteomic Identification of Protease Cleavage Sites Characterizes Prime and Non-prime Specificity of Cysteine Cathepsins B, L, and S. *J. Proteome Res.* **10**, 5363–5373 (2011).

Acknowledgements

We would like to thank the National overseas scholarship by Indian Ministry of Social Justice and Empowerment, the German Academic Exchange Service (DAAD), the Studienstiftung des Deutschen Volkes, and the DFG Research Unit FOR2625 for financial support.

Author contributions

S.P. performed cell culture experiments, mass spectrometry sample preparation and analysis by DDA, data analysis, and edited the manuscript. F.A. performed cell culture experiments, mass spectrometry sample preparation and analysis by DDA, data analysis, and edited the manuscript. J.S. performed DIA mass spectrometry analyses and edited the manuscript. D.W. conceptualized the study, analyzed the data, and wrote the manuscript. All authors reviewed and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to D.W.

Reprints and permissions information is available at www.nature.com/reprints.

8.2 Co-author Publication II:

3 Part three of chapter 1 has been published: “[Mosen, P., Sanner, A., **Singh, J.**, Winter, D. (2021). Targeted Quantification of the Lysosomal Proteome in Complex Samples. Proteomes, 9(1), 4].

- For the online version of the paper, visit: https://doi.org/10.3390/proteomes9010004
- The supporting information including supplementary tables are available at: https://www.mdpi.com/2227-7382/9/1/4/s1
- **Remark:** The third publication within the scope of chapter 1 **SHOULD NOT BE CONSIDERED** for this cumulative doctoral dissertation, as the majority of the work has been done by the first two co-authors.

8.2 Co-author Publication II: Main Article

Reproduced with permission from: Mosen, P., Sanner, A., Singh, J., & Winter, D. (2021). Targeted Quantification of the Lysosomal Proteome in Complex Samples. Proteomes, 9(1), 4. *This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license: http://creativecommons.org/licenses/by/4.0/.*

proteomes MDPI

Article

Targeted Quantification of the Lysosomal Proteome in Complex Samples

Peter Mosen †, Anne Sanner †, Jasjot Singh and Dominic Winter *

Institute for Biochemistry and Molecular Biology, Medical Faculty, University of Bonn, 53115 Bonn, Germany; pmos@uni-bonn.de (P.M.); anne.sanner@uni-bonn.de (A.S.); jsin@uni-bonn.de (J.S.)
* Correspondence: dominic.winter@uni-bonn.de; Tel.: +49-228-73-7081
† These authors contributed equally.

Abstract: In eukaryotic cells, lysosomes play a crucial role in the breakdown of a variety of components ranging from small molecules to complex structures, ascertaining the continuous turnover of cellular building blocks. Furthermore, they act as a regulatory hub for metabolism, being crucially involved in the regulation of major signaling pathways. Currently, ~450 lysosomal proteins can be reproducibly identified in a single cell line by mass spectrometry, most of which are low-abundant, restricting their unbiased proteomic analysis to lysosome-enriched fractions. In the current study, we applied two strategies for the targeted investigation of the lysosomal proteome in complex samples: data-independent acquisition (DIA) and parallel reaction monitoring (PRM). Using a lysosome-enriched fraction, mouse embryonic fibroblast whole cell lysate, and mouse liver whole tissue lysate, we investigated the capabilities of DIA and PRM to investigate the lysosomal proteome. While both approaches identified and quantified lysosomal proteins in all sample types, and their data largely correlated, DIA identified on average more proteins, especially for lower complex samples and longer chromatographic gradients. For the highly complex tissue sample and shorter gradients, however, PRM delivered a better performance regarding both identification and quantification of lysosomal proteins. All data are available via ProteomeXchange with identifier PXDD023278.

Keywords: targeted proteomics; lysosomes; parallel reaction monitoring; data-independent acquisition; label-free quantification

check for updates

Citation: Mosen, P.; Sanner, A.; Singh, J.; Winter, D. Targeted Quantification of the Lysosomal Proteome in Complex Samples. *Proteomes* **2021**, *9*, 4. https://doi.org/10.3390/proteomes9010004

Academic Editor: Jörg Reinders
Received: 23 December 2020
Accepted: 21 January 2021
Published: 26 January 2021

1. Introduction

Lysosomes are membrane-bound organelles, which are well-known as the main degradative compartment of eukaryotic cells [1]. They fulfil a crucial function for the breakdown of a variety of cellular components and the recycling of their building blocks. This is achieved by ~60 hydrolases and ~40 transporters residing in the lysosomal lumen and membrane [2]. The proper function of these hydrolases is crucial for cellular homeostasis, as exemplified by the detrimental consequences of lysosomal enzyme malfunction. Mutations resulting in their altered activity, stability, or subcellular distribution can result in the accumulation of their respective substrates within lysosomes, interfering with the correct function of the organelle. Impaired lysosomal function is the primary cause of a group of ~70 inherited rare genetic diseases, so-called lysosomal storage disorders (LSDs), which frequently result in neurodegeneration, metabolic dysfunction, impaired development, and premature death [3]. To date, therapies exist only for a handful of LSDs and those available are almost exclusively symptomatic [3–5].

While the connection between lysosomal dysfunction and LSDs has been known for decades, altered lysosomal or lysosome-associated proteins have recently been shown in an increasing number of studies to be involved in more common conditions, increasing the public interest in this organelle. This includes, but is not limited to, cancer [6], neurodegenerative disorders [7], and cardiovascular diseases [8]. As part of this development, the view on lysosomes as unregulated cellular waste bags, which persisted for decades, is currently transitioning towards highly mobile organelles that act as a major regulatory hub of cellular metabolism. In recent years, lysosomes have been shown to vary in their properties, to be actively transported, to interact with other organelles, and to respond to various cellular and environmental stimuli with the help of an extensive network of proteins [2,9,10]. This involves several key players regulating cellular growth and energy metabolism, such as the mechanistic target of rapamycin complex 1 (mTORC1) or the AMP-dependent kinase (AMPK), which are activated at the lysosomal surface [11].

These emerging roles of lysosomes have led to an increasing interest in the analysis of lysosomal proteins. For the unbiased characterization of large numbers of proteins, mass spectrometry (MS)-based proteomics is currently the method of choice, as it allows for the identification, quantification, and characterization of thousands of proteins from a given sample [12]. To date, ~740 proteins have been assigned in one way or the other to lysosomes, ~300 of which are either located in the lysosomal lumen, at the lysosomal surface, or directly interact with it [13].

Lysosomal proteins are typically of low abundance and therefore frequently not covered in DDA whole proteome shotgun analyses. The most common way to increase the coverage of lysosomal proteins is lysosome enrichment, resulting in a reduced sample complexity and therefore facilitating their analysis. Several lysosome enrichment methods are currently available, the most common of which are based either on density gradient centrifugation, superparamagnetic iron oxide nanoparticles (SPIONs) in combination with magnetic columns, or the immunoaffinity enrichment of tagged lysosomal proteins [14]. While all of these approaches allow for a certain degree of enrichment, they come with restrictions that limit the samples they can be applied to. While density gradient centrifugation can be performed for virtually any starting material, recovery is low and high amounts of contaminating organelles, mostly mitochondria, are included in the lysosome-containing fractions [14]. The use of SPIONs, which are taken up by unspecific fluid phase endocytosis and delivered to the lysosomal compartment through the endocytic pathway [15], is limited to cells grown in culture which actively perform fluid phase endocytosis. Furthermore, with this approach, only those lysosomes can be isolated that receive cargo from late endosomes, which may be affected when studying models of LSDs with impaired endosome-lysosome fusion. For the immunoaffinity enrichment of lysosomes through tagged membrane proteins [16,17], a fusion protein has to be stably expressed in cells or animals, requiring the generation of the respective organism. Furthermore, only lysosomes expressing the protein of choice are covered, which may result in a selection bias, and the overexpression of the tagged protein may influence lysosomal properties. For all approaches, millions of cells or milligram amounts of tissue are needed as starting material, excluding low-abundant samples from these analyses. As many LSDs affect distinct populations of cells, and the majority of LSDs can only be simulated in animal models, these limitations stall the proteomic investigation of LSDs, as it is frequently not possible to obtain lysosome-enriched fractions in sufficient quantities.

The need for enrichment of lysosomal proteins arises from the limitations of untargeted data-dependent acquisition (DDA)-based acquisition strategies, as highly abundant peptides prevent the fragmentation, and therefore identification, of those originating from low-abundant lysosomal proteins. Therefore, a promising alternative for the characterization of lysosomal proteins from small amounts of complex samples are targeted proteomics strategies. Currently, two major approaches for targeted proteomics are applied. On the one hand, previously defined peptides are fragmented in single, multiple, or parallel reaction monitoring (SRM, MRM, PRM) experiments, and abundance is determined based on the intensity of their fragment ions [18]. On the other hand, unbiased fragmentation of pre-determined m/z windows is performed in data-independent acquisition (DIA) approaches, and the abundance of the respective peptides is determined from unique fragment ions identified in mixed MS/MS spectra [19]. In comparison to DDA-based label-free quantification strategies, PRM and DIA approaches offer increased sensitivity and reproducibility for low-abundant peptides in complex samples [20–22], making them ideal candidates for the analysis of lysosomal proteins from cell or tissue samples without prior enrichment.

So far, to our knowledge, targeted approaches have not been frequently used for the investigation of the lysosomal proteome. PRM was applied in a few studies for the investigation of selected lysosomal proteins (e.g., [23–25]), while only DIA approaches have been used for the analysis of the whole lysosomal proteome, so far solely investigating lysosome-enriched samples [14,16,26]. While it was reported that DIA is able to identify and quantify > 10,000 proteins within a single run [27], the reproducible quantification of lysosomal proteins suffers in highly complex samples and the achievable performance in whole cell lysates is significantly lower compared to lysosome-enriched fractions [14].

In the present study, we compared DIA and PRM for the analysis of the lysosomal proteome from samples of different complexities. We investigated lysosome-enriched fractions, as well as whole cell and liver lysate, and systematically compared the performance of DIA and PRM. While we could detect lysosomal proteins with both approaches in all sample types, and DIA identified higher numbers for most samples, PRM showed a better performance in liver lysate allowing for the detection of quantitative changes which were not identified by DIA.

2. Material and Methods

2.1. Cell Culture Experiments and Sample Lysis

All cell culture experiments were performed under a sterile hood and all solutions were pre-warmed to 37 °C. Mouse embryonic fibroblasts (MEFs) were cultured at 37 °C and 5% CO_2 in Dulbecco's modified eagle medium (DMEM), supplemented with 10% (*v*/*v*) fetal calf serum (FCS), 100 IU/mL penicillin, and 100 µg/mL streptomycin. For the generation of MEF whole cell lysate samples, 1.5×10^6 cells each were seeded on three 15 cm plates and cultivated for 72 h. The cells were washed once with 5 mL of ice-cold 1× phosphate-buffered saline (PBS), scraped in 600 µL of ice-cold PBS, and collected in a 1.5 mL microtube. Cells were pelleted by centrifugation at 1000× *g* and 4 °C for 4 min, the supernatant was discarded, and the cell pellet was re-suspended in 600 µL lysis buffer (4% SDS, 100 mM HEPES pH 7.5). Subsequently, the cell suspension was incubated at 95 °C for 10 min followed by sonication using a Ultrasonics Sonifier 250 (Branson, Danbury, CT, USA) at a duty cycle of 60% and an output of 6 for 90 s. Samples were centrifuged at 20,000× *g* and RT for 30 min and the clear supernatants were transferred to new microtubes.

Lysosome isolation was performed from MEF cells using SPIONs as described elsewhere [26]. In brief, cells were cultivated in DMEM with 2.5% FCS for 72 h (3×10^6 cells per 10 cm dish), 1 mL of magnetite solution (EndoMAG40, Liquids Research, North Wales, UK) was added to each plate, and the cells were incubated for 24 h (pulse period). Subsequently, the cells were washed twice with pre-warmed PBS, fresh DMEM (10% FCS) was added, and the cells were incubated for 24 h (chase period). Cells were washed with ice-cold PBS and harvested using a cell scraper in 2 mL lysosome isolation buffer (250 mM sucrose, 10 mM HEPES pH 7.4, 1 mM $CaCl_2$, 15 mM KCl, 1 mM $MgCl_2$, 1.5 mM MgAc, 1 mM dithiothreitol (DTT), 1x cOmplete EDTA-free protease inhibitor cocktail (Roche Diagnostics, Mannheim, Germany)). Plasma membranes were disrupted using a dounce homogenizer, and lysosomes were enriched using Miltenyi LS columns (Miltenyi Biotech, Auburn, CA) and eluted using a plunger.

2.2. Preparation of Mouse Liver Samples

Mice were handled in accordance with local regulations concerning the welfare of animals. Three months-old male C57BL/6 mice were sacrificed by cervical dislocation, the liver was extracted, and snap-frozen in liquid nitrogen. The frozen tissue was chopped into small pieces using a razor blade, and 1 mL of lysis buffer (4% SDS, 100 mM HEPES pH 7.5) was added. The sample was incubated for 10 min at 95 °C and sonicated using an Ultrasonics Sonifier at a duty cycle of 60% and an output of 6 for 90 s. Subsequently, the samples were again incubated at 95 °C for 10 min, centrifuged at 20,000× *g* and RT for 30 min, and the clear supernatants were transferred to new microtubes.

2.3. Sample Preparation for Mass Spectrometry

The protein concentration of all samples was determined using the DC Protein Assay (Bio-Rad Laboratories, CA, USA). For MEF whole cell lysate and liver samples, 100 μg of protein were used for each replicate while 20 μg were used for lysosome-enriched fractions. Sample volumes were adjusted to 200 μL using HPLC-grade water and proteins were precipitated by addition of 1 mL ice-cold chloroform/methanol (2:1 *v*/*v*), vigorous vortexing, and centrifugation at 20,000× *g*, 4 °C for 1 h. The liquid phases were discarded, the protein pellets washed once with 1 mL of ice-cold methanol, and centrifuged at 20,000 × *g*, 4 °C for 15 min, followed by the removal of methanol. Protein pellets were air-dried and solubilized in 1% RapiGest (Waters, Milford, MA, USA), 0.1 M NH_4HCO_3 pH 7.8 at 95 °C for 10 min. Subsequently, samples were diluted 1 to 5 with 0.1 M NH_4HCO_3 and trypsin (Promega, Mannheim, Germany) was added at an enzyme-to-substrate ratio of 1:500, followed by incubation at 37 °C, 800 rpm in a thermomixer for 45 min. Proteins were reduced using DTT (5 mM final concentration) at 56 °C for 30 min and alkylated with acrylamide (20 mM final concentration) for 30 min at RT, followed by quenching of the reaction through the addition of 5 mM DTT. Finally, trypsin was added at an enzyme-to-sample ratio of 1:50 and the RapiGest concentration adjusted to 0.1% using 0.1 M NH_4HCO_3. Proteins were digested overnight at 37 °C, and on the following day, RapiGest was hydrolyzed by addition of 1% TFA (final concentration) and incubation in a thermomixer at 800 rpm, 37 °C for 30 min. Hydrolyzed RapiGest was precipitated by centrifugation at 20,000× *g*, RT for 10 min and the supernatants were desalted using Oasis HLB cartridges (Waters) as described elsewhere [28]. Briefly, cartridges were equilibrated with 70% ACN, 0.1% acetic acid (AA), washed with 0.1% AA, and the sample was loaded. Subsequently, cartridges were washed with 0.1% AA and peptides were eluted sequentially with 30%, 50%, and 70% ACN, 0.1% AA. Eluate fractions were pooled and the combined samples dried in a vacuum centrifuge. Dried peptides were re-suspended in 5% ACN, and the peptide concentration was determined using the Quantitative Fluorometric Peptide Assay (ThermoFisher Scientific, Waltham, MA, USA), and the peptides were dried again.

2.4. LC-MS/MS Analysis

All analyses were performed using a Dionex Ultimate 3000 nano-UHPLC system coupled to an Orbitrap Fusion Lumos mass spectrometer (both Thermo Fisher Scientific). Analytical columns were produced in-house as follows: spray tips were generated with a P-2000 laser puller (Sutter Instruments, Novato, CA) from 360 μm outer diameter and 100 μm inner diameter fused silica capillaries and packed to a length of 40 cm with 3 μm ReprosilPur AQ C_{18} particles (Dr. Maisch, Ammerbuch-Entringen, Germany). Dried peptides were reconstituted in 5% ACN, 5% formic acid (FA), and 1 μg was loaded together with 750 fmol of internal retention time standards (iRTs, Biognosys, Schlieren, Switzerland) to the analytical column at a flow rate of 600 nl/min with 100% solvent A (0.1% FA in water) for 25 min. Peptides were eluted with 60, 120, and 240 min linear gradients from 5–35% solvent B (90% ACN, 0.1% FA) at a flow rate of 300 nl/min. For parallel reaction monitoring (PRM) measurements, precursor masses were selected from a previously recorded dataset (Table S1, [26]) while the spectral library for data-independent acquisition (DIA) analyses was generated using 240 min data-dependent acquisition (DDA) runs. In these analyses, survey spectra were acquired with a mass range of m/z 350–1200 at a resolution of 60,000 and an AGC target setting of 4×10^5 The most abundant precursor ions (charge states of 2–4) were isolated using the quadrupole (isolation width of m/z 1.6), and fragmented by HCD with a collision energy of 27 in the top speed mode (cycle time of 3 sec). Fragment ion spectra were acquired in the Orbitrap mass analyzer at a resolution of 30,000 and fragmented ions were excluded from further fragmentation for 120 s. For DIA analyses, one MS scan was performed with a mass range of m/z 350–1200, a resolution of 120,000, a maximum injection time of 20 ms, and an AGC target setting of 5×10^5. The MS scan was followed by static DIA MS/MS scans, covering the same m/z range with an overlap of m/z 0.5, with the following gradient lengths/ scan numbers/ isolation windows/ cycle times: 60 min/ 24 scans/ m/z 35.9/ 2.34 s; 120 min/ 36 scans/ m/z 24.1/ 3.44 s; 240 min/ 58 scans/ m/z 15.2/ 5.45 s). The HCD collision energy was set to 27% and DIA MS/MS scans were acquired in the Orbitrap with a resolution of 30,000, a maximum injection time of 60 ms, and an AGC target setting of 1×10^6. For PRM analyses, MS spectra were acquired with a mass range of m/z 300–1500 at a resolution of 60,000, a maximum injection time of 118 ms, and an AGC target setting of 4×10^5. Peptides were isolated in the quadrupole with an isolation width of m/z 1.2 and fragmented by HCD with a collision energy of 27%. MS/MS scans were acquired in the Orbitrap mass analyzer with a mass range of m/z 200–2000, a resolution of 30,000, a maximum injection time of 54 ms and an AGC target setting of 5×10^4.

2.5. Data Analysis

For DIA library generation, DDA *.raw files were analyzed with the Pulsar search engine integrated in Spectronaut (Version: 14.7.20, Biognosys) (1). Uniprot *Mus musculus* (release date: 09.09.2019 with 17,023 entries), in combination with a database containing common contaminants, was used for database searching with Spectronaut standard settings [29]. In brief, cleavage by trypsin with up to two missed cleavage sites was defined, propionamide (cysteine) was set as fixed and oxidation (methionine) as variable modification, and three to six fragment ions were selected for library generation, dependent on the intensity of the respective peptide. The high-precision iRT concept (dynamic) was applied for retention time alignment. Matching of mass tolerances for precursors, fragment ions, as well as peak extraction windows were determined automatically by Spectronaut. Only MS precursor information was utilized for peak detection, and interference correction was enabled. Global normalization was performed for individual runs based on the median abundance. Data were filtered with a 1% FDR cut off on the precursor and protein level (q-value < 0.01) [30]. *p*-value determination and unsupervised clustering were performed with the post-analysis pipeline of Spectronaut applying default parameters (distance metric: Manhattan Distance; linkage strategy: Ward's method; multiple testing correction: Storey's method).

For PRM analyses, a spectral library was generated using a subset of our previously published DDA dataset [26] with Skyline [31], applying a cut-off score of 0.95. Ambiguous peptide matches were excluded, and the library was filtered for peptides which were previously manually selected to be included in the assay (Table S1). For analysis of PRM data, *.raw files were loaded into Skyline daily version 20.2.1.315. Automated fragment ion selection by Skyline was utilized (6 ions/peptide) with the exception of the peptides with the sequence SLQPLYR and GSFSLSVR, for which only 5 fragment ions matched, using the following criteria: maximum mass error of 10 ppm for MS and MS/MS ion trace filtering (centroid mode) and charge states of 1+/2+ for b- and y-ions as well as 2+/3+ for precursor ions. Integration boundaries of iRT peptides were inspected manually and corrected, if necessary. Experimental data were only reviewed when Skyline reported a peak truncation, and peptides with truncated peaks or no MS/MS signal were excluded from further analysis. Peptide-centric reports were exported and further processed in MS Excel. For peptide and protein quantification, the summed area under the curve (AUC) of fragment ions was used. For all analyses, only peptides with quantitative values in all three replicates were considered.

3. Results and Discussion

We showed previously that the analysis of lysosome-enriched fractions with DIA allows for a superior performance compared to DDA measurements in a reduced amount of time [26]. When we investigated the lysosomal proteome in samples of higher complexity (such as whole cell lysates); however, we observed that the number of lysosomal proteins that can be reproducibly identified and quantified was markedly lower, indicating a reduced performance in such samples [14]. This is most likely due to the fact that co-fragmenting peptides increase the complexity of MS/MS spectra, which results in a decreased performance for the quantification of lysosomal proteins, as they are of low abundance relative to the whole proteome. In theory, PRM approaches should be superior in this aspect, as only a small m/z window, that is specific for the individual peptide, is fragmented.

In order to determine which strategy is best-suited for the MS/MS-based quantification of the lysosomal proteome in samples with different complexities, we compared DIA- and PRM-based quantification (Figure 1). Initially, we defined a highly reproducible lysosomal proteome from a dataset generated previously by our group, comprising 39 DDA LC-MS/MS runs of lysosome-enriched fractions from MEFs [26]. From these data, we only considered proteins which were assigned to the lysosomal compartment based on gene ontology (GO) and Uniprot categories, and which were detected in $\geq 75\%$ of LC-MS/MS runs with ≥ 2 unique peptides, resulting in a final list of 374 proteins (Table S1). For the comparison of DIA and PRM, we used a lysosome-enriched fraction from MEFs (LEF) as benchmark samples, as it contains the highest percentage of lysosomal proteins. Furthermore, we used MEF whole cell lysate (MWCL) as well as liver tissue lysate (LTL), representing samples of increasing complexity. We performed all experiments in triplicates with independent experimental replicates for each measurement.

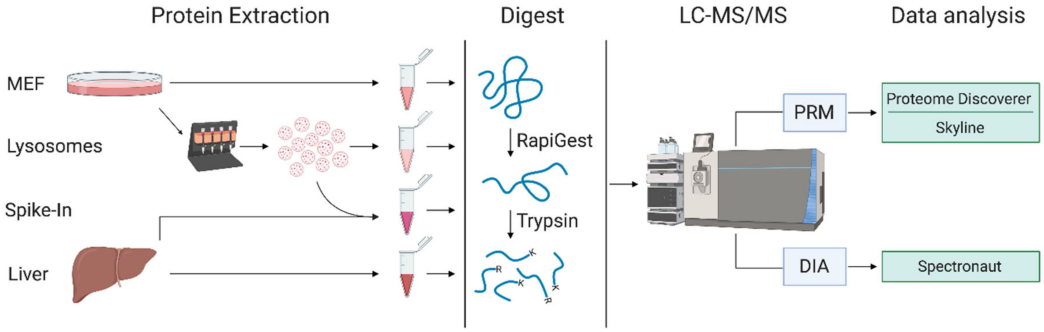


Figure 1. Workflow for sample preparation and analysis. For each sample type, proteins were extracted and digested in three experimental replicates and analyzed by LC-MS/MS using either DIA or PRM with different gradient lengths. MEF: mouse embryonic fibroblasts; DIA: data independent acquisition; PRM: parallel reaction monitoring; K: lysine; R: arginine. Created with Biorender.com.

3.1. Gradient Length and Sample Complexity Affect Lysosomal Protein Quantification by DIA

Initially, we analyzed all three sample types with different gradient lengths (60 min, 120 min, and 240 min) by DIA. We adjusted the width of DIA fragmentation windows depending on the gradient length in order to allow for a similar number of data points across chromatographic peaks of the individual gradients. Consequently, a shorter gradient resulted in a larger m/z window and vice versa, influencing the number of co-fragmented precursor ions. To assess performance of the individual methods, we evaluated both the numbers of total proteins, and those previously reported to be lysosomal (Table S1) that were identified in each run (Table S2).

We found highest total protein numbers in the MWCL, followed by the LEF, and the LTL (Figure 2a). While we observed a steady increase in the number of identified total proteins from 60 min to both 120 min and 240 min gradients for MWCL (increase of 19% and 28%) and LTL (increase of 28% and 46%) samples, the numbers of IDs detected in the LEF only increased from 60 min to 120 min gradients (increase of 30%). When considering only lysosomal proteins, we identified, as expected, highest numbers in the LEF, followed by MWCL and LTL. While the latter two showed a similar correlation of gradient length and protein identifications, the LEF produced virtually constant numbers for all gradients and only CV values improved. The differences in identifications were particularly pronounced when considering lysosomal proteins quantified with < 5% CV in the 60 min gradient analyses, 116 of which were found in the LEF but only 45 in the LTL (Figure 2b).

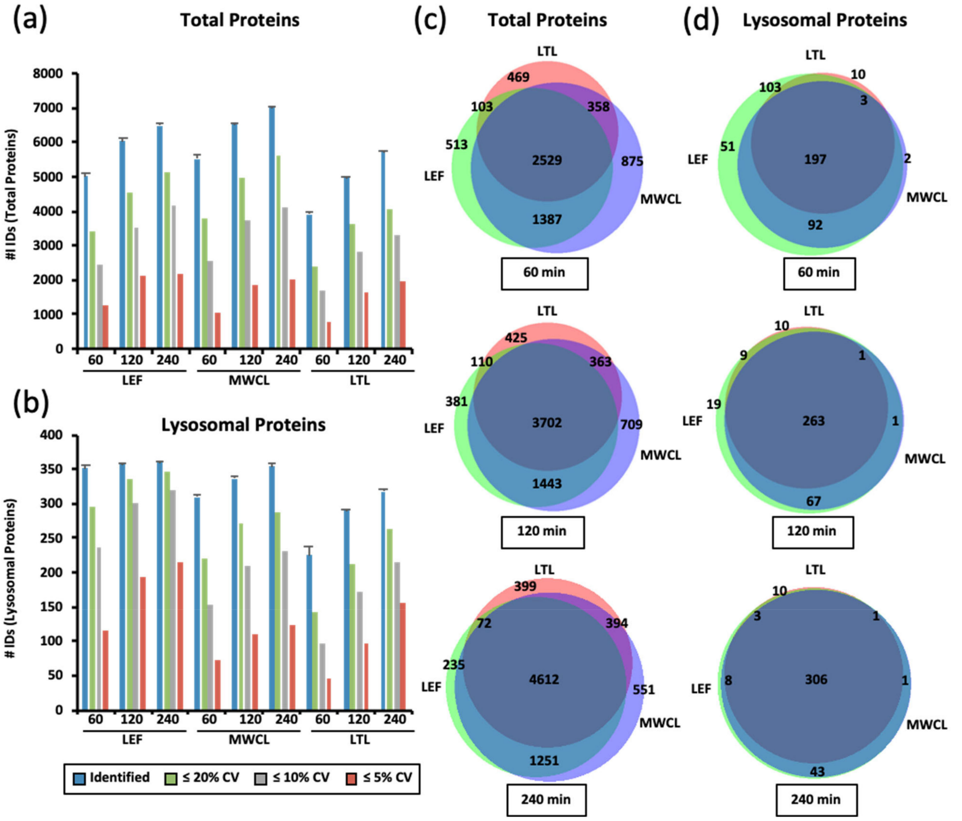


Figure 2. Identification of total and lysosomal proteins in DIA experiments. (**a**,**b**) Numbers of identified total proteins (**a**) and those known to be located at/in the lysosome (**b**). Shown are average values (n = 3) for the analysis of lysosome-enriched fractions of MEF (LEF), MEF whole cell lysate (MWCL), and liver tissue lysate (LTL) with three different gradient lengths. Total IDs as well as those quantified with a CVs ≤ 20%, 10%, and 5% are shown. (**c**,**d**) Overlap in protein identification for total proteins (**c**) and those known to be located at the lysosome (**d**) for proteins identified in all three replicates of LEF, MWCL, and LTL samples analyzed with different gradient lengths. Venn diagrams were generated with the tool BioVenn [32]. LEF: lysosome-enriched fractions from mouse embryonic fibroblasts; MWCL: mouse embryonic fibroblast whole cell lysates; LTL: whole liver tissue lysate; CV: coefficient of variation; ID: number of identified proteins.

It was quite surprising to us that we identified the highest number of proteins in the MWCL, as the LTL should theoretically be the most complex sample. A possible explanation for this observation is that the LTL contains a certain number of highly abundant proteins, which account for a larger percentage of the sample than highly abundant MWCL and LEF proteins. Consequently, in LTL the remaining proteins present a smaller fraction of the total sample. As the C-trap — which is used for ion storage prior to injection into the Orbitrap — has a limited capacity [33], this results in reduced fragment ion intensities for the lower abundant proteins, which are not sufficient for identification/quantification. Furthermore, the highly abundant fragment ions from these proteins dominate the DIA MS/MS spectra, resulting in reduced detection of co-fragmented lower abundant peptides.

When considering the increase in identification of total unique proteins with increasing analysis time, each sample contributed a distinct population (Figure 2c). Lysosomal protein identifications, on the other hand, were very similar between the samples, and for 240 min gradients the majority was identified at least in both the MWCL and the LEF (Figure 2d). These results imply that short gradients suffice to achieve a good coverage of lysosomal proteins in LEFs, while longer gradients are needed when MWCL and LTL are analyzed.

While this confirms that lysosomal proteins are more abundant in lysosome-enriched fractions, which was certainly expected, it also shows that the detection of such lower abundant proteins in DIA analyses suffers from sample complexity. As our DIA analyses were performed with different m/z windows for the different gradient lengths, this effect is most likely related to the number of co-fragmented precursor ions and the resulting MS/MS spectrum complexity. The bigger the fragmentation window is, the more peptides are co-fragmented, and consequently the fragment ions of the lower abundant lysosomal proteins are identified with a lower efficiency.

3.2. Variation of Protein Abundance and Variance between Sample Types in DIA Analyses

To further follow up on this effect, we investigated the protein abundances for the individual samples utilizing median-normalized AUCs (Figure 3a). Confirming our previous assumption, the liver lysate resulted in the highest average abundance (1.6- and 1.8-fold higher compared to the lysosome-enriched fractions and MEF lysate, respectively, for 60 min gradients) and the largest number of highly abundant proteins (36 proteins compared to 11 and 12 with log10 values > 7.5 for LEF and MWCL, respectively). Average protein abundance correlated inversely with the number of protein identifications, with highest values in the shortest gradient, irrespective of the sample type. For lysosomal proteins, we observed highest average abundances in the LEF (1.8- and 2.4-fold higher compared LTL and MWCL for 60 min gradients) and, unlike the total protein identifications, no decrease in abundance with increasing gradient length (Figure 3c). Average CV values, however, behaved similarly for all types of proteins (Figure 3b,d).

To visualize the differences of the individual datasets on a global scale, we generated heatmaps for the average abundances of total and lysosomal proteins, clustered in a row- and column-wise manner (Figure 4a,b). For both analyses, we observed distinct protein populations which formed individual clusters, based on their abundance in the respective sample types and gradient lengths. In most cases, gradient length played a decisive role, while the highest differences existed for the LTL relative to the other samples. For the majority of known lysosomal proteins, we detected a higher abundance in the LEF relative to the MWCL and the LTL, while certain proteins were exclusively identified in the LEF. We also identified, however, some clusters with a higher abundance in MWCL and LTL, implying that either not all lysosomal proteins were recovered efficiently in the lysosome-enrichment step, or that a certain population of these proteins was located in a different cellular compartment.

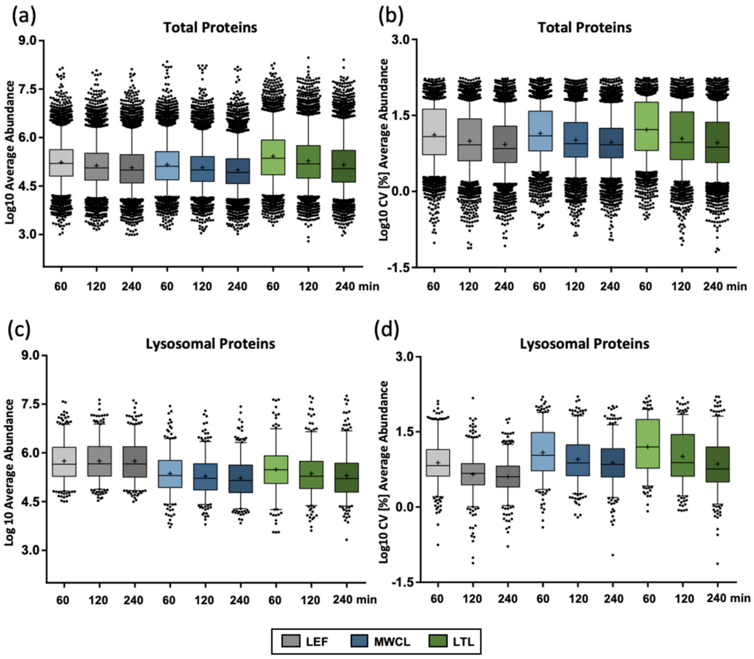


Figure 3. Reproducibility of protein abundance in DIA analyses. (**a**) Protein abundance for total proteins identified in individual samples with different gradient lengths. (**b**) CV values for total proteins identified in individual samples with different gradient lengths. (**c**) Protein abundance for lysosomal proteins identified in individual samples with different gradient lengths. (**d**) CV values for lysosomal proteins identified in individual samples with different gradient lengths. Shown are combined values from 3 replicates, the median is indicated by a line, while the average is marked with a "+". LEF: lysosome-enriched fractions from mouse embryonic fibroblasts; MWCL: mouse embryonic fibroblast whole cell lysates; LTL: whole liver tissue lysate; CV: coefficient of variation.

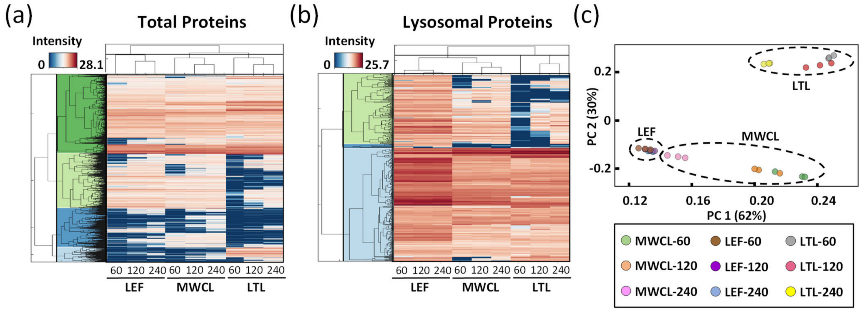
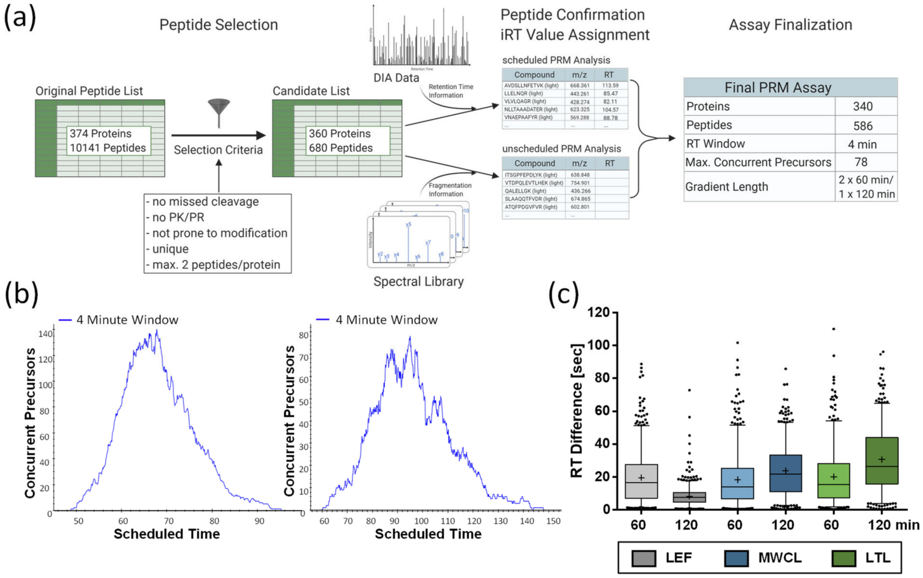


Figure 4. Global analysis of DIA datasets. (**a**,**b**) Unsupervised clustering of average abundances (columns) of LEF, MWCL, and LTL for three different gradient lengths (60, 120 and 240 min) for total proteins identified ((**a**), n = 7145) and lysosomal proteins ((**b**), n = 314). The color code indicates the normalized intensity of the individual proteins. (**c**) Principal component analysis (PCA) for all analyses with two defined standardized principal components (PC1 and PC2). LEF: lysosome-enriched fractions from mouse embryonic fibroblasts; MWCL: mouse embryonic fibroblast whole cell lysates; LTL: whole liver tissue lysate.

Finally, we assessed the global variability between the datasets by principal component analysis (PCA, Figure 4c). The two main principal components (PC1 and PC2), which are responsible for 62% and 30% of the variance in the dataset, allowed for a good separation of the samples. As individual replicates of the same sample type and gradient length clustered closely together, the main variance in the dataset (PC1) can be explained based on the difference of the sample type itself. However, especially for the MWCL, the 240 min gradient data behaved significantly differently than those acquired with other gradients, being actually closer to the LEF. This relates most likely to the fact that the LEF originated from MEFs and that proteome coverage in the 240 min gradient increased to such an extent that very similar proteins were identified (Figure 2c).

3.3. PRM Assay Development

For the 374 proteins included in our lysosomal proteome reference list (Table S1), 10,141 unique peptides were identified in the course of our previous analysis [26]. We narrowed down the list of putative peptides by excluding those identified with variable modifications, missed cleavage sites, or containing the amino acid combination PK or PR (as proline residues interfere with tryptic cleavage). These criteria were fulfilled by 3816 peptides representing 367 proteins. Based on the average signal intensities in this dataset, we considered the two most abundant peptides for each protein, resulting in a final list of 680 peptides, as for some proteins only a single peptide fulfilled our criteria (Figure 5a, Table S1).



scheduled PRM Analysis

Compound	m/z	RT
AVDSLLNFETVK (light)	668.361	113.59
LLELNQR (light)	443.261	85.47
VLVLQAGR (light)	428.274	82.11
NLLTAAADATER (light)	623.325	104.57
VNAEPAAFYR (light)	569.288	88.78
...	...	...

unscheduled PRM Analysis

Compound	m/z	RT
ITSGPFEPDLYK (light)	638.848	
VTDPQLEVTLHEK (light)	754.901	
QALELLGK (light)	436.266	
SLAAQQTFVDR (light)	674.865	
ATQFPDGVFVR (light)	602.801	
...	...	

Final PRM Assay	
Proteins	340
Peptides	586
RT Window	4 min
Max. Concurrent Precursors	78
Gradient Length	2 x 60 min/ 1 x 120 min

Figure 5. Establishment of PRM assay. (**a**) Workflow for the development of the PRM assay. Created with Biorender.com. (**b**) Distribution of concurrent precursor elution in 60 min and 120 min gradients. (**c**) Differences between predicted and experimentally determined retention times for individual measurements. Shown are combined values of three replicates and the mean (+), median (line), and interquartile range are indicated. LEF: lysosome-enriched fractions from mouse embryonic fibroblasts; MWCL: mouse embryonic fibroblast whole cell lysates; LTL: whole liver tissue lysate; CV: coefficient of variation; RT: retention time.

For PRM assay scheduling, we extracted the peptides' retention time information from the 120 min DIA runs of the LEF, followed by the refinement of the assay by PRM analysis of the same sample, including high-precision indexed retention time (iRT) standards [29].

After analysis with an initial scheduled PRM assay (15 min retention time windows), we performed unscheduled PRM runs for those peptides that were not detected in these initial analyses. Finally, we combined the acquired retention times of all peptides that we were able to detect with distinct fragment ion signatures, and created an iRT-normalized library. This resulted in a final assay comprising 586 peptides from 340 lysosomal proteins. For the analysis of acquired PRM data with Skyline [31], we built a reference spectral library from our previously measured DDA dataset of the LEF [26]. Finally, we generated two assays utilizing 4 min retention time windows. In one assay, all peptides were analyzed in a single 120 min gradient, while the other assay consisted of two 60 min gradients. This was necessary, as the high number of concurrent precursors (up to 140) would have drastically reduced the number of data points for chromatographic peaks eluting in the middle of gradient (Figure 5b). For subsequent analyses, we determined iRT correction factors for the different sample types using DDA runs, and adjusted the scheduling accordingly. For data export from Skyline, a minimum number of six data points was defined.

3.4. Gradient Length and Sample Type Affect Data Quality in PRM Analyses

Due to the high number of peptides, we defined parameters for the acceptance of PRM quantification data without manual inspection of each peptide. Initially, we assessed the difference between predicted and experimentally observed retention times. For the 60 min and 120 min gradients, we observed average peak widths of 21 sec and 25 sec, respectively, and an average retention time variability of ±15 sec, with a slightly lower average shift for the 60 min gradients (Figure 5c). LEF analyses with 120 min gradients presented with only ±8 sec an exception, which might be due to the fact that we performed the PRM assay retention time normalization with 120 min gradient measurements of the LEF, while the scheduling for MWCL and LTL was solely based on iRT predictions.

Next, we investigated the quality of acquired fragment ions for the individual peptides utilizing the dot product (dotP) value [34], which allows for correlation between the acquired spectrum and the spectral library (generated from our reference dataset [26]). Especially for the analysis of unfractionated highly complex samples, this allows to identify the impact of interfering ions that may result in false quantification results. Across all analyses, the average dotP value was > 0.85, indicating a good matching of our PRM data with the spectral library (Table S3). While we observed roughly similar dotP values for both the 60 min and 120 min assays, they decreased with sample complexity (Figure 6a). Compared to the LEF, which displayed the least variation, especially the LTL resulted in lower dotP values and higher variability. These findings imply a lower relative abundance of lysosomal proteins and an increase of interfering fragment ions in the MS/MS spectra for the more complex samples, which is also in agreement with the DIA data.

Subsequently, we investigated the correlation of dotP values and numbers of fragment ions used for quantification of the different sample types and gradient lengths (Figure S1). Utilizing three to six fragment ions, we applied different dotP value thresholds and determined the number of peptides passing it. As expected, lower numbers of fragment ions resulted in more peptides passing the threshold at higher dotP values. This was especially true when dotP thresholds ≥ 0.9 were applied, as we observed a clear difference between the peptides identified with 3, 4, 5, or 6 fragment ions. For lower dotP values (0.7–0.8), this effect was far less pronounced. As already indicated by the average dotP values (Figure 6a), an inverse correlation with sample complexity could be observed. Based on these analyses, we defined 6 fragment ions per peptide with a dotP value of 0.7 as cut-off for the acceptance of quantification information from PRM data.

For LEF data, this cut-off resulted in an acceptance rate of 92% of the peptides included in our assay for both gradients. The value for MWCL was 87% for both gradients and for LTL 73% and 78%, for the 60 min and 120 min gradient, respectively. Applying these cut-offs, we exported the data from Skyline and utilized them for all further analyses (Table S4).

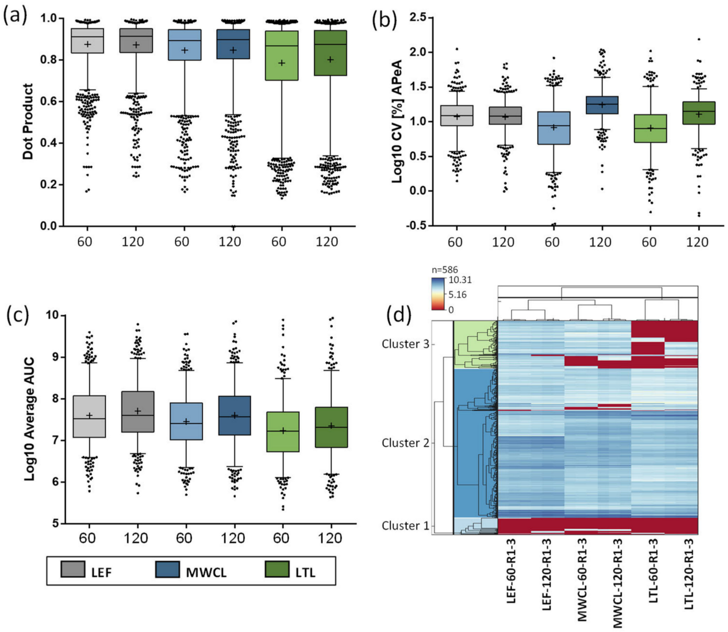


Figure 6. Characterization of data generated by PRM analysis. (**a**) Assessment of data quality in PRM analyses by Dot Product values. (**b**) Log10 CV values for the summed AUC of all targeted peptides. (**c**) Averaged log10-transformed summed peptide AUCs of the replicates (n = 3) across the different sample types and gradient lengths from PRM measurements. (a/b/c) Indicated are mean (+), median (line), and interquartile range. (**d**) Heatmap of the log10-transformed AUCs of all peptides covered by the PRM assay across all measurements. Each column contains data from one measurement and each row represents one peptide. Peptides are clustered if they exhibit similar trends across the samples. LEF: lysosome-enriched fractions from mouse embryonic fibroblasts; MWCL: mouse embryonic fibroblast whole cell lysates; LTL: whole liver tissue lysate; AUC: area under the curve; CV: coefficient of variation; APeA: average peptide abundance.

3.5. PRM Analysis of the Lysosomal Proteome

We initially investigated the reproducibility of quantification (Figure 6b). While CVs of the LEF analyses were similar for both gradients, the 120 min gradient resulted in consistently higher CVs for both MWCL and LTL. Surprisingly, the higher complexity samples resulted in a lower average CV than the LEF for the 60 min runs. A possible explanation for this observation is the lower sample amount utilized for LEF sample preparation (~20 µg) compared to MWCL and LTL (~100 µg), which may have resulted in a higher variability during pipetting and desalting. Subsequently, we calculated the average summed area under the curve (AUC) for each sample type and gradient length (Figure 6c). Interestingly, while we saw a higher summed abundance for the LEF in comparison to the other samples for 60 min gradients, the values were more similar for the 120 min analyses, especially for the comparison of LEF and MWCL. This could be related to the different numbers of data points acquired over the chromatographic peak as well as variances in peak width/shape between gradients.

Finally, we assessed the overall correlation of the data in a heatmap, depicting the signal intensities of individual peptides in each sample and replicate, clustered in a row- and column-wise manner (Figure 6d). In general, the lysosomal peptides formed three distinct clusters. Two clusters showed similar expression levels (general high or low expression) in all samples, while the third cluster contained proteins that were detected with differing levels in the individual samples. In agreement with the DIA data, we observed subsets of peptides that were only found in the LEF, and were not detected in both the MWCL and LTL. Moreover, peptides existed that were detected both in LEF and MWCL, but not in LTL, and a small subset with higher abundance in LTL compared to the other samples.

3.6. Comparison of DIA- and PRM-Based Quantification of the Lysosomal Proteome

In order to correlate the performance of DIA- and PRM-based quantification of the lysosomal proteome, we initially compared the data obtained from the individual datasets. When considering the average abundance and CVs of lysosomal proteins, we observed for all DIA analyses that longer gradients resulted in lower CV values but also lower abundances (Figure 3c,d), while for PRM both intensities and CVs (with the exception of the LEF data) increased with gradient length (Figure 6b,c). When considering how many lysosomal proteins were found in the individual analyses, we identified higher numbers for DIA, with the exception of LTL analyzed with 60 min gradients (Figure S2). To assess to what extend the abundances acquired with the individual approaches correlate, we extracted the AUCs of the peptides included in our PRM assay from the DIA dataset (60 min gradients for both approaches) and performed a direct comparison (Figure 7a). We observed for all three sample types that PRM resulted in higher signal intensities than DIA and that correlation of signal intensities was dependent on the abundance of the respective protein. We observed a good correlation for high-abundant proteins (upper 50% of DIA intensities) in all sample types. For low-abundant proteins (lower 50% of DIA intensities), we only detected a good correlation between DIA and PRM for the LEF. For the more complex samples, however, DIA seemed to underestimate high signal intensities, resulting in poor correlation with the PRM data.

As the main application of both methods is the quantitative comparison of the lysosomal proteome between different states, we performed a spike-in experiment to simulate constitutive upregulation of the whole lysosomal proteome and analyzed the sample by both PRM and DIA, applying 60 min gradients for both approaches. For this purpose, we combined LEF and LTL in a 1 to 5 ratio and compared the data to LTL samples without spike-in (Figure 1). In theory, as LEFs contain higher amounts of lysosomal proteins, this should result in a general increase of intensity for all lysosomal proteins present in the sample.

For both approaches, the number of detected lysosomal proteins increased in comparison to LTL without spike-in, while the increase for PRM was 50% higher compared to DIA (223 to 243 for DIA and 278 to 308 for PRM, Figure 7b). Subsequently, we investigated the fold change ratios for proteins identified in all samples with both approaches (Figure 7c). We detected a median increase of intensity of 1.8 for PRM and 1.2 for DIA. When investigating values for individual proteins, we observed a discrepancy of $\geq$ 30% between fold change values acquired by DIA and PRM for 75% of proteins (average CV for DIA and PRM analysis of LTLs: 16% and 7%). While 142 proteins were detected with a higher value in PRM, only 35 were higher in the DIA data (Figure 7d). Classification of proteins based on their fold change values between the spike-in and the LTL sample further showed that DIA failed to detect any increase in signal intensity for 81 proteins upon spike-in, while this was only the case for 10 proteins in the PRM data (Figure 7e). Subsequently, we investigated if this effect was related to the abundance of individual proteins, as we observed markedly reduced correlation coefficients between DIA and PRM for lower abundant proteins in LTL samples (Figure 7a). Along this line, we grouped all proteins based on their abundance relative to the highest/lowest abundant protein in the respective dataset and plotted the observed fold change ratios for the individual groups (Figure 7f). While we observed highly similar fold change value distributions between LTL and spike-in samples for proteins across the whole range of abundance for the PRM data, a clear shift in the pattern of the DIA data was visible. Relative to the PRM data, DIA reported higher fold change ratios for low-abundant proteins while it resulted in lower values for high-abundant proteins. Taken together, these data indicate that PRM is better suited for the quantification of changes in the lysosomal proteome of LTL, which is mainly related to the better performance for the highest- and lowest-abundant lysosomal proteins in the dataset.

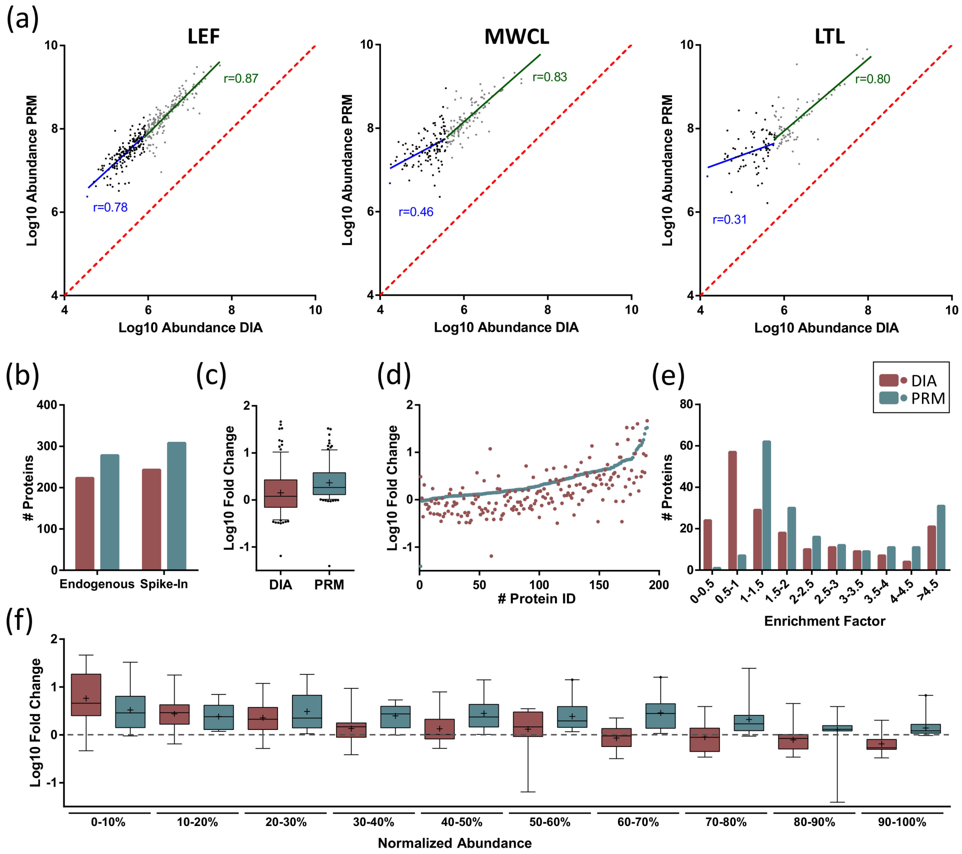


Figure 7. Comparison of DIA and PRM analysis of lysosomal proteins. (**a**) Correlation of normalized peptide signal intensities for DIA and PRM runs. Proteins are sorted based on their intensity in DIA measurements and grouped into two groups based on their intensity (upper/lower 50% of proteins). For each population, a linear regression analysis was performed and the respective correlation coefficient (r) is indicated. (**b**) Identification of proteins in LTL with and without spike-in of LEFs. (**c**,**d**) Fold change values for individual lysosomal proteins in LTL with spike-in of LEFs. (**e**) Frequency of proteins within distinct fold change quantiles for DIA and PRM data for ratios of LTL with/without spike-in of LEFs. (**f**) Protein fold change values for LTL with/without spike-in of LEFs. Proteins are grouped based on their abundance in the respective dataset relative to the highest/lowest-abundant protein. LEF: lysosome-enriched fractions from mouse embryonic fibroblasts; MWCL: mouse embryonic fibroblast whole cell lysates; LTL: whole liver tissue lysate.

4. Conclusions

In the present study, we analyzed the lysosomal proteome in samples of varying complexity by DIA and PRM. While both methods were well-suited for the analysis of lysosomal proteins in all samples, differences between the approaches became apparent that were mostly related to sample complexity. DIA identified more proteins in lower complexity samples and at longer gradients, since it was not limited by a predefined list of peptides, as was the case for the PRM analyses. Furthermore, no assay development was necessary for DIA analyses, thus greatly reducing the amount of time needed. For peptides covered by both approaches, DIA and PRM performed similarly for lower complexity LEFs, while PRM outperformed DIA in both MWCL and LTL. Especially for the quantification of protein level changes in LTL, PRM was able to identify significantly higher numbers of protein level alterations than DIA, which reported no change in abundance for a high number of proteins. Therefore, for the analysis of highly complex samples, such as whole tissue lysates, PRM presents the method of choice. Our developed PRM assay allows for the direct analysis of the lysosomal proteome from small amounts of whole tissue samples, without the need for lysosome enrichment, extending the toolbox for the investigation of the lysosomal proteome in complex samples.

Supplementary Materials: The following are available online at https://www.mdpi.com/2227-7382/9/1/4/s1, Table S1: Reference list of high confidence lysosomal proteins and peptides covered by PRM assay. Supplementary Table S2: DIA Data, direct output from Spectronaut as well as processed data. Supplementary Table S3: PRM data, direct output from Skyline. Supplementary Table S4: PRM data filtered for dot p values > 0.7 and further analyses. Supplementary Figure S1: Dot product threshold determination for the acceptance of PRM data. Figure S2: Overlap of identified proteins from PRM and DIA runs.

Author Contributions: Conceptualization, D.W.; methodology, P.M., A.S., and J.S.; validation, P.M., A.S., and J.S.; formal analysis, P.M., A.S., and J.S.; investigation, P.M., A.S., and J.S.; resources, P.M., and A.S.; writing—original draft preparation, A.S., and D.W.; writing—review and editing, P.M., A.S., J.S., and D.W.; visualization, P.M., A.S., and J.S.; supervision, D.W.; project administration, D.W.; funding acquisition, D.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Deutsche Forschungsgemeinschaft, grant number FOR2625, and the Studienstiftung des Deutschen Volkes.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [35] partner repository with the dataset identifier PXD023278.

Acknowledgments: The authors would like to thank Robert Hardt for helpful discussions, Shiva Ahmadi for help with liver samples, and Fatema Akter as well as Srigayatri Ponnaiyan for help with lysosome-enriched fractions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. De Duve, C.; Wattiaux, R. Functions of Lysosomes. *Annu. Rev. Physiol.* **1966**, *28*, 435–492. [CrossRef] [PubMed]
2. Perera, R.M.; Zoncu, R. The Lysosome as A Regulatory Hub. *Annu. Rev. Cell Dev. Biol.* **2016**, *32*, 223–253. [CrossRef] [PubMed]
3. Platt, F.M.; d'Azzo, A.; Davidson, B.L.; Neufeld, E.F.; Tifft, C.J. Lysosomal Storage Diseases. *Nat. Rev. Dis. Primers* **2018**, *4*, 27. [CrossRef] [PubMed]
4. Ballabio, A.; Gieselmann, V. Lysosomal Disorders: From Storage to Cellular Damage. *Biochim. Biophys. Acta* **2009**, *1793*, 684–696. [CrossRef] [PubMed]
5. Jeyakumar, M.; Dwek, R.A.; Butters, T.D.; Platt, F.M. Storage Solutions: Treating Lysosomal Disorders of The Brain. *Nat. Rev. Neurosci.* **2005**, *6*, 713–725. [CrossRef] [PubMed]
6. Perera, R.M.; Stoykova, S.; Nicolay, B.N.; Ross, K.N.; Fitamant, J.; Boukhali, M.; Lengrand, J.; Deshpande, V.; Selig, M.K.; Ferrone, C.R.; et al. Transcriptional Control of Autophagy-Lysosome Function Drives Pancreatic Cancer Metabolism. *Nature* **2015**, *524*, 361–365. [CrossRef]
7. DePaolo, J.; Goker-Alpan, O.; Samaddar, T.; Lopez, G.; Sidransky, E. The Association Between Mutations in The Lysosomal Protein Glucocerebrosidase And Parkinsonism. *Mov. Disord.* **2009**, *24*, 1571–1578. [CrossRef]
8. Chi, C.; Riching, A.S.; Song, A.K. Lysosomal Abnormalities in Cardiovascular Disease. *Int. J. Mol. Sci.* **2020**, *21*, 811. [CrossRef]
9. Ballabio, A.; Bonifacino, J.S. Lysosomes as Dynamic Regulators of Cell and Organismal Homeostasis. *Nat. Rev. Mol. Cell Biol.* **2019**, *2019*, 1–18. [CrossRef]

10. Settembre, C.; Di Malta, C.; Polito, V.A.; Arencibia, M.G.; Vetrini, F.; Erdin, S.; Erdin, S.U.; Huynh, T.; Medina, D.; Colella, P.; et al. Tfeb Links Autophagy to Lysosomal Biogenesis. *Science* **2011**, *332*, 1429–1433. [CrossRef]
11. Zhang, C.S.; Jiang, B.; Li, M.; Zhu, M.; Peng, Y.; Zhang, Y.L.; Wu, Y.Q.; Li, T.Y.; Liang, Y.; Lu, Z.; et al. The Lysosomal V-Atpase-Ragulator Complex Is A Common Activator for Ampk And Mtorc1, Acting as A Switch Between Catabolism and Anabolism. *Cell Metab.* **2014**, *20*, 526–540. [CrossRef] [PubMed]
12. Aebersold, R.; Mann, M. Mass-Spectrometric Exploration of Proteome Structure and Function. *Nature* **2016**, *537*, 347–355. [CrossRef] [PubMed]
13. Akter, F.; Ponnaiyan, S.; Koegler-Mohrbacher, B.; Bleibaum, F.; Damme, M.; Renard, B.Y.; Winter, D. Multi Cell Line Analysis of Lysosomal Proteomes Reveals Unique Features and Novel Lysosomal Proteins. *bioRxiv* **2020**. [CrossRef]
14. Singh, J.; Kaade, E.; Muntel, J.; Bruderer, R.; Reiter, L.; Thelen, M.; Winter, D. Systematic Comparison of Strategies for The Enrichment of Lysosomes by Data Independent Acquisition. *J. Proteome Res.* **2020**, *19*, 371–381. [CrossRef] [PubMed]
15. Thelen, M.; Winter, D.; Braulke, T.; Gieselmann, V. Silac-based comparative proteomic analysis of lysosomes from mammalian cells using Lc-Ms/Ms. In *Lysosomes: Methods and Protocols*; Öllinger, K., Appelqvist, H., Eds.; Springer: New York, NY, USA, 2017; pp. 1–18.
16. Wyant, G.A.; Abu-Remaileh, M.; Frenkel, E.M.; Laqtom, N.N.; Dharamdasani, V.; Lewis, C.A.; Chan, S.H.; Heinze, I.; Ori, A.; Sabatini, D.M. Nufip1 Is A Ribosome Receptor for Starvation-Induced Ribophagy. *Science* **2018**, *360*, 751–758. [CrossRef] [PubMed]
17. Zoncu, R.; Bar-Peled, L.; Efeyan, A.; Wang, S.; Sancak, Y.; Sabatini, D.M. Mtorc1 Senses Lysosomal Amino Acids Through an Inside-Out Mechanism That Requires the Vacuolar H+-Atpase. *Science* **2011**, *334*, 678–683. [CrossRef] [PubMed]
18. Shi, T.; Song, E.; Nie, S.; Rodland, K.D.; Liu, T.; Qian, W.J.; Smith, R.D. Advances in Targeted Proteomics and Applications To Biomedical Research. *Proteomics* **2016**, *16*, 2160–2182. [CrossRef]
19. Bilbao, A.; Varesio, E.; Luban, J.; Strambio-De-Castillia, C.; Hopfgartner, G.; Müller, M.; Lisacek, F. Processing Strategies and Software Solutions for Data-Independent Acquisition in Mass Spectrometry. *Proteomics* **2015**, *15*, 964–980. [CrossRef]
20. Hoofnagle, A.N.; Becker, J.O.; Oda, M.N.; Cavigiolio, G.; Mayer, P.; Vaisar, T. Multiple-Reaction Monitoring-Mass Spectrometric Assays Can Accurately Measure the Relative Protein Abundance in Complex Mixtures. *Clin. Chem.* **2012**, *58*, 777–781. [CrossRef]
21. Lange, V.; Picotti, P.; Domon, B.; Aebersold, R. Selected Reaction Monitoring for Quantitative Proteomics: A Tutorial. *Mol. Syst. Biol.* **2008**, *4*, 222. [CrossRef]
22. Schmidt, A.; Gehlenborg, N.; Bodenmiller, B.; Mueller, L.N.; Campbell, D.; Mueller, M.; Aebersold, R.; Domon, B. An Integrated, Directed Mass Spectrometric Approach for In-Depth Characterization of Complex Peptide Mixtures. *Mol. Cell. Proteom.* **2008**, *7*, 2138–2150. [CrossRef] [PubMed]
23. Damaghi, M.; Tafreshi, N.K.; Lloyd, M.C.; Sprung, R.; Estrella, V.; Wojtkowiak, J.W.; Morse, D.L.; Koomen, J.M.; Bui, M.M.; Gatenby, R.A.; et al. Chronic Acidosis in The Tumour Microenvironment Selects for Overexpression of Lamp2 in the Plasma Membrane. *Nat. Commun.* **2015**, *6*, 8752. [CrossRef] [PubMed]
24. Sjödin, S.; Öhrfelt, A.; Brinkmalm, G.; Zetterberg, H.; Blennow, K.; Brinkmalm, A. Targeting Lamp2 In Human Cerebrospinal Fluid with A Combination of Immunopurification and High Resolution Parallel Reaction Monitoring Mass Spectrometry. *Clin. Proteom.* **2016**, *13*, 4. [CrossRef] [PubMed]
25. Ivry, S.L.; Knudsen, G.M.; Caiazza, F.; Sharib, J.M.; Jaradeh, K.; Ravalin, M.; O'Donoghue, A.J.; Kirkwood, K.S.; Craik, C.S. The Lysosomal Aminopeptidase Tripeptidyl Peptidase 1 Displays Increased Activity in Malignant Pancreatic Cysts. *Biol. Chem.* **2019**, *400*, 1629–1638. [CrossRef] [PubMed]
26. Ponnaiyan, S.; Akter, F.; Singh, J.; Winter, D. Comprehensive Draft of The Mouse Embryonic Fibroblast Lysosomal Proteome by Mass Spectrometry Based Proteomics. *Sci. Data* **2020**, *7*, 1. [CrossRef]
27. Muntel, J.; Gandhi, T.; Verbeke, L.; Bernhardt, O.M.; Treiber, T.; Bruderer, R.; Reiter, L. Surpassing 10 000 Identified and Quantified Proteins in A Single Run by Optimizing Current Lc-Ms Instrumentation and Data Analysis Strategy. *Mol. Omics* **2019**, *15*, 348–360. [CrossRef]
28. Winter, D.; Steen, H. Optimization of Cell Lysis and Protein Digestion Protocols for The Analysis of Hela S3 Cells by Lc-Ms/Ms. *Proteomics* **2011**, *11*, 4726–4730. [CrossRef]
29. Bruderer, R.; Bernhardt, O.M.; Gandhi, T.; Reiter, L. High-Precision Irt Prediction in The Targeted Analysis of Data-Independent Acquisition and Its Impact on Identification and Quantitation. *Proteomics* **2016**, *16*, 2246–2256. [CrossRef]
30. Rosenberger, G.; Bludau, I.; Schmitt, U.; Heusel, M.; Hunter, C.L.; Liu, Y.; MacCoss, M.J.; MacLean, B.X.; Nesvizhskii, A.I.; Pedrioli, P.G.; et al. Statistical Control of Peptide and Protein Error Rates in Large-Scale Targeted Data-Independent Acquisition Analyses. *Nat. Methods* **2017**, *14*, 921–927. [CrossRef]
31. Pino, L.K.; Searle, B.C.; Bollinger, J.G.; Nunn, B.; MacLean, B.; MacCoss, M.J. The Skyline Ecosystem: Informatics for Quantitative Mass Spectrometry Proteomics. *Mass Spectrom. Rev.* **2020**, *39*, 229–244. [CrossRef]
32. Hulsen, T.; de Vlieg, J.; Alkema, W. Biovenn—A Web Application for The Comparison and Visualization of Biological Lists Using Area-Proportional Venn Diagrams. *BMC Genom.* **2008**, *9*, 488. [CrossRef] [PubMed]
33. Zubarev, R.A.; Makarov, A. Orbitrap Mass Spectrometry. *Anal. Chem.* **2013**, *85*, 5288–5296. [CrossRef] [PubMed]

34. Frewen, B.E.; Merrihew, G.E.; Wu, C.C.; Noble, W.S.; MacCoss, M.J. Analysis of Peptide Ms/Ms Spectra from Large-Scale Proteomics Experiments Using Spectrum Libraries. *Anal. Chem.* **2006**, *78*, 5678–5684. [CrossRef] [PubMed]
35. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D.J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. The Pride Database and Related Tools and Resources In 2019: Improving Support for Quantification Data. *Nucleic Acids Res.* **2019**, *47*, D442–D450. [CrossRef] [PubMed]

8.3 Normalized Beta-Hexosaminidase Activity Assays

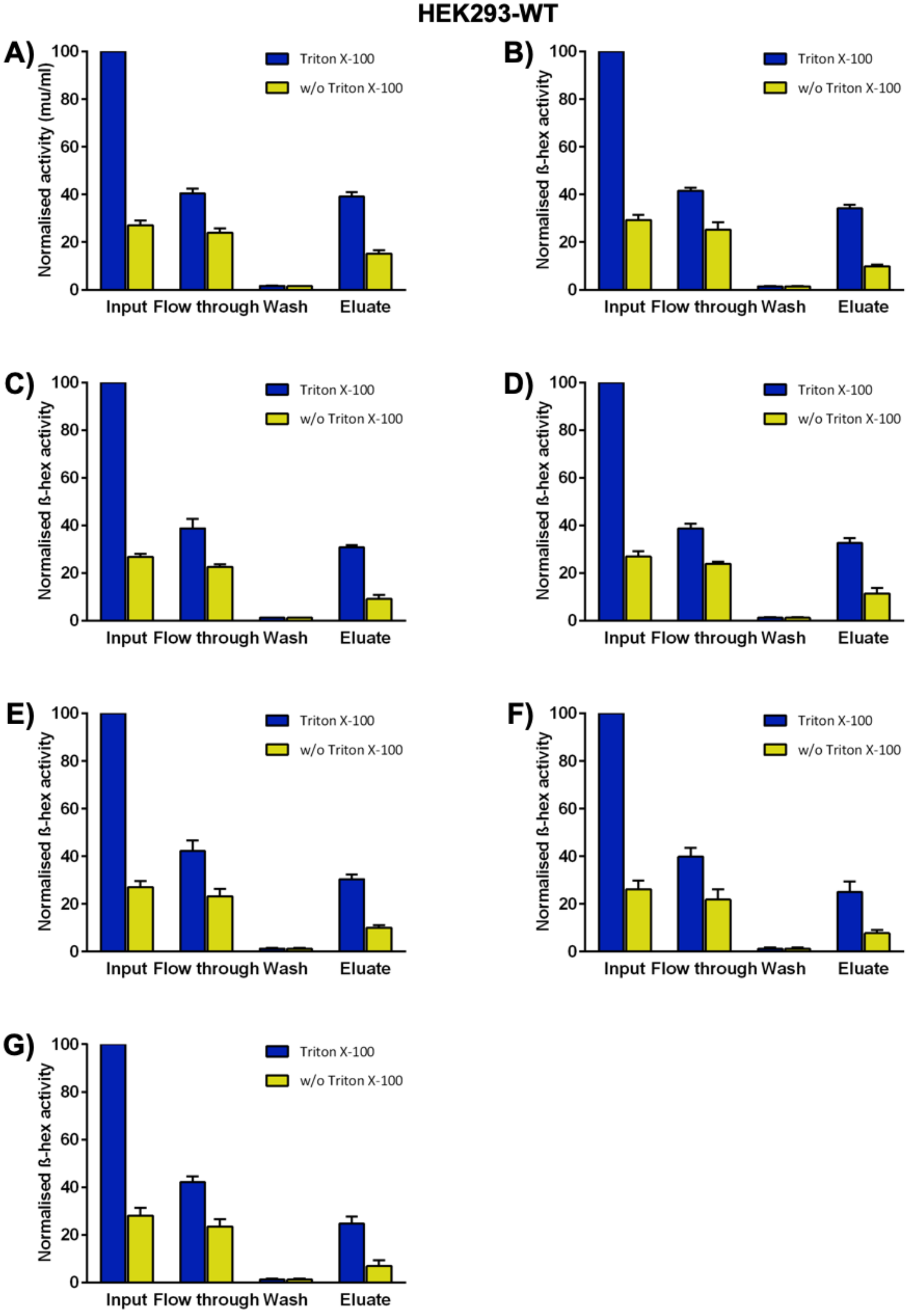


Figure I: Normalized β-hexosaminidase activity assay for monitoring lysosomal enrichment and intact status (HEK293-WT). Shown are samples incubated with **A)** 0h, **B)** 1h, **C)** 2h, **D)** 4h, **E)** 8h, **F)** 16h, **G)** 32h medium containing amino acids labelled with heavy isotopes. Samples treated with Triton X-100 representing broken lysosomes and without representing intact lysosomes. Difference between both fractions represents the intact ratio. Shown are average values (n = 3) and standard deviation. WT: wild-type.

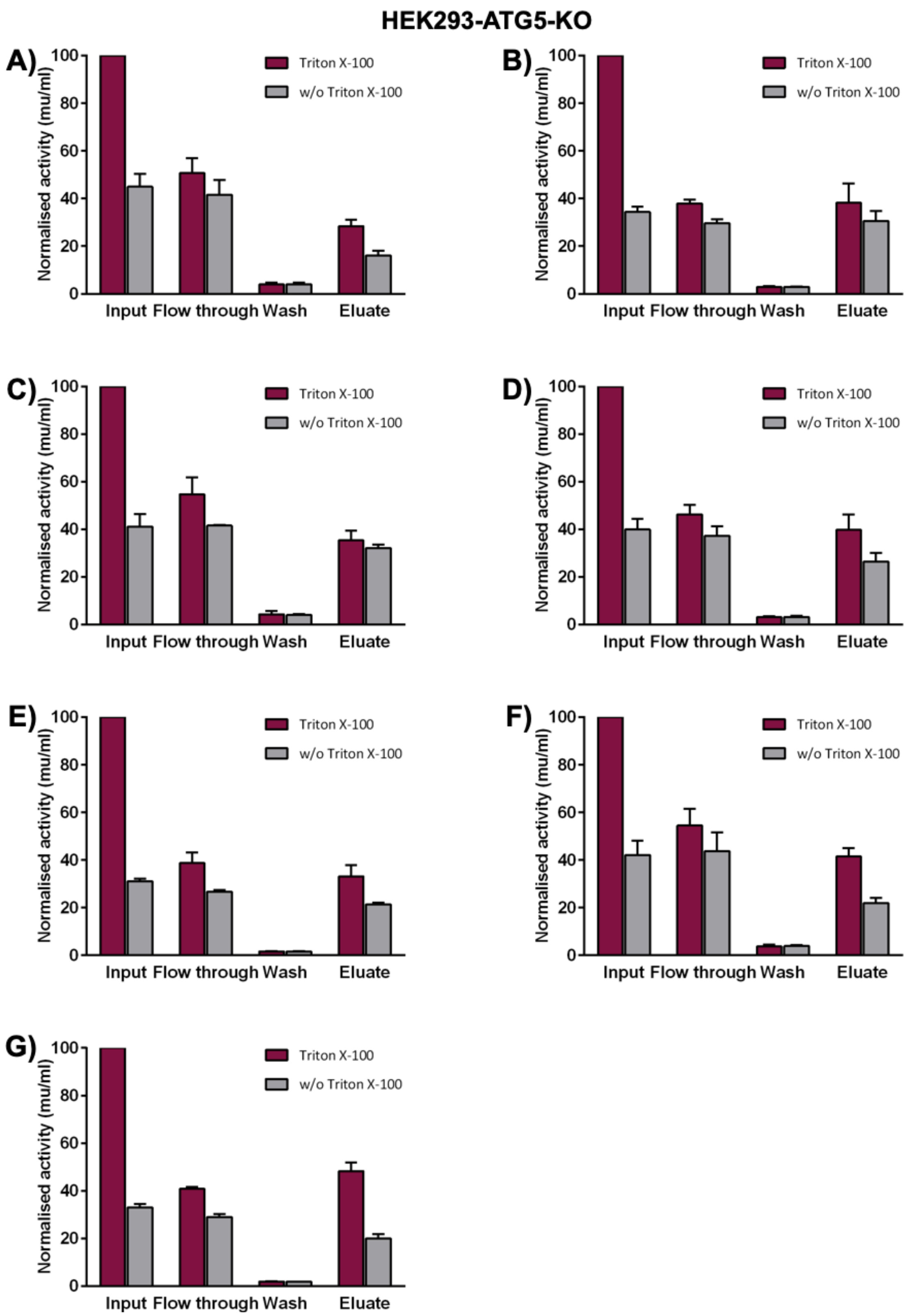


Figure II: Normalized β-hexosaminidase activity assay for monitoring lysosomal enrichment and intact status (HEK293-ATG5-KO). Shown are samples incubated with **A)** 0h, **B)** 1h, **C)** 2h, **D)** 4h, **E)** 8h, **F)** 16h, **G)** 32 h medium containing amino acids labelled with heavy isotopes. Samples treated with Triton X-100 representing broken lysosomes and without representing intact lysosomes. Difference between both fractions represents the intact ratio. Shown are average values (n = 3) and standard deviation. ATG5-KO: autophagy related gene knock out 5.

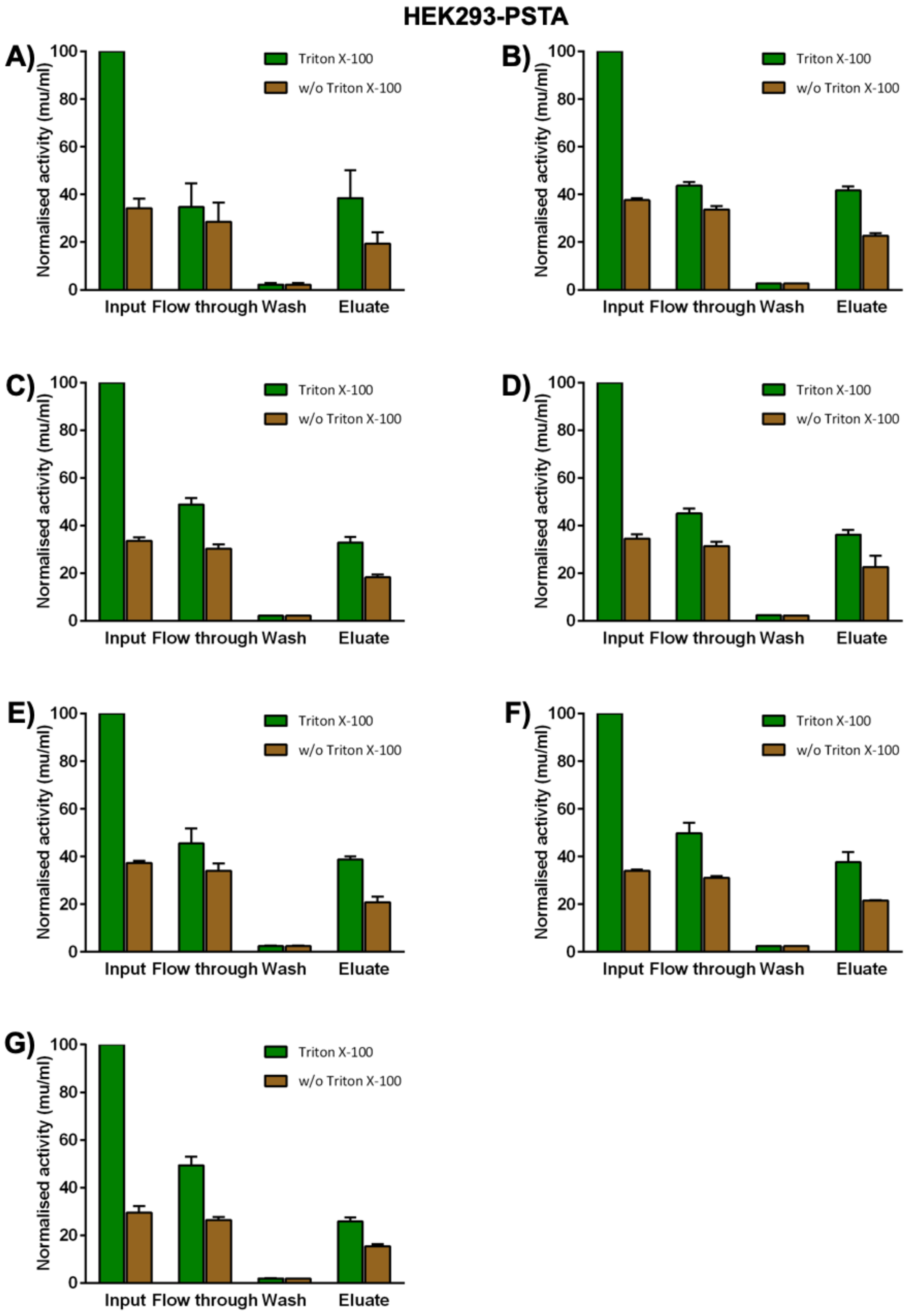


Figure III: Normalized β-hexosaminidase activity assay for monitoring lysosomal enrichment and intact status (Pepstatin A treatment). Shown are samples incubated with **A)** 0h, **B)** 1h, **C)** 2h, **D)** 4h, **E)** 8h, **F)** 16h, **G)** 32h medium containing amino acids labelled with heavy isotopes. Samples treated with Triton X-100 representing broken lysosomes and without representing intact lysosomes. Difference between both fractions represents the intact ratio. Shown are average values (n = 3) and standard deviation. PSTA: Peptstatin A.

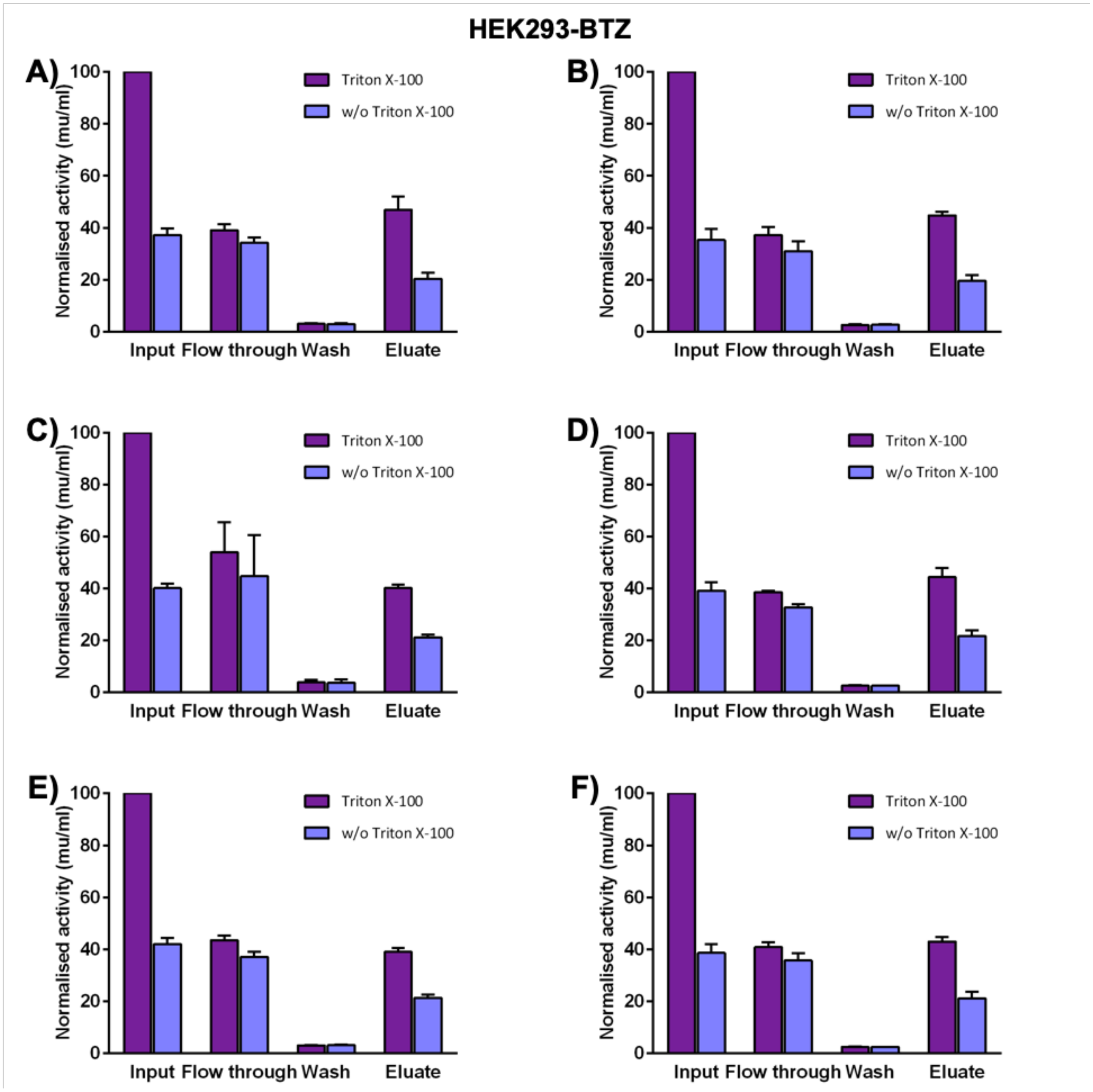


Figure IV: Normalized β-hexosaminidase activity assay for monitoring lysosomal enrichment and intact status (Bortezomib treatment). Shown are samples incubated with **A)** 0h, **B)** 1h, **C)** 2h, **D)** 4h, **E)** 8h, **F)** 16h, **G)** 32h medium containing amino acids labelled with heavy isotopes. Samples treated with Triton X-100 representing broken lysosomes and without representing intact lysosomes. Difference between both fractions represents the intact ratio. Shown are average values (n = 3) and standard deviation. BTZ: Bortezomib.

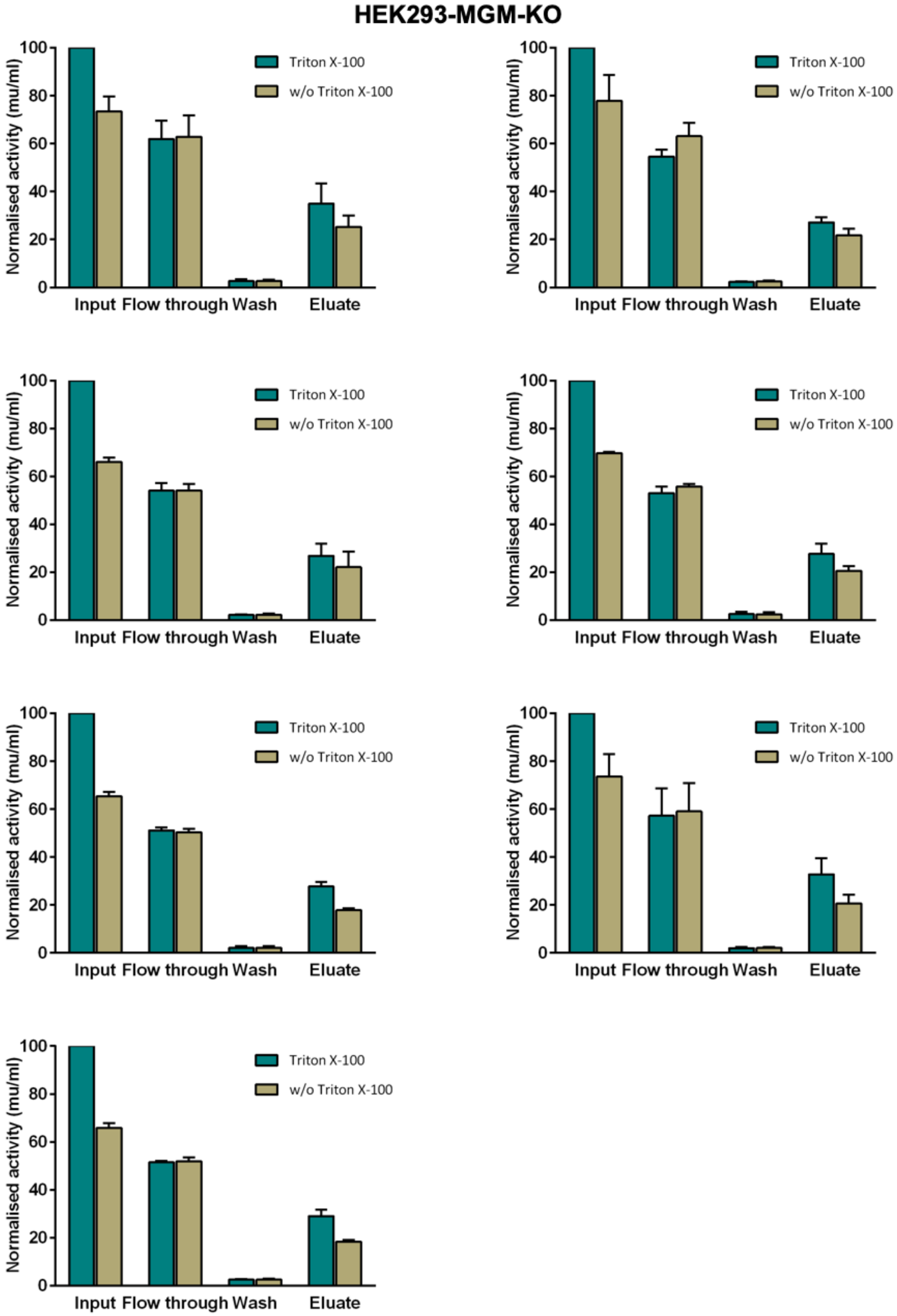


Figure V: Normalized β-hexosaminidase activity assay for monitoring lysosomal enrichment and intact status (HEK293-MGM-KO). Shown are samples incubated with **A)** 0h, **B)** 1h, **C)** 2h, **D)** 4h, **E)** 8h, **F)** 16h, **G)** 32h medium containing amino acids labelled with heavy isotopes. Samples treated with Triton X-100 representing broken lysosomes and without representing intact lysosomes. Difference between both fractions represents the intact ratio. Shown are average values (n = 3) and standard deviation. MGM-KO: multiple glycosylation machinery related genes knock out.

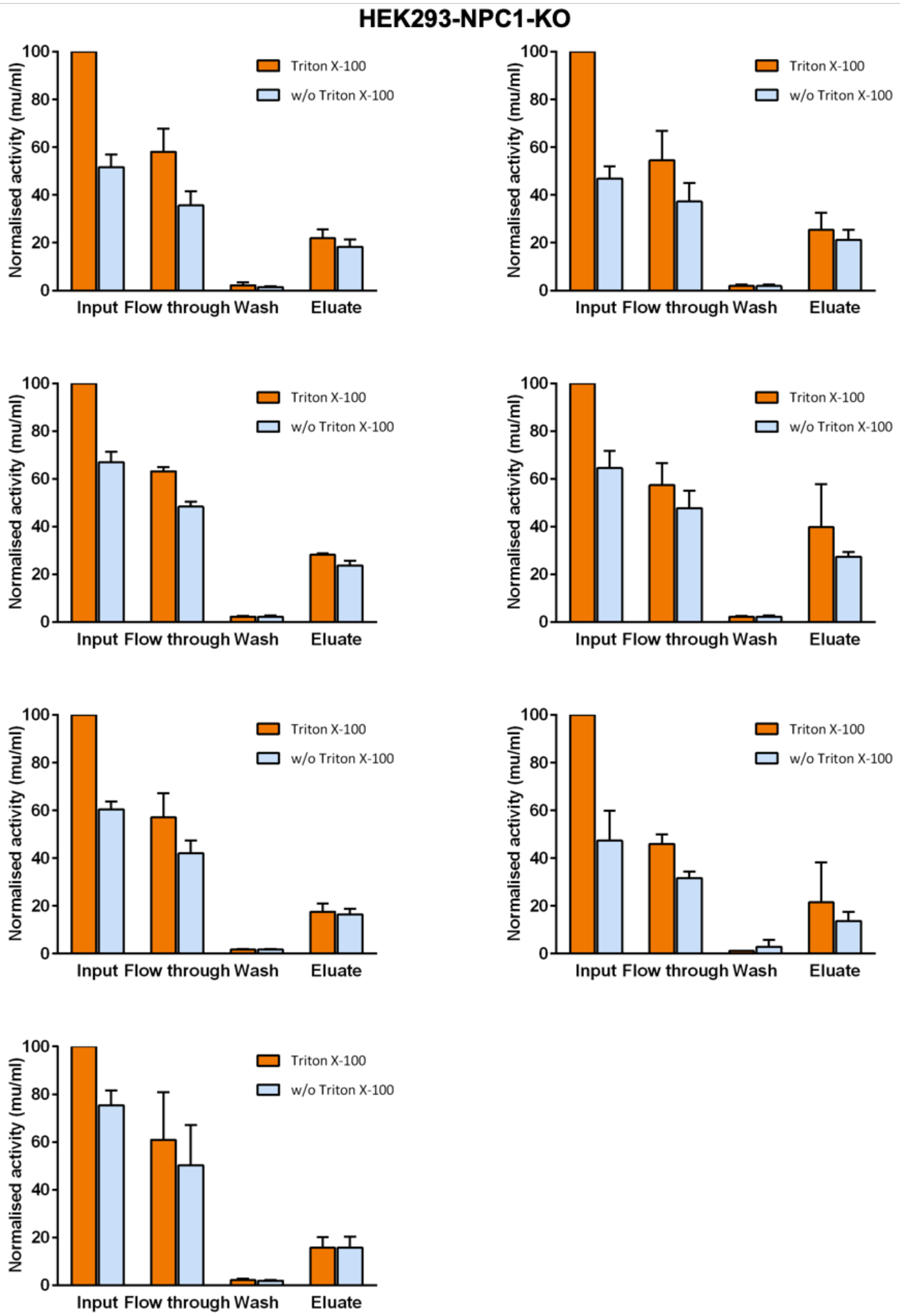


Figure VI: Normalized β-hexosaminidase activity assay for monitoring lysosomal enrichment and intact status (HEK293-NPC1-KO). Shown are samples incubated with **A)** 0h, **B)** 1h, **C)** 2h, **D)** 4h, **E)** 8h, **F)** 16h, **G)** 32h medium containing amino acids labelled with heavy isotopes. Samples treated with Triton X-100 representing broken lysosomes and without representing intact lysosomes. Difference between both fractions represents the intact ratio. Shown are average values (n = 3) and standard deviation. NPC1-KO: Niemann-pick type C1 knock out

Publications

- **Singh, J.** et al. (2022). Cross- linking of the Endolysosomal System Reveals Flotillin Structures and Putative Cargo. bioRxiv
- **Singh, J.,** et al. (2021). Generation of Antibodies Targeting Cleavable Cross-Linkers. *Analytical Chemistry*, 93(8), pp.3762-3769.
- Mosen, P., Sanner, A., **Singh, J.**, & Winter, D. (2021). Targeted Quantification of the Lysosomal Proteome in Complex Samples. *Proteomes*, *9*(1), 4.
- Ponnaiyan, S., Akter, F., **Singh, J.,** & Winter, D. (2020). Comprehensive draft of the mouse embryonic fibroblast lysosomal proteome. *Scientific Data*, 7(1), pp.1-13
- **Singh, J**. et al. (2019). Systematic Comparison of Strategies for the Enrichment of Lysosomes by DIA. *J Proteome Research*, 19(1), pp.371-381
- **Singh, J**. et al. (2016). Toxicological and analytical assessment of e-cigarette refill components on airway epithelia. *Science Progress*, 99(4), pp.351-398
-

Selections of Ph.D. Research on Conferences

- European Study Group on Lysosomal Storage Diseases (**ESGLD-**2019, Spain)
- Deutsche Gesellschaft für Massenspektrometrie (**DGMS**-2020, Germany)
- American Society of Mass-Spectrometry (**ASMS**-2020, USA)
- European Study Group on Lysosomal Storage Diseases (**ESGLD**-2021, UK) - **Award**
- Inaugural International Research Conference on Neurodegenerative Diseases (**IRCND**-2021, USA) – **Award**
- Proteomic Midterm Symposium (**DGPF**-2021, Germany)
- European Proteomics Association (EuPA-2022, Germany) - **Award**