TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Proteomik und Bioanalytik

# STUDYING PEPTIDOFORM-RESOLVED PROTEOME TURNOVER

Jana Zecha

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

| | |
|---|---|
| Vorsitzender: | Prof. Dr. Martin Klingenspor |
| Prüfer der Dissertation: | 1. Prof. Dr. Bernhard Küster |
| | 2. Prof. Dr. Florian Bassermann |
| | 3. Prof. Dr. Matthias Selbach |

Die Dissertation wurde am 05.11.2019 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 09.12.2019 angenommen.

*“Writing a book is an adventure. To begin with, it is a toy and an amusement.*
*Then it becomes a mistress, and then it becomes a master, and then it becomes a tyrant.*
*The last phase is that just as you are about to be reconciled to your servitude,*
*you kill the monster and fling him to the public.”*

**- Winston Churchill**

# ABSTRACT

Cellular proteins exist in a dynamic equilibrium in which they are continuously destroyed and rebuilt. The processes of protein degradation and synthesis determine the rate of protein turnover, and their coordination is fundamental to the regulation of protein abundance and the maintenance of cellular functions. Plenty of research has demonstrated that cancer, neurodegenerative, and age-related diseases are associated with a dysregulation of protein turnover leading to a disruption of protein homeostasis. Further, there is increasing evidence that protein degradation is tightly controlled by post-translational modifications (PTMs). Hence, there is a growing need for methods that facilitate the global investigation of protein turnover and stability in a modification-specific manner.

Today, mass spectrometry-based, bottom-up proteomics technologies allow for the assessment of turnover characteristics for thousands of endogenous proteins in parallel. However, isoform- and modification-specific proteome turnover has so far been mostly neglected due to underlying challenges of the robust quantification of single peptides across multiple conditions. Here, the power of recent technological improvements was harnessed to decipher proteoform-specific protein turnover, and investigate the impact of PTMs on protein turnover for the first time on a proteome-wide scale.

Initially, label-free and tandem mass tag (TMT)-based quantification approaches were systematically assessed for their suitability to quantify specifically single phosphorylated peptides reproducibly and accurately across multiple conditions. MS3-based TMT quantification was considered most expedient for such analyses, and the corresponding data acquisition method was further optimized regarding identification quality and quantification accuracy of phosphopeptides. In addition, an improved TMT labelling protocol was established that reduced the required quantity of expensive labelling reagent by a factor of eight and still achieved complete labelling.

Following method optimizations, the merits of combining pulsed SILAC (stable isotope labelling by amino acids in cell culture) with TMT labelling for the determination of endogenous turnover were evaluated. This novel approach provided enhanced human proteome coverage (>7,000 proteins) and single peptide resolution. The obtained comprehensive data facilitated a global evaluation of the impact of protein properties and functions on their half-life under steady-state conditions and revealed a potential role of oxidative stress in the regulation of protein turnover within the respiratory chain complex I. Moreover, the significance of peptide-resolved turnover measurements was illustrated by several examples of splice variants and post-translationally processed proteins with significantly different turnover.

The combination of the pulsed SILAC and pulsed SILAC-TMT approaches with enrichment methods for acetylated, ubiquitinated, and phosphorylated peptides further enabled the direct measurement of modification-regulated turnover for >30,000 modified sites and disclosed an under-appreciated level of post-translational regulation of protein complexes and metabolic enzymes. Furthermore, an unexpected high fraction of acetylation sites with considerably decreased turnover was observed. Integration with ubiquitin counterpart sites and drug treatment data suggested that part of these sites indicated a protein stabilization, for example by precluding ubiquitination, but another fraction reflected an imbalance in acetylation kinetics caused by the lack of eraser enzyme activity. This demonstrates that differential turnover of modified peptidoforms can uncover cellular mechanisms beyond the stabilization or destabilization of proteins and indicates the tremendous potential of this approach to identify functional relevant modification sites.

In summary, this work provides guidance on how to achieve robust turnover estimates for peptidoforms meanwhile pointing out and tackling challenges in the measurement and data analysis, and compiles the first global atlas of modification-resolved peptide turnover.

# ZUSAMMENFASSUNG

Zelluläre Proteine existieren in einem dynamischen Gleichgewicht, in dem sie kontinuierlich abgebaut und neu synthetisiert werden. Die Prozesse des Proteinabbaus und der -synthese bestimmen die Rate des Proteinumsatzes und ihre Koordination ist von grundlegender Bedeutung für die Regulierung der Proteinabundanz und die Aufrechterhaltung zellulärer Funktionen. Viele Studien haben gezeigt, dass Krebs, neurodegenerative und altersbedingte Erkrankungen mit einer Dysregulation des Proteinumsatzes verbunden sind, die zu einer Störung der Proteinhomöostase führt. Darüber hinaus gibt es immer mehr Hinweise darauf, dass der Proteinabbau durch posttranslationale Modifikationen (PTMs) streng geregelt ist. Daher besteht ein wachsender Bedarf an Methoden, die die globale Untersuchung des Proteinumsatzes und der Proteinstabilität in modifizierungsspezifischer Weise erlauben.

Heute ermöglichen massenspektrometriebasierte Bottom-up-Proteomik-Technologien die Bewertung des Umsatzes für Tausende endogene Proteine gleichzeitig. Der isoform- und modifizierungsspezifische Proteomumsatz wurde jedoch bisher weitgehend nicht untersucht aufgrund der Herausforderungen für die robuste Quantifizierung einzelner Peptide über mehrere Bedingungen hinweg. In dieser Arbeit wurde nun das Potenzial jüngster technologischer Verbesserungen ausgeschöpft, um proteoformspezifischen Proteinumsatzes nachzuweisen und erstmals proteomweit die Auswirkungen von PTMs auf den Proteinumsatz zu untersuchen.

Zunächst wurden markierungsfreie und Tandem-Mass-Tag (TMT)-basierte Methoden systematisch auf ihre Eignung hin bewertet, einzelne phosphorylierte Peptide reproduzierbar und genau über mehrere Bedingungen hinweg zu quantifizieren. Die MS3-basierte TMT-Quantifizierung wurde für solche Analysen als zweckdienlichste angesehen. Die entsprechende Messmethode wurde weiterhin optimiert hinsichtlich der Identifikationsqualität und Quantifizierungsgenauigkeit von Phosphopeptiden. Darüber hinaus wurde ein verbessertes TMT-Markierungsprotokoll erstellt, das die erforderliche Menge an teurem TMT-Reagenz um einen Faktor acht reduzierte und dennoch eine vollständige Markierung erreichen konnte.

Im Anschluss wurde der Vorteil einer Verbindung von pulsed SILAC (stable isotope labelling by amino-acids in cell culture) mit der TMT-Markierung zur Bestimmung des endogenen Proteinumsatzes untersucht. Diese neuartige Methodik sorgte für eine verbesserte Abdeckung des humanen Proteoms (>7.000 Proteine) und einer Messauflösung auf Ebene einzelner Peptide. Die gewonnenen Daten ermöglichten eine globale Untersuchung der Auswirkungen von Proteineigenschaften und -funktionen auf ihre Halbwertszeit und legten eine mögliche Rolle von oxidativem Stress bei der Regulierung des Proteinumsatzes innerhalb des Atemkettenkomplex I nahe. Darüber hinaus wurde die Bedeutung peptidaufgelöster Umsatzmessungen veranschaulicht anhand mehrerer Beispiele von Spleißvarianten und posttranslational prozessierten Proteinen mit signifikant unterschiedlichem Umsatz.

Die Kombination der pulsed SILAC und pulsed SILAC-TMT-Methodik mit Anreicherungsmethoden für acetylierte, ubiquitinierte und phosphorylierte Peptide ermöglichte ferner die direkte Messung des Umsatzes von Peptiden mit >30.000 Modifikationsstellen und legte ein unterschätztes Level an posttransla-tionaler Regulierung von Proteinkomplexen und metabolischen Enzymen nahe. Darüber hinaus wurde ein unerwartet hoher Anteil an Acetylierungsstellen mit deutlich verringertem Umsatz beobachtet. Eine nähere Untersuchung deutete darauf hin, dass ein Teil dieser Modifikations-stellen auf eine Proteinstabilisierung hindeutete, z. B. durch die Verhinderung von Ubiquiti-nierung. Eine andere Fraktion spiegelte allerdings ein Ungleichgewicht in der Acetylierungskinetik hervorgerufen durch fehlende Deacetylierungs-Enzymaktivität wider. Das verdeutlicht, dass ein unterschiedlicher Umsatz modifizierter Peptidoformen zelluläre Mechanismen jenseits der Stabilisierung oder

Destabilisierung von Proteinen aufdecken kann und zeigt das enorme Potenzial dieses Methode, funktionell relevante Modifikationsstellen zu identifizieren.

Zusammenfassend lässt sich sagen, dass diese Arbeit Leitlinien für eine robuste Analyse des Peptidoformumsatzes bereitstellt und dabei Herausforderungen in der Messung und Datenanalyse aufzeigt und bewältigt. Zum Abschluss wird erstmalig ein globaler Atlas des modifizierungsaufgelösten Peptidumsatzes erstellt.

# TABLE OF CONTENTS

Abstract ..... i

Zusammenfassung ..... iii

Table of Contents ..... vii

**I GENERAL INTRODUCTION ..... 1**

**II GENERAL METHODS ..... 37**

**III A CASE FOR ROBUST PEPTIDE QUANTIFICATION ..... 61**

**IV TOWARDS PEPTIDOFORM-RESOLVED TURNOVER ESTIMATES ..... 99**

**V DECIPHERING MODIFICATION-SPECIFIC PROTEIN TURNOVER ..... 129**

**VI GENERAL DISCUSSION AND FUTURE PERSPECTIVES ..... 161**

Bibliography ..... I

List of Abbreviations ..... XXIII

List of Figures ..... XXV

List of Tables ..... XXVII

Appendix ..... XXIX

Danksagungen ..... LIII

List of Publications ..... LV

*“My definition of an expert in any field is a person who knows enough about what’s really going on to be scared.”*

**- P.J. Plauger**

# I GENERAL INTRODUCTION

**1 Expanding life's portfolio ..... 3**

1.1 Gene- and transcript-encoded variations ..... 3

1.2 Co- and post-translational protein modifications ..... 4

1.3 Functional significance of proteoforms ..... 8

**2 Linking protein homeostasis and turnover ..... 11**

2.1 Regulation of protein degradation ..... 11

2.2 Critical maintenance of proteostasis ..... 13

2.3 Methodologies to quantify protein turnover ..... 15

**3 Analysing peptidoforms with bottom-up proteomics technologies ..... 19**

3.1 Sample preparation steps to enhance peptidoform coverage ..... 20

3.2 Principles of tandem mass spectrometry measurements ..... 22

3.3 Analysis of mass spectrometry data ..... 29

**4 Objective and outline ..... 35**

# 1 Expanding life's portfolio

Proteins participate in the control and execution of virtually every process that is involved in the perpetuation of cellular homeostasis thereby defining the functional state of a cell. As stated in the central dogma of molecular biology [1], the blueprint of proteins is encoded as genes in the DNA (deoxyribonucleic acid). The genetic information is transcribed into mRNA (messenger ribonucleic acid) before this is translated into the amino acid sequence of proteins (Figure I-1). In spite of a common genetic basis, cells and tissues of organisms feature a wide range of physiological diversity not least determined and regulated by underlying differences in gene expression patterns [2, 3]. It had been assumed that this extensive, cellular heterogeneity is based on a large number of more than 100,000 human genes [4]. However, the number of estimated genes has dropped ever since the first results of the human genome project were published [5-7]. By now, depending on the database, 19,900 to 21,300 human genes are reported to serve as blueprint for proteins [8]. Likely, the number of genes will yet increase slightly again since small open reading frames (smORFs) of less than 300 nucleotides (100 amino acids) are so far underrepresented due to limits that had to be set for genome annotation to reduce false positives [9, 10]. Nevertheless, the number of genes appears small considering the extraordinary structural and functional diversity of highly specialized cells, tissues, and organs. This suggests that additional important layers of regulated information gain from the gene to protein level exist to enable the development and maintenance of complex living organism. Indeed, different sources of gene-encoded, post-transcriptional, co- and post-translational protein variation are known to shape the complexity of the total protein complement of a genome, commonly termed proteome (introduced in [11]). Such mechanisms go far beyond errors that may occur during transcription and translation (for an overview see [12-16], Figure I-1), and they jointly determine the number of so-called proteoforms ("*different molecular forms in which a protein product of a single gene can be found"* [17]).

## 1.1 Gene- and transcript-encoded variations

**Gene-level variation** – Protein diversity can be encoded at the gene locus itself via somatic mutations [18] or non-synonymous single-nucleotide polymorphisms (SNPs) resulting in single-amino acid polymorphisms (SAPs) [19]. Those determine intra-individual and population (inter-individual) variations in protein sequences, respectively. The somatic, recombinatorial rearrangement of gene regions represents another, unique source of protein alteration. It takes place during the development of bone marrow-derived lymphocytes and provides a basis for antibody and T-cell receptor variety during adaptive immune response [20, 21].

**RNA-level variation** – The most significant contribution to protein diversity before translation, however, occurs at the RNA level. In the late 1970s, the observation of "split genes" and the subsequent discovery of alternative splicing [22, 23] questioned the dogma that one gene encodes one protein (originally described as the one-gene-one-enzyme [24], later expanded to the one-gene-one-polypeptide hypothesis [25]). It became clear that precursor mRNA consists of introns (intragenic, non-coding regions) and exons (expressed regions) [26]. The latter can be spliced together in various combinations to produce mature mRNAs serving as several templates for different proteins derived from the identical gene sequence (Figure I-1). The process of splice variant generation is typically catalysed and regulated by the spliceosome, a ribonucleoprotein complex [27]. The implementation of the RNAseq technology in 2008 [28] enabled the genome-wide, systematic study of splicing events and revealed an unexpectedly high degree of alternative splicing which affected >90 % of human genes [29, 30]. It has been argued that this large number could

also be driven by erroneous splice site choice leading to a multitude of low abundant, likely non-functional splice variants [31]. Yet, in ~86 % of the cases, the minor (non-canonical) splice variants were detected at a frequency of not less than 15 % [29] suggesting at least some biological significance. Transcription itself can also be initiated and terminated at varying sites, which often results in distinct 5' terminal exons and can subsequently change splicing patterns and ultimately protein N-termini [32, 33]. Moreover, RNA editing (a term coined by Benne and colleagues [34]) comprises enzymatically regulated nucleotide conversions, insertions, and deletions in mRNAs and can lead to new start or stop codons, coding frame shifts or SAPs (reviewed in [35, 36]). In humans, adenosine-to-inosine conversions seem to occur most commonly [37], but so far less than 10 % of these modifications have been reported to be translated into actual amino acid substitutions [38].

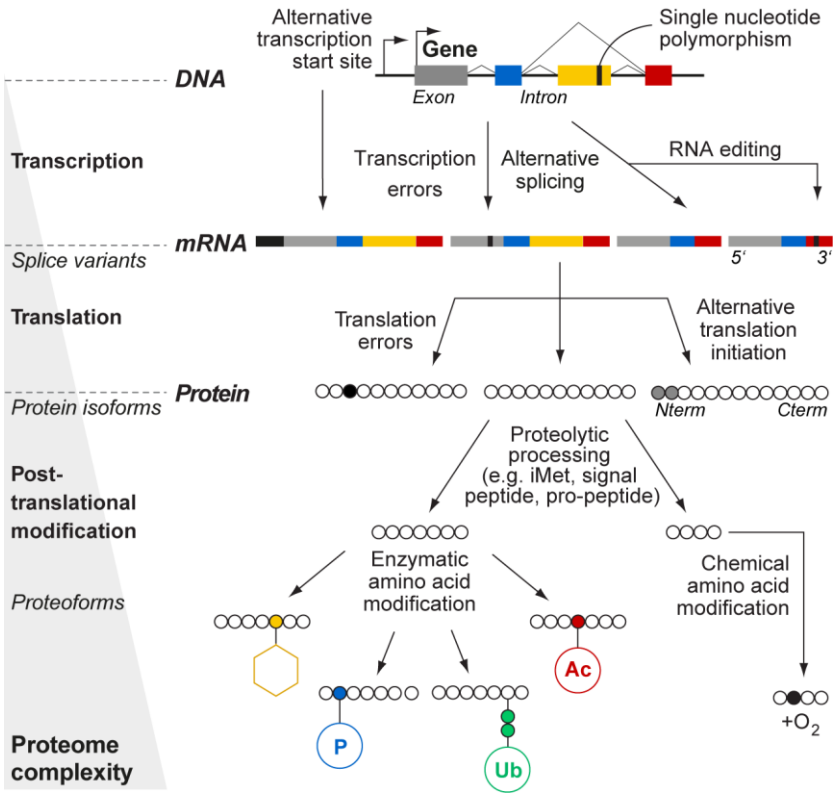



**Figure I-1 | Sources of proteome complexity.** *While non-synonymous genetic polymorphisms that cause amino acid substitutions produce inter-individual variation, transcription errors, alternative transcription initiation, and mRNA processing (alternative splicing, RNA editing) can fuel protein diversity within single cells. This is additionally augmented by potential translation errors and alternative translation initiation sites. Enzymatic and non-enzymatic post-translational modifications like proteolytic cleavages, additions of functional moieties and chemical amino acid modifications further amplify the number of potential protein variants (proteoforms) derived from the same gene locus and contribute to the generation of complex proteomes (Nterm: protein N-terminus; Cterm: protein C-terminus; iMet: initiator methionine; P: phosphorylation, Ub: ubiquitination, Ac: acetylation; adapted with permission from [15]).*

## 1.2 Co- and post-translational protein modifications

**Translation-level variation** – The protein synthesis process at the ribosome, where triplets of mRNA nucleotides are translated into amino acids with the help of transfer RNAs, is another source of proteome complexity. Alternative translation initiation sites (TIS) can produce N-terminal extended and truncated proteins or, in the case of out-of-frame start codons, totally new proteins (outlined in [39]). Sequence analyses have shown that more than a third of human transcripts have AUG start codons in the 5' untranslated region of the annotated coding sequence [40]. Indeed, ribosome profiling revealed that almost half of the transcripts have multiple TISs, frequently also employing non-AUG start codons [41] which complicates a computational annotation of such translation initiation sites. In line with the linear scanning mode of the ribosome, most alternative TIS produce N-terminal extended protein isoforms from start codons upstream of the annotated TIS [13, 41]. However, in non-optimal sequence environments [42], ribosomes can also pass over

start codons (known as leaky scanning), initiate translation downstream of the annotated start codon, and produce shorter protein variants [39].

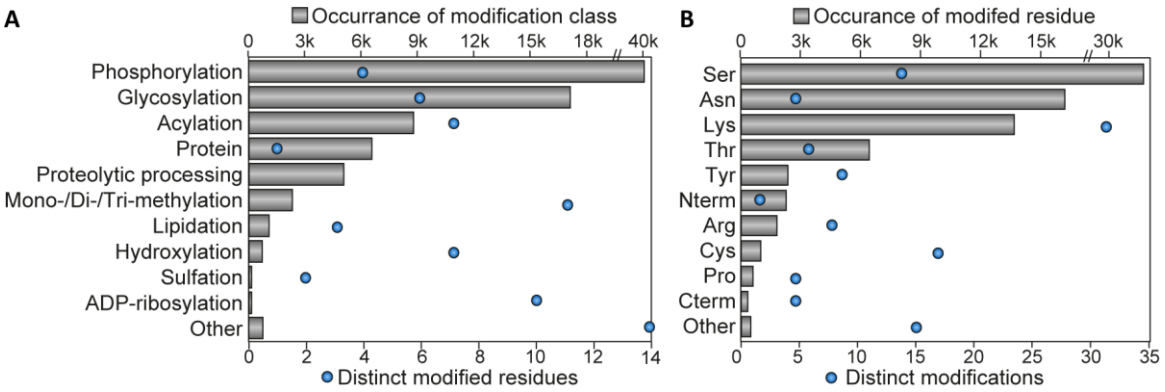



**Figure I-2 | Frequencies of curated post-translational modifications and types of modified residues of human proteins from the Swiss-Prot database** [43]**.** *PTMs were extracted for the 20,431 canonical, human protein sequences in the curated Swiss-Prot database. (A) Phosphorylation is by far the most frequently annotated PTM, but the broadest spectrum of residues can be modified by methylation. 'Glycosylation' comprises N-, O-, and S-linked glycosylation. 'Acylation' includes the most prominent and frequent class member acetylation, and additionally comprehends all common acyl-modifications with chain lengths from one (formylation) to five carbon atoms (glutarylation), and biotinylation. ubiquitination, sumoylation, and neddylation are contained in the category 'Protein', and 'Proteolytical processing' comprises annotated cleavages of signal and pro-peptides. (B) Serine is the amino acid that is annotated to be modified most frequently, but lysine residues feature the highest diversity of possible modifications providing a broad basis for competitive crosstalk. Note that here modifications were not grouped into classes like in (A) but e.g. every acylation type was counted separately.*

**Variation through protein modifications** – The last layer of protein information gain is mediated by a multitude of post-translational modifications (PTMs) causing a combinatorial explosion of potential proteoforms (Figure I-1). It has been known for many decades that proteins are being modified in cells, for instance, through phosphorylation [44] and acetylation [45], and that these modifications can be controlled by enzymes [46, 47]. However, the detailed elucidation of the complexity of the modification landscape and its regulation has just begun in recent years. There is a huge variety of cellular protein modifications (~400 according to http://www.unimod.org) including proteolytic processing, enzymatic addition of small functional groups, sugars, fatty acids or even small proteins, and chemical modifications which can also lead to amino acid conversion (reviewed in [14, 48]). Only few of them have been studied extensively, most prominently phosphorylation, glycosylation, acetylation, and ubiquitination (Figure I-2 A). PTMs can take place on different amino acids or protein N- or C-termini (Figure I-2 B), exhibit a simple or very complex structure, be reversible or irreversible in nature, and occur co- or post-translationally and enzymatically or non-enzymatically. Enzymes that catalyse PTMs account for more than 5 % of the genome in higher eukaryotes [48] evidencing the functional importance of these modification processes. In humans, there are 500 proteases alone which are critically involved in the maturation or degradation of proteins by spatiotemporally regulated proteolysis. A prominent example is the co-translational removal of the initiator methionine (iMet) on proteins by methionine aminopeptidases which is often followed by N-terminal acylation or alkylation [49]. Other proteolytic modifications include the removal of propeptides which is frequently associated with the activation of enzymes [50] and hormones [51] and the cleavage of signals peptides which dictate the extracellular or intracellular organelle localization of proteins [52]. In addition to such irreversible processing events, many reversible post-translational modifications shape proteoform diversity.

Those are often subject to an evolutionary conserved writer-reader-eraser principle where proteins with enzymatic activity add modifications enabling proteins with certain binding domains to interact with modified residues in a specific manner before other enzymes remove the modification again. Examples for such dynamic PTMs are phosphorylation, acetylation, and ubiquitination, which are outlined in the following.

**Phosphorylation** – More than 500 human protein kinases (writers) [53] and 140 phosphatases (erasers) [54, 55] jointly control cellular signalling cascades via protein phosphorylation of serine, threonine, and tyrosine hydroxyl groups (Figure I-3 A). Interestingly, less than 1/5th of kinases, but nearly 2/3rd of phosphatases act on tyrosine residues [53, 54]. On the other hand, almost half of tyrosine phosphatases also hydrolyse phosphorylated serine and threonine residues, whereas protein kinases generally seem to possess selectivity towards either serine/threonine or tyrosine [54]. Several attempts have been made to systematically identify substrates of phospho-writer and -eraser enzymes and in some cases consensus motifs have been identified (reviewed in [56, 57]), but many kinase/phosphatase-substrate relations still remain largely elusive in particular in vivo. The phosphate group transferred in the kinase reaction originates from the phosphate anhydride ATP (adenosine triphosphate) and changes neutral hydroxyl into di-anionic functional groups (Figure I-3 A). Those feature larger hydrated shells and are capable of building stronger hydrogen bonds and salt bridges than the acidic amino acids aspartate and glutamate [58]. Consequently, phosphorylation often entails a change in protein conformation and states but also enables the selective interaction with several 100s of distinct reader-proteins via phosphorylation-specific binding domains (for an overview see [59, 60]). SH2, SH3 (Src-homology-2/3), and PTB (phosphotyrosine-binding) domain-containing proteins, for instance, bind to phosphorylated tyrosine, while FHA (forkhead-associated) domains interact with phospho-threonine and 14-3-3 domains with phosphorylated serine and threonine (Figure I-3 A). It has been assumed that more than 70 % of cellular proteins are being phosphorylated at some point [61] illustrating the broad involvement of this modification in cell signalling. Mostly phospho-serine and -threonine sites (84 and 13-15 %) are being detected in phosphoproteomic studies due to their probably higher frequency and stoichiometry. In contrast, tyrosine phosphorylation usually only accounts for a small fraction in the phosphoproteome (<1-3 %) [62, 63]. This observation has been suggested to be caused by a higher degree of specificity for phospho-tyrosine controlled signalling [64]. Recently, the identification of non-canonical phosphorylation on histidine, aspartate, glutamate, lysine, arginine, and cysteine via mass spectrometry has been reported in a human cells [65]. Although non-canonical phosphorylation has been shown to occur in mammalian cells previously [66], these observations are being highly debated due to specific challenges in phosphoproteomics that can under certain circumstances easily lead to mislocalization of phosphosites (see also pp. 29).

**Acetylation** – As mentioned above, α-amino groups of protein N-termini can undergo acylation which most commonly emerges as acetylation and is estimated to affect 80 to 90 % of cellular proteins irrespective of iMet cleavage [49, 67]. The modification is catalysed by N-terminal acetyltransferases (NATs) which feature diverse specificities and utilize acetyl coenzyme A (ac-CoA) as an acetyl donor. So far, no Nα-acetylation-erasing enzyme has been identified for which reason this modification is considered irreversible [49, 68]. In contrast, the second type of N-acetylation is invertible and occurs on ε-amino group of lysine residues neutralizing the positive charge of their side chains. Lysine acetylation has first been identified on histones in the 1960s [45] and soon afterwards its involvement in the regulation of chromatin structure and transcriptional activation was suggested [69]. This dominated the research on acetylation and its writer and eraser enzymes as transcriptional regulators for decades. Only within recent years, non-histone acetylation and

its functional relevance for a multitude of cellular processes has gained attention [70]. The transfer of acetyl groups from ac-CoA onto the ε-amino group of lysine residues is catalysed by histone/lysine acetyl transferases (HATs/KATs) and reversed by deacetylases (HDACs/KDACs, Figure I-3 B). Thirteen functionally and structurally diverse KATs have been well characterized, but the existence of additional ones has been proposed (for a detailed overview see [68, 70]). Similarly, the number of proteins with potential KDAC activity is growing. The 18 commonly accepted members of the KDACs family are distinguished into zinc2+-dependent KDACs and NAD+ (nicotinamide adenine dinucleotide)-dependent sirtuins. They exhibit different subcellular localizations and are often part of multi-enzyme complexes. Substrate specificities of KATs and KDACs are largely unknown, but considering the comparably small number of identified enzymes it is assumed that substrates are mainly determined by localization and protein interaction partners instead of substrate sequence motifs [70]. Importantly, it has been shown decades ago that protein acetylation can also occur non-enzymatically in the presence of ac-CoA [71], but only recently a link to energy metabolism has been made [72, 73]. It is assumed that non-enzymatic acetylation predominantly takes place in mitochondria favoured by elevated ac-CoA concentrations and a slightly higher pH compared to other cell compartments [73]. Lysine acetylation is recognized by less than 50 bromodomain (BRD)-containing proteins [74] which appears to be a small number compared to phosphorylation-reader proteins. Recently a new recognition domain for acetylated lysine (YEATS) has been identified [75] suggesting that additional acetyl-reader domains so far may have been evading discovery. Importantly, many acetyl-lysine binding proteins are members of large protein complexes or have catalytic activities themselves, which lays the foundation for acetylation signal transmission and integration [74].

**Ubiquitination** – The formation of an (iso)peptide bond between the C-terminus of the 76 amino acid small protein ubiquitin (Ub) and the ε-amino group of a lysine (rarely the N-terminal α-amino group) of a target protein is catalysed by at least three enzymes (reviewed in [76, 77], Figure I-3 C). Initially, Ub is activated under consumption of ATP and covalently linked to a cysteine within the E1 (Ub-activating) enzyme via a high-energy thioester bond. Next, the Ub is transferred to a second cysteine in the active site of the E2 (Ub-conjugating) enzyme in a trans-thiolation reaction. Finally, Ub is either directly ligated to lysine residues of the target protein aided by the E3 (Ub-ligating) enzyme or first covalently attached to the E3 ligases and then transferred to the substrate. The number of proteins in the three different classes increases in each reaction step providing the basis for a universally applicable and still specific attachment of ubiquitin to target proteins. In humans there are 2 Ub-specific E1, ~40 E2, and more than 600 E3 enzymes [77]. The latter class of enzymes determines the specificity and is divided into three mechanistic categories, namely RING (really interesting gene)/Ubox, HECT (homologous to E6-associated protein C terminus), and RBR (RING-between-RING) ligases. Mono-ubiquitinated proteins can be further modified by attachment of additional Ub molecules on one of the eight amino groups of the already linked Ub. This results in a myriad of potential poly-Ub assemblies with various chain lengths and linkage types, which have been described to feature differing cellular functions [76]. Most prominently, lysine 48-linked poly-Ub chains with at least 4 Ub molecules typically mark proteins for proteasomal degradation [48]. In many cases, however, the consequences of different poly-ubiquitinations are still not completely elucidated or vary between modified proteins [76]. Ubiquitin-binding domains (UBDs) with differing affinities towards mono- and differentially linked poly-Ub are being identified in a growing number of proteins (> 150) suggesting a still underappreciated involvement of Ub in cell signalling beyond degradation processes [78]. UBDs are structurally as diverse as the modifications they bind to but are broadly categorized into four subclasses, namely

α-helix, zinc finger, plekstrin homology and ubiquitin-conjugating-like domains. Notably, UBDs have also been identified in deubiquitinases (DUBs). Approximately 80 DUBs can either cleave (iso)peptide bonds irrespective of the Ub-linkage type or act in a linkage-specific manner [79]. They erase mono and poly-Ub from substrate proteins and recycle Ub. In addition, they are responsible for releasing Ub from translation products generated from genes which encode multiple, fused ubiquitin sequences or fusions of Ub with ribosomal proteins [79]. Importantly, DUBs can also edit Ub modifications by trimming poly-Ub chains and thereby modify Ub signals.

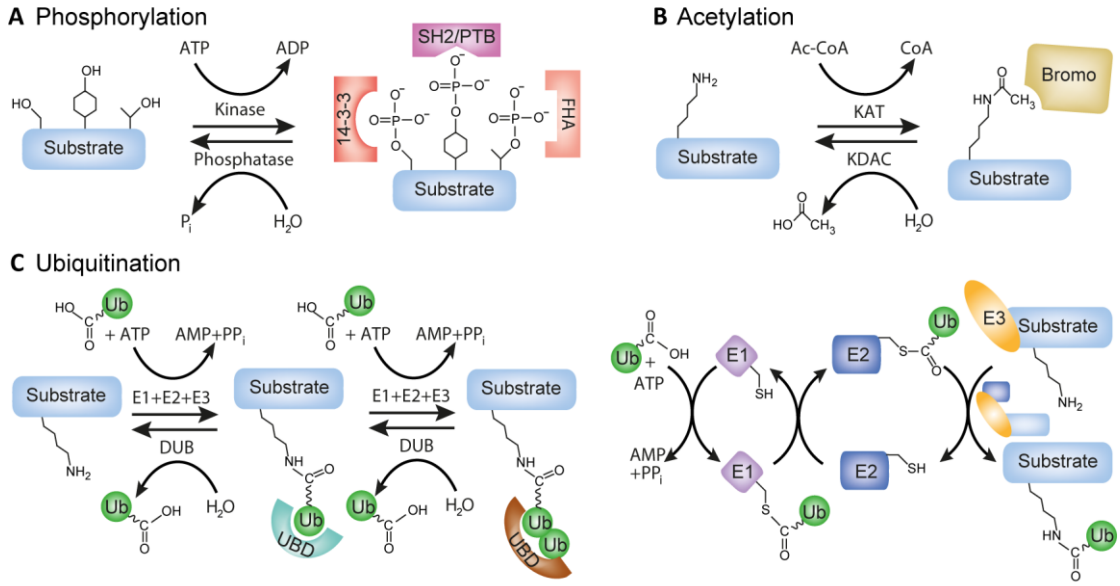



**Figure I-3 | Writer-reader-eraser principle.** *Reversible reactions of protein phosphorylation on serine, tyrosine, and threonine residues (A), acetylation on lysine residues (B), and mono- and poly-ubiquitination on lysine residues (C) are depicted. Writing and erasing enzymes are indicated (KAT: lysine acetyl transferase; KDAC: lysine deacetylase; DUB: deubiquitinase; E1: Ub-activating enzyme; E2: Ub-conjugating enzyme; E3: Ub-ligating enzyme). For acetylation, the reaction catalysed by zinc2+-dependent enzymes is displayed. Examples of PTM-specific binding domains that recognize modified residues are shown (SH2: Src-homology-2; PTB: phosphotyrosine-binding; FHA: forkhead-associated; UBD: Ub-binding domain). For ubiquitination, the 3-step process of the transfer of ubiquitin (Ub) to its substrate is illustrated (adapted with permission from [14]).*

## 1.3 Functional significance of proteoforms

**Constraints on the proteoform inventory** – All the above specified sources of proteome complexity may lead to an exponential increase in the protein portfolio within a biological system. Yet, the number of observable proteoforms lies well below the number of theoretically possible protein variants [80]. As an example, only 74 of ~3 million possible combinatorial modification codes could be detected on the N-terminal tail of histone H4 [81]. There are several potential, technical, and biological explanations for this discrepancy. First, even the most sophisticated technologies available today are likely not sensitive enough to detect all of the expressed but rare proteoforms on a proteome-wide level [80]. Hence, the true biological dimension of proteoform diversity cannot be determined with complete certainty. Second, proteoform complexity is limited simply by the number of expressed proteins, i.e. the maximal possible number of proteoforms derived from a single gene equates the number of its protein copies [16]. Last, and maybe most significantly, there is a high degree of enzymatic control imposed on proteome complexity confining its biological magnitude. Expression of proteoforms arising from transcription and translation errors or al-

ternative splicing and TIS utilization may effectively be diminished by cellular surveillance pathways such as nonsense-mediated mRNA decay or ribosome quality control mechanisms, especially when they exhibit a reduction or loss in function [82, 83]. Moreover, as outlined above, post-translational defined proteome diversity is regulated by the spatiotemporally controlled activity of PTM writers and erasers and governed by the enzymes' specificities and affinities for different substrates. Despite these potential constraints on proteoform expression and detection, there is a large body of experimental evidence for a multitude of structurally diverse protein variants, and many have been linked to altered functionalities of gene products (Figure I-4).

**Biological implications of protein isoforms** – Transcriptome studies have indicated that tissue specific functions are associated with the expression of different splice variants [29, 30] and even stronger with alternative transcription start and termination events [33]. In addition, protein isoforms have been shown to feature vastly different interaction profiles with most isoform pairs sharing less than 50 % of their interaction partners [85]. Importantly, differences across protein isoforms were found to be potentially as big as for proteins encoded by entirely different genes. Likewise, alternative translation initiation has been linked to altered protein localization [86]. As an example, translation of ornithine decarboxylase-antizyme can start from two different TIS. The longer isoform is translated less efficiently and transported to mitochondria, whereas the shorter isoform is missing the mitochondrial transit peptide and thus resides in the cytoplasm [87]. Besides, alternative translation initiation can also lead to protein products with totally different function, as exemplified by the secreted osteogenic growth peptide which is translated from the same mRNA as the histone protein H4 [88]. These examples illustrate the possibility to adapt and broaden the functional repertoire of a living system by tailoring gene expression patterns.

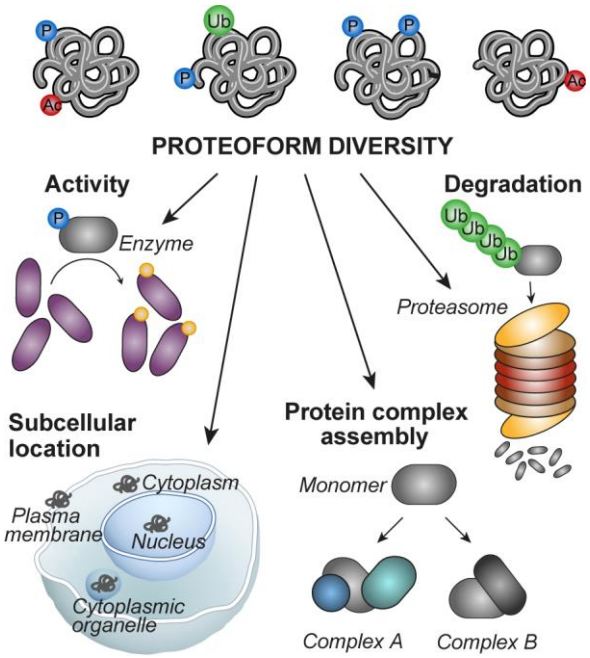



**Figure I-4 | Impact of proteoform diversity on gene function.** *The different sources of proteome complexity illustrated in Figure I-1 produce a multitude of proteoforms which can feature great structural and functional diversity potentially influencing protein activity, localization, interaction partners and stability (adapted with permission from [84]).*

**Regulation of protein function via PTMs** – Post-translational modifications are often reversible and thereby highly dynamic which offers great potential for the flexible and invertible regulation of cellular functions as a response to environmental cues. As molecular switches, reversible PTMs facilitate an initially even more rapid and energy-efficient way to control protein function temporarily than a change in protein synthesis may achieve. Consequently, PTMs play a pivotal role for effectively every cellular process. The modification of amino acid residues in proteins can remove or add charges, alter protein conformations locally or even globally, and establish or mask surfaces for protein binding (reviewed in [14, 48, 89]). Accordingly, protein properties such as interactions with other proteins, activity, localization, or stability can be modulated (Figure I-4). As an example, enzymatic activity and substrate binding of many protein kinases is regulated by

phosphorylation of one or more positions in their so-called activation loop. Upon modification, the loop changes its conformation uncovering the site for substrate binding. Further, the proper orientation of a conserved aspartate within the active site is facilitated which is required for catalysis [90]. This principle mechanism of kinase activation via an upstream protein kinase creates signalling cascades that enable signal amplification and transduction [91]. Similarly, activity of transcription factors is known to be regulated by PTMs. Mono-ubiquitination of FOXO4, for instance, has been described to prompt nuclear translocation of the cytoplasmic transcription factor and induce expression of target genes [92]. This illustrates that a prolonged change in the modification landscape will ultimately pass into modulated gene expression patterns. Another well-studied example of a transcription factor with more than 300 different modification sites is the tumour suppressor p53 [93]. While ubiquitination in the C-terminal domain marks it for degradation, acetylation on the same residues has been suggested to stabilize p53 [94] and increase binding to DNA [95]. This evidences the potential for PTM crosstalk, in this case in a competitive manner by different modifications that occur on the identical amino acid. Lysine offers the greatest opportunity for such competitive crosstalk (Figure I-2 B). Negative and positive crosstalk can also take place sequentially when one modification promotes another or in concert when different modifications emerge independently of each other but in combination induce an altered response compared to their exclusive presence (reviewed in [96]). This enables signal integration and generation of complex structural and regulatory networks that are usually controlled by more than one type of PTM [14]. Concisely, protein modifications allow for a fast adaption of protein functionalities and fine-tuning of cellular signalling. Together with pre-translationally controlled proteome diversity, they provide an incredible plasticity for living organisms.

# 2 Linking protein homeostasis and turnover

Proteome abundance and activity are tightly controlled and continually adjusted in response to internal and external stimuli. To this end, the coordination of protein synthesis and degradation (summarized under the term turnover) are fundamental processes that regulate the abundance of functional proteins and maintain cellular protein homeostasis (proteostasis). Even under steady-state conditions when the overall abundance of cellular proteins and their PTMs remains constant over time, proteins still exist in a dynamic state in which they are continuously destroyed and rebuilt [97]. While this default turnover may appear unfavourable from an energy efficiency point of view, it ensures proper proteostasis by replacing damaged, malfunctioning, or even harmful proteins and contributes to a cell's capability to respond rapidly to a changing environment. Proteome integrity is maintained by a multitude of cellular machineries and pathways many of which were only recently discovered and are still under investigation (reviewed in [83, 98-100]). Protein degradation plays a pivotal role in many of these dynamic responses. In the following, a brief overview of the regulation of cellular degradation with a focus on the proteasome and its contribution to the critical maintenance of proteostasis is given and methods for measurement of protein turnover are outlined.

## 2.1 Regulation of protein degradation

**Degradation machineries** – Spatiotemporal regulated, intracellular protein degradation is enabled by two major systems, the autophagosome-lysosome and the ubiquitin-proteasome system (UPS; reviewed in [101-103]). Autophagy involves the sequestering of cytoplasmic components via membrane vesicles, subsequent fusion with the lysosome, and protein degradation facilitated by more than 60 acid hydrolases [101, 104]. Originally, it was thought to occur in a non-specific manner, but there is increasing evidence for targeted ways of protein clearance via autophagy (outlined in [105]). While autophagy is considered mainly responsible for the removal of protein aggregates, long-lived proteins, and damaged organelles, the UPS is understood as the major route for selective protein degradation in mammalian cells particularly for dysfunctional and regulatory proteins [102].

Protein degradation via the 26S proteasome, a huge multi-subunit protein complex, is strongly energy dependent [106], although peptide hydrolysis is in principle an exergonic reaction. This underpins the high degree of regulation underlying this type of protein breakdown. As described above, the common first step is the attachment of poly-ubiquitin to a substrate. Substrates are then bound to the 19S regulatory subunit of the 26S proteasome either directly via the attached Ubs or indirectly mediated by an Ub- and proteasome-binding adaptor protein [107] (Figure I-5). Additionally, ATP-dependent binding of a loosely folded, degradation-initiation region in the substrate is required [108] before proteasome-associated DUBs hydrolyse UB-chains thereby recycling Ub. This is followed by substrate unfolding via the action of ATPases at the 19S subunit and proteolysis in the 20S catalytic chamber via β1 (caspase-like), β2 (trypsin-like), and β5 (chymotrypsin-like) subunits [103, 109]. Proteasomal degradation results in small polypeptides (95 % within a range of 3 to 22 amino acids [110]) which are further digested into single amino acids by cytosolic peptidases [102].

Diverse alterations of the 26S proteasome itself have been described to influence its substrate preference including a varying composition and modifications of proteasomal subunits (reviewed in [102, 103]). The selectivity of 26S proteasomal degradation is also collectively mediated by the activity and substrate specificity of E3 ligases, the availability of an E3 ligase recognition motif, an

acceptor site for Ub, and a disordered structure for proteasome interaction on the substrate [107] (Figure I-5). On the other hand, also ubiquitin-independent degradation via solitary 20S proteasomes that are missing the 19S regulatory subunit has been observed for intrinsically disordered proteins [111]. Besides, ubiquitination is increasingly recognized as signal for selective autophagosomal-lysosomal degradation [112]. This exemplifies the existent crosstalk and reciprocal regulation between the two proteolytic pathways via shared components and substrates (outlined in [105]). New regulatory mechanisms are being constantly reported which demonstrates that the intricacies of regulation of protein degradation have only just begun to be elucidated.

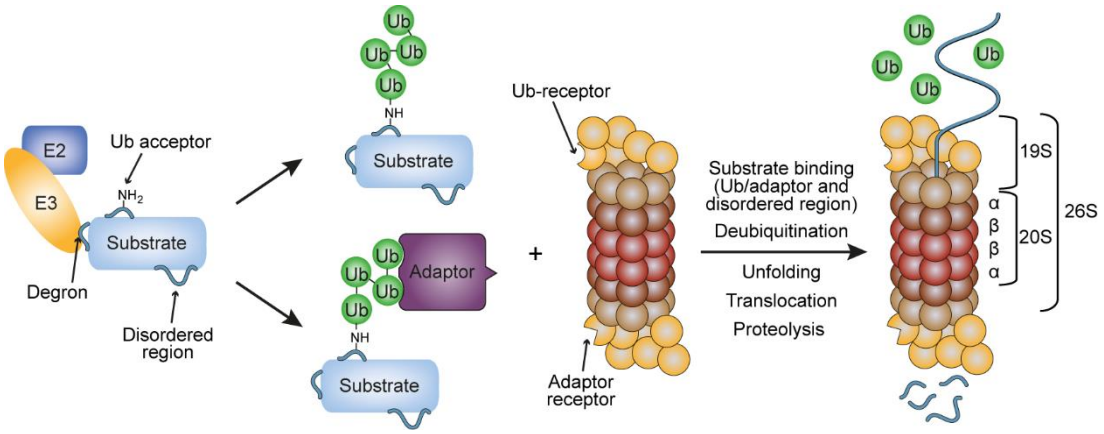



**Figure I-5 | The ubiquitin-proteasome system (UPS).** *UPS substrates must contain an E3 ligase recognition element (also termed degron), an ubiquitin (Ub) acceptor site, and a loosely folded region for interaction with the proteasome. Ubiquitinated substrates initially bind to the 19S regulatory subunit either directly via ubiquitin or indirectly via an adaptor protein. Only upon additional interaction of the disordered region with the proteasome, deubiquitination is initiated. This is followed by unfolding of the substrate, translocation into the 20S catalytic core unit and digestion into small peptides by proteolytically active β-subunits (adapted with permission from [107]).*

**Degrons** – Half-lives of cellular proteins have been determined to range between minutes and more than 1,000 hours [113, 114] demonstrating the high degree of regulation that is imposed on protein degradation. As noted above, selectivity is mediated through several mechanisms, one being the recognition of specific binding motifs by ubiquitinating E3 ligases. These universal (i.e. protein independent), minimal recognition elements for the proteolytic apparatus are commonly termed degrons [115]. The first degradation signals were identified upon permutation of the first amino acid in β-galactosidase resulting in vastly differing protein stabilities [116]. This lead to the formulation of the N-end rule predicting a destabilization for basic and bulky hydrophobic amino acids. Following further investigations, this was later expanded to elaborate N-end rule pathways in which protein stability was additionally influenced by N-terminal modifications such as acetylation, deamidation, and arginylation [117]. Importantly, N-degrons must be accessible to be effective, which means that N-termini that are by default buried within a properly folded protein or protein complex escape the N-end rule [118]. Noteworthy, the erroneous exposure of N-degrons in misfolded proteins offers one mechanism how dysfunctional proteins can be selected for degradation.

Like N-degrons, most hitherto identified degradation signals are short linear motifs and they seem to occur more frequently on N- or C-termini than in the middle of protein sequences [107, 119]. As an example, the stability of the tumour suppressor p53 is regulated by its E3 ligase MDM2 (murine double minute 2) which binds to a degron consisting of three hydrophobic residues (Phe, Trp, and Leu at positions 19, 23, and 26) located in an N-terminal α-helix [120]. Interestingly, the interaction can be blocked by phosphorylation of a threonine residue at position 18 resulting in

stabilization of p53 [121]. The effectiveness of degrons can also be impacted by modifications of the Ubiquitin acceptor lysine as exemplified by the acetylation-ubiquitination competition that was described above for the C-terminal domain of p53 [94]. This illustrates a common principle of degrons already insinuated above for N-degrons: They are usually conditional in nature and can depend on the presence of PTMs and other structural constraints. This means degrons can often be switched on and off adding another layer of sophisticated regulation. This is equally true for degradation motifs that are only created by modifications such as the phosphodegron within the sequence of the cell cycle inhibitor p27 [122]. Another prominent example is the oxygen level-dependent degron in the hypoxia inducible factor (HIF) 1α which contains a degradation-inducing, hydroxylated proline [123]. Hydroxylation of this transcription factor only occurs when oxygen is available in sufficient concentrations. Via this mechanism, the foundation is laid for oxygen level sensing and adaption of gene expression in hypoxia, which results in HIF 1α stabilization and activation.

It is noteworthy that many degrons are located within intrinsically disordered sequence stretches [119] which often are characterized by a high proportion of hydrophilic amino acids. Interestingly, disordered regions themselves have been associated with shorter half-lives [124, 125]. However, it is not clear whether they represent an inherent degradation signal or whether this association rather reflects the higher possibility of interactions with other proteins and modifications within disordered structures, which may create conditional degrons. Contrarily, larger hydrophobic protein regions have been identified to serve as inherent degradation signals [126, 127]. Usually they are buried within correctly folded protein structures, but in misfolded proteins they are solvent exposed and can be bound by chaperones, the cellular protein-folding assistants. Those can further mediate the interaction with E3 ligases that subsequently ubiquitinate the misfolded chaperone cargo and mark it for degradation [126]. Similarly, assembly interfaces of protein complexes can serve as degradation signals resulting in a removal of supernumerary subunits [127].

The degradation response can further be fine-tuned via a range of additional regulations, for instance via multiple degrons working in cooperation and exhibiting differing E3 ligase affinities, or when a preceding modification by a different enzyme is required before the actual degron can be induced through another modification (reviewed in [119]). It becomes clear that the creation of an active degron is a multifactorial process. To date, only a small fraction of the ~600 E3 ligases has a known target sequence [119]. The strong influence of structural accessibility of degrons complicates their systematic and global identification and their often PTM-dependent nature adds another layer of complexity. This illustrates that cellular stability of endogenous proteins must be determined in different modification states to enable elucidation of regulatory principles underlying protein degradation.

## 2.2 Critical maintenance of proteostasis

**Proteostasis players and their duties** – The three tightly regulated processes of protein synthesis, protein folding, and protein degradation are main contributors to a balanced protein flux and, as outlined above, they are executed by the ribosome, chaperones, and the proteasome or in the lysosome, respectively (Figure I-6). They function in a concerted manner building an intertwined proteostasis network to ensure proper quality control in every single phase of a protein's lifetime. It has been reported that already during protein biogenesis 12-15 % of nascent chains are poly-ubiquitinated and marked for proteasomal degradation [128]. This co-translational ubiquitination can affect polypeptide chains on ribosomes that are stalled, for instance due to mRNA

with rare codons or inhibitory secondary structures, but more frequently it occurs on active translation complexes [83, 128]. A connection to protein misfolding has been suggested for the latter phenomenon, which was underpinned by a 50 % increase of co-translational ubiquitination upon chaperone inhibitor treatments that block co-translational folding [128]. Although the direct proof that ubiquitinated proteins in active ribosome complexes are incorrectly folded is still missing, this presumption would be in line with several other known protein misfolding-induced degradation pathways. Secretory and membrane proteins that cannot fold or assemble properly in the endoplasmic reticulum (ER) are retro-translocated into the cytosol, ubiquitinated and degraded (ER associated degradation [129]). Cytosolic proteins that feature a non-functional conformation state resulting from either misfolding or protein age-related damage underlie similar degradation principles [99].

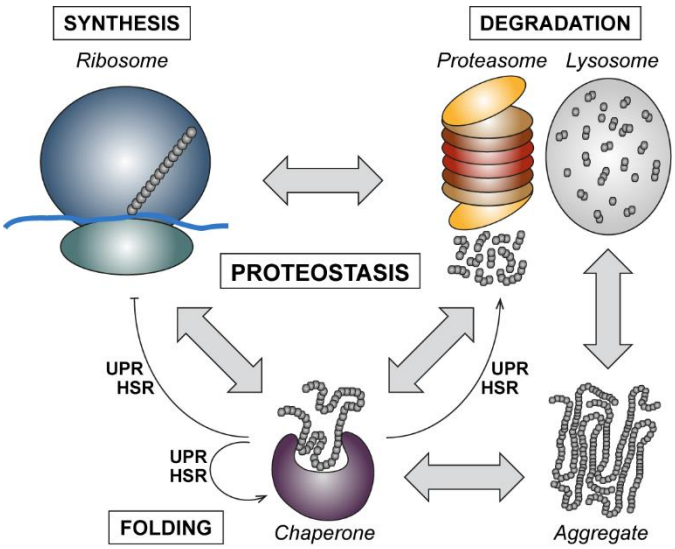



**Figure I-6 | The cellular proteostasis network.** *The strongly regulated processes of protein synthesis, folding, and degradation, and their executors (ribosome, chaperones, proteasome, lysosome) built an intertwined network to ensure protein homeostasis. In case of insufficient protein folding or degradation, misfolded or damaged proteins may accumulate and protein aggregates may be formed. As a result, chaperones mediate the initiation of specialized stress pathways such as the unfolded protein response (UPR) or heat shock response (HSR). This leads to a re-establishment of proteostasis via inhibition ( ⊥ ) of protein synthesis, and enhancement ( ↑ ) of protein folding and degradation capacities.*

As mentioned above, malfunctioning proteins are mainly cleared by the UPS in close collaboration with chaperones such as the heat-shock proteins (HSPs). Chaperones have a certain buffer capacity in case of a sudden increase in faulty proteins, act as sensors for increased proteotoxic stress, and regulate the balance between protein anabolism and catabolism [84, 130, 131]. The unfolded protein response [100] and the heat shock response [98], for instance, are activated in a chaperone-dependent manner upon accumulation of insufficiently folded proteins in the ER or of damaged proteins in the cytosol. Both responses result in an attenuation of protein synthesis and a concomitant elevation of protein degradation and folding capacities. If proteostasis networks are nevertheless overwhelmed with proteins in non-native conformations, potentially toxic aggregates consisting of two or more protein molecules can arise [131]. These protein aggregates can be removed either by dissociation and targeting to the UPS with the help of certain chaperones or by autophagy resulting ultimately in lysosomal protein degradation [130, 131]. Apart from the removal of malfunctioning proteins, the UPS and lysosome have also an important role in signal abrogation or initiation via degradation of signalling and regulatory proteins, respectively. Prime examples are the 26S proteasome-dependent, periodic degradation of cell cycle regulatory proteins via phosphodegrons to enable progression of cell proliferation [132] and the elimination of activated receptor tyrosine kinases from the cell membrane via endocytosis and subsequent lysosomal degradation [133]. In summary, a properly functioning proteostasis network minimizes the formation and aggregation of misfolded proteins, eliminates potentially toxic protein aggregates, replaces damaged proteins, and has essential regulatory functions in cellular signalling.

**Proteome imbalance in diseases** – Evidence suggest that in a healthy state there is an excess of protein folding and degradation capabilities that can buffer increased proteostasis stress [84]. In steady-state hippocampal neurons, for instance, only 20 % of available proteasomes were shown to be occupied with substrates, whereas the remaining proteasomes exhibited the substrate-accepting ground state [134]. However, many diseases are marked by proteome imbalances that are caused by a lasting increase in translational output and/or a decrease in protein folding activity and degradation capacity and often result in an accumulation of dysfunctional proteins. Indeed, protein aggregates are increasingly recognized as the base for cytotoxicity in neurodegenerative diseases like Alzheimer's, Huntington's, and Parkinson's disease (outlined in [135]). While the molecular mechanisms underlying the aetiopathogenesis of these aggregation diseases require further investigation, many have been linked to disruptions of the proteostasis network. This is exemplified by a certain type of early onset Parkinson's disease that is associated with mutations in the Parkin gene. This encodes an E3 ligase that has been reported to contribute to the degradation of potential neurotoxic proteins by the UPS [136, 137] and of defective mitochondria by mitophagy [138, 139]. Interestingly, ageing itself has been linked to a decline in proteostasis [140] providing an explanation for the age-related risk of many neurological diseases. Many metabolic diseases, which are also associated with age, likely compromise cellular proteostasis as well. Type 2 diabetes, for instance, has been connected to a prolonged, unresolvable unfolded protein response which ultimately triggers β-cell death [141]. In line with an overwhelmed molecular chaperone system, a supply of chemical chaperones has been observed to mitigate diabetes- and obesity-related symptomatology in mouse models [142].

The development of cancer represents another example for a disease tightly linked to proteome imbalance. Crucial regulators of the proteostasis network such as E3 ligases and other signalling proteins are commonly mutated during cancerogenesis [119, 143]. Further, oncogenic transformation is generally accompanied with a broad range of alterations in overall proteome homeostasis [84]. On the one hand, oncogenic transformation is characterized by an amplified, uncontrolled cell proliferation [144] which must require an elevated protein synthesis and also increased protein folding capabilities. On the other hand, it has been demonstrated that cancer cells are at the same time highly dependent on an enhanced proteasomal degradation activity [145]. This may be explained by a higher demand to degrade cell cycle regulators to proceed in cell division, faulty proteins caused by increased genome mutation and translation error rates, or potentially aggregation-prone complex subunits that may be produced in excess due to gene duplications, translocations, and aneuploidy [84, 146]. It has been suggested that the observed oncogenic addiction on proteasome function is a result of an exhausted buffering capacity for proteotoxic stress in cancer cells [145]. Indeed, inhibition of the proteasome and certain chaperones like HSP90 have shown promising clinical response presumably based on the induction of a proteotoxic crisis which culminates in the desired tumour cell death [147, 148]. However, the efficacy of these therapies is largely dependent on the combination with other drug molecules and the tumour background which demonstrates the need for a detailed and cancer-dependent, molecular investigation of protein turnover, drug effects and their interplay.

## 2.3 Methodologies to quantify protein turnover

Since the discovery of the dynamic state of body constituents in the late 1930s [97], a multitude of methodologies have been employed to study protein turnover and alterations thereof under different cellular conditions. Many approaches are based on the incorporation or loss of a tracer that is added to culture medium or infused into animals [116, 149-151]. Subsequently, turnover

rates of proteins can be inferred from the time-dependent change of the tracer abundance in individual proteins or the whole protein pool. Alternatively, changes in protein abundances upon inhibition of translation have been used to estimate degradation behaviour [152]. Even cleverly devised fluorescence-tagging strategies have been utilized to determine relative and absolute protein turnover [153-155]. The different strategies provide differing information contents and performance profiles and are briefly discussed in the following.

**Bulk and single protein approaches** – For the first time, protein turnover was revealed in mice by quantifying the differences in a fed and excreted, stable heavy isotope of nitrogen ($^{15}N$) using early mass spectrometers [149]. Later, amino acids that were labelled with radioactive isotopes of hydrogen ($^{3}H$), carbon ($^{14}C$), or sulphur ($^{35}S$) were utilized in a pulse-chase experimental setup and protein synthesis and degradation were estimated from tracer incorporation or loss kinetics measured by scintillation counting (Figure I-7 A). Generally, essential amino acids that were largely metabolically isolated (methionine, leucine, phenylalanine, arginine, valine) were chosen to ensure efficient incorporation and avoid labelling artefacts due to amino acid conversions. Such methods were applied to measure bulk turnover (i.e. turnover of all expressed proteins at once [156, 157]) or stability of single proteins after immunopurification [116]. Coupling of radioactive pulse-chase experiments to two-dimensional gel electrophoresis (2DE) increased the number of investigated proteins up to a couple of hundreds but without identifying them [158, 159]. In contrast, utilization of subsequent tandem mass spectrometry (MS) or western blotting enabled protein identification, but in turn reduced the throughput to less than 20 proteins [160, 161]. While these radioisotope-based approaches are very sensitive because even a marginal fraction of labelled proteins can be detected, they are associated with radioactivity hazards. Moreover, an unambiguous identification of individual quantified proteins is necessary if differences on protein-level shall be resolved. Bulk measurement and even 2DE approaches do not provide this resolution since 2DE spots often contain several proteins.

Another strategy that is often used to assess or validate degradation rates of single proteins utilizes translation inhibitors like cycloheximide and time-dependently measures abundance changes after drug administration using western blots [162, 163] (Figure I-7 B). However, antibody-based approaches can suffer from cross-reactivity and are inherently restricted to a predefined set of proteins. Moreover, quantification on western blots provides only limited quantitative accuracy and can therefore only be used to resolve larger turnover differences. Besides, treatments with translation inhibitors have been shown to induce cellular stress [164] and thus can trigger undesired signalling responses that may affect the proteostasis network and bias the readout.

**Proteome-wide strategies** – Cyclohexamide experiments coupled to anti-TAP (tandem affinity purification)-tag western blots have also been proven feasible for higher throughput methods using thousands of TAP-tagged yeast strains covering a large fraction of the proteome [152]. Another tagging approach, named global protein stability (GPS) profiling, circumvents the translation inhibitor treatment using a fluorescence-based assay and flow cytometry. With this method, relative stability of ~8,000 proteins in HEK-293T cells was estimated based on the fluorescence ratio of a dye fused to the protein (and hence being degraded together with the protein) to another dye translated independently (thus not being degraded) but from the identical mRNA [153] (Figure I-7 C). Alternatively, bleach-chase experiments in cells expressing fluorophore-tagged proteins can extrapolate absolute protein half-lives. Here, fluorophores are bleached with a short pulse of light and the difference in fluorescence between bleached and non-bleached cells measured with time-lapse fluorescence microscopy is used to estimate the protein removal rate. While this approach

offers the potential for a similar throughput as GPS profiling, it has so far only been applied to 100 proteins [154]. Likewise, tandem fluorescence timers, which feature a time-dependent change in colour, have been utilized as a proxy for protein age, degradation and synthesis rates for tens to one-hundred proteins [155, 165]. Each tandem fluorescence timer consists of two fluorescent proteins with differing maturation kinetics, which are fused to a protein of interest. The ratio of their fluorescence reflects protein age within a pre-defined time-range (i.e. the fluorophore maturation times ranging from minutes to several hours), and a shift in ratios indicates an alteration in synthesis or degradation rates. While all these fluorophore-based approaches provide high sensitivity and a striking single-cell resolution, protein tagging at the N- or C-terminus entails the risk of disrupting the functional structure, binding properties, and localization, which all can alter protein stability. Moreover, incomplete degradation of certain fluorescence tags has been reported [166]. Since this distorts stability estimations, dyes have to be chosen carefully. Further, the ectopic overexpression of proteins may affect proteostasis and consequently protein half-lives, and impair the transferability of measured turnover rates to endogenously expressed proteins. Finally, establishing clonal libraries of tagged proteins can consume a lot of time and resources and is only possible in genetically tractable systems thereby limiting their feasibility for studies on the scale of proteomes.

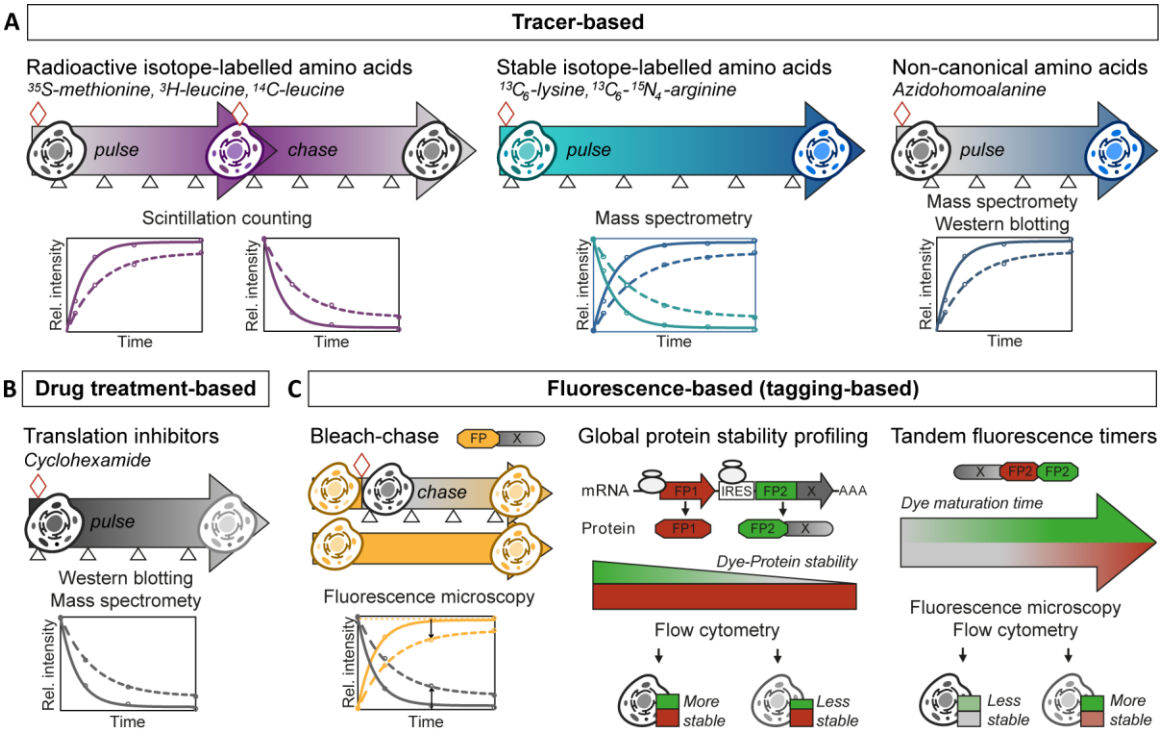



**Figure I-7 | Methods to quantify protein turnover.** *Different tracer-based (A), drug treatment (B) and fluorescence tag-based approaches (C) can be employed to estimate absolute or relative protein turnover. Examples of tracers or treatments are indicated in italics. For details of methods please see text (Δ: sampling time-point, ◊: addition of amino acids, washout start, treatment start, or bleach pulse; FP: fluorescence protein; X: protein of interest; IRES: internal ribosome entry site).*

In recent years, pulsed SILAC (stable isotopic labelling of amino acids in cell culture [167]) approaches in conjunction with modern MS-based technologies have emerged as a powerful alternative for protein turnover measurements. Similar to radioisotope-based methods, they employ isotopically labelled amino acids (mostly lysine, arginine, leucine) incorporated into newly synthesized proteins, but in this case using stable isotopes ($^{2}H$, $^{13}C$, $^{15}N$) and coupling protein identification

and turnover quantification within the mass spectrometric readout (Figure I-7 A). Such pulsed SILAC (pSILAC; also called dynamic SILAC [150]) approaches enable, in theory, the unbiased, hypothesis-free, and proteome-wide assessment of turnover of endogenous proteins expressed at physiological levels [113, 168-170]. Since the mass spectrometer can quantify both, isotopically labelled and non-labelled protein version, experimental setups only require a pulse and no chase to determine label incorporation and loss simultaneously. Improvements in sample processing, MS measurement, and data analysis techniques now enable the quantification of endogenous turnover behaviour for >6,000 up to 8,500 proteins in a single biological sample [114, 170, 171]. However, accurate quantification can be difficult for relatively low label incorporation within a high background of non-labelled proteins as it is often observed for very short pulses or in non-dividing cells. In this context, BONCAT (bio-orthogonal non-canonical amino acid tagging) methodologies have been proven valuable for enrichment of trace levels of newly synthesized proteins. They utilize amino acids that carry small reactive groups and thus allow for a reaction and enrichment with modified affinity tags. The concomitant depletion of non-labelled proteins reduces background signals significantly and ultimately facilitates the detection of labelled proteins. In this manner, azidohomoalanine (AHA) and homopropargylglycine (HPG) have been applied successfully for the detection of newly synthesized proteins [151, 172, 173]. Noteworthy, unspecific binding of non-labelled proteins in the enrichment process can distort quantification and thus needs to be evaluated diligently. This can also be controlled for by adding isotopically labelled SILAC amino acids in parallel to non-canonical ones and restricting the quantification to isotopically labelled proteins [172, 174]. More importantly, the introduction of non-canonical amino acids into biological systems has been reported to induce protein-folding pathways and downregulate synthesis-associated proteins [174] resembling an induction of the UPR pathway. This effect was apparently minimized by adjusting the AHA-to-methionine ratio [174] but still reveals that the potential impact of non-canonical amino acids on proteostasis has to be assessed thoroughly and AHA-pulse times should be kept short.

Based on mentioned limitations and advantages, the method used for determination of protein turnover should be chosen carefully and depending on the research question and cellular system at hand. To this end, the design of conventional dynamic SILAC experiments has been adapted in many ways, for example to improve detection of slow label incorporation [169], to compare different conditions within one measurement [175], or to simultaneously acquire protein abundance changes during the time-course of the pulse [170, 176]. Pulsed SILAC coupled to advanced MS-based proteomics technologies promises the analytical coverage necessary to assess proteoform-specific turnover globally, and is currently considered the only suitable method to study isoform- and modification-specific turnover of endogenous, un-tagged proteins.

## 3 Analysing peptidoforms with bottom-up proteomics technologies

In terms of depth, throughput, and sensitivity, mass spectrometry-based approaches provide a unique applicability for studies of cellular functions on a proteome-wide scale (reviewed in [177]). They are divided into 'top-down' and 'bottom-up' methodologies. Historically, top-down proteomics referred to all methods that included an initial separation of proteins irrespective of the subsequent sample processing before MS data acquisition (i.e. it could include a protein digestion step). Now it is generally understood as the MS analysis of whole proteins [178]. Thus, it is the only MS-based way to truly identify proteoforms. Due to the largely overlapping isotope clusters of big molecules, however, it is so far limited to more abundant and lower molecular weight proteins (<50 kDa) in rather uncomplex samples [179]. On the contrary, bottom-up (also called shotgun) proteomics characterises all workflows that result in the measurement of peptides derived from the digestion of proteins usually using cleavage-specific proteases. While this way most sensitivity issues of top-down analyses can be overcome, it poses several challenges for the analysis of proteoforms. It is difficult to distinguish highly homologous proteoforms since they share most of their peptides with one another. For about 18 % of all splice variants annotated in the Swiss-Prot database [43], no tryptic peptide (within the size requirements of a bottom-up MS measurement) does exist to identify them unambiguously. One-third of the residual non-canonical splice variants feature only three or less of such unique peptides (Figure I-8 A). This issue is exacerbated in actual measurements due to differing isoform abundances and detectability of peptides. Besides, the relation between different PTMs within a proteoform is lost upon digestion and modified peptides cannot be reassembled without ambiguity, which makes an inference of entire proteoforms that actually exist in a sample often impossible [180] (Figure I-8 B). Nevertheless, bottom-up proteomics is still by far the most convenient approach for large-scale quantification of proteomes [177], and there are several methods to improve peptide coverage and consequently the basis to infer information about proteoforms. In the following, the principal steps of a bottom-up proteomics workflow will be covered with a focus on procedures that have been used and are relevant for this thesis (Figure I-9). Since (modified) peptides and not proteoforms are the entity

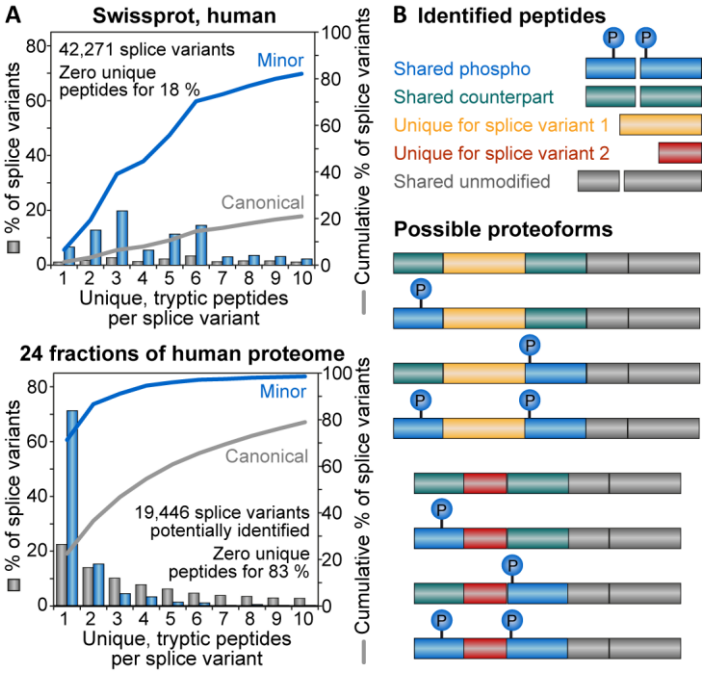



**Figure I-8 | Challenges in analysis of proteoforms with bottom-up proteomics.** *(A) The human Swiss-Prot database [43] was digested in-silico using trypsin cleavage specificity and allowing up to 2 missed cleavages. Unique peptides that unambiguously identify main (canonical, grey) or minor (non-canonical, blue) splice variants were plotted (upper panel). The number of identified unique peptides for splice variants is even lower in actual measurements (representative sample in the lower panel). Of the minor splice variants, 70 % can only be identified with a single peptide. (B) Upon digestion of proteins, the overall sequence context of modifications is lost. This results in many combinatorial possible proteoforms of which only a subset may actually be present in the sample.*

that is being measured, the term 'peptidoform' [181] will be used to describe all versions of a peptide that can be derived from the same backbone amino acid sequence.

## 3.1 Sample preparation steps to enhance peptidoform coverage

**Protein extraction and digestion** – A wide range of sample processing workflows for bottom-up proteomics exists, all of which have a specific purpose, advantages and limitations, and are ideally tailored to the research question at hand [177, 182]. Common to all of them is the initial extraction of proteins from the source material. The choice of extraction buffer and lysis procedure depends on several considerations (reviewed in [183]) such as what type of sample is used or whether native protein structures need to be preserved. Extracted proteins can optionally be fractionated based on physicochemical properties or size [160], or specific subproteomes can be enriched [184]. Before protein digestion, disulphide bridges in proteins are usually reduced and free cysteines alkylated [185]. Subsequently, peptides are generated in most cases using sequence-specific proteases. Trypsin is the most commonly utilized enzyme because it generates peptides with C-terminal lysine or arginine residues with advantageous properties for MS measurement and identification [186]. However, to improve sequence coverage and the number of unique peptidoforms, other enzymes such as GlucC (cleaves C-terminal of glutamate), chymotrypsin (C-terminal of hydrophobic amino acids), and AspN (N-terminal of aspartate) have been proven useful [187-189]. Nevertheless, overall identification rates are lower with alternative proteases, thus trypsin remains the enzyme of choice for most standard applications.

**Off-line peptide fractionation** – Digestion produces incredible complex peptide mixtures that cannot be captured by mass spectrometry comprehensively without prior separation. Hence, off-line fractionation methods are often employed to reduce sample complexity and facilitate peptidoform identification. A wide range of techniques is available to separate peptides according to their physicochemical properties (charge, polarity, hydrophobicity, size) and depending on the resultant, reciprocal interactions with the different stationary and mobile phases (reviewed in [190]). Fractionation can be conducted utilizing high-performance liquid chromatography (HPLC) systems [191] or self-packed stop-and-go-extraction (STAGE) tips [192]. The number of fractions is always a compromise between proteome depth and expenditure of MS measurements time. To fully capitalise measurement time, the off-line fractionation approach should exhibit a good separation power and orthogonality to low pH reversed-phase (RP) chromatography, which is typically coupled on-line to the mass spectrometer to further decreases complexity and separates peptides according to hydrophobicity [193] (see p. 22). As an example, strong anion exchange (SAX) is highly orthogonal, as the underlying separation principle is based on charge. It depends on electrostatic interactions of negatively charged peptides and positive functional groups on the stationary phase under alkaline conditions, and peptides are eluted consecutively using salt gradients. In contrast, high pH RP fractionation separates analytes according to hydrophobicity similar to low pH RP chromatography albeit at differing pH values. Hence, it is not perfectly orthogonal which is usually accounted for by discontinuous pooling schemes after fractionation [194].

Even for in-depth fractionated samples, the identification of PTMs is often still difficult. This is a result of the, for most parts, sub-stoichiometric nature of modified peptides that can hamper their detection through suppression of their potentially weak MS signal by high abundant, non-modified peptides. Hence, bottom-up proteomic workflows frequently include enrichment procedures for PTM subproteomes that are based on some form of specific interaction with the modified amino acid side chain [195].

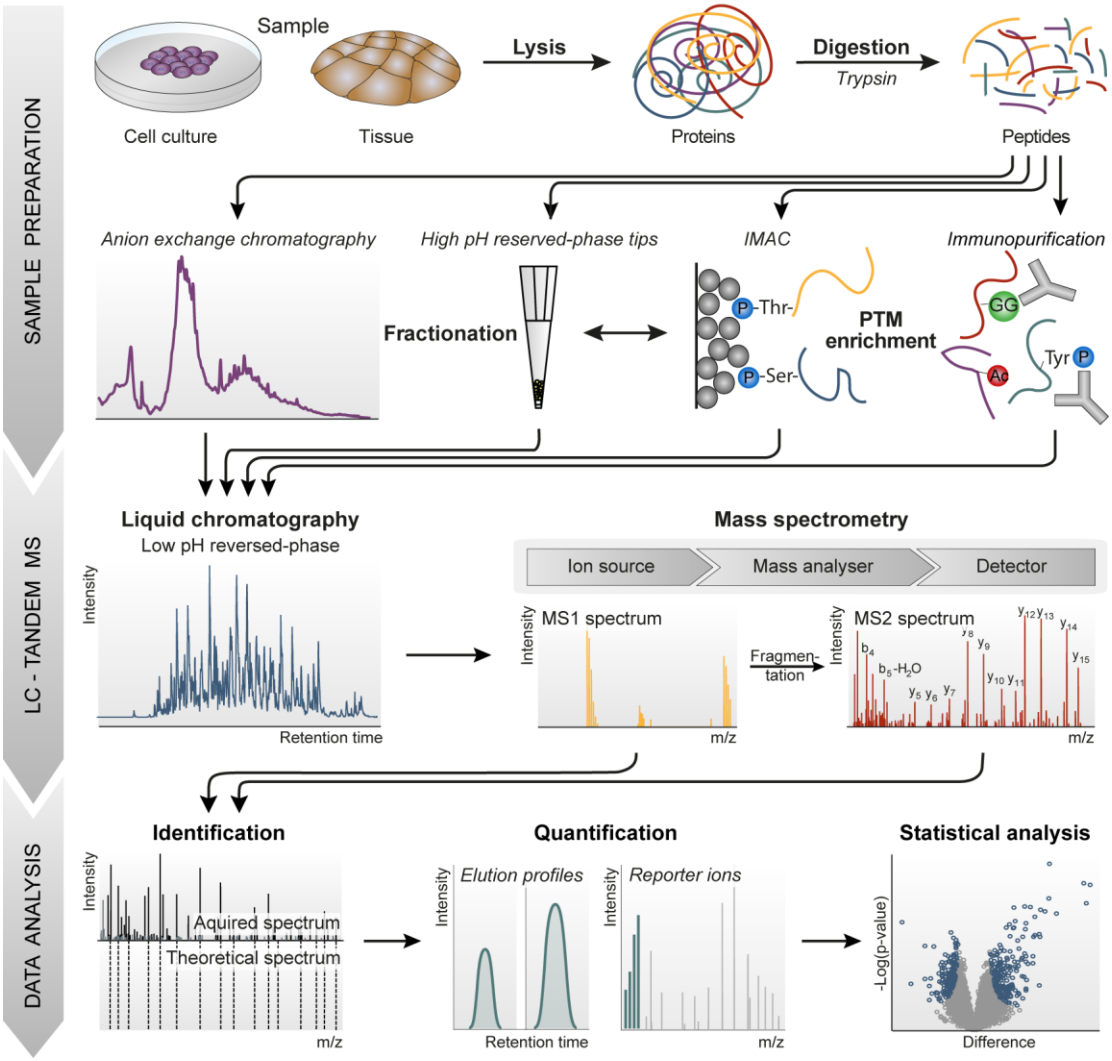



**Figure I-9 | Typical bottom-up proteomics workflow.** *Samples are lysed and extracted proteins are digested. Subsequently, peptides can be fractionated and modified peptidoforms can be enriched utilizing a variety of approaches. These steps can also be conducted consecutively and in interchangeable order. Directly before mass spectrometry (MS) analysis, peptides are separated employing a low pH reversed-phase liquid chromatography (LC) that is coupled on-line to a mass spectrometer. After peptide ionization, spectra of peptides and peptide fragments are acquired. Those are subjected to a database search for peptide identification. Following quantification, data is further processed and analysed, for instance performing statistical tests to assess significant differences between samples (adapted with permission from [196]).*

**Enrichment of phosphorylation** – A variety of techniques has been developed for the enrichment of phosphorylation on protein and peptide level, including immunoaffinity purification, chemical derivatization, charge-based fractionation, and metal-based affinity methods (reviewed in [197]). Nowadays, immobilized iron(III) metal affinity chromatography (Fe-IMAC) is one of the more frequently used approaches for efficient phosphopeptide enrichment and is available in column- [198] and batch-format [194]. Originally, IMAC with divalent ions such as Zn2+ and Cu2+ has been developed for purification of histidine- and cysteine-containing proteins [199], but later IMAC using trivalent Fe3+-ions revealed to also enrich phosphorylated proteins [200]. The enrichment is based on the high-affine coordination of negatively charged phosphate groups (at acidic pH) to positively charged trivalent metal ions that are immobilized on a stationary phase via chelating agents such as iminodiacetic acid [197] (Figure I-10 A). Elution can be performed with alkaline solvents (e.g. ammonium hydroxid) or phosphate buffers [194, 201]. IMAC in conjunction with

off-line fractionation and modern MS technologies has enabled the detection of up to 40,000 phosphorylation sites in a single sample [62, 194]. Efforts to improve enrichment strategies and optimize protocols are still ongoing, and many are focussed on lowering required input amounts [202] or improving the coverage of chronically underrepresented tyrosine phosphorylation sites [203]. To this end, affinity purification based on an engineered phosphotyrosine binding domains has been proven more efficient for the specific enrichment of phosphorylated tyrosine peptides than the commonly used antibody-based approaches [204].

**Enrichment of lysine acetylation** – In comparison to phosphorylation, lysine acetylation often features an even lower stoichiometry on proteins [63, 205] making an enrichment for global analysis indispensable. Currently, immunoaffinity purification using antibodies raised against the acetyl moiety and coupled to Protein A agarose beads is the only widely used enrichment strategy for acetylated peptides [195]. Elution is commonly achieved in acidic conditions leading to the denaturation of the antibodies. As for all antibody-based approaches, the performance depends strongly on the efficacy of the antibody. Moreover, batch-to-batch variation is common, and differing sequence specificities may affect the identification of acetylation sites [206]. Part of these challenges has been tackled by using a mix of different monoclonal antibodies to minimize the sequence bias [206, 207]. In combination with various fractionation techniques, this allowed for the detection of more than 10,000 acetylation sites, but required large protein inputs of 10 to 20 mg [207, 208].

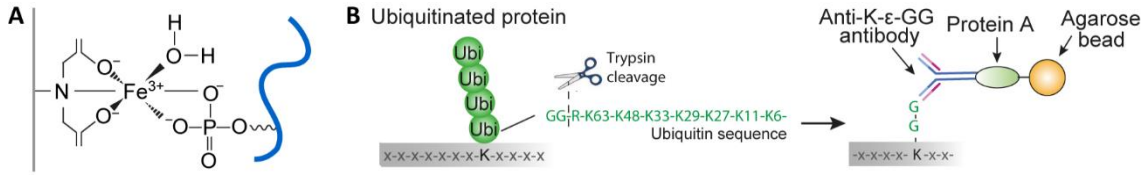



**Figure I-10 | Enrichment approaches for modified peptides.** *(A) In IMAC, negative phosphate groups are coordinated with iron(III) ions that are immobilized using iminodiacetic acid. (B) The di-glycine remnant produced after tryptic digestion of ubiquitinated peptides can be recognized by specific antibodies (adapted with permission from [209, 210]).*

**Enrichment of ubiquitination** – Initially, ubiquitination surveys were based on genetically engineered ubiquitin that carried a histidine or streptavidin tag, and ubiquitinated proteins were enriched via IMAC or using biotin-based affinity matrices [211, 212]. Besides the potential interference with ubiquitin function, this approach resulted in a high number of unmodified peptides upon digestion of the enriched sample, which limited the depth of the PTM analysis. Hence, similar to enrichment of lysine acetylation, antibody-based strategies on peptide level are the method of choice for global ubiquitination surveys today. Tryptic digestion of ubiquitinated proteins creates a di-glycine remnant on modified lysine side chains that can be recognized by antibodies [213] (Figure I-10 B). Noteworthy, other ubiquitin-like protein modifications such as neddylation feature the identical di-glycine remnant and can be co-enriched. However, it has been suggested that more than 94 % of identified modification sites originate from ubiquitination [214]. Optimized workflows, including a peptide fractionation step prior to di-glycine immunopurifications, these days facilitate the identification of up to 25,000 ubiquitination sites from 15 mg of protein input following proteasome inhibition [215].

## 3.2 Principles of tandem mass spectrometry measurements

Following sample processing, enriched and fractionated peptides are subjected to MS analysis. Mass spectrometers function as very accurate molecular scales that can determine the mass of

charged molecules in the gas phase by harnessing their manipulability in electric or magnetic fields (reviewed in [216]). They consist of an ion source generating gas phase ions, and at least one mass analyser and detector enabling the determination of mass-to-charge (m/z) ratios of peptides and peptide fragments. Prior to a segregation according to their m/z values, peptides are typically separated via an additional on-line coupled liquid chromatography (LC).

**On-line peptide separation** – Due to its outstanding peak capacity and solvent compatibility with MS analysis, RP HPLC at acidic pH is the most commonly used technique for on-line separation of peptides [193, 217]. It is based on a non-polar stationary phase (often hydrophobic, octadecyl alkane chains on silica beads) and an aqueous mobile phase containing an amphiphilic ion-pairing reagent (e.g. formic acid). Hydrophobic amino acids of peptides can bind directly to the stationary phase. Retention and thereby resolution is further increased by indirect interactions of polar peptide parts mediated by the ion-pairing reagent [190]. Consecutive elution of absorbed peptides is achieved by gradually increasing the concentration of organic solvent (usually acetonitrile) in the mobile phase resulting in separation of peptides according to their hydrophobicity. Subsequently, early eluting polar and later eluting non-polar peptides are delivered to the mass spectrometer via the ion source interface.

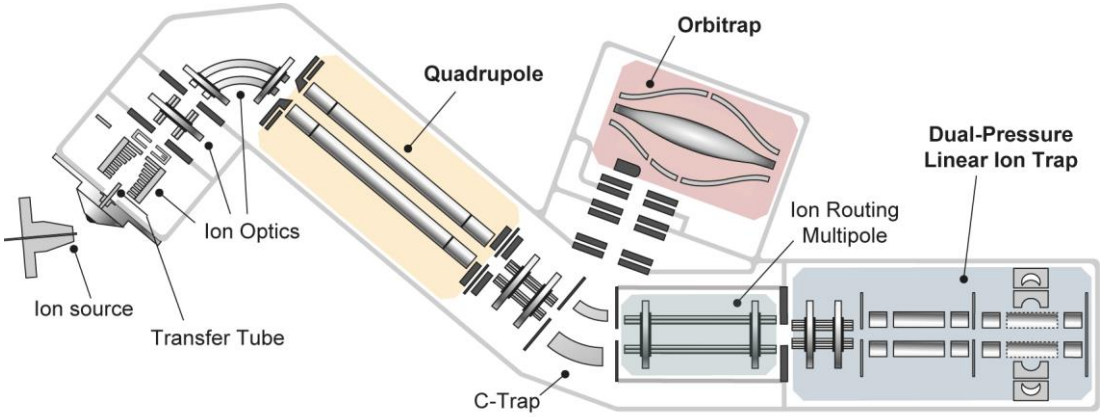



**Figure I-11 | The Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer.** *After ionization, gas-phase peptide ions are focussed and propelled towards mass analysers via the ion optics. The quadrupole acts as mass filter, and peptide and peptide fragment spectra can be acquired either in the Orbitrap or in the ion trap mass analyser. The instrument allows for a variety of fragmentation and acquisition modes such as higher-energy collisional dissociation that is executed in the ion routing multipole (with permission from Thermo Fisher Scientific).*

**Electrospray ionization** – A separation based on m/z values by mass analysers requires the transfer of peptides into the gas phase and their ionization. For on-line LC-MS setups, this is achieved by electrospray ionization (ESI) [218]. This 'soft' ionization technique enables the vaporization of large and polar peptides and even proteins without physical destruction [219]. The analyte solution exits the RP column through a thin conductive capillary (emitter), to which a voltage is applied. The liquid forms the so-called Taylor cone that carries positively charged peptide molecules. At its tip, Coulomb forces cause the solution to disperse into small, multiply charged droplets. During migration through the electrostatic field and towards the mass spectrometer, the LC solvent evaporates leading to an increased surface charge. As soon as Coulomb forces exceed the surface tension (Raleigh limit), aerosol droplets disintegrate further in a process named Coulomb explosion or fission. Final ionization of peptides is hypothesized to occur via an active ion emission from droplets with radii less than 10 nm due to field desorption (ion evaporation model) and/or complete solvent evaporation, where residual droplet charges finally remain on the analyte

(charged residue model) (outlined in [219]). Complete drying is latest accomplished in the heated transfer tube through which positively charged peptide ions enter the vacuum of the mass spectrometer (Figure I-11).

**Mass-to-charge analysis** – Peptide m/z ratios can be measured with a myriad of different mass analysers and corresponding detectors (reviewed in [216, 220]). Common to all of them is the manipulation of ion trajectories *in vacuo* by direct or alternating currents (DC or AC), and the subsequent inference of m/z values by measuring the peptides' responses to electromagnetic forces imposed by applied currents in comparison to known m/z standards. Obtained information is recorded in mass spectra, which display ion currents as intensities at certain m/z ratios (Figure I-9). Due to their different measurement principles, various types of mass analysers provide differing mass accuracy, speed, sensitivity, and resolution (ability to differentiate molecules with very close m/z ratios). Nowadays, different mass analysers are often combined within one instrument platform allowing for flexible data acquisition. One example of such hybrid instruments is the Fusion Tribrid series of mass spectrometers (Fusion and Fusion Lumos, Figure I-11). They incorporate a quadrupole, a dual-pressure linear ion trap, and an Orbitrap mass analyser [221]. Since this was the instrument platform mainly utilized for data acquisition, its major functionalities will be described in the following.

Peptide ions entering a mass spectrometer are initially passing through the ion optics. They consist of lenses and rod-shaped electrodes that focus ions via application of ACs. With the help of DC offsets, which create an electric potential gradient, ions are directed towards the quadrupole [216]. In general, the motion of analyte ions depends on the effective force that acts on them. The force, in turn, is dictated by their mass and net-charge as well as the strength of the applied electromagnetic field regulated by the voltage of applied DC and AC currents. This principle can be harnessed to deflect, store, separate, and detect analyte ions according to their m/z ratio. The quadrupole in the Fusin Lumos, for example, is used for mass filtering by removing ions below or above certain m/z values (lower and upper mass filter). This is achieved by applying certain superimposed DCs and ACs on the four rods of the quadrupole (Figure I-12 A) [216, 220]. The lower mass filter is defined by an AC, also named main radio frequency (RF), that confines peptide ions radially and prompts them to adopt a 'secular' (cork-screw like) motion when passing through the quadrupole. Since the extent of the motion depends on the main RF characteristics as well as the size of the ions with smaller ions featuring larger amplitudes, smaller m/z species can be removed from the quadrupole laterally by increasing the main RF [216]. The upper mass filter depends on a quadrupolar DC generated by voltages that have an equal amplitude but pairwise opposite signs on the four rods. Positively charged peptide species will be attracted to the negatively charged rods, and if they have too much inertia, the forces generated by the radially confining AC will not be strong enough to push them back onto a stable trajectory [216]. Consequently, large ions with higher inertia will either escape between the rods or bump into them. By choosing different AC and DC voltages, varying m/z ranges are allowed to pass through the quadrupole. This can include a range of several hundreds of m/z or the isolation of a single ion species. With the advanced quadrupole technology in the Fusion Lumos, masses can be isolated with a minimal window (isolation width) of down to 0.4 m/z.

In contrast to the quadrupole in the Tribrid instrument, the linear ion trap can be used to determine the m/z value of peptide ions but also to store, isolate, and fragment analyte ions. Ion traps are composed of four rod-shaped electrodes each of which is cut into three segments. Two of the four rods, named exit rods, contain slits through which ions can escape for detection (Figure I-12 B). Initial ion storing is achieved by two forces. Similar to quadrupoles, a quickly oscillating

main RF that is applied to pairs of electrodes restricts the ions' movements radially and induces a secular ion motion. In addition, an offset between DCs applied to central and outer rod segments traps ions axially [216]. Subsequently, ions can be removed from the stored ion pool in two different ways. As in quadrupoles, ramping of the main RF will increase amplitudes of secular ion motions steadily and at some point cause ions to leave their stable trajectories with smaller ions being lost from the ion pool first. Alternatively, an additional AC can be applied to the exit rods in order to eject ions in a directed way towards a detector. If the frequency of the exit rod AC equates the frequency of the secular motion of a peptide ion, resonance will be induced, which will eventually cause the ion to leave the trap through the exit slits (resonance ejection). The time it takes the ion to exit can be controlled by the applied voltage (i.e. the amplitude) of the exit rod AC. Ion selection within an ion trap is achieved at a constant main RF by altering the exit rod AC to remove unwanted ion species. The selection process can be accomplished very quickly via simultaneous resonance ejection [216]. To this end, all AC waves that match frequencies of unwanted m/z values are superimposed to an isolation waveform and only those waves that match ion species that are to be isolated are excluded. This way, it is possible to isolate several analytes of distinct m/z in parallel, a process called synchronous precursor selection (SPS) [222].

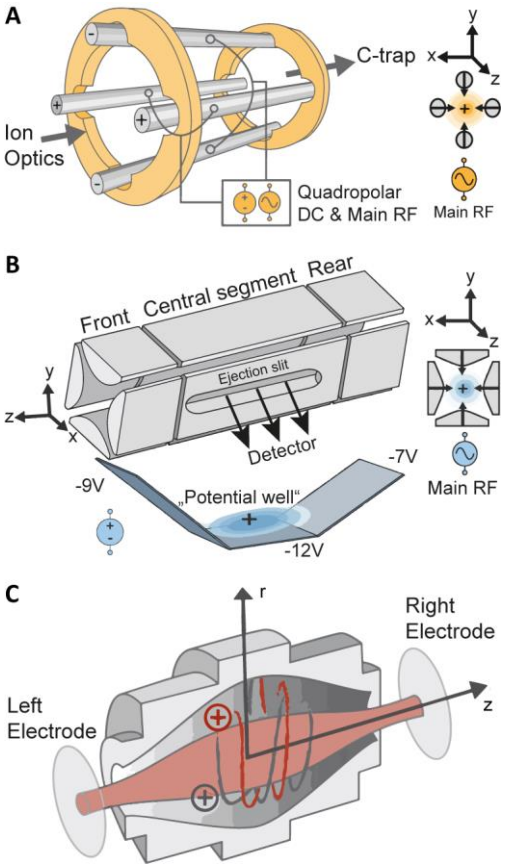



**Figure I-12 | Mass filter and analyzers in the Fusion Lumos mass spectrometer.** *(A) A quadrupole can filter ions via the interplay of a quadrupolar direct current (DC) and an alternating current (main RF). (B) An ion trap can additionally store analytes enabled by a potential well caused by differing DCs between the central and outer segment. Spectra can be acquired by ramping the main RF causing an ejection of ions through the slits, which then can be recorded with an electron multiplier. (C) An Orbitrap mass analysers detects ions via the current that is induced in the outer electrodes by their circular motion of ions around the central, rod-shaped electrode (adapted with permission from [216]).*

Unlike ion selection, spectra acquisition in the ion trap is performed at a fixed frequency of the exit rod AC, while the main RF is ramped up. Peptide ions with increasing masses are ejected sequentially and in a controlled way as soon as their motions' escalating frequencies match the frequency of the exit rod AC [216]. Accordingly, the stream of resonance-ejected ions can be quantified with an electron multiplier. When a single ion hits the surface of such a detector, several secondary electrons are released which themselves will prompt emission of more electrons as soon as they strike the detector surface [223]. In this way, an electron cascade is induced that multiplies the signal until it is collected by an anode. This signal amplification together with the ion accumulation capability, renders ion traps highly sensitive at relatively high scan speeds but with comparably poor mass accuracy and resolution [220, 224].

Superior mass accuracy and resolution is provided by the high-field Orbitrap mass analyser in the Fusion Lumos [225, 226]. It consists of a coaxial, spindle-shaped electrode that is surrounded by

two barrel-like outer electrodes (Figure I-12 C). Since Orbitraps cannot collect or store ions themselves, they are combined with trapping devices. To perform an Orbitrap scan, collected ion packets are injected from the C-trap into the Orbitrap with an offset from the central electrode [227]. A voltage ramp applied during injection squeezes ions towards the spindle-shaped electrode with smaller m/z species moving closer to it than larger ones. This prevents different ion species from colliding when they start to orbit around and oscillate along the central electrode [220]. This movement is caused by a quadro-logarithmic, electromagnetic field that is created by a DC potential between the conical inner and outer electrodes. The frequencies of the harmonic oscillations in the axial dimension depend on the m/z ratios of ions with smaller ions featuring higher frequencies [225]. Based on this known relation, image currents that are induced in the outer electrodes by oscillating ions can be utilized to deduce their m/z ratios. Accordingly, superimposed, highly complex currents of all moving ions are recorded in the time-domain, Fourier-transformed into the frequency domain, and subsequently converted into mass spectra. While Orbitraps are not as sensitive as ion traps, they offer sub-ppm mass accuracy [227] and an extraordinary resolution that increases with oscillation time and can reach up to 1 million [228].

**Peptide fragmentation and tandem mass spectrometry** – If the resolution and mass accuracy of the m/z analysis are high enough, characteristic isotope patterns of peptides that arise from the incorporation of stable, heavy isotopes (in proteins most frequently $^{13}C$ and $^{15}N$) can be determined. The knowledge that the isotope peaks must differ by the mass of one neutron can be utilized to derive the protonation state from the spacing of isotope peaks in the m/z scale. Via subsequent calculation of its accurate neutral mass from the monoisotopic peak, the exact elemental composition of a peptide can be determined [229]. Yet, often there are still many isobaric, but structurally different peptides to which a certain mass could match. Therefore, bottom-up proteomics approaches narrow down potential peptide matches by additionally identifying (at least partially) the sequence of peptides through a tandem MS approach [230]. This refers to the acquisition of peptide spectra (precursor, MS1, or (full) MS spectra), an isolation of an individual peptide ion species (e.g. in a quadrupole or ion trap), its subsequent fragmentation, and recording of m/z ratios of peptide fragment ions in fragment spectra (MS2 or MS/MS spectra; Figure I-9). While the Tribrid series offers a multitude of fragmentation options employing either physical force or chemical reactions [221, 231, 232], most commonly peptides are fragmented using collision-induced dissociation (CID) [233], or higher-energy collisional dissociation (HCD) [234]. Both fragmentation modes are characterised by an excitation of ions and a collision with inert gas molecules like nitrogen, helium, or argon, and only differ in their way of ion excitation. Following clashes with gas molecules, the kinetic energy of peptide ions will be transformed into vibrational energy, which will accumulate from repeated impact and ultimately lead to the breakage of chemical bonds within the analytes.

In the Fusion Lumos, CID is performed in the high-pressure cell of the linear ion trap filled with inert gas molecules [221]. Similar to the process of controlled resonance ejection, resonance of the analyte ion of interest is induced by setting the frequency of the exit rod AC to match the frequency of their secular motion. For fragmentation, however, a smaller voltage for the exit rod AC is set allowing the ions to stay in resonance longer to allow them to collide with the gas molecules causing the peptides to fragment [216]. The resonance time (also called activation time) determines the extent of fragmentation. Small fragment ions are often lost in this ion trap-based fragmentation mode since they cannot be kept on stable trajectories with the main RF that is needed to keep precursor ions long enough in resonance for efficient fragmentation [216]. In contrast, HCD is performed in a quadrupolar or octopolar collision cell, which in the Tribrid series is

called ion routing multipole (Figure I-11). Here, ions gain kinetic energy while they are accelerated from a neighbouring mass analyser into the gas-filled collision cell applying a DC offset [234]. This leads to higher kinetic energies and shorter impact times compared to CID fragmentation. Since in HCD the excitation and fragmentation processes are physically separated, the main RF in the collision cell can be adjusted to retain also small peptide fragment ions without compromising fragmentation efficiency.

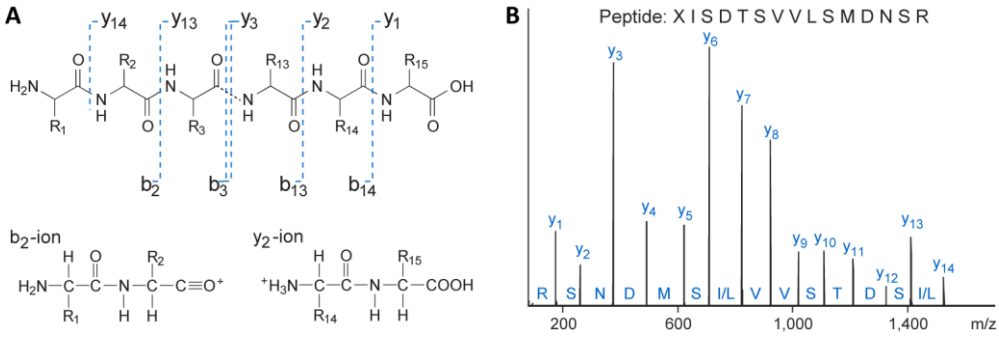



**Figure I-13 | Peptide fragmentation upon collision-type dissociation methods**. *(A) Via collisions with inert gas molecules, CID and HCD typically lead to the cleavage of peptide bonds producing b- and y-ions. (B) In an ideal case, fragment ions differ by one amino acid allowing the direct determination of the complete peptide sequence from mass differences between fragment peaks. For the sake of simplicity, only y-ions are displayed (adapted with permission from [234]).*

Both collision-type fragmentation modes mostly result in fragment ions with an intact amino- and carboxy-terminus that are named b- and y-ions, respectively [235] (Figure I-13). However, CID and HCD spectra still are still somewhat different due to their difference in precursor excitation. Since HCD does not suffer from a lower mass cut-off, it enables the robust detection of smaller fragment ions such as immonium ions that can be diagnostic for the presence of certain amino acids or modifications [234]. At the same time, HCD yields less of the larger b-ions due to higher energies applied and secondary fragmentation taking place. In contrast, activation in CID is comparably slow and leads to a preferential breakage of lower energy bonds, which produces neutral loss ions more frequently than in HCD. They arise from the precursor ion when neutral fragments like water or ammonia are lost [236]. For phosphoserine and -threonine containing peptides a loss of phosphoric acid (-$H_3PO_4$) commonly dominates CID spectra and impedes successful phosphopeptide identification [237]. This can be overcome by a second resonance excitation of the generated neutral loss ion in the ion trap, a fragmentation method termed multistage activation (MSA, also known as pseudo-$MS^n$ [238]). Although this fragmentation mode is slower, it improves identification rates for phosphopeptide spectra significantly because it increases their information content.

**Examples of data acquisition strategies on hybrid instruments** – Besides several fragmentation modes, the instrument architecture and software options on the Fusion Lumos also allow for various data acquisition regimes. Shotgun proteomics approaches often employ a so-called data-dependent acquisition (DDA) strategy. Here, the instrument automatically switches between recording of MS1 and MS2 spectra, meanwhile consecutively isolating typically the N most abundant peptides in the preceding MS1 scan for fragmentation. A dynamic exclusion prevents repeated fragmentation of the same m/z value for a defined time and thus increases the amount of sampled peptidoforms. On the Lumos, peptide scans are obtained in high resolution and high mass accuracy mode utilizing the Orbitrap, whereas fragment spectra can either be acquired sequentially in the Orbitrap or in parallel to the MS1 scan in the ion trap, which enables much faster scan rates [221]. The DDA approach is very convenient since no prior knowledge about the sample is

required. However, the stochastic selection of precursors for fragmentation can lead to irreproducible (i.e. missing) identification and quantification across different samples, for example for lower abundant or strongly regulated peptidoforms. Methods targeting a predetermined set of analytes of interest can overcome this reproducibility issue, but are not hypothesis-free and require prior knowledge. In such targeted measurements like parallel reaction monitoring (PRM [239]), the mass spectrometer only targets predefined m/z values at specified retention times for fragmentation without any dynamic exclusion. Consequently, compared to DDA methods, a much smaller number of peptides is analysed but with superior sensitivity and reproducibility across runs [240].

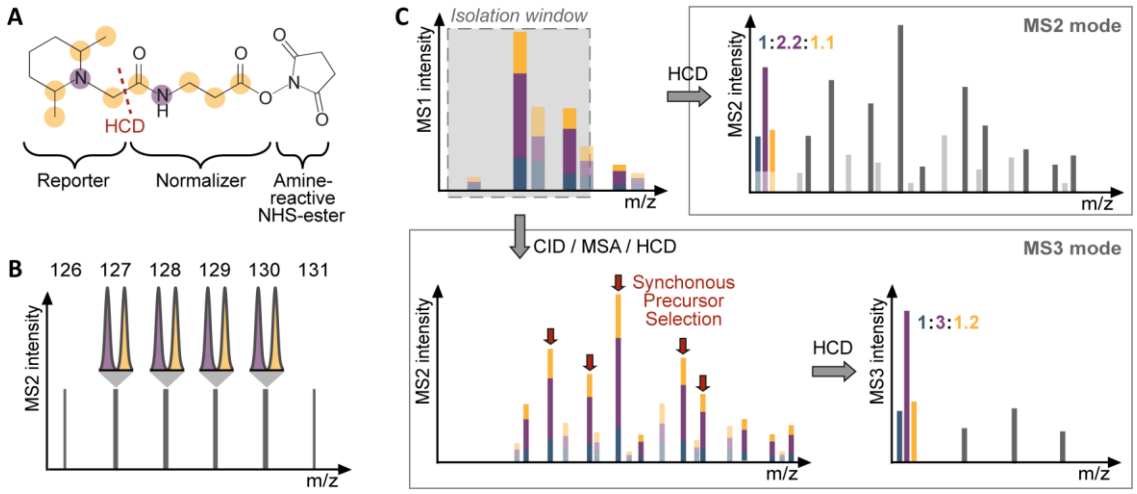



**Figure I-14 | Isobaric tags and data acquisition in MS3 mode.** *(A) The structure of primary amine-reactive tandem mass tag (TMT) is displayed. In each tag, 5 of the marked carbon (yellow) or nitrogen atoms (purple) are replaced with heavy isotopes (NHS: N-hydroxysuccinimide). (B) Upon fragmentation, sample-specific reporter ions are released. For TMT10-plex, high-resolution scans are required to differentiate slightly lighter reporter ions that contain $^{15}N$ (purple) from those that contain the same number of heavy isotopes but exclusively $^{13}C$ (yellow). (C) Co-isolation and co-fragmentation of interfering peptides (lighter colour) can result in ratio compression in MS2 mode measurements. This can be minimized with an MS3 acquisition mode, where a second isolation and fragmentation step of peptide fragments is performed.*

Isobaric labelling and subsequent sample multiplexing present another approach for more reproducible (but untargeted) identification and quantification across different samples. Isobaric tags carry heavy isotopes of carbon or nitrogen on different positions of the molecule and consist of a reactive group for protein or peptide labelling, a mass normalizer, and a cleavable mass reporter (an example of amine-reactive tandem mass tags (TMTs [241]) is illustrated in Figure I-14 A). After labelling and pooling of different samples, the intensity signal of a peptidoform will be summed up for all multiplexed samples in the MS1 spectrum since all isobaric tags feature identical physicochemical properties and thus behave the same during sample chromatography. Only after peptide fragmentation, sample-specific reporter ions will be released in the low m/z range with different masses (owing to differing numbers of $^{13}C$ and $^{15}N$ atoms, Figure I-14 B). Their intensities can be used as a proxy for the relative abundance of peptidoforms in different samples. One substantial shortcoming of quantification with isobaric tags is the frequently observed distortion of true ratios that can arise from co-isolation and co-fragmentation of differentially regulated peptides [242-245] (Figure I-14 C). The quantification accuracy of such peptide tagging strategies can greatly benefit from advanced MS3 methods that are available on the Tribrid instrument series [222, 246]. In MS3-mode, peptide identification and quantification are decoupled by an additional

selection and fragmentation step that reduces interfering ions and significantly reduces ratio distortion in the subsequent MS3 scan. To this end, an MS2 spectrum for peptide identification is acquired followed by re-fragmentation of the same peptide and isolation plus high-energy fragmentation of its most abundant peptide fragment(s) [246]. The resulting TMT reporter ions are recorded in an MS3 scan, which typically constitutes a better reflection of true ratios. Importantly, in so-called MultiNotch MS3 measurements, the ion trap selects several of the top N most abundant peptide fragments simultaneously (SPS, see p. 25) thereby increasing overall reporter intensities in the MS3 spectrum and improving sensitivity [222] (Figure I-14 C). The MS3 scan is commonly acquired after HCD at high energies to ensure an efficient release of the reporter ions from peptide fragments. Noteworthy, isobaric tags such as TMT10-plex exist that yield reporter ions that do not differ in the absolute number of heavy isotopes but only in whether a $^{13}C$ or $^{15}N$ atom is present (i.e. whether the extra neutron is located within the carbon or nitrogen atom) [247]. For these tags, the quantitative scan needs to be acquired in the Orbitrap at high resolution to resolve the small mass difference of ~6.3 mDa resulting from the difference in nuclear binding energies. Apart from these terms, fragmentation modes, number of peptide fragments for the SPS, and mass analysers employed for the MS2 scan can be chosen flexibly on the Fusion Lumos, offering great potential for custom-made methods, for example for PTM-enriched samples [248-250].

## 3.3 Analysis of mass spectrometry data

Subsequent to data acquisition, qualitative and quantitative analyses of raw data provides information about peptides and, accordingly, proteins. In principle, peptide masses can be derived from MS1 spectra (utilizing the m/z value and spacing between isotope peaks), and corresponding amino acid sequences can be deduced from fragment spectra (mass difference between fragments). In reality, however, fragmentation patterns are often incomplete and intervening peaks, for instance from co-isolated peptides, may complicate peptidoform identification [230]. Moreover, modern mass spectrometers produce thousands of spectra in a short time, which precludes manual annotation. Therefore, a multitude of software solutions has been developed for peptide and protein identification and quantification, and the computational pipelines are typically specifically adapted for certain data acquisition approaches and research questions at hand.

**Peptidoform and protein identification** – Except for *de novo* sequencing algorithms [251] (which will not be covered here), peptide identification is commonly based on some type of matching of acquired fragment spectra either to *in silico* generated spectra from a database (sequence database approach [252]) or to a collection of previously acquired, experimental spectra (spectral library matching [253]). For database searching, many search engines such as Mascot [252], Andromeda [254] (integrated in MaxQuant [255]), and Spectrum Mill are available. Spectra are usually pre-processed to improve signal-to-noise and clean up the m/z space. This can include but is not limited to centroiding, de-isotoping and charge state deconvolution [256]. Afterwards, observed fragment masses are matched to a plausible set of theoretical spectra that are obtained from *in silico* digestion of known and putative protein sequences derived from genomic or transcriptomic data (Figure I-15 A). Theoretical spectra need to resemble the specificity of the enzyme used for digestion and should account for expected fixed (such as carbamidomethylation of cysteines) and variable modifications (such as N-terminal acetylation) or optionally introduced labels. Further, the fragmentation mode should be specified to factor in the preferential occurrence of different fragment ions and provide the most likely fragment ion series for matching. Subsequently, probability-based search engines like Mascot and Andromeda compute probabilities that

a peptide spectrum match (PSM) may be a random event meanwhile accounting for the number of matched fragments amongst other criteria. Probabilities are converted into scores using a negative log transformation [252]. Consequently, PSMs with high scores have a low probability of being wrong, and the highest scoring PSM will be reported as the (most likely) identified peptide. Since probability metrics often differ between algorithms, identical spectra can lead to different, highest ranked PSMs using different software solutions [257]. Furthermore, false matches may arise from poor spectra quality, sub-optimal search parameters (e.g. when neglecting a frequently occurring modification), or simply by random chance. To control for erroneous matching, generally a false discovery rate (FDR) is estimated using target-decoy approaches [257]. To this end, experimental spectra are re-searched against an equally sized decoy database that contains reversed or scrambled sequence information. Hits to this decoy database are wrong by definition. Assuming that false matches in the target database follow a similar distribution than hits in the decoy database, an FDR can be estimated (e.g. number of decoy hits/number of target hits). Subsequently, an (arbitrary) FDR cut-off is set (usually ≤ 1 %) to improve the general identification quality [182].

Variable PTMs pose a particular challenge for database searches since all the possible permutations of modified sequences vastly increase the search space, which can potentially compromise identification quality [258, 259]. In addition to sequence identification, the position of a modification has to be correctly determined [260]. This can be especially difficult in phosphoproteomics experiments where multiple possible sites can exist in a given peptide sequence. In theory, neighbouring fragment ions (so-called site determining ions) are indicatory of a modified amino acid, but they can be absent or of low abundance making a distinction from noise difficult. Several algorithms have been established to estimate the positional certainty of modifications (reviewed in [260]). They are typically based on an estimation of the likelihood that a site-determining ion is present by chance [261] or on the difference in spectrum matching likelihood between different positional modification isomers [262]. Owing to these increased identification challenges, modified peptides are often filtered through additional cut-offs for identification scores or delta scores (the score difference to the next best hit in the database). In addition, identifications of modified peptides are frequently further filtered for localization probabilities to gain confidence on the modification position prior to functional analyses [263].

Following quality assessment and filtering of PSMs, identified peptide sequences can be assigned to corresponding proteins, which is not a trivial step due to the protein inference problem [264]. Proteins can only be identified without doubt by peptides that are unique to them, but often peptides are shared between proteins (commonly termed razor peptides, Figure I-8 and Figure I-15 B). Therefore, search algorithms usually report groups that include all proteins that share peptides with other identified proteins in the group and are not identified with any unique peptide. Consequently, splice variants will often end up in the same protein group and all proteins that are potentially but not necessarily present in a sample are listed which clearly complicates proteoform analysis with bottom-up proteomics.

In contrast to DDA-type of experiments, targeted analyses are hypothesis-driven and generally rely on high-quality spectral libraries for method development as well as affirmative identification after data acquisition [240]. Such approaches assess the similarity of experimental spectra to previously acquired, high-confidence PSMs that can be obtained from DDA-type of experiments or measurements of synthetic peptides (Figure I-15 A). They are based on the assumption that fragmentation patterns are highly deterministic given similar measurement parameters. To ensure direct comparability, spectral libraries are usually not de-charged and de-isotoped. Matching

quality can be assessed with specialized software solutions such as Skyline [265]), which extracts elution profiles of fragment ions at the expected retention times and demands them to be present in the expected intensity ratios. Accordingly, similarity scores are provided which do not only necessitate the presence of certain fragment peaks but also account for characteristic intensity patterns [266]. Ultimately, targeted analyses commonly include careful manual inspection of ion traces and potential interfering peaks to ensure high confidence identification and subsequent robust quantification [240] (Figure I-16).

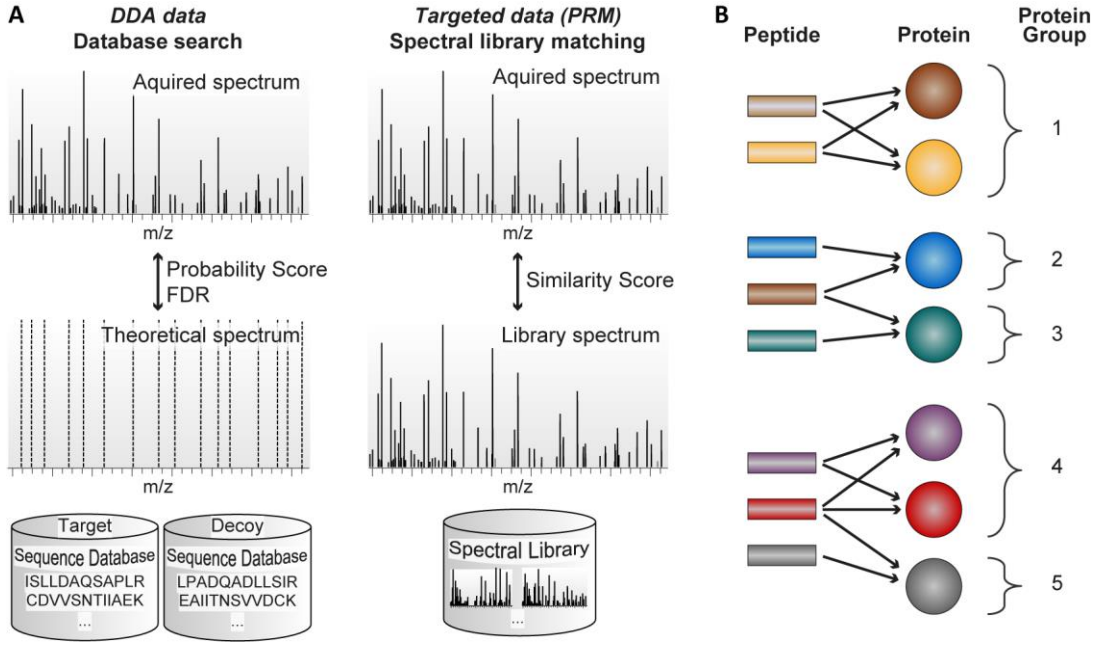



**Figure I-15 | Peptide and protein identification.** *(A) Spectra can be matched to peptides employing a target-decoy database search strategy or via comparison to previously acquired, spectral libraries. (B) Protein inference from peptide-spectrum matches is complicated by shared (razor) peptides. While typically all proteins to which a peptide matches are listed in a protein group, they may not actually be present in the sample (e.g. the red protein could be absent since purple and grey protein are sufficient to account for the purple, red, and grey peptide).*

**Peptidoform and protein quantification** – In addition to identification, quantification is an integral part of mass spectrometry-based proteomics allowing an estimation of abundances for thousands of analytes in parallel. Early quantitative assessments were based on counting of PSMs under the assumption that proteins that are more abundant in one sample yield more of corresponding spectrum matches compared to another sample [267]. Spectrum counting approaches are straightforward, but they are highly dependent on measurement parameters such as dynamic exclusion and require a certain number of data points [268] rendering it inapplicable for quantification of peptidoforms. Today, quantification is more commonly based on measuring the MS signal response of peptides, which is proportional to their abundance over four orders of magnitude [269], hence providing a means to accurately quantify single peptides. Importantly, only intensities of identical peptides can be compared across samples since the response of different peptides may differ vastly, for example owing to unequal losses (e.g. absorptive losses of hydrophobic peptides) or differing ionizability. A multitude of quantification approaches is available depending on the sample preparation and data acquisition method employed (reviewed in [267, 270, 271], for examples see Figure I-16). Since respective data processing algorithms are often integrated in the computational pipelines that are also utilized for identification purposes, in most cases, quantitative values can be obtained easily in an automatic fashion [255, 265]. The MaxQuant software, for

example, was originally developed for SILAC samples [272], but now supports all kinds of label-based and label-free quantification approaches [255].

For metabolic labelling methods such as SILAC [167] (see also pp. 15), differentially labelled and treated samples are combined earliest in the workflow thereby minimizing potential variations introduced during sample processing. Peptides that contain heavy isotope-labelled amino acids have identical physicochemical properties as non-labelled ones and can only be distinguished in the mass spectrometer based on their mass difference. This facilitates accurate relative quantification within one MS run which is achieved by comparing summed MS1 intensities of isotope cluster as a function of retention time (Figure I-16). Based on the availability of heavy isotope-labelled amino acids, typically no more than three samples are combined in a SILAC experiment in practice. Importantly, the number of MS1 features scales linearly with the number of channels. This leads to more complex MS1 spectra, increased co-isolation resulting in chimeric MS2 spectra, and repeated sampling of the same peptide in different labelling states, all of which reduce the overall number of distinct peptide identifications. This can be prevented when employing an MS1 intensity-based label-free quantification (LFQ) approach. The underlying quantification principle is the same as for SILAC samples, but here intensities of elution profiles are compared across samples measured in different runs (Figure I-16). LFQ allows for comparison of a practically infinite number of samples and is applicable to any cellular system, also to those that are not (easily) accessible with metabolic labelling approaches (e.g. human tissues). Since samples are only combined at the level of data analysis, a reproducible and robust sample processing workflow is key to prevent an accumulation of large variations that can bias quantification. As mentioned above (see pp. 28), the stochastic nature of peak picking in DDA-type of experiments can quickly increase the number of irreproducible, missing identifications and quantifications in large LFQ sample sets. This equally applies to SILAC data when only one of the isotope labelling duplets or triplets is being identified via a PSM. Such missing values complicate statistical analyses and can obscure important biological information. To overcome this issue, algorithms have been developed to transfer identifications from one labelling isotope pattern or label-free run to another [255, 272, 273]. In brief, corresponding peptide elution profiles are aligned and matched utilizing accurate m/z and retention time information, and subsequent quantification of matched profiles reduces missing values. It is important to note that a reproducible performance of the on-line coupled chromatography system is inevitable for a successful utilization of such match-between-runs approaches. Furthermore, these algorithms cannot eliminate missing values completely. MS1 based quantification approaches also need to balance acquisition of MS1 and MS2 spectra carefully to allow for confident run alignment. This is generally important to facilitate feature detection in the MS1 space, achieve accurate quantification of elution profiles, and at the same time permit a reasonable proteome depth via a sufficient quantity of potential PSMs.

An alternative to MS1-based strategies is offered by quantification using isobaric tagging approaches (e.g. TMT, see also pp. 28 and Figure I-14). They rely on reporter ion intensities extracted from fragment spectra and their quantification reproducibility is typically less sensitive to performance variations of the LC-MS systems than for label-free measurements. In addition, the high multiplexing capability of TMT reagents (recently increased to 16-plex) reduces measurement time and enables a deep proteome coverage for multiple samples within a reasonable timescale. Together, these advantages likely explain why isobaric tagging approaches are increasingly popular for large-scale and cross-laboratory studies [194, 274, 275] although the reagents are fairly expensive. Noteworthy, the circumvention of missing values additionally makes them attractive for peptidoform studies that depend on robust identification and quantification of single

peptides across conditions. However, as mentioned above (see pp. 28), accuracy of isobaric tag-based quantification can be compromised by ratio distortion. In addition to above described MS3-based acquisition approaches [222, 246], several strategies have been suggested to tackle this issue. Those range from adjusting sample preparation workflows (extended fractionation [244]) and optimizing data acquisition parameters (gas-phase purification of isolated peptides [276], fragmentation at elution peak maximum, narrowing isolation windows [277], utilizing ion-mobility [278, 279]) to customizing data processing (computational removal of co-isolated intensities [245], utilizing complement reporter ions [280]). While most of the proposed methods have proven great merit for the reduction of ratio compression, they frequently come at costs of decreased throughput, or they necessitate specialized instrumentation and data processing workflows, which can limit their overall applicability.

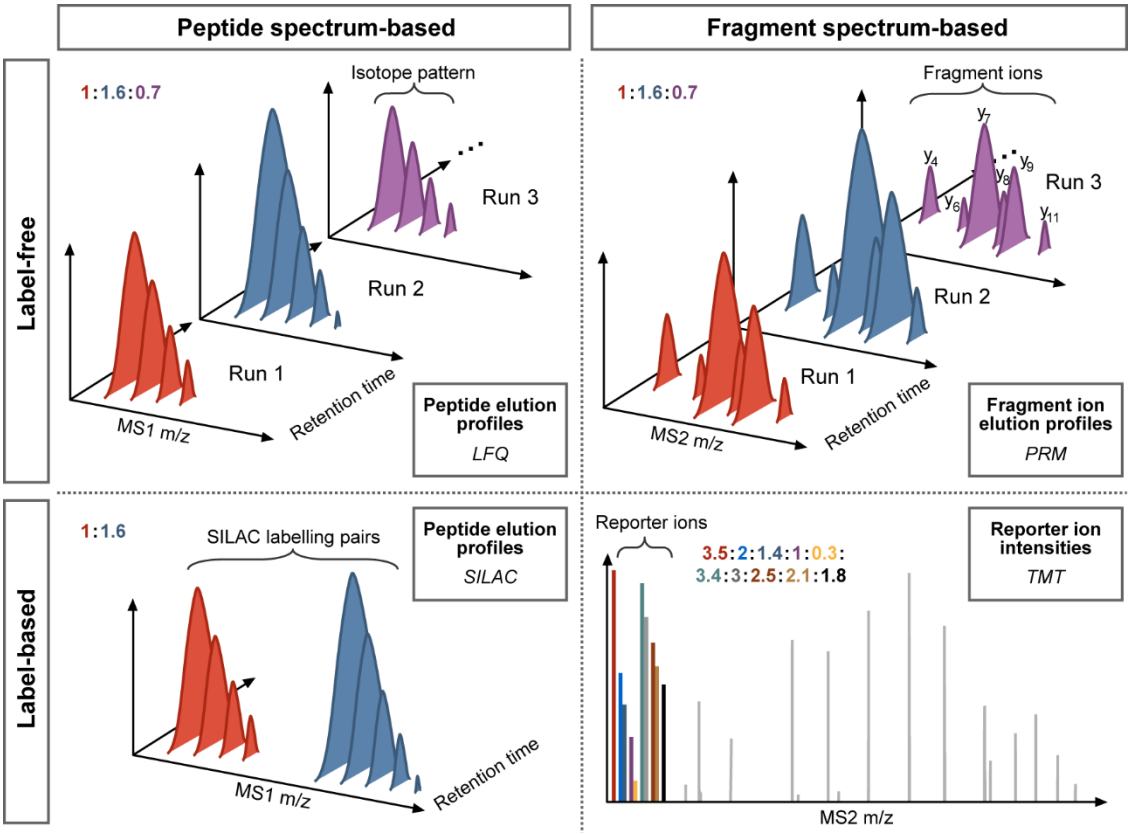



**Figure I-16 | Common quantification approaches in bottom-up proteomics**. *Peptide quantification can be performed label-free or label-based and on peptide precursor or fragment level. Many methods integrate elution profiles to obtain area-under-the-curve values as proxy for peptide abundances. Isobaric tag-based approaches, however, directly utilize reporter ion intensities. Examples for the different quantification approaches are indicated in boxes.*

Generally, peptidoform quantification often already requires some form of integration of quantitative information, for example from peptide elution profiles recorded in differing charge states or detected in multiple off-line fractions of the same sample. This is often achieved by simple signal summation, but can also include more sophisticated data integration steps [281]. Owing to the protein inference problem, consolidation of quantitative peptide information on protein level can pose a substantially greater challenge. Protein group intensities can be influenced considerably depending on how razor peptides are handled. They may be derived from all peptides matching to a certain group (including group-razor peptides) or only from those that exclusively match to this group. Group-razor peptides may also be assigned based on the principle of

parsimony (Occam's razor), which means they are allocated solely to the protein group with the strongest evidence (i.e. the one with the most peptides [272]). Frequently, additional filters for peptides that are included for protein quantification are applied. As an example, post-translationally modified peptides and their unmodified counterparts are commonly excluded, and at least two quantified peptides can be demanded per protein. Subsequently, protein intensities can be computed, for example by summation or averaging of all peptide signals, sometimes only including a fixed number of the highest intense peptide features [282]. Alternatively, quantitative protein fold changes (FCs) can be computed as median of all peptide FCs [272, 281]. This increases robustness of quantification and minimizes the influence of outliers that may amongst other thing result from ambiguous assignment of razor peptides.

Following quantification, typically data normalization on peptide or protein level is essential to optimize quantification performance and correct for systematic biases (e.g. introduced by losses during sample processing). Normalization procedures may be provided by database search algorithms [281], but their applicability should be evaluated on a case-by-case basis. In some instances, a customized normalization procedure that accounts for the specific experimental design by making reasonable assumption about the expected data structure can be advisable.

**Data interpretation** – Noteworthy, measured and normalized intensities are not readily comparable across proteins because peptide signal responses may differ and accumulated protein intensities are highly dependent on additional parameters like protein length and the number of theoretically observable peptides. Different strategies that normalize intensities for these biases such as iBAQ (intensity-based absolute quantification [113]) have been proposed to obtain information on relative expression levels and enable a direct comparison of different proteins. In addition, computation of absolute protein copies for deeply fractionated proteomes by harnessing histone protein intensities has shown promising concordance with spike-in standards for absolute quantification [283]. However, these approaches rely on a sufficient number of peptides per protein (group) and are therefore mostly incapable of estimating absolute abundances of different protein isoforms. Likewise, absolute copy numbers of proteins carrying a certain modification cannot be obtained directly from the intensity of a specific peptidoform alone. Instead modification stoichiometries (also called occupancies) can be calculated, for example when quantitative information of a modified peptide, its unmodified counterpart, and the whole protein (i.e. several other peptides) is available in at least two different sample conditions [61, 63]. The knowledge about occupancies can provide another layer of helpful information for the elucidation of mechanisms underlying the regulation of cell signalling events and proteostasis [205]. Owing to the sheer amount of data generated by MS-based proteomics methodologies, an efficient biological interpretation of experimental results is further often critically dependent on prior functional annotations [284]. Numerous resources are available to guide data exploitation, such as general protein knowledge databases like UniProt [43] and Gene Ontology [285], or specialized repositories for signalling pathways [286], protein classes [287], domains [288], interactions [289], complexes [290], subcellular localizations [291, 292], structures [293], and modifications [294].

It becomes clear that a multitude of intricacies in proteomic workflows renders the comprehensive study of proteoforms quite challenging. However, MS-based proteomics is currently the only methodology to study cellular functionalities beyond the gene level and with high throughput. Therefore, it offers an unprecedented opportunity to observe and investigate complex biological processes and ultimately to understand the processes that govern life.

# 4 Objective and outline

The dynamic equilibrium of protein turnover contributes to a cell's capability to respond rapidly to external stimuli by altering the abundance of functional proteins. It becomes increasingly apparent that dysregulations of processes that regulate protein turnover, such as post-translational modifications, are involved in a multitude of disease states. Thus, there is a growing need for methods that enable the global determination of protein turnover on the level of peptidoforms to elucidate factors implicated in its post-transcriptional and post-translational control. Recent advances in mass spectrometry-based proteomics suggest that global measurements of proteoform-specific turnover are feasible in principal, but so far, they have been widely neglected due to technical challenges that accompany such measurements. The objective of this work was to harness the power of recent technological improvements to unravel proteoform-specific protein turnover and, in particular, investigate the impact of acetylation, ubiquitination, and phosphorylation on protein stability on a proteome-wide scale.

Initially, different quantification approaches were benchmarked to identify a strategy that enables robust quantification of single, modified peptides across multiple conditions. The most promising approach was further optimized regarding costs, robustness, and accuracy of measurements (Chapter III). Gained insights were utilized to establish a novel method for the determination of turnover rates with improved proteome coverage and single peptide resolution. This facilitated the examination of global and proteoform-specific determinants of protein turnover in human cells (Chapter IV). Finally, this new approach was combined with the analyses of the post-translational modification state to discover modification-specific protein turnover, which revealed the tremendous potential of this approach to identify functional relevant modification sites (Chapter V).

# II GENERAL METHODS

**1 Protein and peptide sources ..... 39**

1.1 Cell cultivation and treatment ..... 39

1.2 Protein extraction ..... 40

1.3 Synthetic phosphopeptides ..... 40

**2 Sample preparation for mass spectrometry analyses ..... 41**

2.1 Protein digestion ..... 41

2.2 Sample clean-up by solid-phase extraction ..... 41

2.3 Tandem mass tag labelling ..... 41

2.4 Metal ion affinity purification of phosphorylated peptides ..... 42

2.5 Immunoaffinity purification of ubiquitin-remnant and acetylated peptides ..... 42

2.6 Peptide fractionation ..... 43

**3 Mass spectrometry data acquisition, processing and analysis ..... 45**

3.1 Data dependent LC-MS measurements ..... 45

3.2 Database searching ..... 46

3.3 Data processing and analysis for comparison of quantification approaches ..... 47

3.4 Processing of time-course pulsed SILAC-TMT data ..... 48

3.5 Processing of single time-point pulsed SILAC data ..... 52

3.6 Turnover computation and underlying kinetic model ..... 52

3.7 Parallel reaction monitoring assay for NADH dehydrogenase complex ..... 54

3.8 Functional and statistical data analyses ..... 55

# 1 Protein and peptide sources

Details about the protein source, number of cell culture replicates, and treatment and pulse time-points for certain experiments can be found in the experimental procedure sections of the respective chapters (see p. 64, pp. 103, pp. 133).

WHIM2 and WHIM16 basal and luminal breast cancer PDX (patient-derived xenograft) models were generated as previously described [295] and PDX and Jurkat peptides were kindly provided by the Broad Institute of the Massachusetts Institute of Technology (MIT) and Harvard in Cambridge, Massachusetts. Peptides from murine liver tissue were kindly provided by the Max Delbrück Center for Molecular Medicine in the Helmholtz Society in Berlin, Germany. All other samples were generated in-house.

## 1.1 Cell cultivation and treatment

For method evaluation and optimization experiments, HeLa, K562 and Jurkat cells were cultured at 37 °C and 5 % $CO_2$ in DMEM (Dulbecco's Modified Eagle's Medium), IMDM (Iscove's Modified Dulbecco's Medium) and RPMI-1640 medium, respectively. The media were supplemented with 10 % FBS (Gibco™) and RPMI-1640 medium additionally with 1 % antibiotic, antimycotic solution (Sigma). The haploid *Saccharomyces cerevisiae* cell line Y187 was cultured at 30 °C on yeast peptone dextrose (YPD) agar plates or in YPD broth on a shaker at 350 rpm.

**Steady-state pulsed SILAC experiments** – For pSILAC experiments, HeLa cells were cultured in SILAC DMEM (Thermo Fisher Scientific) supplemented with 10 % dialyzed FBS (GibcoTM), 1 % antibiotic, antimycotic solution (Sigma) and 1.74 mM L-proline (≥ 99 %, Sigma). L-lysine and L-arginine were added to a final concentration of 0.798 mM and 0.398 mM, respectively, in either light (Lys-12C614N2/K0, isotope purity ≥ 99%; Arg-12C614N4/R0 ≥ 98 %, Sigma) or heavy form (Lys-13C615N2/K8, isotope purity ≥ 99 %; Arg-13C615N4/R10, isotope purity ≥ 99 %, Cambridge Isotope Laboratories). Completely heavy labelled cells for label swop experiments were obtained after cultivating light-labelled cells in medium containing K8 and R10 for at least 10 cell doublings. For time-course pSILAC experiments, HeLa cells were seeded at 5.7e3 cells/cm2 to avoid overgrowth during the time-course of the experiments. Pulses were started 40 h after cell seeding by removing old medium, washing light (or heavy) labelled cells twice using PBS (phosphate-buffered saline) with Mg2+/Ca2+, and adding K8R10 (or K0R0) medium. For determination of cell doubling times during pulse experiments, HeLa cells were seeded in 96 well plates at the same density and, after medium exchange, counted in six replicates every 12 h.

**Cell viability assays and treatments** – Cell viability assays were conducted using alamarBlue reagent (Thermo Fisher Scientific) according to the manufacturer's protocol. Tanespimycin (17-AAG, Selleckchem) treatment was started 24 h after seeding of 1e3 K562 cells in 96-well. Cell viability was read-out before treatment start and 72 h after incubation with increasing drug concentrations (0, 3, 10, 30, 100, 300 nM, 1, 3, 10, 30 µM). The lowest 17-AAG concentration that showed a full cytostatic effect (1 µM) was applied in time-dependent treatments (0, 0.5, 1, 2, 4, 6, 8, 16, 24, 32 h) of K562 that were seeded 24 h prior to treatment at 5.1e5 cells/ml.

For the investigation of the effects of rotenone induced, oxidative stress on turnover of respiratory chain complex I (NADH dehydrogenase) proteins, HeLa cells were seeded in K0R0 medium at 3.3e4 cells/cm$^2$. Following 40 h of cultivation, 1 µM rotenone (≥ 95 %) in DMSO and 5 mM L-(-)-malic acid and L-glutamic acid (Sigma) were added to the cells. Cells treated with DMSO plus 5 mM L-(-)-malic acid and L-glutamic acid or DMSO only served as control. After 30 min, K0R0 medium

was removed, cells were washed twice using PBS with $Mg^{2+}$/$Ca^{2+}$, and K8R10 medium containing the above stated treatment or respective control supplements was added to cells.

## 1.2 Protein extraction

HeLa and Jurkat cells were lysed at 80 to 90 % confluence or after indicated pulse or treatment time-points (corresponding to a lower cell density). Yeast cells were harvested in the early log-phase at OD 600 = 1. Cells lines and tissues were lysed after 2 PBS washes in 8 M urea in 40 mM Tris-HCl, pH 7.6 (Hela, yeast) or 75 mM NaCl and 50 mM Tris, pH 8.0 (Jurkat, mouse liver, PDX) containing protease inhibitor (cOmplete™ Mini, Roche) and phosphatase inhibitor cocktails (prepared in-house according to the formula of Phosphatase Inhibitor Cocktail 1, 2, and 3 from Sigma). Lysate buffer volumes were adjusted to cell density and cell culture vessel to reach concentrations of at least 1 mg protein per ml buffer (typically 2 to 5 mg/ml). Lysates were incubated on ice for 10 to 30 min. Yeast cells were, in addition, mechanically disrupted through four cycles of bead beating for 5 min followed by 2 min of cooling on ice. The yeast lysate was centrifuged for 10 min at 13,500 g and 4 °C to sediment the glass beads and the supernatant was collected. All lysates were stored at - 80 °C.

## 1.3 Synthetic phosphopeptides

Crude, combinatorial synthetic phosphopeptides for spike-in experiments originated from the reference library previously published by Marx *et al.* [296]. Two pools were selected based on an initial assessment of synthesis success via the ratio of peptides detected by LC-MS measurements to peptides that theoretically should be present in the peptide pool. Besides, larger pools with 2,400 peptides were excluded due to a disadvantageous increase in sample complexity and quantity after spiking them into a complex human background. This lead to physical overloading of LC-MS columns and, consequently, to a biased loss of hydrophilic peptides in samples with high spike-in ratios which distorted a proper evaluation of label-free spike-in experiments.

After the initial assessment, peptide pool 14 and 61 were selected. Each pool was based on a human, tryptic peptide sequence with prior evidence of a phosphorylation site in at least three large-scale phosphoproteomics studies (14: ATPGNLGSSVLApSK; 61: pYRSPEPDPYLSYR). To obtain a diverse set of (phospho)peptides, the phosphorylation site of these seed sequences had been replaced with serine, threonine, tyrosine and their phosphorylated versions, and, in addition, neighbouring, non-C-terminal amino acids had been permutated with all 20 naturally occurring amino acids. This resulted in 60 different phosphopeptides (ATPGNLGSSVL[X]*p*[STY]K, *p*[STY][X]SPEPDPYLSYR), and their respective unmodified counterparts in each pool.

# 2 Sample preparation for mass spectrometry analyses

Specifics about used protein and peptide amounts, technical replicates, differing parameters during optimization procedures, and detailed experimental workflows can be found in the experimental procedure sections of the respective chapters (see p. 64, pp. 103, pp. 133).

## 2.1 Protein digestion

Before digestion, thawed lysates were cleared at 20,000 g for 10 to 20 min at 4 °C and the protein concentration of the supernatant was determined in at least triplicate measurements using the Pierce™ Coomassie or BCA Protein Assay Kit (Thermo Scientific) according to the manufacturer's instructions. Disulfide bridges were reduced on a shaker at 600 rpm using 10 mM DTT at 30 °C for 20 to 30 min (HeLa, yeast) or 5 mM DTT at 37 °C for 1 h (Jurkat, mouse liver, PDX). Alkylation was performed at RT (room temperature) in the dark using 50 mM chloroacetamide for 20 to 30 min (HeLa, yeast) or 10 mM iodoacetamide for 45 min (Jurkat, mouse liver, PDX). Lysates were diluted to < 2 M urea using 40 mM Tris-HCl, pH 7.6 (HeLa, yeast) or 50 mM Tris-HCl, pH 8.0 (Jurkat, mouse liver, PDX). Digestion was performed by either adding trypsin (Promega) at a 1:50 enzyme-to-substrate ratio and incubating overnight at 37 °C on a shaker at 600 rpm (HeLa, yeast) or by performing a double digestion at 25 °C using 1:50 LysC (Wako) for 2 h and 1:50 trypsin overnight (Jurkat, mouse liver, PDX). Digests were acidified by addition of neat FA (formic acid) to 1 %, centrifuged to pellet insoluble matter and supernatants were stored on ice until sample clean-up.

## 2.2 Sample clean-up by solid-phase extraction

Depending on the peptide quantities, different sorbent weights were employed for sample desalting via solid-phase extraction (SPE). For more than 2.5 mg protein digest, 200 mg tC18 RP (reversed-phase) solid-phase extraction cartridges were used, whereas smaller amounts were desalted on 50 mg tC18 RP cartridges (both Waters) using a vacuum manifold. Peptide quantities below 100 µg were cleaned up using self-packed StageTips [192]. Per 10 µg of protein digest, one plug (Ø 1.5 mm) of C18 material (Octadecyl Extraction Disk, 3M Empore™) was packed into a 200 µl pipette tip and solvents were centrifuged through the tips. Volumes of solvents were adapted to sorbent weights using typically 3 ml, 1 ml, and 220 µl for 200 mg cartridges, 50 mg cartridges, and StageTips, respectively. C18 material was conditioned with 100 % ACN (acetonitrile) and desalting solvent B (0.1 % FA or 0.07 % TFA (trifluoroacetic acid) in 50 % ACN), followed by equilibration applying desalting solvent A (0.1 % FA or 0.07 % TFA). Before loading, samples that were devoid of organic solvent were acidified with neat FA or TFA to a pH of 2 to 3. Samples that contained organic solvents were dried down and reconstituted in desalting solvent A. After slow loading, bound peptides were washed twice with solvent A. Elution was performed by applying two times 250, 150, or 25 µl of solvent B depending on the sorbent weight. For desalting of enrichments of di-glycine remnant and acetylated peptides, 0.1 % FA in 40 % ACN was used as elution solvent. For instantaneous phosphopeptide enrichment, eluates were adjusted to a final ACN concentration of 30 % by addition of 0.7 % TFA. Otherwise samples were frozen, dried by vacuum centrifugation, and stored at - 80 °C until further processing.

Tandem mass tag labelling

For TMT labelling, peptide concentrations were determined after desalting either before dry down using a NanoDrop™ 2000 spectrophotometer (Thermo Fisher Scientific) or after dry down and reconstitution in 0.1 % FA using the Pierce™ BCA Protein Assay Kit and desired peptide quantities were dried by vacuum centrifugation. Standard parameters for the labelling reaction comprised

100 µg of peptides per channel that were labelled with 100 µg of TMT10plex or TMT11plex reagents (Thermo Fisher Scientific) at a peptide, TMT and ACN concentrations of 4 µg/µl, 11.6 mM, and 20 %, respectively. In detail, peptides were dissolved in 20 µl of 50 mM HEPES (pH 8.5), and 5 µl of a 58.1 mM TMT stock dissolved in 100 % anhydrous ACN were added. The peptide-TMT mixture was incubated for 1 h at 25 °C on a shaker at 400 rpm, and the labelling reaction was stopped by addition of 5 % hydroxylamine to a final concentration of 0.4 % and incubation for 15 min at 25 °C and 400 rpm. Peptide solutions were pooled and acidified using 20 µl of 10 % FA. Reaction vessels in which the labelling took place were rinsed with 20 µl of 10 % FA in 10 % ACN, and the solvent was added to the pooled sample. The pools were frozen at - 80 °C and dried by vacuum centrifugation. Differing quantities, volumes, and concentrations were used in experiments in which TMT-to-peptide quantities were titrated, and those are indicated in the respective chapter (see pp. 66).

## 2.3 Metal ion affinity purification of phosphorylated peptides

**Column-based IMAC** – In general, with the exception below, column-based IMAC (immobilized metal ion affinity chromatography) of unfractionated samples was employed for phosphopeptide enrichments as described previously [201]. In brief, a ProPac IMAC 10 column (4 x 50 mm, Thermo Fisher Scientific) connected to an Aekta HPLC system (GE Healthcare Life Sciences) was charged with iron(III) ions, washed with IMAC elution solvent (0.315 % NH4OH), and equilibrated in IMAC loading solvent (0.07 % TFA in 30 % ACN). Samples were dissolved in 0.5 ml of the loading solvent, loaded onto the column and washed for 5 min at a flow rate of 0.2 ml/min meanwhile collecting the flow-through containing non-phosphopeptides. Bound phosphopeptides were eluted and collected applying a two-step gradient from 0 % to 12 % elution solvent within 1.5 min at 0.55 ml/min and from 12 % to 26 % elution solvent within 5 min at 3 ml/min. Flow-through and eluted fraction were dried by vacuum centrifugation and stored at - 20 °C.

**Batch format-based phosphopeptide enrichment** – For the in-depth phosphoproteome analysis to benchmark the optimized TMT protocol, phosphopeptides were enriched from fractionated peptides using NTA (nitrilotriacetic acid) superflow agarose beads (Qiagen) as described recently [297] (performed at the Broad Institute of MIT and Harvard). In brief, peptide fractions were dissolved in 0.1 % TFA in 80 % ACN at a concentration of 0.5 µg/µL. Fractions were added to 10 µl of iron(III) loaded beads in a 1 : 1 : 1 : 1 slurry of beads : ACN : methanol : 0.01 % acetic acid and incubated for 30 min at 25 °C on a shaker at 1,000 rpm. Beads were settled for 1 min at 1,000 g, supernatants were removed, and beads were re-suspended in 200 µl of 0.1 % TFA in 80 % ACN. The slurry was transferred to conditioned and equilibrated StageTips (2 plugs, see p. 41), and beads were washed twice using 50 µl of 0.1 % TFA in 80 % ACN and once using 50 µl of 1 % FA. Phosphopeptides were eluted from the beads onto the C18 material in three iterations using 70 µl of 500 mM potassium phosphate buffer. After washing with 100 µl of 1 % FA, phosphopeptides were eluted from the C18 plugs using 60 µl of desalting solvent B. Eluates were dried by vacuum centrifugation and stored at - 20 °C.

## 2.4 Immunoaffinity purification of ubiquitin-remnant and acetylated peptides

Acetylated and di-glycine remnant peptides were enriched via immunoaffinity purification (IAP). For all washing steps of ubiquitin-remnant antibody beads (PTMScan® Ubiquitin Remnant Motif (K-ε-GG) Kit #5562, Cell Signaling Technology) or acetyl antibody beads (PTMScan® Acetyl-Lysine Motif [Ac-K] Kit #13416, Cell Signaling Technology), 1 ml of solvent was added swirling up beads,

followed by centrifugation for 1 min at 2,000 g and 4 °C and removal of supernatants. Unless otherwise stated, all applied solvents were ice-cold and enrichment steps for acetylated and di-glycine remnant peptides were identical.

**Antibody crosslinking** – Before enrichment of the di-glycine modified peptides, Ubiquitin-remnant antibodies were cross-linked to agarose beads. Acetyl antibodies were not cross-linked. UbiScan beads were washed three times with 100 mM sodium tetraborate decahydrate (Borax), pH 8.8. During the last washing step, 5 µl of the bead slurry were removed. Antibodies were covalently linked to beads at RT for 30 min with end-over-end rotation applying freshly prepared 20 mM dimethyl pimelimidate in 100 mM Borax, pH 8.8. The reaction was stopped by washing beads twice with 200 mM ethanolamine, pH 8.0, and then incubating with 200 mM ethanolamine at 4 °C for 2 h with end-over-end rotation. The capping solution was removed by washing beads three times with IAP buffer (50 mM MOPS, pH 7.2, 10 mM Na2HPO4, 50 mM NaCl). At the third washing step, again 5 µl of the bead slurry were removed to check the coupling efficiency. Beads were re-suspended in PBS containing 0.02 % (w/v) sodium azide and stored at 4 °C.

**Evaluation of coupling efficiency** – Free antibodies in the 5 µl aliquots of bead slurry were eluted at RT applying two times 50 µl of 0.15 % TFA for 5 min. Eluates were dried in by vacuum centrifugation and reconstituted in 20 µl of 2x NuPAGE™ LDS sample buffer (Thermo Fisher Scientific) containing 20 mM DTT. After incubation at 70 °C for 10 min, samples were separated for 40 min at 200 V using NuPAGETM 4-12 % Bis-Tris protein gels in 1 x NuPAGETM MOPS SDS running buffer (Thermo Fisher Scientific). After protein fixation for 1 h in 2 % acetic acid and 40 % methanol, the gel was stained for 20 min in 7.5 % Coomassie (Roth) in 20 % methanol. Cross-linking would be considered efficient, if band intensities of the light and heavy antibody chains decreased by roughly 90 % after the coupling process.

**Enrichment of ubiquitin-remnant and acetylated peptides** – During the optimization of the immunopurification protocol, differing peptide and bead quantities and wash and elution steps were tested, and those are indicated in the respective chapter (see pp. 133). Generally, 10 µl settled beads were employed to enrich ubiquitin-remnant and acetylated peptides from 2 mg of protein digest. First, bead storage solution was removed, and 80 µl settled beads were washed three times with IAP buffer and re-suspended in IAP buffer. Desalted and dried peptides were reconstituted in 1 ml IAP buffer and centrifuged at 20,000 g for 10 min at 4 °C to pellet insoluble matter. Supernatants were added to 130 µl of the bead slurry (~10 µl settled beads) and incubated for 1 h at 4 °C with end-over-end rotation. Then, beads were settled for 1 minute at 2,000 g and 4 °C, and the supernatant was retained as flow-through. After 2 washes with IAP buffer (only 1 wash for acetyl beads) and 2 washes with PBS, modified peptides were eluted twice at RT using 50 µl of 0.15 % TFA for 5 min.

## 2.5 Peptide fractionation

Depending on the input amount, StageTips (>50 ug protein digest) or columns (>150 ug protein digest) connected to off-line HPLC (high-performance liquid chromatography) systems were employed for peptide fractionation. For deep-scale (phospho)proteome analyses to benchmark the optimized TMT protocol, peptides were fractionated on a high-pH (basic) RP (bRP) column. TMT labelled, pulsed SILAC samples were separated via hSAX (hydrophilic strong anion exchange) chromatography. All other samples were fractionated on bRP StageTips (see p. 41). Samples were always cleaned-up (see p. 41) prior to fractionation. Peptide separations were performed according

to previously published, detailed protocols [191, 194, 298] and are described briefly in the following.

**High-pH RP tip fractionation** – After conditioning and equilibration with bRP tip solvent (25 mM ammonium formate, pH 10), StageTips were loaded with peptides dissolved in 50 µl of bRP solvent. Alternatively, peptides bound to StageTips after washing during the desalting procedure were directly re-buffered in 50 µl of bRP tip solvent omitting the elution, dry down, and re-constitution step in between. Peptides were sequentially eluted using 40 to 60 µL of bRP tip solvent containing increasing concentrations of ACN (scheme 1: 5, 7.5, 10, 12.5, 15, 17.5, and 50 % ACN for TMT labelled peptides; scheme 2: 5, 10, 15, 17.5, and 50 % ACN for SILAC-labelled whole proteome digests; scheme 3: 2.5, 7.5, 12.5, 50 % for SILAC-labelled phosphopeptides). The sample flow-through was combined with the 17.5 % ACN eluate (scheme 1 and 2) or the flow through (scheme 3) and the 5 % ACN fraction with the 50 % ACN fraction (scheme 1 and 2), resulting in a total of six or four fractions, respectively. Fractions were dried by vacuum centrifugation and stored at - 20 °C until LC-MS measurement.

**Column-based high-pH RP chromatography** – Peptides were reconstituted in bRP solvent A (4.5 mM ammonium formate, pH 10, in 2 % ACN) and loaded onto a Zorbax 300 Extend-C18 column (3.5 µm, 4.6 × 250 mm; Agilent) coupled to an Agilent 1100 HPLC system. Samples were separated into 96 fractions in 96 min at 1 ml/min, first washing with bRP solvent A for 7 min, followed by increasing bRP solvent B (4.5 mM ammonium formate, pH 10, in 90 %) up to 60 % in several steps (to 16 % in 6 min, 40 % in 60 min, 44 % in 4 min, 60 % in 5 min). In the end, the column was washed with 60 % bRP solvent B for 14 min. Collected fractions 13 to 90 were pooled discontinuously (i.e. fraction 13+37+61, 14+38+62, and so forth) into 23 fractions and fraction 91 to 96 yielded the 24th pooled fraction. Pooled fractions were acidified using 10 % (v/v) of neat FA and 5 % of each fraction was kept for in-depth full proteome analysis, whereas the residual 95 % were further pooled into 12 fractions (fraction 1+13, 2+14, and so forth) for phosphoproteome enrichment using Fe(III) loaded NTS beads (see p. 42).

**Hydrophilic strong anion exchange chromatography** – For hSAX chromatography, samples were reconstituted in hSAX solvent A (5 mM Tris-HCl, pH 8.5) and loaded onto an IonPac AS24 strong anion exchange column (2 x 250 mm) equipped with an IonPac AG24 guard column (2 x 50 mm) and coupled to a Dionex Ultimate 3000 HPLC system (Thermo Fisher Scientific). Samples were washed using 100 % hSAX solvent A for 3 min at 250 µl/min and subsequently eluted in a two-step gradient by raising hSAX solvent B (1 M NaCl in 5 mM Tris-HCl, pH 8.5) to 25 % within 24 min followed by an increase to 100 % within 13 min. During separation, 40 fractions (1 minute each) were collected. Peptide solutions were acidified with 5 µl neat FA and less complex, early and late fractions were pooled (1-4, 5-7, 8-9, 26-27, 28-30, 31-33, 34-35, 36-40). The resulting 24 fractions were desalted on StageTips (see p. 41) before LC-MS measurement.

# 3 Mass spectrometry data acquisition, processing and analysis

Depending on the sample type and research question at hand, differing LC-MS and database search parameters, and data processing and analysis strategies were developed and applied. Highly experiment-specific or varying approaches can be extracted from experimental procedure sections of the respective chapters (see p. 64, pp. 103, pp. 133), while methods that are more generic or remain unchanged within this thesis and their underlying rationales are described in the following.

## 3.1 Data dependent LC-MS measurements

With exception for samples measured at the site of collaboration partners (inter-laboratory evaluation of the optimized TMT protocol, see p. 66), nanoflow LC-ESI-MS measurements were performed using an Ultimate 3000 RSLCnano system coupled to a Fusion Lumos Tribrid or Q-Exactive HF-X mass spectrometer (Thermo Fisher Scientific). Samples were re-constituted in 0.1 % FA except for bRP-fractionated peptides, which were dissolved in 0.5 % FA. For phosphopeptides enriched using the IMAC column, the sample solvent was supplemented with 50 mM citrate.

**Liquid chromatography** – Generally, 1 to 1.5 µg protein digest (corresponding to roughly 500 to 750 ng peptides) were injected onto a trap column (75 µm x 2 cm, packed in-house with 5 µm C18 resin; Reprosil PUR AQ, Dr. Maisch) and washed for 10 min with 0.1 % FA at a flow rate of 5 µl/min. Subsequently, peptides were transferred to an analytical column (75 µm x 45 cm, packed in-house with 3 µm C18 resin; Reprosil Gold, Dr. Maisch) at 300 nl/min and separated within 20, 50, 80 or 100 min using linearly increasing gradients of LC solvent B (0.1 % FA, 5 % DMSO in ACN) in LC solvent A (0.1 % FA in 5 % DMSO). LC solvent B was increased from 4 to 32 % for TMT-labelled phospho-peptides and label-free or SILAC-labelled whole proteome, acetyl- and di-glycine peptides. In contrast, a two-step gradient from 2 to 15 % in 60 % of gradient time and then up to 27 % was employed for unlabelled and SILAC-labelled phospho-proteomes to achieve optimal separation of the more hydrophilic phosphopeptides. TMT-labelled whole proteome samples, acetyl- and di-glycine peptides were separated using 4 to 32 % LC solvent B for 100 min gradients and 8 to 34 % for 20 to 60 min gradients to account for their generally higher hydrophobicity due to the TMT tag.

**Mass spectrometry** – Mass spectrometers were operated in positive ionization and data dependent acquisition (DDA) mode. MS1 spectra were recorded in the Orbitrap from 360 to 1300 m/z (1500 m/z for TMT-labelled samples) at a resolution of 60K, using an automatic gain control (AGC) target value of 4e5 (Lumos) or 3e6 (HF-X) charges and a maximum injection time (maxIT) of 25 to 50 ms. For non-MS3 methods, MS2 spectra were obtained in the Orbitrap at 15K or 30K resolution after HCD fragmentation using 26/28 % (HF-X/Lumos) or 33 % normalized collision energy (NCE) for SILAC-labelled or label-free and TMT-labelled samples, respectively. The AGC target value was set to 1e5 to 2e5 charges, the maxIT to 22 to 120 ms, depending on expected sample complexity and peptide abundance. The first mass was fixed to 100 or 120 m/z and isolation windows were set to 0.7 to 1.7 m/z applying narrower windows for TMT-labelled peptides to reduce co-isolation. The number of MS2 spectra was limited either by a cycle time of 2 s or 3 s or a top12 to top25 method, depending on gradient length and maxITs. In MS3-based methods, cycle time was limited to 2 or 3 s or up to 10 peptide precursors were allowed for isolation (window 0.7 m/z) and fragmentation via CID, MSA, or HCD. MS2 spectra were recorded either in the ion trap in rapid scan mode or in the Orbitrap at 15K or 30K resolution. For each peptide precursor, an

additional MS3 spectrum for TMT quantification was obtained in the Orbitrap at 50 or 60K resolution (scan range 100-1,000 m/z, AGC of 1e5 or 1.2e5 charges, maxIT of 110 or 120 ms). For this, the precursor was fragmented the same way as for the preceding MS2 scan, followed by synchronous selection of typically the 10 most intense peptide fragments in the ion trap and further fragmentation via HCD using a NCE of 55 %. Dynamic exclusion was adjusted according to gradient length (20 to 90 s).

## 3.2 Database searching

**MaxQuant** – Peptide and protein identification and quantification for most experiments was performed using MaxQuant with its built in search engine Andromeda [254, 255]. Depending on the sample type, tandem mass spectra were searched against the human, mouse and/or yeast reference proteome (UP000005640 / 74,468 proteins, UP000000589 / 55,197 proteins, UP000002311 / 6,049 proteins, downloaded Dec 2018), the phosphopeptide library, or the human Swiss-Prot database (42,145 entries including splice variants, downloaded on Feb 2016). In addition, common contaminants were included. Unless stated otherwise, MaxQuant's default parameters were applied. These included trypsin/P as the proteolytic enzyme with up to two missed cleavage sites allowed, carbamidomethylation of cysteine as fixed modification, oxidation of methionine and N-terminal protein acetylation as variable modifications, Andromeda score and delta score cut-offs for modified peptides of 40 and 6, respectively, precursor tolerance of ±4.5 ppm, and fragment ion tolerance of ±20 ppm in the Orbitrap or 0.5 Da in the ion trap. Labels (K8/R10, TMTzero, TMT10, TMT11), experiment types (standard, reporter ion MS2/MS3), and additional variable modifications (phosphorylation on STY, acetylation/GlyGly on K, TMTzero/TMT10 on K and peptide N-termini, TMTzero/TMT10 on H/STY, K8, R10, acetylation on K8) were customized to the sample type. The match-between-runs option was enabled for samples necessitating MS1-based quantification. All results were adjusted to 1 % PSM and protein FDR, employing a target-decoy approach using reversed protein sequences. Isotope impurities of TMT batches were specified in the configuration of modifications to allow MaxQuant the automated correction of TMT intensities.

**Spectrum Mill** – Database searches of in depth fractionated (phospho)proteomes for benchmarking the optimized TMT labelling protocol were conducted with Spectrum Mill suite vB.06.01.202 (Broad Institute and Agilent Technologies). Raw files were searched against the human and mouse RefSeq database (20160914, 37,592 human and 27,289 mouse entries) complemented with common contaminants. Briefly, a 4-cycle fixed/mix modifications search strategy was employed that ran 4 consecutive searches with different sets of modifications in each round and then produced a single integrated output. The four cycles were as follows: all unmodified, both peptide N-termini and lysines labelled, only lysines labelled, and only peptide N-termini labelled. Carbamidomethylation of cysteines and selenocysteines was set as additional fixed modification. N-terminal protein acetylation, oxidation of methionine, de-amidation of asparagine, hydroxylation of proline (when followed by glycine), cyclization of peptide N-terminal glutamine and carbamidomethyl cysteine to pyroglutamic acid (pyroGlu) and pyro-carbamidomethyl cysteine, respectively, and TMT overlabelling of serine, threonine, and tyrosine (limited to histidine containing peptides) were set as variable modifications. For phosphoproteome analysis, phosphorylation of serine, threonine, and tyrosine were allowed as additional variable modifications, while de-amidation of asparagine was restricted to N followed by glycine, and TMT overlabelling and hydroxylation of proline were not allowed. Trypsin Allow P was specified as the proteolytic enzyme with up to 4 missed cleavage sites allowed. For proteome analysis, the allowed precursor mass shift

range was -18 to 262 Da to allow for pyroGlu and up to one additional TMT modification and 2 methionine oxidations per peptide. For phosphoproteome analysis, the range was expanded to -18 to 272 Da to allow for up to 3 phosphorylations and 2 methionine oxidations per peptide. Precursor and product mass tolerances were set to ±20 ppm with PSM-level FDR < 1 % employing a target-decoy approach using reversed protein sequences. To better dissect proteins of human and mouse origin, the subgroup-specific (SGS) protein grouping option in Spectrum Mill was enabled, details of which were previously described [299].

**Mascot** – For generation of a spectral library for a parallel-reaction-monitoring (PRM) assay of members of respiratory chain complex I, a HeLa sample was searched using Mascot Distiller (v2.6.1.0). Trypsin/P was specified as enzyme with up to 2 missed cleavages allowed, Carbamidomethyl (C) of cysteine was required as fixed modification, and Acetyl (N-term) and Oxidation (M) were allowed as variable modifications. Peptide tolerance was set to 10 ppm and MS/MS tolerance to 0.05 Da. Allowed peptide charge states included 2+, 3+, and 4+ precursors and the decharging option was disabled.

## 3.3 Data processing and analysis for comparison of quantification approaches

Decoys and potential contaminants were removed from all datasets and redundant modified sequences in evidence.txt files were merged into one entry with summed intensity.

**Synthetic peptide spike-in experiments** – Since synthetic peptide quantities, used to spike into the complex HeLa background, spanned more than 5 orders of magnitude, the prerequisites underlying the MaxLFQ algorithm [281] for normalization were not fulfilled. Hence, raw intensities were used for analysis of label-free samples. Except for the endogenous seed sequence of both peptide pools (see p. 40), peptides were not naturally occurring in the human proteome and could thus be readily discriminated from the HeLa background. Therefore, normalization across samples was instead performed solely based on endogenous human peptides by shifting the median of their log5 intensity ratios to zero for both quantification approaches.

**HeLa-yeast mixed sample** – For the MS3 standard sample, identified peptides that are shared in an in-silico digest of the human and yeast proteome were excluded from analysis. The constant human background was used for total-sum normalization of TMT channels to account for mixing errors. TMT intensity of yeast peptides in the outermost channels disclosed co-isolation, whereas intensity ratios of the other channels were utilized to estimate ratio distortion by comparing expected to measured ratios to the median of the three channels with the lowest yeast quantity.

**Estimation of amounts of functional groups in protein digests** – An in silico digest of the human reference proteome was performed using the Protein Digestion Simulator released by the Pacific Northwest National Laboratory (https://omics.pnl.gov/software/protein-digestion-simulator). To obtain a conservative estimate of primary amines on peptide N-termini, cleavage was set to trypsin/P with no missed cleavage sites allowed. Minimum and maximum fragment masses were set to 400 and 6,000 Da, and duplicated sequences for given proteins were included. The average of monoisotopic peptide masses $\overline{m}$ was used to calculate amounts of functional groups in 100 µg digest. Quantity $N$ of primary amines (peptide N-termini and ε-amino groups of lysine residues), for instance, was calculated as follows ($r(K)$: fraction of lysine residues containing peptides):

$$\mathrm{N(primary\ amines)} = \frac{100\ \mu\mathrm{g}}{(1+\mathrm{r(K)}) \cdot \overline{\mathrm{m}}} \qquad \textit{(eq. 1)}$$

**Under- and overlabelling analysis of TMT titration experiments** – To estimate underlabelling, samples from TMT titration experiments were searched with TMT as variable modification on lysine residues and peptide N-termini. Peptide sequences which were modified with TMT on all lysine side chains and free (i.e. not acetylated) peptide N-termini were counted as 'fully labelled'. Peptides that did not bear any TMT were annotated as 'not labelled', whereas peptides that contained at least one TMT but were not fully labelled were classified as 'partially labelled'. N-terminal acetylated arginine peptides were excluded from the underlabelling analyses. In an overlabelling search, TMT was specified as a label on lysine residues and peptide N-termini and, in addition, as variable modification on serine, threonine, or tyrosine residues. Peptides that were identified to be labelled with TMT on at least one serine, threonine, or tyrosine were counted as 'overlabelled'.

**TMT-labelled deep-scale (phospho)proteomes** – For analysis of fractionated PDX (phospho)proteomes, reporter ion signals were corrected for isotope impurities, and only human and mouse proteins identified with at least 2 unique peptides were considered for analysis. Relative abundances of proteins and phosphorylation sites were determined using the median of TMT reporter ion intensity ratios from all PSMs matching to the protein or phosphorylation site. PSMs lacking a TMT label, having a precursor ion purity < 50% or a negative delta forward-reverse score (half of all false-positive identifications) were excluded. To normalize quantitative data across TMT10-plex experiments, TMT intensities were divided by the median intensity of all 10 TMT channels for each phosphorylation site and protein. Ratios were further normalized by median centring and median absolute deviation scaling.

## 3.4 Processing of time-course pulsed SILAC-TMT data

Decoys and potential non-human contaminants were removed from the pSILAC-TMT dataset. For calculation of normalization factors, also potential human contaminants were removed.

**Pulsed SILAC-TMT data extraction** – Database searching of SILAC and, at the same time, TMT-labelled samples was performed specifying TMT10plex as a label and SILAC amino acids as variable modifications. Consequently, TMT intensities provided in MaxQuant's proteinGroup.txt output file are derived from the sum of both heavy and light peptides, rendering this quantitative information unfeasible for the assessment of the decay or increase of either of both labels. Hence, TMT data extracted from the evidence.txt output file that discriminated K0/R0 and the K8/R10 labelled peptides was utilized for quantitative analyses. Peptides that could not be assigned to either of both turnover types were removed for turnover analyses. This included missed cleaved peptides that contained both a light and a heavy version of lysine or arginine and C-terminal peptides that did not contain any lysine or arginine residue.

**Pulsed SILAC-TMT data normalization** – Data normalization was conducted under the premise that the total protein amount (i.e. light plus heavy labelled protein) was equal across TMT channels since identical protein amounts were digested and TMT labelled for all pulse time-points. As a result, the principle underlying total sum normalization procedures should be applicable to pulsed SILAC-TMT samples obtained under steady-state conditions. However, in this regard, two additional factors must be considered: (i) Depending on the time-points chosen, intensities of light and heavy SILAC peptides can exhibit globally differential distributions leading to a preferential picking of the overall more intense of both labels for fragmentation in a DDA type of experiment; (ii) TMT intensities do correlate with the MS1 intensity of the peptide and are further highly dependent on the time in the elution profile at which the peptide was fragmented. Consequently, a

normalization based on a simple summation of TMT intensities would be dependent on selected pulse time-points and is likely biased towards synthesis or degradation curves. On account of this, a so-called row-wise normalization (i.e. for each peptide [300]) was employed before the total-sum normalization procedure. This step compensated for differences in overall TMT intensity levels for the corresponding light and heavy peptides due to differing MS1 intensities and times of picking for fragmentation. In detail, the following steps were performed using whole proteome information and for each cell culture replicate separately: First, evidence entries containing channels with zero intensity in between channels with non-zero intensities or exhibited zero intensities in more than 3 channels were removed. Subsequently, intensities of the same TMT channel were summed up for all evidence entries in the same labelling state matching to an identical peptide sequences (oxidized and non-oxidized versions of the same peptide sequence were considered identical). Then, sequence entries were filtered for those for which information on both synthesis and degradation behaviour was available, i.e. which were quantified in both SILAC labelling states. Accordingly, a row-wise normalization factor $\mathrm{NF_{row}}$ was calculated for all remaining peptide sequences and SILAC labelling states. The calculation was based on the TMT reporter intensities $\mathrm{RI}$ in the first $(0\ \mathrm{h})$ and last $(\mathrm{inf.\ h})$ channel of peptides representing degradation $(\mathrm{deg})$ and synthesis $(\mathrm{syn})$ behaviour, respectively:

$$\mathrm{NF_{row}}(\mathrm{pep_{deg}}) = \frac{\mathrm{RI_{deg}}(0\ \mathrm{h}) + \mathrm{RI_{syn}}(\mathrm{inf.\ h})}{2 \cdot \mathrm{RI_{deg}}(0\ \mathrm{h})} \qquad \textit{(eq. 2)}$$

$$\mathrm{NF_{row}}(\mathrm{pep_{syn}}) = \frac{\mathrm{RI_{deg}}(0\ \mathrm{h}) + \mathrm{RI_{syn}}(\mathrm{inf.\ h})}{2 \cdot \mathrm{RI_{syn}}(\mathrm{inf.\ h})} \qquad \textit{(eq. 3)}$$

These row-wise normalization factors equalize TMT reporter intensities representing completely labelled peptides, i.e. the first channel of decreasing and the last channel of increasing curves, thus allowing for a total sum normalization. They were multiplied with all TMT channel intensities of the particular peptide sequence in the respective labelling state. Thereafter, all intensities belonging to the same TMT channel were summed up for all sequences in both labelling states and a total sum normalization factor $\mathrm{NF_{sum}}$ was computed for each TMT channel (time-point t):

$$\mathrm{NF_{sum}}(\mathrm{RI(t)}) = \frac{\text{median of all } \sum_{\text{all sequences}} \mathrm{RI(t)}}{\sum_{\text{all sequences}} \mathrm{RI(t)}} \qquad \textit{(eq. 4)}$$

These 10 normalization factors normalize for TMT sample mixing errors and were applied to respective TMT channels of all entries of MaxQuant's evidence output table.

***In-silico*** **removal of ratio compression** – Depending on the SILAC label of the peptide fragmented, TMT intensities in pSILAC-TMT experiments are expected to show a constantly increasing or decreasing behaviour reflecting label incorporation or loss, respectively. Consequently, co-isolation and fragmentation of peptides with opposing quantitative characteristics can result in severe ratio distortion and failure to pass filter criteria for turnover rate estimation or adulteration of turnover rate estimations. To tackle this quantitative bias in the pSILAC-TMT dataset including PTMs (see Chapter V), ratio distortion was corrected in silico by subtraction of the average synthesis curve from all degradation curves and the average degradation curve from all synthesis curves in the respective subproteome samples. This was implemented with the help of the outermost TMT channels that, owing to the experimental design, indicated the degree of co-isolation and ratio distortion. More precisely, for adjustment of degradation curves, TMT intensities of all evidence entries representing synthesis and exhibiting less than 10 % ratio compression (calculated by the ratio of the first (0 h) to the last TMT channel (inf. h)) were summed up per channel. Then summed intensities were divided by the summed intensity of the last TMT channel to obtain

average synthesis intensity ratios $\mathrm{IR}_{\mathrm{syn\_average}}$ that represented the average synthesis curve normalized to a range of 0 to 1. Subsequently, adjusted reporter intensities $\mathrm{RI\_RC}$ were calculated for each pulse time-point t and each evidence entry illustrating degradation by subtracting the fraction of intensity corresponding to the average synthesis curve:

$$\mathrm{RI\_RC}_{\mathrm{deg}}(\mathrm{t}) = \mathrm{RI}_{\mathrm{deg}}(\mathrm{t}) - \mathrm{RI}_{\mathrm{deg}}(\mathrm{t}) \cdot \mathrm{IR}_{\mathrm{syn\_average}}(\mathrm{t}) \qquad \textit{(eq. 5)}$$

TMT reporter intensities of evidence entries showing synthesis were corrected accordingly by utilizing the average degradation intensity ratios $\mathrm{IR}_{\mathrm{syn\_average}}$ derived from degradation evidence entries with less than 10 % ratio compression (calculated by the ratio of the last (inf. h) to the first TMT channel (0 h)):

$$\mathrm{RI\_RC}_{\mathrm{syn}}(\mathrm{t}) = \mathrm{RI}_{\mathrm{syn}}(\mathrm{t}) - \mathrm{RI}_{\mathrm{syn}}(\mathrm{t}) \cdot \mathrm{IR}_{\mathrm{deg\_average}}(\mathrm{t}) \qquad \textit{(eq. 6)}$$

Intensities with removed ratio compression where then subjected to normalization as described above (see pp. 48).

**Correction for peptide abundance variations** – The kinetic model underlying the computation of turnover rates assumes steady-state conditions (see pp. 52). However, the abundance of individual (modified) proteins may still change over the time course of a pSILAC experiment. To improve turnover estimations, intensity ratios to which turnover curves were fitted were corrected for such abundance fluctuations in the pSILAC-TMT dataset including PTMs (see Chapter V). As an example, if abundance increased steadily over the course of the experiment, ratios of later time-points were appropriately scaled down to account for the gain in total abundance. The assessment of abundance changes was based on the premise that the sum of the MS1 intensity of the light and heavy version of a peptide at a certain time-point corresponds to its total abundance at this respective time-point. Therefore, abundance was only corrected for peptides for which MS1 intensity information of both versions was available. Further, at least 4 quantified MS1 peak elution profiles in any replicate or, alternatively, 2 quantified MS1 peak elution profiles in at least 2 replicates were required to improve robustness and avoid artefacts during abundance correction. In the same manner as for the calculation of normalization factors (see pp. 48), evidence entries containing too many zero TMT intensities were removed and MS1 intensities and TMT intensities were summed up for identical peptide sequences in the same labelling state and replicate. Then, the fraction of the MS1 intensity ($\mathrm{MS1\ intensity\ fraction}$) corresponding to a peptide at a certain time-point t was calculated from ratio compression adjusted reporter intensities, separately for the two labelling states corresponding to degradation and synthesis:

$$\mathrm{MS\ intensity\ fraction}_{\mathrm{deg}}(\mathrm{t}) = \frac{\mathrm{RI\_RC}_{\mathrm{deg}}(\mathrm{t})}{\sum_{\mathrm{all\ time-points}} \mathrm{RI\_RC}_{\mathrm{deg}}} \cdot \mathrm{MS1\ intensity}\left(\mathrm{pep}_{\mathrm{deg}}\right) \qquad \textit{(eq. 7)}$$

$$\mathrm{MS\ intensity\ fraction}_{\mathrm{syn}}(\mathrm{t}) = \frac{\mathrm{RI\_RC}_{\mathrm{syn}}(\mathrm{t})}{\sum_{\mathrm{all\ time-points}} \mathrm{RI\_RC}_{\mathrm{syn}}} \cdot \mathrm{MS1\ intensity}\left(\mathrm{pep}_{\mathrm{syn}}\right) \qquad \textit{(eq. 8)}$$

The sum of the MS intensity fractions of degradation and synthesis curves was utilized as a proxy for the abundance at a certain time-point in each replicate:

$$\mathrm{Abundance}(\mathrm{pep}(\mathrm{t})) = \mathrm{MS\ intensity\ fraction}_{deg}(\mathrm{t}) + \mathrm{MS\ intensity\ fraction}_{syn}(\mathrm{t}) \qquad \textit{(eq. 9)}$$

Factors $\mathrm{F}$ for the adjustment of intensity ratios for degradation and synthesis curve fitting were calculated as the abundance changes relative to the first and the last channel, respectively:

$$\mathrm{F}_{\mathrm{deg}}(\mathrm{t}) = \frac{\mathrm{Abundance}(\mathrm{pep}(\mathrm{t}))}{\mathrm{Abundance}(\mathrm{pep}(0\mathrm{h}))} \qquad \textit{(eq. 10)}$$

$$\mathrm{F_{syn}(t)} = \frac{\mathrm{Abundance(pep(t))}}{\mathrm{Abundance(pep(inf.h))}} \qquad \textit{(eq. 11)}$$

The median of these factors across replicates was utilized to obtain abundance adjusted intensity ratios (IR_Ab) to which degradation and synthesis curves were fitted:

$$\mathrm{IR\_Ab_{deg}(t)} = \frac{\mathrm{RI\_RC_{deg}(t)}}{\mathrm{median(F_{deg}(t))} \cdot \mathrm{RI\_RC_{deg}(0h)}} \qquad \textit{(eq. 12)}$$

$$\mathrm{IR\_Ab_{syn}(t)} = \frac{\mathrm{RI\_RC_{syn}(t)}}{\mathrm{median(F_{syn}(t))} \cdot \mathrm{RI\_RC_{syn}(inf.h)}} \qquad \textit{(eq. 13)}$$

**Estimation of protein copy numbers** – In principle, copies of a protein can be determined from its mass, molecular weight (MW) and the Avogadro constant ($\mathrm{N_A}$):

$$\mathrm{Copies(protein\ x)} = \frac{\mathrm{Mass\ (protein\ x)}}{\mathrm{MW\ (protein\ x)}} \cdot \mathrm{N_A} \qquad (\textit{eq. 14})$$

Usually the mass of a protein in a sample or a single cell is not known. However, based on the assumption that, for deep-scale proteome analyses, a protein's MS signal intensity as a fraction of the total MS signal is an appropriate measure for the proportion of its mass to the total protein mass [301], it can be approximated as:

$$\mathrm{Mass(protein\ x)} = \frac{\mathrm{MS\ intensity\ (protein\ x)}}{\mathrm{Total\ MS\ intensity}} \cdot \mathrm{Total\ protein\ mass} \qquad \textit{(eq. 15)}$$

Accordingly, protein copy numbers per cell can be derived from ***eq. 14*** and ***15*** as:

$$\mathrm{Copies/cell(protein\ x)} = \frac{\mathrm{MS\ intensity\ (protein\ x)}}{\mathrm{Total\ MS\ intensity}} \cdot \frac{\mathrm{Cell\ protein\ mass}}{\mathrm{MW\ (protein\ x)}} \cdot \mathrm{N_A} \qquad \textit{(eq. 16)}$$

For determination of total protein mass per HeLa cell, protein amounts were quantified with the Bradford method (Coomassie (Bradford) Protein Assay Kit, Thermo Fisher Scientific) for increasing cell numbers (3e4, 6e4, 1.2e5, 2.5e5, 5e5, 7.5e5, 1e6) from three different HeLa cell batches in three replicates each. Linear correlation of cell numbers with corresponding protein amounts resulted in 251.4 pg protein per single HeLa cell.

Since TMT intensities are critically dependent on the time in the elution profile at which the precursor was picked for fragmentation, TMT intensities cannot directly be used for the calculation of intensities as needed in eq. 16. However, due to the experimental design, the first or the last TMT channel always represented the completely light or heavy labelled peptide, which could be harnessed to obtain meaningful single protein to total protein intensity ratios. To do so, the fraction of the MS1 intensity (MS1 intensity fraction) corresponding to completely labelled peptides was calculated in the same manner as for the correction of abundances:

$$\mathrm{MS\ intensity\ fraction_{deg}} = \frac{\mathrm{RI_{deg}(0\ h)}}{\sum_{\mathrm{all\ time-points}} \mathrm{RI_{deg}}} \cdot \mathrm{MS1\ intensity(pep_{deg})} \qquad \textit{(eq. 17)}$$

$$\mathrm{MS\ intensity\ fraction_{syn}} = \frac{\mathrm{RI_{syn}(inf.\ h)}}{\sum_{\mathrm{all\ time-points}} \mathrm{RI_{syn}}} \cdot \mathrm{MS1\ intensity(pep_{syn})} \qquad \textit{(eq. 18)}$$

Accordingly, protein copies per cell were determined by using the estimated protein amount per cell and by deriving the protein intensity from the sum of intensities of peptides assigned exclusively to the respective protein group:

$$\mathrm{Copies/cell(protein\ x)} = \frac{\sum_{\mathrm{all\ peptides\ for\ protein\ x}} \mathrm{MS\ intensity\ fraction}}{\sum_{\mathrm{all\ peptides}} \mathrm{MS\ intensity\ fraction}} \cdot \frac{2.514 \cdot 10^{-7}}{\mathrm{MW\ (protein\ x)}} \cdot \mathrm{N_A} \qquad \textit{(eq. 19)}$$

Since the MS intensity fraction can be calculated for both light and heavy labelled peptides, protein copies per cell could be computed twice for each cell culture replicate. Median values were taken for correlation analysis with protein half-lives.

## 3.5 Processing of single time-point pulsed SILAC data

Hits to the reverse database and potential non-human contaminants were removed from the pSILAC dataset. Heavy and light peptide and protein intensities and new-to-old (N/O) peptide and protein ratios were calculated based on intensities and heavy-to-light (H/L) ratios specified in the MaxQuant's evidence.txt output file. Intensities were summed up for all evidence entries in the same labelling state and pulse time-point corresponding to an identical peptide or protein group. H/L ratios were log transformed and the median of all evidence entries matching to the same peptide sequences or protein group was calculated. Oxidized and non-oxidized versions of the same peptide sequence were combined and only protein group unique peptides were included to compute protein intensities and H/L ratios. For replicates for which the cell culture medium was switched from K0R0 to K8R10, H/L ratios corresponded to N/O ratios, while the inverse value of H/L ratio was used as N/O ratio for replicates starting with K8R10 labelled cells and exchanging K8R10 with K0R0 medium.

## 3.6 Turnover computation and underlying kinetic model

**Kinetic turnover model** – Briefly, the applied model was based on the assumptions that (i) the probability of a protein being degraded is the same for pre-existing and newly synthesized proteins and stays constant over the life-time of these proteins (i.e. the degradation rate is constant), (ii) protein synthesis occurs at a constant rate, and (iii) cells are in steady state implying that the average abundance of a (modified) protein per cell doesn't change during the course of the experiment (i.e. the synthesis rate of a protein equals its degradation rate). Hence, protein turnover described as the loss of originally available molecules over time ($\mathrm{N(t)}$) ideally follows first-order kinetics:

$$\mathrm{N(t)} = \mathrm{e}^{-\mathrm{K}\cdot\mathrm{t}} \qquad \textit{(eq. 20)}$$

In order to account for experimental constraints such as amino acid recycling and ratio compression, this simple model was adjusted for curve fitting to measured label incorporation and loss data as previously described by Boisvert *et al.* [170] and Welle *et al.* [302] (see *eq. 21* and *22*).

**Curve fitting to time-course data for turnover rate estimation** – Before curve fitting, intensity ratios $\mathrm{IR}$ were computed to enable the implementation of global, intensity-independent curve fitting constraints. For pSILAC time-course data the ratio of the light or the heavy label to the sum of both labels was calculated. For pSILAC-TMT experiments, TMT intensities were divided by the intensity of the in last (inf. h) TMT channel for peptides representing label incorporation, whereas, for peptides illustrating label loss, ratios were calculated relative to the intensity of the first (0 h) TMT channel. Intensity ratios (IR) were fitted to following exponential equations:

$$\mathrm{IR_{syn}(t)} = \left(\mathrm{B_{syn}} - \mathrm{A_{syn}}\right) \cdot \mathrm{e}^{-\mathrm{K_{syn}}\cdot\mathrm{t}} + \mathrm{A_{syn}} \qquad \textit{(eq. 21)}$$

$$\mathrm{IR_{deg}(t)} = \left(\mathrm{A_{deg}} - \mathrm{B_{deg}}\right) \cdot \mathrm{e}^{-\mathrm{K_{deg}}\cdot\mathrm{t}} + \mathrm{B_{deg}} \qquad \textit{(eq. 22)}$$

$\mathrm{IR_{syn}(t)}$ is the proportion of protein molecules (or peptides) which have incorporated the newly provided label and $\mathrm{IR_{deg}(t)}$ is the fraction of protein molecules still bearing the old label at each

time-point t in hours. $K_{syn}$ and $K_{deg}$ are the rate constants of label incorporation and loss, respectively. As they should be the same by definition of steady-state conditions, they are referred to as turnover rate $K$ in the following. The parameter $A$ refers to the maximum of the curve (i.e. the normalized total protein amount) and should be 1 in an ideal case. The parameter B accounts for a potential curve offset which ideally should be zero. Offsets bigger than zero could either be attributed to the recycling of amino acids or ratio compression. Consequently, $(A - B)$ represents the amplitude of the fitted curve. Turnover rates, curve maxima and offsets were obtained for each evidence entry via performing a nonlinear least square (NLS) optimization in R (version 3.3.3, function "nls") [303]. Peptide and proteins rates were obtained from a combined fit including all evidence entries filtered for criteria stated below and belonging to the respective peptide or protein sequence. In addition, only protein group unique peptides were allowed for determination of whole protein (group) turnover.

For the comparison of pSILAC-TMT and classical time-course pSILAC data on protein level, samples were derived from identical protein digests and raw files of the two approaches were searched together in separate experimental groups in the MaxQuant software. For the MS3 data, TMT intensities were normalized as described above and summed up for all peptides with the same labelling state belonging to the same protein group. Protein TMT or SILAC intensity ratios for label incorporation and loss were calculated and subjected to curve fitting as described above. For the curve fitting of SILAC data, missing quantitative data for up to 3 out of 6 time-points were allowed, meanwhile counting zero intensities resulting from quantification of only 1 SILAC channel as valid (i.e. non-missing) values.

**Curve filter criteria** – To establish adequate filter criteria and remove poor quality quantitative data, different constraints were applied for curve variables K, A and B, and for the curve fitting quality parameter R2. Curves that were at the border of passing filter criteria were manually inspected, and correlation of turnover rates across replicates was assessed. Final filtering criteria were based on the goal to filter out PSMs which lead to irreproducible turnover rate estimations based on replicate correlations, showed a high variation of data points along the fitted curve (A and R2), a high ratio compression (B), or resulted in turnover rates (K) which simply could not be determined accurately based the pulse time-points that were chosen in the experimental design.

**Estimation of amino acid recycling** – The degree of amino acid recycling was estimated harnessing quantitative information on missed cleaved peptides that were quantified in the label state representing synthesis and the mixed labelling state (containing both a light and a heavy version of lysine and/or arginine residues) in the pSILAC-TMT dataset. First, the fractions of the MS1 intensities were calculated for synthesis and mixed-label peptides for each pulse time-point according to eq. 7 and 8. Second, the recycled fraction at certain pulse time-points t was calculated as the ratio of the MS1 intensity fraction of the mixed-label peptide to the sum of MS1 intensity fractions of mixed and the synthesis-representing peptide:

$$\text{Recycled fraction}(t) = \frac{\text{MS intensity fraction}_{\text{mixed}}(t)}{\text{MS intensity fraction}_{\text{mixed}}(t) + \text{MS intensity fraction}_{\text{syn}}(t)} \qquad \textit{(eq. 23)}$$

Since it was impossible to remove ratio compression for mixed-label peptides, non-adjusted TMT reporter intensities were used for the computation of the recycled fraction.

**Single pulse time-point analysis** – The equation for computing the turnover rate K from single time-point pSILAC data was obtained by conversion of eq. 22 and approximation of the curve amplitude by 1:

$$\mathrm{K_{deg}} = -\frac{\ln(\mathrm{IR_{deg}(t)} - \mathrm{B})}{t} \qquad \textit{(eq. 24)}$$

The intensity ratio for curves that indicate label loss at a certain time-point ($\mathrm{IR_{deg}(t)}$) further can be re-written using the fraction of newly synthesized and pre-existing protein or peptide:

$$\mathrm{IR_{deg}(t)} = \frac{\text{Old fraction(t)}}{\text{Old + New fraction(t)}} = \frac{1}{\frac{\text{New fraction(t)}}{\text{Old fraction(t)}} + 1} = \frac{1}{\mathrm{N/O(t)} + 1} \qquad \textit{(eq. 25)}$$

Combining ***eq. 24*** and ***25*** results in:

$$\mathrm{K} = -\frac{\ln\left(\frac{1}{\mathrm{N/O(t)} + 1} - \mathrm{B}\right)}{t} = = -\frac{\ln\left(\frac{1}{\mathrm{N/O(t)} + 1} - \text{Recycled fraction(t)}\right)}{t} \qquad \textit{(eq. 26)}$$

Hence, the turnover rate for pSILAC data from single time-points was calculated using the computed N/O ratios and the recycled fraction estimated from pSILAC-TMT data.

**Computation of turnover times and half-lives** – The time at which half of total protein molecules have lost the "old" label or incorporated the new label is known as 50 % turnover time ($\mathrm{T50\%}$). $\mathrm{T50\%}$ was derived from eq. 20 as the time at which half of the molecules have been degraded:

$$\mathrm{T50\%} = \frac{\ln(2)}{\mathrm{K}} \qquad \textit{(eq. 27)}$$

In order to estimate actual protein and peptide degradation (or synthesis) rates ($\mathrm{k}$) and half-lives ($\mathrm{T1/2}$), the influence of sheer cell doubling on labelling kinetics needs to be taken into account. Cell doubling parameters were determined via fitting an exponential growth equation to cell counts monitored over time during the pulse experiment. Then, labelling rates were corrected for cell doubling rates ($\mathrm{k_{cd}}$) to obtain degradation rates:

$$\mathrm{k} = \mathrm{K} - \mathrm{k_{cd}} \qquad \textit{(eq. 28)}$$

Accordingly, half-lives were calculated as:

$$\mathrm{T1/2} = \frac{\ln(2)}{\mathrm{K} - \mathrm{k_{cd}}} \qquad \textit{(eq. 29)}$$

## 3.7 Parallel reaction monitoring assay for NADH dehydrogenase complex

**Spectral library construction** – For acquisition of spectra for a spectral library generation for members of respiratory chain complex I (NADH dehydrogenase), single time-point SILAC pulse samples obtained upon rotenone and control treatment and non-treated, non-pulsed HeLa sample were analysed in DDA mode on a Lumos Fusion mass spectrometer employing 100 min gradients as described above (see pp. 45). In addition, a scheduled m/z inclusion list for peptides of respiratory chain complex I proteins was specified in the MS method, utilizing information on charge states and retention times from the ProteomeTools project [231]. MS2 spectra of peptides from the targeted m/z list were recorded with scan priority 1 in the Orbitrap at 15K using an AGC target value of 2e5 charges, a maxIT of 100 ms, and a dynamic exclusion of 30 s. Non targeted m/z values were recorded with scan priority 2 using an AGC target value of 1e5, a maxIT of 22 ms and a dynamic exclusion of 60 s. The spectral library was constructed using the Skyline 3.7.0 software [265] importing a MaxQuant derived msms.txt from the search of pulsed SILAC samples and a Mascot derived DAT file from the search of the non-pulsed sample. Uniqueness of NADH dehydrogenase peptides was checked in Skyline using the canonical Swiss-Prot database as background proteome. Precursor charge states and transitions were automatically chosen from the spectral library resulting in 38 and 93 peptide entries for the MaxQuant and the Mascot derived library. A

scheduled inclusion list containing mass and charge information for light and heavy labelled peptides with 6 min retention time windows for monitoring was exported for the PRM LC-MS method.

**Parallel reaction monitoring** – PRM measurements of rotenone treated and control, pulsed SILAC samples were performed on a Lumos Fusion mass spectrometer as described above (see pp. 45) with following modifications: Per injection, 100 fmol of PRTC retention time calibration mixture (Pierce) were spiked into samples and the MS method was set up to switch between two separate experiments after each duty cycle. The first experiment consisted of a full scan MS1 spectrum recorded in the Orbitrap at a resolution of 15K. The second experiment consisted of a tMS2 PRM scan triggering peptide isolation and fragmentation based on the scheduled inclusion list containing m/z and charge information. MS2 spectra were recorded in the Orbitrap from 100 to 2000 m/z at a resolution of 15K using an AGC target value of 2e5 and a maximum injection time of 200 ms.

**PRM data filtering and processing** – RAW files were imported into Skyline for data inspection and filtering. The transitions were extracted specifying Orbitrap as mass analyser with 15K resolution and allowing precursor charges 2, 3 and 4, and y-ion types. Peaks were integrated using the automatic peak finding function followed by manual curation of all peak boundaries and transitions to remove fragment ions exhibiting interferences. At least 3 transitions that showed robust elution profiles and a dot product larger than 0.8 were required in at least one of the two SILAC channels in at least one condition. The summed area under the fragment ion traces and heavy-to-light (H/L) ratios were exported for every peptide to perform data normalization in Microsoft Excel. Since cell treatments can lead to a change in growth behaviour of cells that would globally shift SILAC ratios in a pulse experiment and bias interpretation of results, H/L ratios derived from the PRM experiment were normalized based on a median centring of all H/L ratios across all measured samples of each pulse time-point. For this, the DDA MS analyses for spectral library generation were utilized to calculate a normalization factor that shifted the median of log transformed H/L ratios of all samples of one pulse time-point to the same value. This normalization factor was then applied to log transformed H/L ratios of the PRM experiment. It should be noted that such normalization procedures will only work properly if ratio distributions of the different conditions have similar variances. This usually only applies for rather short treatment times and/or treatments which only affect a small fraction of the proteome.

## 3.8 Functional and statistical data analyses

**Integration with public data** – For integration of turnover data with protein properties and functions, the median of log transformed turnover rates or half-lives from all four replicates was utilized. Functional annotations of UniProt Keywords [43], protein complexes [290, 304], domains [288], and modifications [294] were performed using the based on the first UniProt identifier for each protein group or peptide sequence. In contrast, protein class [287] and localization information [291, 292] was on available on gene level and was therefore matched based on the first gene name. All other literature data (turnover [114, 169, 170, 302, 305], protein copy numbers [306, 307], thermal protein stability [308], enzyme-substrate information [70, 208, 309, 310]) were first matched based on the first IPI/UniProt identifier and second, for still unmatched data or for data for which no protein identifier was available, on the first gene name entry for each protein group or peptide sequence. Acetylation sites were identified as targets of KDACs and KATs when inhibitor treatments or enzyme transfections induced at least a 2-fold change in PTM abundance. Degrons [311, 312] were annotated based on (modified) peptide sequences. Prediction of

secondary protein structure was performed employing the s2d method [313]. Crystal structures of respiratory chain complex I were obtained from the RCSB protein databank website (http://www.rcsb.org [293]; PDB ID: 5XTD [314]).

**Statistical analyses and data visualization** – The Perseus software suite (v.1.5.6.0 or 1.6.2.3) [315] was utilized to perform correlation analyses, functional 1D enrichment analyses [316], ANOVA (analysis of variance), Fisher's exact and Student's t-tests using log-transformed, normal-distributed intensities, turnover rates or new-to-old protein or peptide ratios. Expected linear relationships were analysed using Pearson's correlation coefficients (R), otherwise Spearman rank correlation coefficients (ρ) were computed. Statistical tests were corrected for multiple testing applying a permutation based or Benjamini-Hochberg FDR calculation at 1 or 5 % as indicated. S0 was computed in R (v3.4.1 or 3.6.0, function "samr" [317]) for each statistical group comparison separately. This constant is based on the significance analysis of microarrays [318] and adjusts the significance cut-off of statistical analyses on the fold-change level while accounting for differing variances across the range of measured values and groups. Mann-Whitney tests performed in GraphPad were used to check whether two non-Gaussian distributions were significantly different. Significant differences in multiple non-Gaussian distributions were assessed using a Kruskal-Wallis test, and a post hoc Dunn's multiple comparison test was employed to pinpoint the specific, significantly different distribution(s). ROC (receiver operator characteristic) curves were also computed in GrapPad. Motif enrichment analyses were performed using the motifX algorithm [319] via the R package rmotifx [320]. Sequence windows with 11 amino acids around sites that showed either significantly faster or slower turnover were specified as foreground, and sequence windows of all other identified sites with quantified turnover were used as background. Duplicates were removed, a minimum of 20 sequences was required per motif, and significance cut-off was set to 2.5e-4 corresponding to an actual alpha-value of <0.05 after Bonferroni correction [321]. Amino acid distributions in sequence windows of sites with significantly slower or faster turnover were plotted using pLogo [322]. Protein crystal structures were visualized in PyMOL and all other data using R, GraphPad, Excel, or Tableau.

*„Quality means doing it right when no one is looking."*

**- Henry Ford**

# III A CASE FOR ROBUST PEPTIDE QUANTIFICATION

**1 Introduction and summary ..... 63**

**2 Experimental designs ..... 64**

2.1 Comparison of label-free and TMT-based phosphopeptide quantification ..... 64

2.2 Generation of a TMT-labelled standard ..... 65

2.3 Optimization of the TMT-labelling protocol ..... 66

**3 Results ..... 69**

3.1 A compromise between accurate dynamic range and reproducible coverage ..... 69

3.2 A versatile TMT standard ..... 73

3.3 A robust and cost-efficient TMT-labelling protocol ..... 77

**4 Discussion and conclusion ..... 85**

4.1 Confining the dynamic range of peptide quantification ..... 85

4.2 Negotiating quantification approaches for (modified) peptides ..... 87

4.3 Managing constraints of TMT-based (phospho)peptide quantification ..... 89

4.4 Characterizing critical parameters for TMT labelling ..... 92

4.5 Conclusion ..... 95

The following chapter is in part based on data generated by Florian P. Bayer during his research internship "*Comparison of label-free versus tandem mass tag based quantification for phosphoproteomic analysis*" and his Master Thesis "*Evaluation of a phosphoproteomic workflow for in cellulo drug selectivity and residence time analysis*". Both were conducted under the author's continuous supervision at the Chair of Proteomics and Bioanalytics, TUM. In addition, large parts of the publication "*TMT labeling for the masses: A robust and cost-efficient, in-solution labeling approach*" [323] (published in Molecular & Cellular Proteomics on April 9) are included in the following chapter.

**Author contributions for** [323]

J.Z. developed the methodology and wrote the paper with input from S.S., K.R.C., P.M., S.A.C., and B.K.; J.Z., S.S., P.M., S.A.C., and B.K. designed experiments; J.Z., S.S., T.K., S.C.A., and M.H.K. performed experiments; J.Z. and S.S. analysed the data and prepared the figures.

**Supplementary material for** [323]

Additional supplementary tables are freely available for download at the publisher's website (https://www.mcponline.org/content/early/2019/04/09/mcp.TIR119.001385/tab-figures-data). The MS proteomics raw data and complete MaxQuant search results have been deposited to the ProteomeXchange Consortium (http://www.proteomexchange.org/) via the PRIDE [324] partner repository with the data set identifier PXD012703.

# 1 Introduction and summary

The basis of biological interpretation of bottom-up proteomics data is the robust identification and quantification of peptides across multiple samples. To determine quantitative differences in the abundance of (modified) peptides, a variety of strategies can be followed (reviewed in [267, 270, 271]). Among those, TMT-based and label-free quantification are two commonly employed approaches. Both have their own strengths and limitations with regard to dynamic range, accuracy and precision of quantification, reproducibility of peptide detection across samples, costs of sample processing and measuring, and sample throughput and achievable proteome coverage (see pp. 15). Although there have been several studies comparing label-free (or other MS1-based quantification strategies) with isobaric tagging approaches [245, 325-328], these reports have often only included a limited number of samples (≤ 3) and mainly focussed on quantification performance on protein level. In this case, quantitative information of several peptides mapping to the same protein are integrated thereby improving quantitative robustness. Consequently, missing quantification of individual peptides in single conditions may have little adverse effects. However, analysis of proteoforms with bottum-up proteomics is critically dependent on accurate and precise peptide level quantification since quantitative variation usually cannot be countervailed by integrating information of a multitude of different peptides (see also Figure I-8). In addition, missing values can accumulate substantially the more conditions are included. Modification proteomics involves the additional challenge of correct site localization of the modified amino acid side chain. This is often especially difficult for phosphoproteomic analyses, and ambiguous identifications will affect quantification performance.

In order to identify the most suitable quantification strategy for peptidoform analyses across multiple conditions, the robustness of label-free and TMT-based quantification was investigated specifically on peptide level. To this end, different experimental approaches were employed including 10 to 11 different samples. A dilution row of synthetic, non-endogenous phosphopeptides in a complex human background demonstrated the superior dynamic range of label-free quantification, but also confirmed its considerable limitations with respect to missing quantification across conditions. This was corroborated in a treatment time course experiment using an HSP90 inhibitor, which also confirmed that this disadvantage could be overcome by TMT-labelling. The latter also showed benefits with regard to precision and phosphorylation site localization. Additional analyses of a human-yeast mix sample indicated that the drawback of ratio compression for the TMT-based strategy can be significantly reduced by utilizing advanced MS3-mode measurements ultimately rendering this strategy an appropriate compromise for robust peptidoform quantification. Further, the species interference sample was also used for optimization of MS3 data acquisition where, for instance, CID-MSA fragmentation for MS2 spectra and their read-out in the Orbitrap proved beneficial for phosphopeptide identification and quantification. Further, the high costs of the employed stable isotope reagents were tackled by systematically evaluating the impact of labelling reaction parameters and establishing a robust and efficient TMT-labelling protocol that achieves complete labelling of primary amines in peptides using 8-times less TMT reagent than recommended by the vendor. Finally, inherent limitations of the studied quantification approaches and how they can be managed are discussed, and guidance on the adoption of the optimized labelling protocol for different peptide quantities is provided.

# 2 Experimental designs

The following sections provide an overview of the experimental designs for the assessment of label-free versus TMT-based quantification, establishment of a TMT-labelled standard, and the optimization of the TMT labelling protocol. Details on individual experimental steps and data analyses are specified in chapter II (see pp. 37)

## 2.1 Comparison of label-free and TMT-based phosphopeptide quantification

To evaluate the linear dynamic range, identification and quantification reproducibility of the two approaches, two label-free and TMT experiments were conducted. The first experiment, a dilution series of synthetic phosphopeptides was spiked into a constant background of HeLa peptides, and expected and measured spike-in ratios were compared (Figure III-1 A). The chosen peptide pools contained phosphopeptides and their unmodified counterparts (see p. 40) and were thus first subjected to Fe(III) IMAC. This should also deplete synthesis by-products such as non-phosphorylated, truncated peptide versions or non-phosphorylated peptides that still carry the protection group. Hence, it was expected to improve the robustness of TMT labelling process. The enriched phosphopeptide fraction was mixed with 1 mg of HeLa protein digest in 9 one-to-five dilution steps starting with an estimated amount of 1,000 pmol per phosphopeptide down to 2.6 fmol. Additionally, one sample without any synthetic spike-in was included. For the label-free quantification approach, 70 % of each sample were separately IMAC enriched and desalted, while 30 % were labelled with TMTs and the pool was subjected to a single phosphoproteome enrichment. In order to achieve more comparable LC-MS measurement times for both quantification approaches, the TMT-labelled phosphopeptides were further fractionated in a bRP StageTip. In the second experiment, K562 cells were treated with the HSP90 inhibitor 17-AAG (tanespimycin) for up to 32 h and per time-point 2.2 mg proteins were processed similar to synthetic spike-in samples, subjecting 90 % of the samples to the label-free and the rest to the TMT-based workflow (Figure III-1 B).

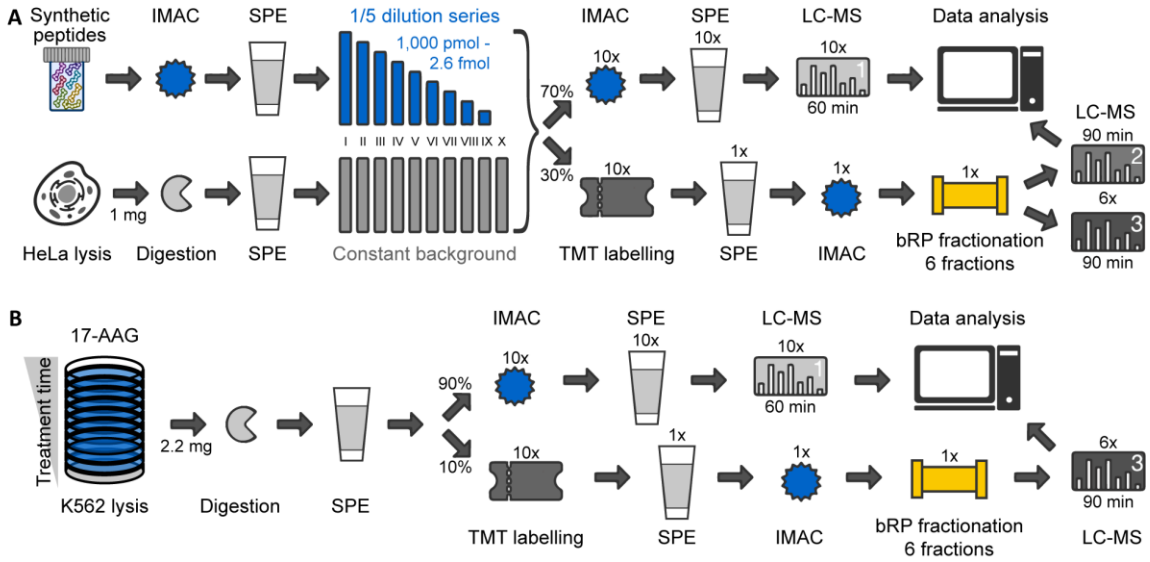



**Figure III-1 | Experimental designs for the comparison of label-free and TMT-based quantification.** *(A) Synthetic phosphopeptides were spiked into a constant HeLa background in 1/5 dilution steps including one sample without spike-in. Sample aliquots were either processed as label-free, phosphoproteome single-shots or subjected to a TMT-labelling workflow including peptide fractionation. (B) K562 cells were treated with 17-AAG for up to 32 h and digests were processed as described in (A).*

All samples were measured on a Fusion Lumos mass spectrometer and 1/5$^{th}$ of the TMT-labelled fractions and 1/7$^{th}$ the label-free single shots were injected. TMT-labelled fractions of spike-in experiments were measured employing both an 80 min MS2 and MS3 method for quantification of TMT reporter ions. In the MS2 method, peptides were fragmented with HCD (NCE of 33 %) and MS2 spectra were required at 50K resolution using 1e5 charges and a maxIT of 86 ms. For MS3 data acquisition, first a CID MSA fragmentation (35 % NCE) was performed to obtain a spectrum for identification in the Orbitrap at 30K resolution (AGC of 5e4 charges, maxIT of 60 ms). Then, the same precursor was fragmented again and the 10 most abundant fragments were subjected to HCD fragmentation (55 % NCE, AGC of 1e5 charges, maxIT of 120 ms) to obtain quantitative reporter ion information. For both methods, an isolation window of 0.7 m/z, a cycle time of 3 s and a dynamic exclusion of 90 s was employed. Label-free samples were measured using an effective gradient of 50 min resulting in a total of 12.5 h turn-around time (10x 75 min for total MS method length) versus 10.5 h for the TMT-based measurements (6x 105 min for total MS method length). Cycle time was set to 2 s and dynamic exclusion to 25 s. Peptides were fragmented via HCD (28 % NCE, isolation window of 1.2 m/z) and MS2 resolution, AGC and maxIT were set to 15K, 1e5 charges, and 50 ms, respectively. Following parameters were changed for measurement of samples from 17-AAG treated K562 cells: For non-labelled peptides, MS2 resolution and maxIT were increased to 30K and 120 ms, and a top15 method was used. For the MS3 method for TMT-labelled phosphoproteomes, peptides were fragmented using CID and the 10 most abundant precursors and MS3 maxIT was set to 120 ms. Raw data were searched using MaxQuant v1.6.2.10 (spike-in experiment) or v1.5.6.5 (17-AAG treatment) and the human reference proteome supplemented with sequences of the synthetic peptides for spike-in experiments.

## 2.2 Generation of a TMT-labelled standard

For the systematic assessment of TMT-ratio compression, MS methods and machine performance, a TMT-labelled, human-yeast-mix sample was designed analogous to the two-proteome standard established by Ting et al. [246]. As human background, equal quantities of HeLa-derived peptides (300 µg protein digest) were used in each channel (Figure III-2). Yeast peptides were added in increasing amounts in a 1:2:8 ratio in triplicates (10, 20, and 80 µg protein digest), while the first and last channel of the 11-plex sample did not contain yeast peptides. After TMT-labelling and pooling, the total ratio of yeast:human peptides was 1:10. Whole cell proteomes were utilized for comparison of MS2- and MS3-based quantification and for tracking of instrument performance using 50 min gradients, whereas phosphopeptides were employed for evaluation of different MS3 methods employing 80 min gradients.

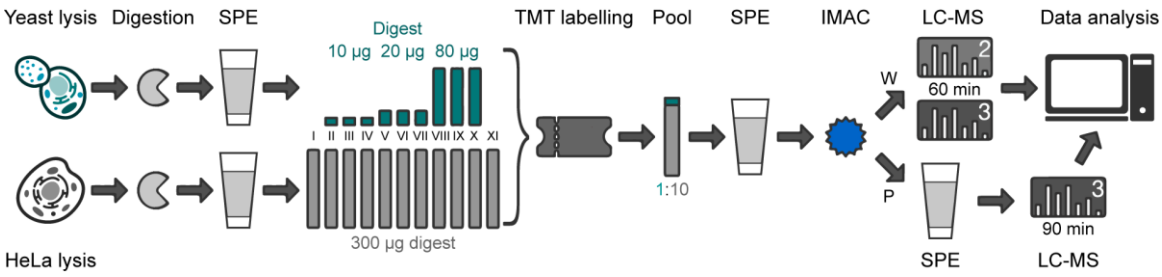



**Figure III-2 | Generation of an MS3 standard**. *Yeast peptides were spiked in low quantities as triplicates into a constant, dominating HeLa background. Outermost channels were left empty. Phospho- and whole cell proteomes were measured with varying parameters to optimize MS methods and assess ratio compression, quantification accuracy, and precision (W: whole cell proteome; P: phosphoproteome).*

MS3 method optimizations for phosphopeptides included amongst others the variation of the width of the MS2 isolation window (0.4 to 3 m/z) and the number of SPS notches (2 to 20) to assess their influence on ratio compression. Further, Orbitrap and ion trap read-out for MS2 spectra were compared using different fragmentation modes (CID, CID-MSA, and HCD) and excluding or allowing the neutral loss for subsequent SPS and fragmentation (Table III-1). The neutral loss exclusion was achieved via the precursor ion exclusion option that was, for example, set to disallow m/z values for SPS that are within a -50 to +5 window around a doubly charged precursor. For all MS3 methods for phosphopeptides, the cycle time was limited to 3 s, maxITs for MS2 and MS3 were set to 60 ms and 120 ms, respectively, and up to 1e5 charges were collected for the MS3 spectrum that was recorded at 50K resolution. MS3 mode measurements of whole proteomes were conducted applying the same parameters with following modifications: Up to 1e4 charges were collected for MS2 spectra and peptides were fragmented using CID and read out in the ion trap. For MS2 mode measurements, MS2 AGC and maxIT were increased to 1e5 charges and 86 ms and acquisition took place in the Orbitrap at 50K resolution after HCD fragmentation (NCE 33 %). Raw files were searched against the human and yeast reference proteomes using MaxQuant v1.6.2.10 as specified in general methods but setting the ion trap fragment match tolerance to 0.4 Da.

**Table III-1 | Parameters altered during optimization of MS2 acquisition in MS3 mode measurements of phosphopeptides.** *Only parameters that differed in at least one of the displayed methods are shown. For all CID(-MSA) methods, activation time was set to 10 ms and activation Q to 0.25 (OT: Orbitrap; IT: ion trap, NL: neutral loss).*

| | MSA OT | HCD OT | CID OT | CID OT NL | MSA IT | HCD IT | CID IT | CID IT NL |
|---|---|---|---|---|---|---|---|---|
| **Activation type** | CID | HCD | CID | CID | CID | HCD | CID | CID |
| **Collision energy** | 35 | 33 | 35 | 35 | 35 | 33 | 35 | 35 |
| **Multistage activation** | True | --- | False | False | True | --- | False | False |
| **NL Mass [m/z]** | 97.9673 | --- | --- | --- | 97.9673 | --- | --- | --- |
| **Detector type** | Orbitrap | Orbitrap | Orbitrap | Orbitrap | Ion trap | Ion trap | Ion trap | Ion trap |
| **Resolution (OT) or scan rate (IT)** | 30 k | 30 k | 30 k | 30 k | Rapid | Rapid | Rapid | Rapid |
| **AGC target** | 5e4 | 5e4 | 5e4 | 5e4 | 2e4 | 2e4 | 2e4 | 2e4 |
| **NL exclusion for SPS** | True | True | True | False | True | True | True | False |

## 2.3 Optimization of the TMT-labelling protocol

For evaluation of economically optimized TMT labelling conditions, peptides from different protein sources were labelled using various peptide and reagent quantities, and concentrations and TMT under- and overlabelling of peptides was investigated (Figure III-3). An overview of labelling conditions including quantities, volumes, and concentrations of labelling reactants, buffers, and solvents is provided in Table III-2, and, in more detail, in the supplemental information of the online publication ([323], suppl. table 2). All peptide amounts are based on quantification after digestion and clean-up.

The rationale of single optimization experiment series is described in more detail in the respective results sections. In brief, in three independent experiments, increasing peptide quantities (12.5 to 800 µg) were labelled using the same TMT concentration and quantity (100 or 800 µg). To this end, 11 conditions were included as technical, intra-laboratory duplicates or triplicates. Moreover,

17 samples were labelled in three experiment series as singlicates applying different TMT (40 to 400 µg) and peptide (40 or 200 µg) quantities and concentrations to explore the impact of these parameters on labelling performance and to examine the adaptability of optimized protocol parameters to lower peptide quantities. To assess inter-laboratory robustness, four labelling experiments, in which the TMT quantity was titrated (50 to 400 µg) against a constant peptide amount (100 µg), were carried out as 7 replicates of which 2 or 3 were performed in three independent laboratories. After stopping the reaction, peptide solutions were either acidified using 45 % (vol/vol) of 10 % FA in 10 % ACN or directly frozen at -80 °C and dried by vacuum centrifugation. All experiments for method optimization were measured as single-shot LC-MS/MS runs (for measurement parameter see Appendix Table 0-1) and searched against the human and/or mouse reference proteome using MaxQuant v1.6.3.3.

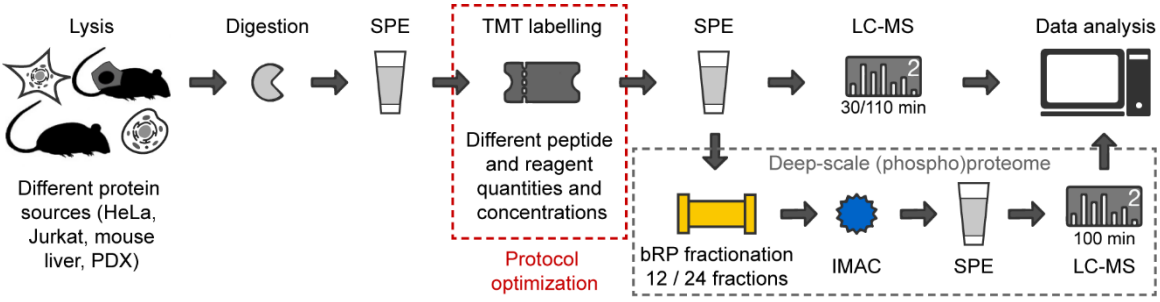



**Figure III-3 | Workflow for the optimization of the TMT-labelling protocol**. *Varying peptide amounts from different protein sources were labelled using varying reagent quantities and concentrations (red box) and measured either as single shots or after fractionation (24 fractions) and phosphopeptide enrichment (12 fractions, grey box). Labelling efficiency, overlabelling of serine, threonine, tyrosine and histidine and reproducibility were analysed to assess the performance of the differing labelling protocols.*

Further, to evaluate the utility of the final, optimized labelling protocol to highly fractionated samples, a deep-scale (phospho)proteome analysis was performed as previously described [194] but using the optimized protocol (i.e. using 8x less TMT reagent) and comparing the result to the original labelling protocol. Briefly, peptides derived from digests of basal (B) and luminal (L) breast cancer PDX models (WHIM2 and WHIM16) were labelled in 5 replicates within a TMT10-plex experiment (TMT channels: B-L-B-B-L-B-L-L-B-L), fractionated using high pH reversed-phase (RP) chromatography, and subjected to IMAC enrichment. Finally, 24 whole proteome and 12 phosphoproteome fractions were measured by LC-MS/MS (Figure III-3, Appendix Table 0-1) and analysed using Spectrum Mill vB.06.01.202. Sample preparation and data analysis of the deep scale experiment were performed in the proteomics facility of the Broad Institute of MIT and Harvard, Massachusetts, USA.

**Table III-2 | Overview of labelling conditions of different experiment series for optimization of TMT labelling.** *Concentrations (conc.) refer to concentrations during the labelling reaction ($NH_2OH$: hydroxylamine).*

| | VENDOR | DOWN-SCALED 1 | DOWN-SCALED 2 | JURKAT | PDX | LOW INPUT | INTER-LAB | DEEP-SCALE |
|---|---|---|---|---|---|---|---|---|
| **Peptide source** | Mouse liver | Mouse liver | HeLa | Jurkat | PDX | PDX | PDX | PDX |
| **TMT reagent** | Mix of TMT10 | Mix of TMT10 | TMTzero | TMTzero | TMTzero | TMTzero | TMTzero | TMT10 |
| **TMT / peptide (wt/wt)** | 8to1 - 1to1 | 8to1 - 0.5to1 | 8to1 - 0.5to1 | 2to1 - 0.5to1 | 2to1 - 0.5to1 | 8to1 - 1to1 | 4to1 - 0.5to1 | 8to1 and 1to1 |
| **Peptide (µg)** | 100 - 800 | 100 - 800 | 12.5 - 200 | 200 | 200 | 40 | 100 | 300 |
| **TMT (µg)** | 100 | 100 and 800 | 100 | 100 - 200 | 100 - 200 | 40 - 320 | 50 - 400 | 300 and 2400 |
| **TMT stock conc. (mM)** | 56.7 | 56.7 | 59 | 59 | 59 | 59 | 28.1-224.6 | 56.7 |
| **Hepes (µl)** | 100 | 20 | 20 | 20, 40 | 40 | 16-80 | 20 | 60, 300 |
| **TMT stock (µl)** | 41 | 5 | 5 | 5, 10 | 2-20 | 2, 4 | 5 | 15, 123 |
| **Total labelling volume (µl)** | 141 | 25 | 25 | 25-50 | 45-60 | 18-96 | 25 | 75, 423 |
| **TMT conc. (mM)** | 16.5 | 11.3 | 11.8 | 6.6-19.7 | 6.6-19.7 | 1.4-29.5 | 5.6-44.9 | 11.3, 16.5 |
| **Peptide conc. (g/l)** | 0.7-5.7 | 4, 8 | 0.5-8 | 4-8 | 3.3-4.4 | 0.4-2.2 | 4 | 1, 4 |
| **ACN conc. (%)** | 29 | 20 | 20 | 11-33 | 11-33 | 2-29 | 20 | 20, 29 |
| **Quenching** | 8 µl Tris, pH 8 | 8 µl Tris, pH 8 | 2 µl $NH_2OH$ | 2-4 µl $NH_2OH$ | 3-5 µl $NH_2OH$ | 1-7 µl $NH_2OH$ | 2 µl $NH_2OH$ | 6, 32 µl $NH_2OH$ |
| **Conditions** | 4 | 2 | 5 | 4 | 3 | 10 | 4 | 2 |
| **Replicates** | 3 | 3 | 2 | 1 | 1 | 1 | 2 or 3 | 1 |

# 3 Results

In order to assess label-free and TMT-based workflows qualitatively and quantitatively for peptide-level analysis, various experimental setups (Figure III-1, Figure III-2) were employed and the performance of the different quantification methods was evaluated for phosphopeptides.

## 3.1 A compromise between accurate dynamic range and reproducible coverage

A 5-fold diluted series of two synthetic peptide pools, featuring phosphorylation sites on either the most N-terminal amino acid or the most C-terminal amino-acid before lysine, were spiked into a constant amount of HeLa peptides and measured employing a label-free or a TMT-based quantification strategy in MS2 or MS3 mode (Figure III-1 A). Apart from the respective base peptide sequence in each pool, all other 59 peptides with permutated amino acids at and around the phosphorylation site represented non-endogenous phosphopeptides. This setup allowed for examination of the dynamic range of accurate, relative quantification in the presence of a non-changing background and the assessment of the correct site localization.

**Dynamic range, accuracy and precision of different quantification approaches** – In the label-free dataset, the sample with the highest amount of peptide spike-in (1,000 fmol) showed selectively decreased intensities of hydrophilic peptides and a shift in the peptide elution profile (Appendix Figure 0-1 A). This unequal loss of hydrophilic peptides that was likely a result of trap column overloading was also apparent by worse correlations of HeLa peptide intensities to all other samples (average R=0.85) compared to other pairwise correlations (average R=0.95, Appendix Figure 0-1 B). The sample containing the second highest quantity of synthetic peptides (200 fmol) also showed a slight, less obvious decrease in correlation coefficients (average R=0.93). In addition, both samples exhibited notably lower MS2 spectra identification rates (10.4 and 16.2 % compared to 19.5 % in other samples, Appendix Figure 0-1 C). This may be caused by the high proportion of spike-ins, interfering with and hampering the identification of less abundant peptides. To circumvent any potential quantitative bias, ratios for label-free samples were therefore calculated based on the sample with the third highest amount of synthetic phosphopeptides. For the TMT-labelled sample, ratios were computed relative to the channel with the second highest quantity of non-endogenous spike-in peptides.

A total of 93 synthetic phosphopeptides were identified with the phosphorylation on the designated amino acid in any of the three quantification approaches. Additional 26 and 39 were detected with the phosphorylation annotated in the wrong position. Despite false localizations, all of these peptides should exhibit the expected five-fold decrease in signal intensity from sample to sample and were thus included in the comparison of expected and measured ratios across samples. Data from TMT-labelled samples acquired in MS2 mode quantified most synthetic phosphopeptides in a single condition, followed by the MS3-based and label-free approach (118, 114, and 99 peptides). Label-free quantification, however, featured the highest dynamic range (Figure III-4 A). Accurate relative quantification in this experimental setup could be obtained for phosphopeptides differing up to roughly 625-fold, only showing a deviation from the anticipated ratios for very low and very high spike-in amounts. For the latter, also the background signal was lower than expected. This was in line with above mentioned observations in chromatograms and HeLa peptide correlations (Figure III-1 A-B) and suggested a bias in the samples with the highest spike-in amounts that hindered the approximation of the upper limit of the dynamic range. The proportion of quantitative values arising from the match-between-runs function in MaxQuant increased linearly with decreasing, synthetic phosphopeptide quantities. From the sixth dilution

step onward, quantification was solely based on few matches across runs and started to diverge from expected intensities (Figure III-4 A). In the sample void of synthetic spike-ins, only one phosphopeptide was spuriously matched. MS2-based TMT quantification performed poorly compared to label-free quantification with a dynamic range of only, at best 25-fold (Figure III-4 B). Even within this range, the average of the log ratio distributions deviated slightly from the anticipated log ratios likely due to ratio distortion caused by co-isolation, co-fragmentation, and co-quantification of non-changing HeLa peptides. This also resulted in the detection of reporter ion intensities in the, by experimental design, empty channel for 58 % (69) of all quantified phosphopeptides. For measurements in MS3 mode, this fraction dropped to 11 %, and the dynamic range increased 5-fold to roughly 125-fold (Figure III-4 C). Generally, the TMT-based strategies exhibited a higher precision indicated by the smaller error bars of log ratios for background peptides and a lower median CV of 16 % (MS2) and 14 % (MS3) compared to 19 % for label-free quantification of HeLa phosphopeptides (excluding the two samples with highest spike-in quantities).

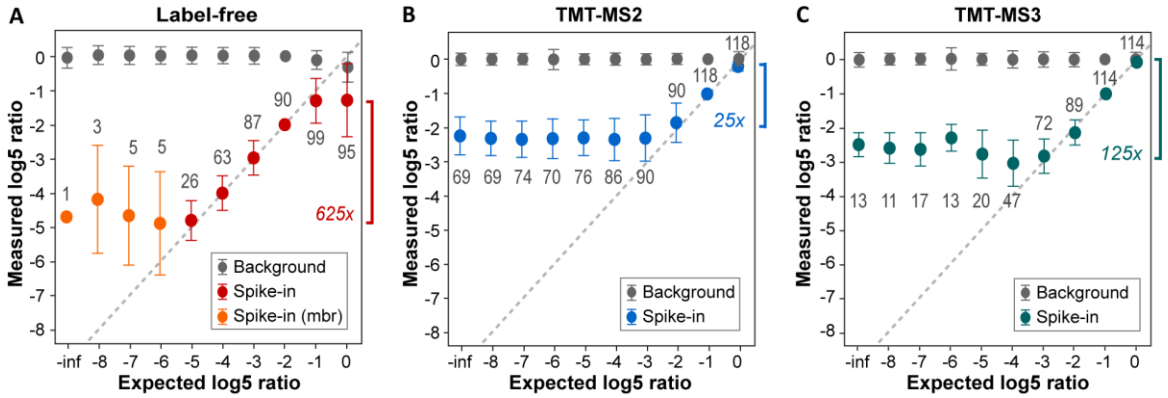



**Figure III-4 | Linear dynamic range of label-free and TMT-based phosphopeptide quantification examined with synthetic peptide spike-ins.** *A comparison of the expected and measured ratios of synthetic phosphopeptides (spike-in) mixed in increasing quantities into a complex HeLa proteome (background) illustrates the accuracy, precision (error bars), and dynamic range obtained by label-free (A), and TMT-based quantification in MS2 (B) and MS3 (C) mode. Estimates of the dynamic ranges are illustrated by brackets. Numbers for peptides with non-zero intensities included in the analysis are indicated. All phosphorylated peptides apart from the endogenous base sequence were considered, irrespective of the phosphorylation site localization (error bars: standard deviation; mbr: exclusively identified by the match-between-runs algorithm).*

**Reproducible peptide detection and correct localization of phosphorylation sites** – In addition to the dynamic range of accurate quantification, the reproducible detection of peptides was of interest. This was examined including quantitative information about the constant background. The label-free and MS2-based TMT approach quantified comparable numbers of distinct, modified peptide sequences (12,465 and 13,629) in any of the ten samples, whereas expectedly less peptides were identified in the MS3-based measurements (11,748, Figure III-1Figure III-5 A). The overlap of identified peptide sequences among TMT-labelled samples was bigger than for the label-free and the TMT approaches suggesting that the two different quantification strategies tend to identify in part different subsets of the phosphoproteome likely due to a change in peptide properties. Despite comparable total identifications, the label-free approach suffered severely from missing values showing zero intensities in at least one of the samples for 82 % of identified peptides (Figure III-5 B). This fraction decreased to 71 % when the sample with the highest spike-in amount was excluded. On the contrary, only 2.1 % and 3.3 % of peptides had zero intensities in any of the TMT-labelled samples in MS2 and MS3 mode, respectively. While label-free peptides with zero and one missing value exhibited on average higher intensities, this negative correlation of missing values and intensity could not be detected for 3 or more missing values (Figure III-5 C).

Moreover, non-labelled peptides with more missing values exhibited significantly lower localization probabilities for phosphorylation sites (Kruskal-Wallis p-value < 0.001 for spike-in peptides, Figure III-5 D). This was also apparent from the higher fraction of class I sites (localization prob. > 0.75) in peptides with fewer missing values and an increasing fraction of class III sites (localization prob. < 0.5) for peptides with 7 or more missing values (Figure III-5 D).

With a ground truth dataset of synthetic phosphopeptides at hand, the performance of the three quantification methods could also be examined regarding the correct localization of the phosphorylation. The peptide pool bearing the phosphorylation at the N-terminus of the peptides (pool 61, see p. 40) generally showed a higher proportion of uncertain or mislocalized sites (15-29 % vs. 7-15 % for peptide pool 14, Figure III-5 E). Strikingly, this was most profound for label-free phosphopeptides for which only 71 % identified spectra contained a class I site. In contrast, TMT-labelled peptides measured in MS3 mode outperformed the other two approaches and featured the highest percentage of correctly localized sites in both peptide pools (93 and 85 %, Figure III-5 E).

**Comparison of label-free and TMT-labelling approaches in a real-world scenario** – To examine whether similar advantages and limitations of label-free and TMT-based quantification as in the spike-in experiment can be observed in a biological setting, a time-dependent treatment of K562 cells with the HSP90 inhibitor 17-AAG was chosen as a real-world scenario (Figure III-1 B). Here, instead of increasing amounts of synthetic peptides, equal total peptide quantities were present, but an overall higher percentage of phosphopeptides was expected to exhibit changes across conditions as compared to the

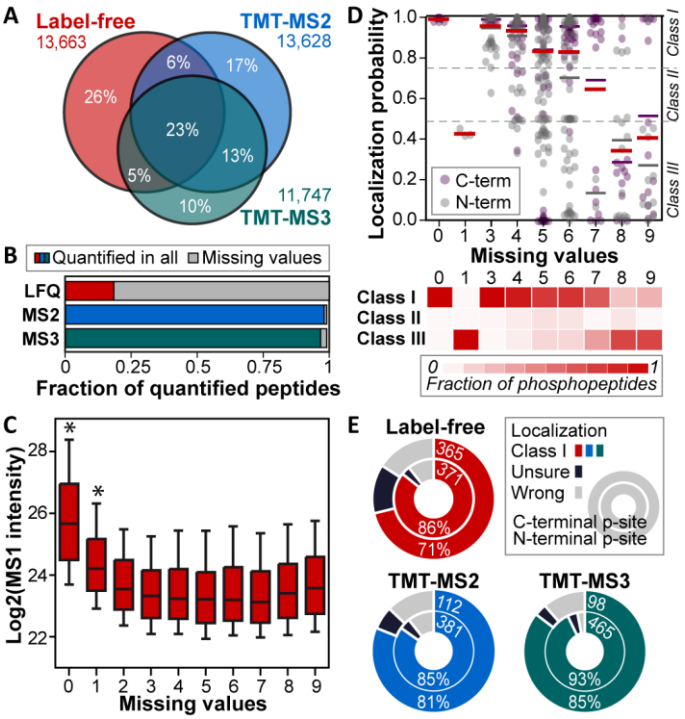



***Figure III-5 | Missing values and localization probabilities for label-free and TMT-based phosphopeptide quantification in synthetic peptide spike-in experiments.*** (A) The overlap of total identified, distinct modified peptides is shown. (B) The fraction of peptides exhibiting a zero intensity (missing value) in at least one sample is displayed. (C) Boxplots reveal intensities of peptides with increasing numbers of missing values in the label-free dataset. Asterisk mark distributions that are significantly different from all other distributions (Kruskal-Wallis and Dunn's post hoc test, p-value <0.001; box: $25^{th}$-$50^{th}$-$75^{th}$ percentile; whiskers: $10^{th}$ and $90^{th}$). (D) A dot plot and heat map illustrate localization probabilities and proportions of localization certainty categories (Class I, II, and III as indicated) for label-free peptides dependent on the number of missing values. Horizontal lines indicate the average per missing value category (purple: average of pool 14 with a C-terminal phosphorylation; grey: average of pool 61 with an N-terminal phosphorylation, red: total average). (E) Doughnut diagrams indicate the proportions of PSMs that feature different localization certainties within the two different peptide pools. Total number of spectra and fraction of class I sites are indicated (p-site: phosphorylation site; class I: localization probability >0.75; unsure: loc. prob. between 0.75 and 0.5 or identical to a different site; wrong: loc. prob. smaller than on another site).

spike-in experiment. Based on the previously determined, limited dynamic range and accuracy of MS2-based quantification, TMT-labelled peptides were only measured in MS3 mode. Similar to observations in spike-in experiments, a higher total number of modified peptide sequences was identified using the label-free (19,216) compared to the TMT-based approach (16,805, Figure III-6 A). However, missing values accumulated quickly with every additional, label-free sample, while TMT-labelled peptides showed virtually no missing values. Only 3,652 (19 %) or 6,701 label-free peptides (35 %) were quantified across all conditions with or without the match-between-runs function, compared to 16,422 TMT-labelled peptides (98 %). Moreover, a higher fraction of TMT-labelled peptides was phosphorylated compared to label-free peptides (90 vs. 79 %, Appendix Figure 0-2 A) which further improved the coverage of phosphopeptides across conditions in the TMT (14,747) compared to the label-free dataset (5,278, Figure III-6 A). Of note, the spectra identification rate was much lower after TMT-labelling (19 % vs. 33 %, Appendix Figure 0-2 A) suggesting that the depth of the TMT-labelled phosphoproteome could be additionally enhanced when low identification rates are tackled. Identical to the spike-in experiment, only initially decreasing intensities but constantly declining localization probabilities were observed with increasing numbers of missing values in the label-free dataset (Kruskal-Wallis p-value < 0.001, Appendix Figure 0-2 B, Figure III-6 B). Often the localization was uncertain when several potential phosphorylation sites were present in close proximity. In fact, ambiguous allocation of modifications could lead to a higher number of missing values, and artificially increase the number of total, quantified peptides since every new, inconsistently localized phospho-isomer would result in an additional modified peptide identification (Figure III-6 C).

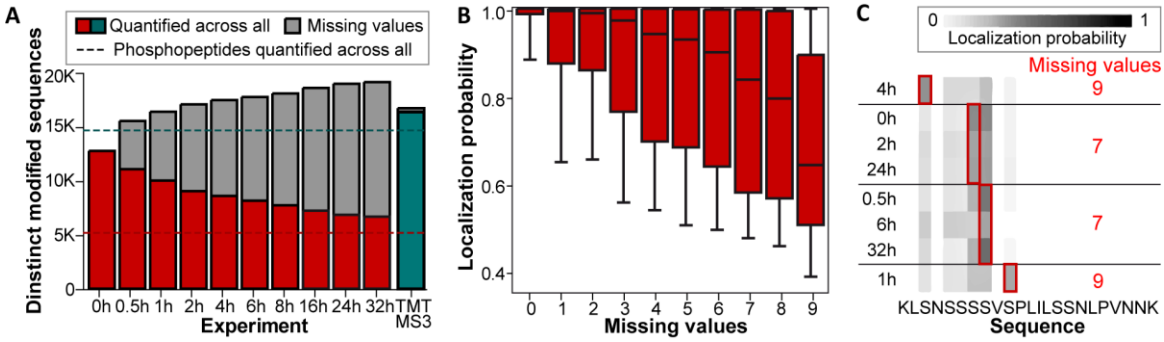



**Figure III-6 | Comparison of label-free and MS3-TMT based phosphopeptide identification and quantification for time-dependent 17-AAG treatments.** *(A) The cumulative number of quantified peptides and missing values (zero intensity) is displayed for label-free (match-between-runs enabled) and TMT-labelled phosphoproteomes. The latter were measured in MS3 mode. (B) Boxplots illustrate the distribution of localization probabilities for peptides with increasing numbers of missing values in the label-free dataset (box: 25$^{th}$-50$^{th}$-75$^{th}$ percentile; whiskers: 10$^{th}$ and 90$^{th}$ percentile). (C) An example of a phosphopeptide with several uncertain site allocations in different label-free samples reveals how ambiguous site allocation can increase the number of total phosphopeptide identifications and missing values.*

In summary, label-free quantification outperformed TMT-based quantification with regard to accurate dynamic range, but suffered from missing values that considerably reduced the number of quantifiable peptides across several conditions. This issue was aggravated by ambiguous site localizations in label-free phosphoproteomic analyses and resulted in a superior performance of TMT-based approaches when it came to reproducible detection of the same peptide across conditions. However, the accuracy and dynamic range of MS2-based TMT quantification was limited due to ratio compression. In contrast, measurements in MS3 mode exhibited an improved and seemingly acceptable dynamic range and accuracy. A more detailed comparison of the two different TMT quantification strategies should provide further clues about when MS2 mode analyses

may be adoptable and when measurements in MS3 mode are inevitable and thus was conducted as a next step.

## 3.2 A versatile TMT standard

To generate a standard sample for a more systematic assessment of peptide co-isolation and subsequent ratio distortion in MS2 and MS3 mode analyses, small quantities of yeast peptides were mixed into a high, constant human background and labelled with TMT (Figure III-2). Single shot measurements of the phospho- and whole proteome of this species interference sample were further utilized for MS3 method optimization and tracking of machine performance.

**Quantitative performance of measurements in MS2 and MS3 mode** – Due to the high ratio of the human:yeast mix (10:1), the proportion of identified yeast peptides was much smaller than for human peptides (Table III-3). For the whole cell proteome, roughly 11 % of identified peptides were derived from yeast (977 and 498), whereas only 5 % of peptides could be assigned to yeast in the results the phosphoproteome measurements (302 and 192, Figure III-7 A, Appendix Figure 0-3 A). This was likewise reflected by high human-to-yeast MS1 intensity ratios of 9.5 and 37.8 for unmodified and phosphorylatd peptides, respectively. Phosphorylated yeast peptides displayed a smaller overlap across measurement modes than yeast peptides in the full proteome sample (Figure III-7 A, Appendix Figure 0-3 A). As expected, MS3 mode analyses detected overall less peptides showing a 22 % and 45 % decrease in identifications for phospho- and full proteomes, respectively (Table III-3). However, yeast peptides quantified in MS3 mode exhibited much less co-isolation of human peptides indicated by the overall lower intensities and the higher proportion of actual zero intensities in the two, by experimental design, empty TMT channels (Figure III-7 B, Appendix Figure 0-3 B). This, in turn, also resulted in a significant reduction of ratio distortion and shifted measured ratio distributions closer to their expected values (e.g. median log2 ratio of 2.68 vs. 1.58 for phosphopeptides and expected log2 ratio of 3, Figure III-7 C, Appendix Figure 0-3 C). In MS2 mode, the phosphopepetide quantification appeared to be overall more compressed compared to peptides in the whole cell proteome measurements. However, this difference in ratio compression was diminished by the additional fragment ion selection step in the MS3 methods rendering the ratio distributions of phospho- and unmodified peptides virtually indistinguishable (Figure III-7 C, Appendix Figure 0-3 C).

**Table III-3 | Numbers of quantified, modified peptides sequences identified in MS2 and MS3 mode in the whole cell and phosphoproteome of the TMT standard and their species assignment.**

| | MS2 MODE | MS3 MODE |
|---|---|---|
| **Whole cell proteome** | **8,415** | **4,625** |
| Human | 6,841 (87.9 %) | 4,083 (88.3 %) |
| Yeast | 882 (11.3 %) | 498 (10.7 %) |
| Human/yeast | 58 (0.7 %) | 44 (1.0 %) |
| **Phosphoproteome** | **5,693** | **4,444** |
| Human | 5,386 (94.6 %) | 4,250 (95.6 %) |
| Yeast | 302 (5.3 %) | 192 (4.3 %) |
| Human/yeast | 5 (0.1 %) | 2 (0.05 %) |

Further, the precision of quantification was evaluated by calculating the CVs within each ratio group for yeast peptides and across all channels for human peptides. Generally, the latter exhibited slightly smaller CVs compared to yeast peptides with exception of MS2 mode measurements of whole cell proteomes, which showed comparable CVs (Figure III-7 D, Appendix Figure 0-3 D). While MS3 measurements significantly lowered precision for unmodified peptides indicated by increased CVs, phosphorylated peptides surprisingly showed no such reduction in the quantification reproducibility, but even significantly decreased in CVs for HeLa peptides acquired

in MS3 mode (Mann-Whitney p-value < 0.0001). In general, more then 90 % of CVs were below 20 %, indicating an excellent precision in all cases which can be beneficial to identify statistical significant differences across groups irrespective of the accuracy of intensity ratios.

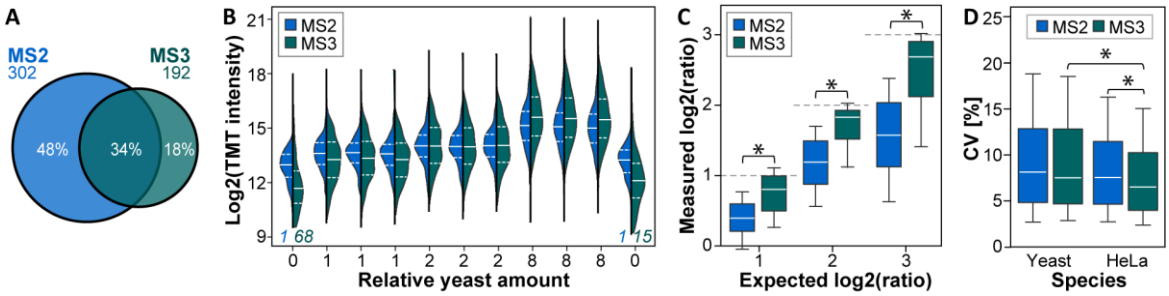



**Figure III-7 | Evaluation of TMT-based phosphopeptide quantification in MS2 and MS3 mode using a yeast-human species-mix sample**. *(A) The overlap of identified yeast phosphopeptides is shown. (B) TMT intensity distributions of yeast phosphopeptides are displayed for different channels containing different amounts of yeast sample. Intensities detected in the first and the last channel arise from co-isolation of human peptides. The number of peptides with zero intensities in these channels is indicated. (C) Distributions of measured intensity ratios of phosphopeptides (calculated from intensity averages of each group) are illustrated in comparison to the expected ratios for different yeast amounts. (D) Coefficients of variation (CV) are shown for HeLa and yeast phosphopeptides. Asterisks denote significantly different distributions (Mann-Whitney p-value < 0.0001).*

**Statistical recovery of differential abundances of yeast peptides** – Next, it was assessed to which extent the known differences in the triplicate yeast channels could be recovered by statistical tests using MS2 and MS3 data. One-sided t-tests were performed for all group-wise comparisons (exemplified for phospho data in Figure III-8 A). Then, true positive and true negative rates (sensitivity and specificity), and false positive and false negative rates were calculated based on the fraction of yeast and HeLa peptides that were identified to be significantly different (Figure III-8 B, Appendix Figure 0-4 A). Further, ROC curves were computed using corrected p-values (q-values) to assess the overall diagnostic ability of the two quantification strategies to discriminate between upregulated yeast and non-changing HeLa peptides (Figure III-8 C, Appendix Figure 0-4 B).

For expected 4-fold and 8-fold differences (log2 of 2 and 3), the sensitivity of detecting a quantitative change through the statistical test was very high in phospho- and whole cell proteome data and for both quantification approaches (> 94.7 % for phosphopeptides and > 96.7 % for unmodified peptides, Figure III-8 B, Appendix Figure 0-4 A). In turn, this resulted in small false negative rates (< 5.3 % and < 3.3 %). Due to the enhanced (phospho)proteome depth, the absolute number of significant yeast peptides was much higher for MS2 data (286/287 vs. 182/181 for 4-fold/8-fold differences of phosphopeptides, Figure III-8 A). However, MS3 acquisition led to an improved specificity of detection of these larger differences (> 98.3/94.2 % vs. < 93.9/87.7 % for phospho-/unmodified peptides) and a likewise decreased false positive rate (Figure III-8 B, Appendix Figure 0-4 A). In general, MS3-based TMT quantification showed a much smaller false postive rate with maximal 5.8 % for 8-fold differences in the full proteome data. In contrast, MS2 data entailed up to 14.1 % human peptides that were falsly identified as significantly upregulated despite the 5 % FDR that was applied in the statistical test. This is likely a result of yeast peptides being co-isolated with human peptides and distorting their technically unchanged abundances. This is in line with negligible false positive rates for the smallest ratio of 2 (<1.2 %). For this, instead, false negative rates increased substantially (not less than 21.7 %), especially for MS2-based quantification of phosphopeptides (66.6 %) resulting in an accordingly decreased

sensitivity of detecting these smaller changes (33.4-78.2 %, Figure III-8 B, Appendix Figure 0-4 A). Consequently, MS3-based acquisition even identified more significant yeast phosphopeptides among 2-fold upregulated ones compared to MS2 (145 vs. 101, Figure III-8 B). These observations were also resembled by computed ROC curves. They revealed that both quantification approaches performed fairly well in identifying specifically yeast peptides as significantly different (in at least 97.3 % of the cases for an at least 4- and 8-fold change, as implied by the areas under the ROC-curves). However, a performance drop was detectable for the comparison of the two groups which differed by the smallest factor of 2 (max. 91.1 % of the cases, Figure III-8 C). In a nutshell, MS3 data yielded improved or at least comparable sensitivity and specificity for all tested ratios and, hence, also minimized false negative and false positive rates, but identifed less sigificant peptides in absolute terms for larger ratios. Of note, in our experimental setup, especially phosphorylated peptides benefited from MS3-based quantification (compare Figure III-8 B-C, Appendix Figure 0-4 A-B).

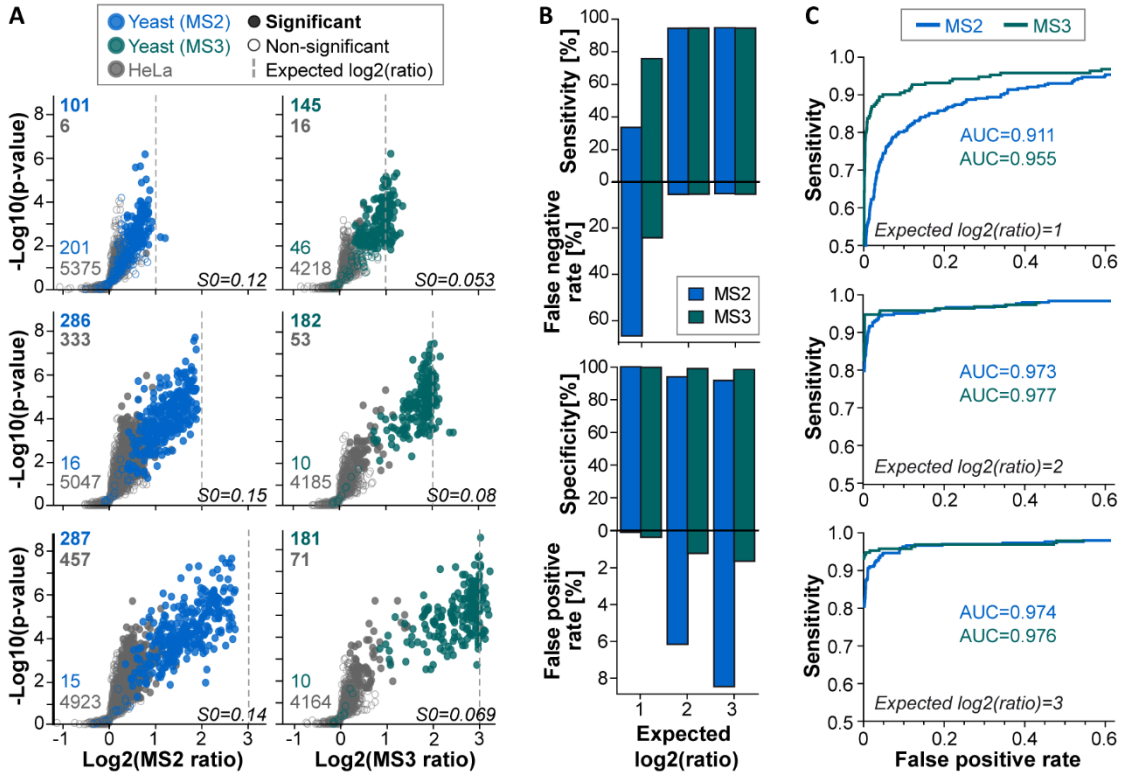



**Figure III-8 | Statistical recovery of differential phosphopeptide abundances in the yeast-human species-mix sample.** *(A) The results of one-sided student's t-tests comparing groups of TMT channels that contain different yeast quantities are plotted for MS2 and MS3 based TMT quantification. A 5% permutation based FDR and S0 (as indicated) cut-off were applied. (B) Sensitivity (significant yeast fraction) and specificity (non-significant human fraction) and corresponding false negative and positive rates are shown for t-tests conducted in (A). (C) ROC curves computed from corrected p-values obtained from t-tests in (A) are shown (AUC: area under the curve).*

**MS method optimization** – In addition to a global comparison of MS2 and MS3-based MS methods, the species interference sample allowed for a systematic evaluation of the impact of individual method parameters on identification and quantification performance. First, its suitability to assess ratio compression was confirmed in measurements with varying isolation windows. As expected, broadening the MS2 isolation window stepwise from 0.4 to 3 m/z significantly suppressed expected ratios for the MS2-based quantification approach (Kruskal-Wallis p-value < 0.0001, e.g.

decrease of median log2 ratio from 2.56 to 2.09 for expected log2 ratio of 3, Appendix Figure 0-5 A). At the same time, TMT reporter intensities increased continuously with wider isolation windows and lead to a concomitant improvement of precision (Appendix Figure 0-5 B-C). More surprisingly, similar results were observed for MS3 data despite the additional selection step of fragment ions before TMT acquisition (Appendix Figure 0-5 A-C). Yet, this trend was overall less pronounced (median log2 ratio of 2.85 to 2.59 for expected log2 ratio of 3, Appendix Figure 0-5 C) and not significant for isolation windows between 0.4 and 1.2 m/z. Likewise, stepwise increasing the number of fragment ions (SPS notches) used for the MS3 spectrum from 2 to 20, increased ratio distortion, TMT intensities, and precision of quantification, albeit to a lesser extent (median log2 ratio of 2.87 to 2.73 for expected log2 ratio of 3, Appendix Figure 0-5 D-F). Other parameters (injection times or amounts, MS1 resolution, advanced peak detection, inject beyond for MS2 spectra) had no measurable effect on ratio compression or general performance of MS3-based methods.

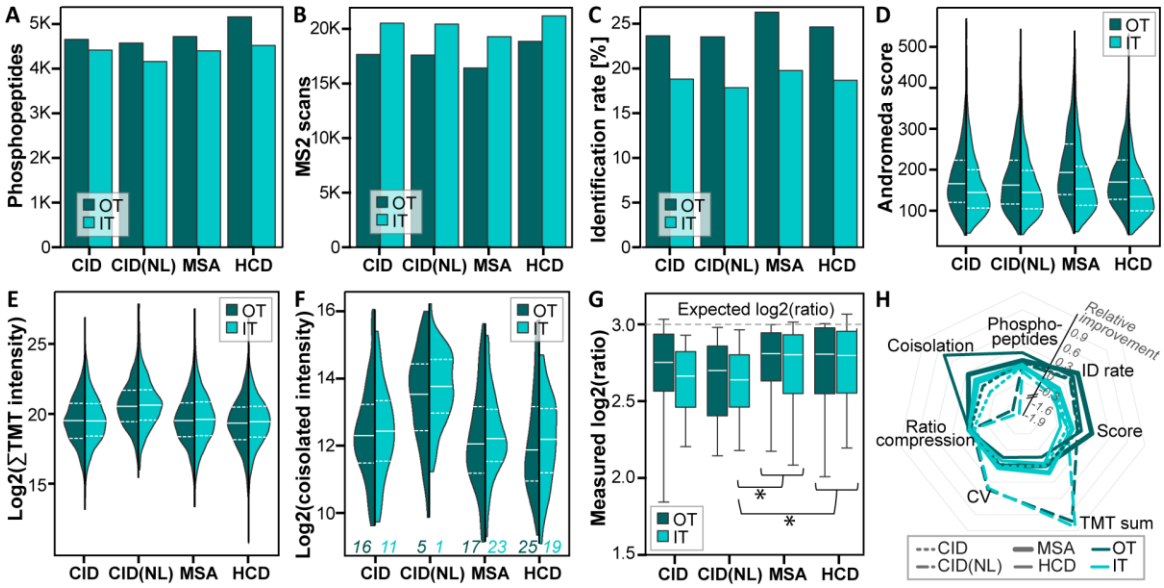



**Figure III-9 | Optimization of MS2 acquisition for MS3 mode measurements of phosphopeptides.** *Bar charts indicate (A) numbers of quantified phosphopeptides, (B) acquired MS2 scans corresponding to the speed of the different methods, and (C) the spectra identification rate. Violin plots show the distributions of (D) Andromeda scores, (E) summed TMT intensities for all phosphopeptides, and (F) TMT intensities for yeast phosphopeptides in the first and last, theoretically empty channels illustrating co-isolation of HeLa peptides. (G) Boxplots display the distributions of log ratios (calculated from intensity averages of each group) which would be expected to be 3. For all distribution plots, only phosphopeptides that were present in results of all 8 methods were included (n=46). (H) The radar plot summarizes results of the optimization runs with outward facing spikes showing an improvement and inward directed spikes indicating a decline in performance for certain methods relative to the other methods. For specifics of different methods, see also* ***Table III-1****.*

The phosphopeptide enrichment of the TMT standard was further utilized to optimize MS2 data acquisition and phosphopeptide identification specifically for the MS3 method. Depending on the fragmentation mode, neutral losses of the phosphate group can be common and lead to sparse ion series, which makes the identification of phosphopeptide sequences often more challenging compared to unmodified peptides. Therefore, following fragmentation methods were combined with either Orbitrap (OT) or ion trap (IT) readout and examined for their performance for phosphopeptide identification and quantification (Table III-1): CID while excluding or allowing the neutral loss in the subsequent fragment ion selection for the MS3 spectrum, CID with an additional multi-stage activation (MSA) step targeting the neutral loss, and HCD fragmentation. In general,

OT methods identified 6 to 14 % more phosphopeptides (on average 4,781 vs. 4,375) despite less acquired MS2 spectra (17,650 vs. 20,454) but explained by vastly improved identification rates (24.8 vs. 19 %, Figure III-9 A-C). This was also connected to improved Andromeda scores, especially for the MSA and HCD method (median of 133 vs. 114, Figure III-9 D). As expected, TMT intensities in the MS3 spectrum were unaffected by the analyser used for MS2 spectra acquisition. Surprisingly, though, ratio distortion appeared to be somewhat increased after IT readout implied by slightly higher intensities of yeast peptides in the, by experimental design, empty TMT channels and log2 ratios that were farther off the expected ratio, in particular for the CID method (2.67 vs. 2.76 for expected log2 ratio of 3, Figure III-9 E-F). These differences, however, were not statistically significant (according to group-wise Mann-Whitney tests). With respect to fragmentation modes, HCD with OT readout yielded the highest number of phosphopeptide identifications (5,163) which was mostly connected to its enhanced speed compared to CID-based methods (Figure III-9 A-C). Although the CID-MSA-OT method is comparably slow due to the additional fragmentation of the neutral loss peak, it provided the second highest number of phosphopeptides (4,724), better scores, and identification rates based on the improved ion series (Figure III-9 A-D). Including the neutral loss after CID fragmentation for generation of the MS3 spectrum, naturally resulted in the highest TMT intensities, but also the largest ratio compression, although this was only significant for the comparison to HCD and MSA methods (Mann-Whitney p-value < 0.05, Figure III-9 E-H). Taken together, CID-MSA and HCD spectra acquired in the Orbitrap performed best for peptide identification in MS2 with CID-MSA potentially providing an additional advantage for more confident localization of phosphorylation sites reflected by the higher Andromeda scores.

## 3.3 A robust and cost-efficient TMT-labelling protocol

After TMT-based quantification proved advantageous in overcoming missing values across multiple samples and drawbacks such as ratio compression were shown to be manageable with specialized data acquisition methods, the high costs of TMT workflows were to be tackled as a next step. For this purpose, different peptide quantities were labelled using decreasing reagent quantities at various concentrations and labelling efficiency was examined.

**Reducing reagent-to-peptide ratios for high protein quantities** – To assess the minimally required amount of TMTs necessary for sufficient labelling of a certain peptide quantity, corresponding amounts of primary amines were estimated based on an in silico digest. One-hundred µg of human peptides yielded ~116 nmol free primary amines. The TMT-labelling protocol provided by the manufacturer recommends adding 800 µg labelling reagent, which equates to 2.32 µmol of TMT (2.36 µmol in case of TMTzero), to peptides originating from a digest of 25 to 100 µg protein. Hence, the standard protocol uses at least a 20-fold molar excess of the labelling reagent. Even if a certain degree of reagent hydrolysis and overlabelling on hydroxyl and imidazole groups is taken into account, the TMT reagent is still applied in great excess. These theoretical considerations suggest that considerably higher quantities than 100 µg peptides could be labelled using 800 µg TMT reagent.

**Table III-4 | Theoretical amount of functional groups in a complete digest of 100 µg of a human proteome.** *Estimations are based on average peptide length, and pKa values were taken from literature* [329, 330].

| FUNCTIONAL GROUP | AMOUNT [nmol] | PKA |
|---|---|---|
| α-amine on N-term | 78 | 7.7 ± 0.5 |
| ε-amine on Lys | 38 | 10.5 ± 1.1 |
| ***Primary amines*** | ***116*** | |
| Hydroxyl on Tyr | 25 | 10.3 ± 1.2 |
| Hydroxyl on Ser | 76 | ~16 |
| Hydroxyl on Thr | 49 | ~16 |
| ***Hydroxyl groups*** | ***150*** | |
| ***Imidazole on His*** | ***24*** | 6.6 ± 1.0 |

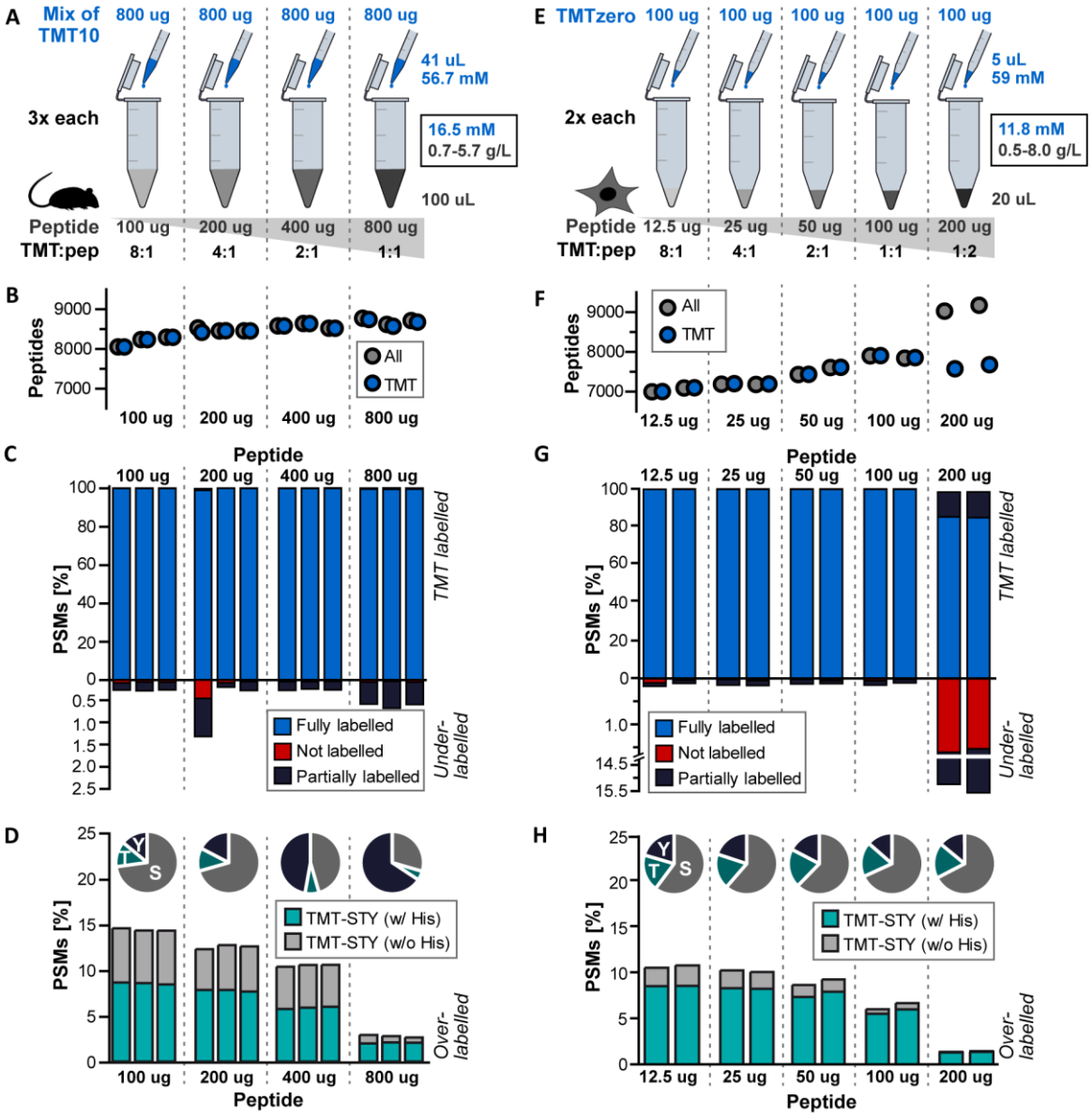



**Figure III-10 | Peptide titration experiments using the vendor recommended (A-D) and a down-scaled (E-H) TMT-labelling protocol.** *(A) Quantities and concentrations of a mix of TMT10-plex reagents (blue) and peptides (grey) are shown for increasing peptide amounts in labelling volumes recommended by the TMT vendor (pep: peptide). The TMT reaction was quenched using 50 mM Tris, pH 8. (B) The numbers of all peptides (grey circles, including fully, partially and non-labelled sequences) and only fully TMT-labelled peptides (blue circles) are displayed for the labelling experiment series illustrated in (A). (C) PSMs identifying underlabelled and fully labelled peptides are depicted for intra-laboratory replicates using the labelling protocol displayed in (A). (D) The number of PSMs assigned to overlabelled, O-acylated peptides and the distribution of serine, threonine, and tyrosine in these spectra are shown for the peptide titration row displayed in (A). (E) Same as (A) but using TMTzero and smaller peptide quantities in decreased volumes (pep: peptide). The TMT reaction was quenched using 0.4 % hydroxylamine. (F) Same as (B) but for the labelling experiment series shown in (E). (G) Same as (C) but for the labelling protocol displayed in (E). (H) Same as (D) but for the peptide titration row depicted in (E).*

To test this hypothesis, a peptide titration experiment was performed in triplicates using amounts ranging from 100 to 800 µg of murine liver peptides resulting in TMT-to-peptide ratios (wt/wt) of 8:1 up to 1:1. Across labelling reactions, the total reaction volume and thus TMT concentration

was kept constant (16.5 mM TMT during labelling), whereas the protein concentration consequently increased with increasing peptide input amounts (Figure III-10 A). The reaction was stopped by adding Tris, pH 8 to a final concentration of 50 mM.

Single shot LC-MS/MS analysis led to the identification of 8,081 to 8,807 peptide sequences with a slight increase with decreasing TMT-to-peptide ratios (Figure III-10 B). Across the entire range of tested peptide quantities, at least 98.7 % of PSMs corresponded to peptides that were fully labelled (Figure III-10 C). Consequently, few non-labelled or partially labelled peptides (where either the lysine side chain or peptide N-terminus was not labelled) were observed (less than 0.7 % of PSMs for all but one outlier sample corresponding to 200 µg peptides). Furthermore, most of the underlabelled PSMs (77-100 %) contained at least one TMT modification. The corresponding overlabelling analysis revealed that 10.4 to 14.6 % of PSMs contained at least one TMT-labelled serine, threonine, or tyrosine residue, when the labelling reaction was conducted using 100 to 400 µg peptides (Figure III-10 D). Interestingly, in the 800 µg peptide samples, the fraction of PSMs assigned to overlabelled peptides decreased to less than 3 % with only a very small concomitant increase in partially labelled PSMs. For lower peptide quantities, overlabelling primarily affected serine residues (up to 74.2 % of overlabelled PSMs), whereas tyrosine residues were overrepresented when using lower TMT-to-peptide ratios (up to 67.3 % of overlabelled PSMs, Figure III-10 D). Noteworthy, 55.5 to 78.1 % of overlabelled PSMs contained a histidine residue. To exclude that this observation was an artefact created by false TMT localization, the data were researched allowing TMT as a variable modification on histidine. Only 1.5 % of the spectra were assigned to peptides containing a TMT labelled histidine and up to 95.7 % of these contained a serine, threonine, or tyrosine residue. This indicates that false TMT localization in the overlabelling search is not a substantial issue. Intensity distributions of overlabelled peptides were comparable to correctly labelled peptides, while underlabelled peptides showed significantly lower intensities (Appendix Figure 0-6 A). Taken together, this indicates that the recommended quantity of 800 µg TMT reagent can label at least 4 to 8 times more peptides than what the vendor protocol suggests with a concomitant reduction in overlabelling of undesired amino acid residues.

**Downscaling TMT quantities using optimized labelling parameters** – Encouraged by the above findings, it was subsequently examined whether smaller peptide quantities can be efficiently labelled using less TMT reagent than recommended by the vendor (for a detailed overview of all performed experiments and results, see Suppl. Table 2 of the online publication). From chemical reaction kinetics and the law of mass action, it follows that the efficiency of the labelling reaction depends not only on the absolute quantities of tagging reagent used but, more importantly, on the molar concentrations of the reactants, i.e. TMT and peptides or, more precisely, relevant functional groups on peptides. Hence, in order to keep conditions similar to the initial peptide titration experiment, in addition to decreasing TMT and peptide quantities, also the reaction volume was reduced to maintain relatively high concentrations.

Initial experiments were performed using 100 µg TMT reagent and between 12.5 and 200 µg HeLa peptides while decreasing reaction volumes by a factor of 5.6 (Figure III-10 E). Consequently labelling took place at TMT concentrations of 11.8 mM and the reagent-to-peptide ratio varied from 8:1 to 1:2. This time, the reaction was stopped by adding hydroxylamine to a final concentration of 0.4 %. Replicate analyses demonstrated that up to 100 µg peptides were efficiently labelled resulting in 7,005 to 7,906 fully TMT labelled peptides (Figure III-10 F) and a PSM labelling efficiency of 99.8 to 99.9 % (Figure III-10 G). Notably, the lower the TMT-to-peptide ratio was, the higher were the peptide identifications obtained. For 200 µg peptides (1:2 ratio of TMT-to-pep-

tide), the proportion of PSMs corresponding to partially or non-labelled peptides sharply increased to an average of 14.9 % (Figure III-10 G), affecting ε-amines of lysine residues more than peptide N-termini (18 % of all lysine residues vs. < 4 % of all N-termini). The MS1 intensities of underlabelled peptides were again found to be always considerably lower than those of correctly labelled peptides, whereas overlabelled peptides showed comparable intensities (Appendix Figure 0-6 B). TMT-labelled serine, threonine, and tyrosine residues were present in on average 10.8 % of identified spectra for a TMT-to-peptide ratio of 8:1 (Figure III-10 H). This fraction decreased to 6.3 % and 1.5 % for a reagent-to- peptide ratio of 1:1 and 1:2, respectively. Serine accounted for about two thirds of the overlabelled amino acids for all peptide quantities used. Again, most of the overlabelled peptides contained a histidine (Figure III-10 F). In accordance with the first experiment series, a search allowing histidine to be labelled by TMT assigned, on average, 4.6 % of the PSMs to peptides with a TMT labelled histidine residue, and up to 99 % of these peptides also comprised at least one serine, threonine, or tyrosine residue.

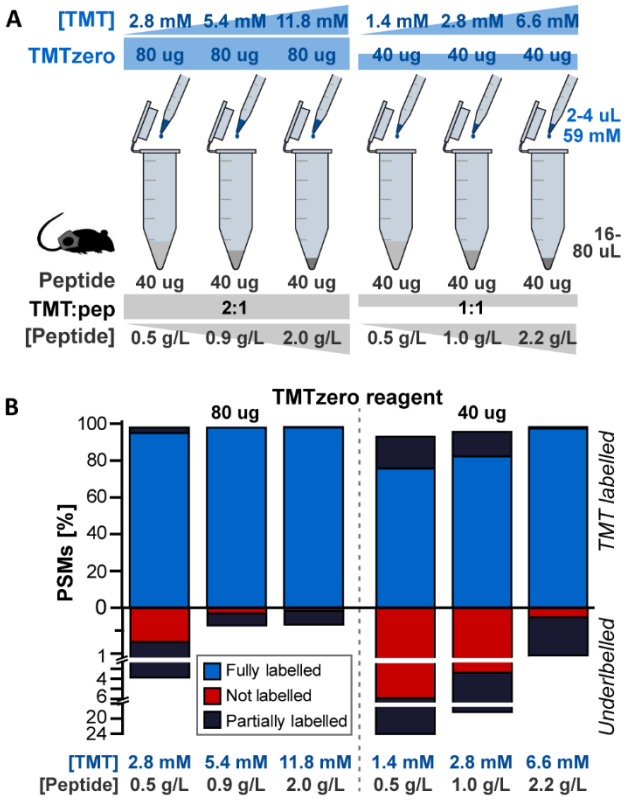



**Figure III-11 | Selection of peptide titration experiments using smaller peptide quantities.** *(A) Quantities and concentrations of TMTzero reagent (blue) and peptides (grey) are illustrated for a titration of the absolute labelling volume employing smaller peptide quantities of 40 ug and 40 or 80 ug of TMTzero reagent (pep: peptide). The TMT reaction was quenched using 0.4 % hydroxylamine. (B) PSMs identifying underlabelled and fully labelled peptides are depicted for labelling experiments displayed in (A).*

The above findings were corroborated in independent experiments using murine liver tissue. Triplicate experiments using 100 µg TMT revealed fully labelled peptides in on average 99.7 % of PSMs for 100 µg peptides and 42.7 % of PSMs for 200 µg peptides. Additional experiments using 200 µg peptides from Jurkat cell and PDX digests showed complete labelling using 200 µg TMT at the same TMT concentration of 11.8 mM (> 99.4 % of PSMs identified correctly labelled peptides). Together, it can be concluded that, for 100 µg or higher peptides quantities, peptides can be efficiently labelled at a TMT-to-peptide ratio of 1:1 and at TMT and peptide concentrations of 11.8 mM and 4 g/L, respectively.

Another series of labelling experiments was performed using a smaller peptide quantity of 40 µg PDX peptides. TMT-to-peptide ratios ranging from 8:1 to 1:1 and different TMT (1.4 to 29.5 mM) and peptide concentrations (0.5 to 2.2 g/L) demonstrated the importance of maintaining sufficient TMT and peptide concentrations during labelling (see Figure III-11 A for experiments using smaller TMT quantities). Spectra labelling efficiencies of > 99.6 % were obtained in all experiments employing a TMT-to-peptide ratio of at least 2:1 For a TMT-to-peptide ratio of 1:1 at 6.6 mM TMT and 2.2 g/L peptides, 98.2 % of PSMs identified fully labelled peptides, but this fraction dropped substantially to 82.1 % and 75.7 % with lower TMT and peptide concentrations

(Figure III-11 B). This illustrates that for peptide quantities below 100 µg, less than 100 µg TMT can be used if the TMT and peptide concentrations are adapted accordingly.

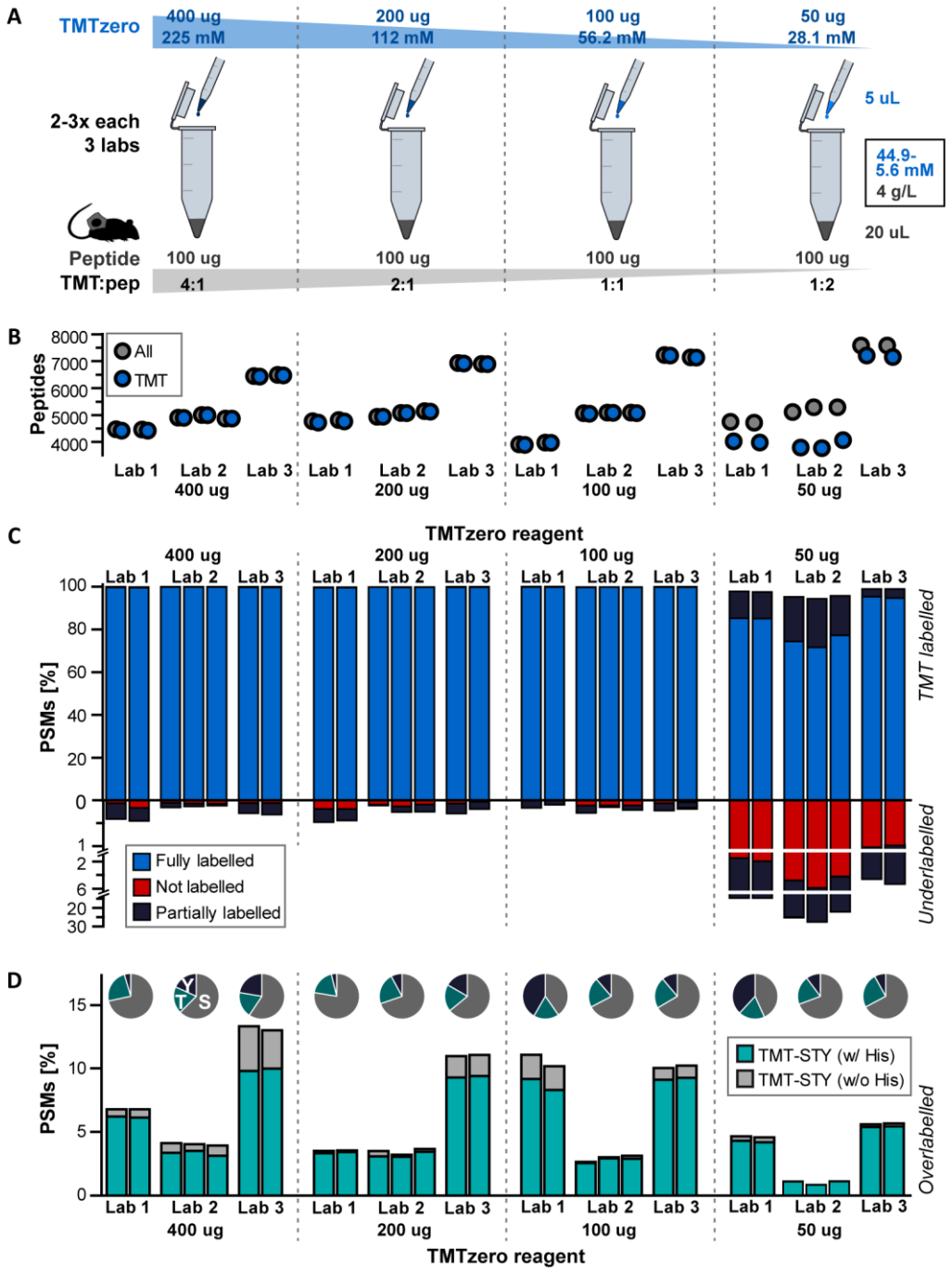



**Figure III-12 | TMT titration experiments using the downscaled TMT labelling strategy across laboratories.** *(A) Quantities and concentrations of TMTzero reagent (blue) and peptides (grey) are illustrated for increasing TMT quantities in constant labelling volumes (pep: peptide). The TMT reaction was quenched using 0.4 % hydroxylamine. (B) PSMs identifying underlabelled and fully labelled peptides are shown for intra- and inter-laboratory replicates following the protocol depicted in (A). (C) The number of PSMs assigned to overlabelled, O-acylated peptides and the distribution of serine, threonine, and tyrosine in these spectra are displayed for the workflow shown in (A).*

**Assessing inter-laboratory reproducibility of the optimized labelling protocol** – Having established that a TMT-to-peptide ratio of 1:1 (wt/wt) is sufficient to label a proteome efficiently,

it was to be demonstrated that reproducible labelling efficiencies can be achieved in different laboratories when using identical labelling workflows. To accomplish this, peptides from digests of cryopulverized patient-derived breast cancer xenograft tumours were distributed to three laboratories and replicates of 100 µg peptide aliquots were labelled with 50 to 400 µg TMTzero reagent spanning TMT-to-peptide ratios from 4:1 to 1:2 while maintaining a constant reaction volume (Figure III-12 A). This time, the TMT amount and concentration was increased to be able to assess if doing so would result in differences in over- or underlabelling compared to the previous peptide titration experiments. The labelling reaction was stopped by adding hydroxylamine to a final concentration of 0.4 %.

Despite differences in the overall numbers of identifications between laboratories due to different LC setups and LC/MS instrument performance (3,877 to 7,197 modified peptide sequences, Figure III-12 B), on average 99.7 % of PSMs consistently identified fully labelled peptides in all reactions using a TMT-to-peptide ratio of 4:1 to 1:1 (Figure III-12 C). Moreover, the percentage of underlabelling in these experiments was < 0.5 % of PSMs. However, reducing the TMT-to-peptide ratio to 1:2 led to significant underlabelling of between 4.7 and 28.4 % of PSMs depending on the laboratory (Figure III-12 C). The fraction of identified spectra assigned to overlabelled peptides also differed between laboratories and ranged from 2.6 to 13.3 % in efficiently labelled samples Figure III-12 D). This fraction dropped by a factor of 2 to 3 in experiments using only 50 µg TMT for 100 µg peptides. Again, serine was the predominantly O-acylated amino acid, though discrepancies in the fraction of TMT labelled serine and tyrosine residues were observed among overlabelled peptides in single experiments (Figure III-12 D, Figure III-10 D-H). Despite evaluation of several potential parameters that could influence overlabelling (see pp. 92), a well-founded explanation for these differences could not be established. As already observed in the peptide titration experiments, up to 98 % of overlabelled peptides also contained a histidine in the sequence (Figure III-12 D). Consistent with the prior observations, underlabelled peptides exhibited consistently lower MS1 intensities compared to correctly labelled peptides, while overlabelled peptides showed comparable to marginally higher signals (Appendix Figure 0-6 C). No apparent difference in the under- or overlabelling trend caused by higher TMT concentrations compared to higher peptide concentrations could be determined when comparing the TMT titration to the peptide titration experiments.

**Benchmarking the optimized protocol for deep (phospho)proteomes** – After it had been established in several lines of experiments and across laboratories that a TMT-to-peptide ratio of 1:1 is sufficient to achieve high labelling efficiency judged by single-shot LC-MS/MS analyses, the optimized protocol was evaluated for deep-scale (phospho)proteome studies. Here, peptides from five replicates of basal and luminal breast cancer PDX models were combined into a TMT10-plex experiment and separated into 24 whole cell proteome and 12 phosphoproteome fractions. The same workflow as described in Mertins et al. [194] was employed but the TMT labelling step was adjusted such that only 1/8th of the recommended amount of TMT reagents was used. Specifically, 300 µg of TMT reagent and 300 µg of peptides were labelled per channel in a final volume of 75 µL, and results were benchmarked against the original protocol using 2,400 µg TMT reagents to label 300 µg of peptides in a total volume of 423 µL (Figure III-13 A). These samples were generated and analysed at the Broad Institute.

Not surprisingly, the observed labelling efficiency was slightly lower for both fractionated TMT10-plex experiments (Figure III-13 B) compared to the single-shot analysis described above because the fractionation step enabled identification of more of the lower abundant and underlabelled peptides. The overall numbers of collected MS2 spectra, PSMs, distinct (phospho)peptides and

labelling efficiencies were comparable between the two labelling protocols (Figure III-13 B). Underlabelling at peptide N-termini was 2 % for the standard protocol and 3 % for the reduced TMT protocol, while underlabelling of lysines was 0.5 % and 0.6 %, respectively. Overlabelling on histidine containing peptides and fractions of overlabelled serine, threonine and tyrosine residues were also comparable between the standard (11.4 %) and the optimized protocol (11.6 % of PSMs). More than 12,000 proteins were identified in both experiments, of which > 8,400 were of human origin, and protein identifications showed a large overlap (> 90 %) between experiments (Figure III-13 C). On average, ~42,000 phosphorylation sites were detected (> 35,000 of human origin), and three quarters of these were identified in both workflows. An excellent intra-plex correlation (Pearson > 0.8) of human and murine proteins and phosphopeptides was observed across luminal and basal quintuplicates for both labelling protocols (Figure III-13 D). Similarly, proteins and phosphopeptides correlated well (Pearson > 0.7) between the two workflows. Importantly, this inter-workflow correlation was comparable to the inter-plex correlation reported previously

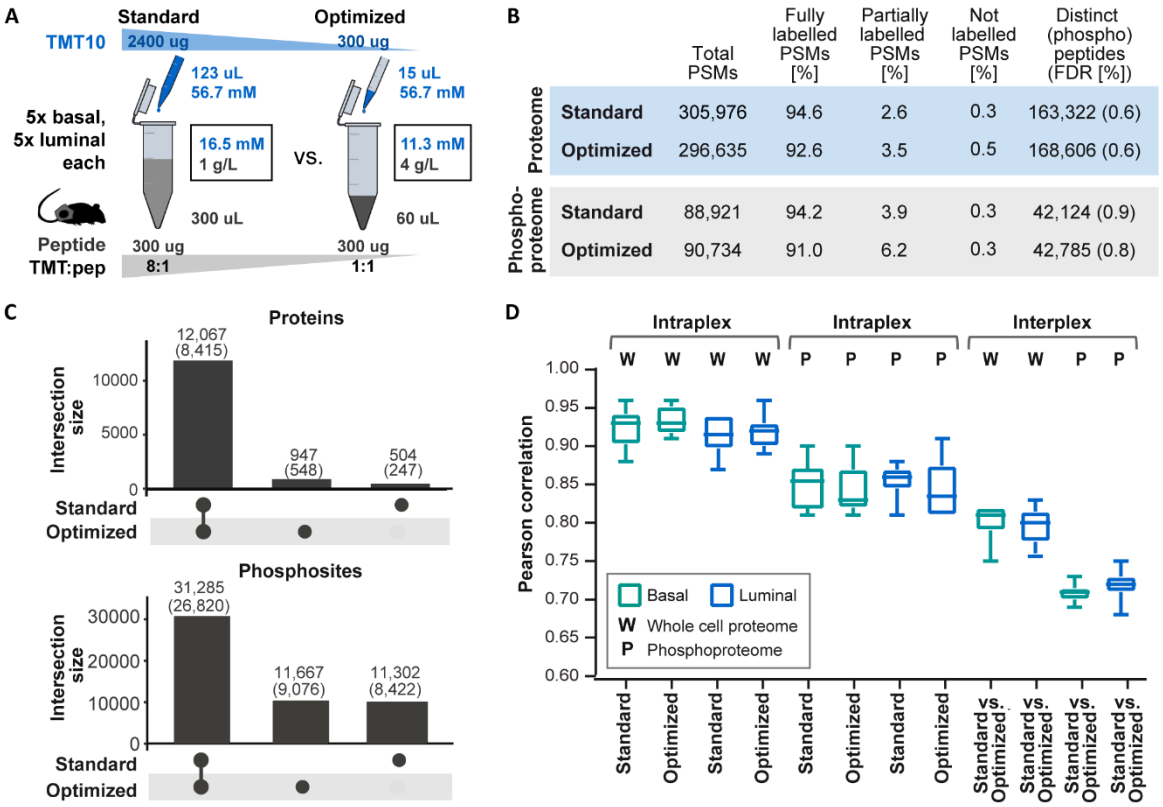



***Figure III-13 | Benchmarking the optimized protocol for deep-scale (phospho)proteomic analysis.*** (A) TMT10-plex experiments were performed using five replicates each of peptides derived from basal and luminal breast cancer PDX models and following the two different labelling protocols displayed here. Quantities and concentrations of TMT10-plex reagents (blue) and peptides (grey) used per channel are shown for the standard and the optimized labelling protocol (pep: peptide). (B) The table lists the number of total PSMs, PSMs identifying fully and partially labelled peptides, and distinct (phospho)peptides for the whole cell and phosphoproteome analyses following the labelling protocols displayed in (A). (C) Bar charts illustrate proteins (upper panel) and phosphosites (lower panel) that were identified for both or only one of the two labelling workflows depicted in (A). Proteins and phosphorylation sites mapping to the human database are reported in brackets. (D) Pearson correlation coefficients are plotted for correlations within TMT10-plex experiments (intra-plex) and between TMT10-plex experiments (inter-plex, i.e. inter-workflow) following the protocols depicted in (A).

for two identical TMT10-plex experiments using the vendor recommended amount of TMT reagent [194]. In summary, this demonstrates the utility of the optimized TMT protocol employing 1/8th of the original amount of TMT for deep-scale proteomics and phosphoproteomic studies.

# 4 Discussion and conclusion

While limitations and advantages of label-free and isobaric-tagging strategies for quantitative proteomics have been extensively reviewed (see [270, 271]), the primary goal here was to quantitatively explore their dynamic range, accuracy, and precision specifically on peptide level. Phosphoproteomic analyses of up to 11 samples were chosen as a particularly challenging example because they require not only peptide sequence identification and robust quantification, but also reproducible assignment of the modified residue(s) within the peptide sequence and across all conditions.

## 4.1 Confining the dynamic range of peptide quantification

Synthetic phosphopeptides, which were spiked into a constant human background and spanned in total almost 6 orders of magnitude, quickly demonstrated vastly different dynamic ranges for label-free and TMT-based quantification. The TMT methods exhibited a 5- (MS3) to 25-fold (MS2) reduction in linear dynamic range compared to the label-free approach. The big difference between the two TMT measurement modes is in accordance with an earlier investigation of the dynamic range of iTRAQ [243]. TMT-MS3 achieved two orders of magnitude (corresponding to iTRAQ without background spiked in), while TMT-MS2 featured a dynamic range of only one order of magnitude (as iTRAQ in presence of a complex background). For label-free data, the actual dynamic range is supposably even larger as determined in the presented experiment since the observed upper limit was not a consequence of MS signal saturation but rather due to an overloaded trap and analytical column resulting in a selective loss of hydrophilic peptides. The overloading could have been circumvented by reducing the number of different synthetic peptide sequences spiked in and, thus, the total mass loaded onto the trap column. This, however, may in turn have limited the significance of the whole analysis. Alternatively, an increased trap capacity or direct injection of the unlabelled samples onto the analytical column can provide information on how far the linear dynamic range can be extended for label-free phosphopeptide quantification and when an actual upper limit of quantification is reached.

TMT-based quantification is unlikely to show any upper limit of quantification reflecting a detector saturation with the MS parameters used, even if one would increase the amount of spike-in. Instead, the highest TMT reporter ion will determine the maximum of the linear dynamic range and the ratios to other reporter ions in the spectrum will dictate its lowest possible limit that can be detected (Appendix Table 0-2). This is due to the inherently compositional nature of TMT quantification [331] resulting from the readout in a single spectrum and the AGC which restricts the maximum number of collected charges. In general, the maximal possible dynamic range within one Orbitrap spectrum is defined by the number of charges injected into the mass analyser, the minimum number of charges that induce a signal, and the number of ion species (i.e. peptides or fragments) in the spectrum. For Orbitrap analyser the intra-scan dynamic range has been reported to span 4 to 5 orders of magnitude when only two analytes were present [332, 333]. In the case of TMTs (and assuming that the AGC target is always reached), the number of charges is distributed among the 6 to 16 reporter ions for any given peptide. If the abundance in one sample increases and, consequently, the fraction of ions for the corresponding TMT reporter scales up, then the intensity fraction of other reporters must proportionately decrease [331]. Taking the applied AGC target value into account (1e5) and assuming that a minimum of 10 charges is necessary to induce a detectable signal in the Orbitrap [332], the conducted 5-step dilution experiment could

at best detect a dynamic range of ~3,100. Increasing or decreasing the dilution steps would accordingly change the maximum detectable dynamic range (Appendix Table 0-2). This reveals a general drawback of such an analysis for isobaric tagging approaches: results depend on the experimental design, and the actual dynamic range in a biological experiment may differ from peptide to peptide. Nevertheless, the comparison of the two TMT acquisition approaches is valid and reveals the setback of MS2 relative to MS3-based quantification. The dynamic range of TMT-MS2 suffers not only from a higher number of ion species (i.e. peptide fragments) in the quantitative spectrum, but it is also reduced by peptide co-isolation and ratio compression.

The above exemplified connection further implies that increased multiplexing capacity of TMTs (6-, 10-, 11-, and 16-plex) naturally comes at some expense for the dynamic range, and that higher AGC values can improve the dynamic range. However, the potential of the latter is restricted by the maximum number of ions that do not lead to space charging in the Orbitrap (or the C-trap for that matter) which will cause high abundant ion species to coalesce [334, 335]. Instead of two distinct, nearby peaks, ion coalescence results in a single peak with the average m/z value of the two original species. It is more likely to occur, the more charges are collected, the closer and more abundant the two ion species are, and the higher their m/z [336]. Although rather unexpected due to the small mass of TMT reporter ions, it has been reported that AGCs larger than 2e5 can cause coalescence of proximate TMT10-plex reporter ions on a QExactive mass spectrometer [337] which precludes accurate quantification.

In theory, MS1-based quantification is also compositional since similar AGC constraints are applied, but the total intensity in a spectrum is usually distributed among many more ion species (i.e. peptides). This confines the influence of an increase of a single ion species on the other species. Consequently, the effects of a restricted number of allowed charges for MS1-based quantification within the same spectrum (e.g. SILAC) is in most cases insignificant [331]. Importantly, the relative quantification for label-free experiments is not performed within the same spectrum. Thereby, the possible dynamic range is enhanced via the automatic adjustment of injection times for different scans resulting in an improved inter-scan versus intra-scan dynamic range [333]. Label-free quantification also differs globally from TMT quantification in ion injection times for the quantitative scan, which are usually in the low ms range due to the much higher ion current in full scan mode. This can more quickly result in an apparent detector saturation for vastly increased injection amounts. Saturation is determined by the highest number of charges that the electrometer responsible for AGC determination can accurately measure and the smallest possible ion injection time that can be accomplished which together dictate the ability to inject the correct number of charges defined by the AGC. In addition, the accuracy of calculations of actual intensities from the AGC target value will drop with extremely small ion injection times, and injection of too many charges can again lead to ion coalescence hampering accurate quantification [335].

Concisely, label-free and TMT-based quantification exhibit large differences in dynamic range. While part of the smaller dynamic range of isobaric tagging strategies can be explained by ratio compression, they are also simply more affected by inherent constrains of instrument architecture and MS measurement principles than the MS1-based, label-free approach. This implies that even with a complete removal of ratio compression, TMT-based quantification would still exhibit a more confined dynamic range than (inter-scan) label-free methods.

## 4.2 Negotiating quantification approaches for (modified) peptides

Inseparable connected to the dynamic range, is quantification accuracy. In this regard, the label-free method was clearly superior to both TMT-based approaches. Nevertheless, accuracy of TMT quantification significantly improved to an acceptable level when the synthetic peptide spike-in and the species mix sample were measured in MS3 mode. This is in line with studies reporting diminished ratio compression owing to the additional selection step during MS3 measurement [222, 246]. Besides accuracy, precision is another parameter to consider, especially for statistical analysis of quantitative differences. As reported by others [250, 325, 327, 328], precision of TMT-based quantification was improved compared to the label-free quantification strategy. Commonly, the lower precision of label-free data is explained by the independent sample processing during which inhomogeneous biases can potentially be introduced [267, 271], whereas samples after TMT labelling are processed together and affected uniformly by sample losses preserving precision. However, these biases would equally influence accuracy of label-free quantification that is generally considered excellent for cautiously processed samples. This makes it very likely that the difference in precision is largely coupled to the difference in data acquisition and post-acquisition processing. This appears plausible especially because other MS1-based quantification methods like SILAC and dimethyl show a similar precision to label-free data [250, 327] even though samples are combined much earlier in the workflow. Hence, the superior precision of isobaric tag approaches is mainly connected to the improved signal-to-noise ratios in MS2 versus MS1 spectra and the less complicated extraction of intensity information from MS2 spectra [271]. MS1-based quantification approaches require sophisticated algorithms for detection, extraction, tracking, and integration of peptides features along retention time [272] and the performance of the algorithms is highly dependent on the number of MS1 spectra acquired across the peptide elution profile which overall limits precision of quantification.

Noteworthy, it has recently been shown that, due to the enhanced precision, ratio distortion and a subsequently declined accuracy are less critical for calling significant differences between conditions when replicates are included within the same TMT experiment [250, 328, 338]. When TMT-labelled and label-free data comprising known abundance differences were compared, both approaches overall demonstrated an excellent performance with only minor differences in sensitivity and false positive rates [250, 326, 328]. Meanwhile, it often depended on the extent of the true ratio between conditions which of both strategies showed slightly superior ability to flag known differences correctly. Importantly though, TMT-based quantification outperformed label-free quantification reliably with regard to the total number of repeatedly quantified proteins and peptides across replicates. This in turn resulted in an overall higher number of proteins/peptides that were available for statistical testing and subsequently identified as significantly different. This illustrates the importance of another parameter for evaluation of quantification approaches, which is the occurrence of missing values across conditions.

The unmatched dynamic range and accuracy of label-free quantification was countervailed by a much higher number of missing values in both the spike-in and the 17-AAG treatment experiment. Indeed, as already observed by others [250, 325, 328], the label-free method quantified similar total numbers of peptides as the MS2-based TMT approach, and even much higher numbers as corresponding MS3 data when comparable measurement times were employed. However, about two thirds of the peptides showed at least one zero intensity in any of the ten samples. This was the case although match-between-runs [255, 281] was enabled which allowed for the inter-run transfer of PSMs using accurate retention time and m/z information followed by quantification of

corresponding peptide elution profiles [255, 281]. Ultimately, 2.5 to 3 times more peptides were reproducibly quantified in all conditions with MS3-based quantification of TMT-labelled as compared to label-free samples, which is in line with recent reports [250, 328]. Notably, the fraction of missing values among peptides is higher than commonly observed for proteins (~40-50 %) [328, 339] which emphasizes the aggravation of this issue on peptide level. Noteworthy, the advantage of less sparse quantification matrices on peptide level for TMT-based approaches only comes into effect properly as long as the number of samples does not exceed the available number of isobaric plexes. More samples can still be compared using isobaric tags when several plexes are analysed including a common reference sample (see e.g. [274, 275, 299, 340]). Then, however, similar challenges with regard to missing values will have to be tackled during data analysis [341] and these cannot be overcome by elution profile matching across runs.

For the label-free quantification method, match-between-runs has proven to be powerful in improving the number of peptides without missing values by more than 80 % (similar to what was reported in [250]). Still, the label-free quantification matrix was vastly incomplete illustrating the limitation of such algorithms especially for very complex samples. It was also observed that matching of very low abundant peptides in the spike-in dilution experiment led to inaccurate intensities essentially setting the lower limit of the dynamic range. It should be noted, though, that there was no difference in quantification accuracy between PSMs identified by an actual MS2 spectrum and matched intensities for higher abundant peptides corroborating earlier observations [328]. Consequently, the match-between-runs algorithm offers a considerable improvement for label-free quantification in reducing missing values. One could even argue that the less accurate intensities of extremely low abundant species are nothing else than automatically imputed values for missing quantifications.

It is commonly accepted that missing values in MS1-based quantification approaches occur more often for low abundant peptide species since they are less reliably picked in every run [250, 328, 342]. The spike-in experiment (without the biased, highest spike-in sample) was in particular suited to examine the relationship between intensities and missing values in more detail. In this setup, a constant amount of HeLa peptides was present in every run and, thus, every missing value must be connected to an MS2 spectrum of an equally abundant peptide in another run. While it was true that peptides with no or only one missing value had, on average, significantly higher intensities, all other peptides showed very similar intensity distributions and their intensities spanned 4 orders of magnitude irrespective of the number of missing values. This supports earlier findings [343] and suggests that the relationship between intensities and missing values is not strongly linear. Hence, a higher number of missing values has little predictive value with regard to a supposedly low intensity of a peptide. Missing values among high abundant precursors may be explained by overlapping isotope patterns that complicate the reliable assignment of charges states and subsequent selection for fragmentation by the instrument software. A recent improvement of the algorithm that determines peptide precursor charge states from isotope clusters (APD: advanced peak determination) indeed showed improved sampling of precursors in highly populated m/z regions and also for peptides in the high abundance range [344]. Consequently, treating missing values in label-free data as indicator for low abundance can be misleading and they must rather be understood as a true missing value where no quantitative information is available. Conversely, zero intensities in TMT-based quantification approaches are inherently more informative since they reflect a quantitative value that is truly below the limit of quantification in the recorded scan. The poor correlation of intensity and missing values in label-free data puts the broad application of missing value imputation assuming low abundance of missing quantification

under question. It has even been reported recently that such imputation of label-free data did not improve the detection of known regulated proteins but even increased the number of false negatives instead [328]. Clearly, imputation of proteomics demands more sophisticated algorithms and a careful evaluation of the influence of any imputation strategy on the label-free datasets at hand [343].

Specifically for phosphoproteomics, additional advantages of TMT over label-free quantification became apparent. First and unexpectedly, IMAC enrichments from TMT-labelled samples showed consistently better selectivity towards phosphorylated peptides. It was hypothesized that this might be connected to a decreased contamination with non-phosphorylated peptides possessing multiple acidic residues which are known to be frequently co-enriched [345]. However, this could not be confirmed by an examination of non-phosphorylated peptides in label-free and TMT data since they showed virtually the same distribution of acidic residues in both datasets. Hence, the reason for this discrepancy remains elusive. Second, challenging site localizations (e.g. at neighbouring sites) led to ambiguous and varying allocation of the modification in different label-free runs. Hence, the number of total phosphopeptide species and, at the same time, missing values was artificially increased. In contrast, TMT quantification within the same spectrum assured that the identical peptide is compared, even if localization probability was low. Third, TMT-labelled phosphopeptides exhibited an improved localization, particular for N-terminal phosphorylation as determined in the synthetic peptide spike-in experiment. It can be hypothesized that this is mediated by the commonly observed increase in the complimentary b-ion series for TMT-labelled peptides [346, 347], which presumably increases the likelihood of observing a site determining ion. The fact that the gain in b-ions is generally less profound in the higher mass range [347] may explain why this boost in localization certainty is not observed for peptides bearing the phosphorylation at the C-terminus. Interestingly, TMT-MS3 provided the highest fraction of correctly localized sites. This may be due to the use of CID-MSA fragmentation for corresponding MS2 spectra, which has been shown to improve correct phosphosite localization [238, 249]. In addition, this could be a result of the comparably slow MS3 method leading to preferential sampling of more abundant peptides that, in turn, give rise to more informative spectra. The connection of TMT labelling and phosphosite localization needs to be further investigated including larger sets of label-free and TMT-labelled peptides with varying positions of the phosphorylation site and using different fragmentation modes. Current efforts to systematically acquire such data are ongoing [231] and will provide insight on whether and how TMT-labelling globally improves site localization in the future.

## 4.3 Managing constraints of TMT-based (phospho)peptide quantification

The capability to repeatedly quantify (phospho)peptides across many conditions without introducing missing values that are difficult to impute or even spurious was considered the major advantage of TMT above label-free quantification. This made the isobaric approach especially attractive for peptidoform analysis. Hence, it was evaluated in more detail how TMT acquisition can be optimized and substantial drawbacks such as ratio distortion and decreased identification rates can be managed most efficiently.

It became clear that improved accuracy as effectuated by measurements in MS3 mode comes at the expense of proteome coverage. Empirically, MS3 methods lead to a 30-40 % decrease in identifications on peptide and modification site level and 15-20 % on protein level [250, 348]. The compromised proteome coverage is a result of the additional fragmentation and scan recording steps that decelerate the acquisition. This becomes increasingly relevant the shorter the applied

gradients are, and likely explains the even more profound reduction in peptide identifications in more complex, whole proteome single shots of the species mix sample compared to the less complex phosphoproteome measurements (45 % loss for 60 min vs. 22 % for 90 min). As mentioned above, MS3 measurements at the same time significantly reduced ratio compression by minimizing co-isolation as visible from the considerably increased number of true zero intensities in outermost channels of the yeast-human mix sample. Notably, the species mix sample was designed to emulate a worst-case scenario for quantification by mixing yeast peptides into human background in very low quantities and measuring single shots without any fractionation. This promoted co-isolation and a substantial contribution of human background to yeast intensities resulting in ratio distortion. In the phosphoproteome, the fraction of yeast peptide was even lower (about 4 times less as deduced from MS1 peptide intensities) which explains the even more profound ratio distortion of phosphopeptides in MS2 mode measurements. Interestingly, this difference between unmodified and phosphorylated peptides disappeared upon MS3 acquisition suggesting that such kind of advanced measurements are equally powerful for higher and lower abundant, and more and less compressed peptides.

Next, it was assessed how differences in accuracy influence statistical test results specifically for the two TMT measurement modes. As the conducted statistical tests indicated, ratio distortion in MS2 mode impeded the identification of true regulations when they were small (below 4-fold). The argument has been made that larger regulations are biologically more interesting and that thus the lower sensitivity on smaller ratios is justifiable [250]. On the other hand, it may as well be argued that rewiring of protein expression or induction of PTMs (which is limited by their initial stoichiometry) may often only lead to smaller, but potentially widespread changes that may be of functional importance. In such cases, MS3 measurements are beneficial, as they have been proven to considerably increase the sensitivity for detecting smaller changes for very low abundant yeast phosphopeptides in the presented setup. In this context, it is important to note that the sensitivity of detecting such changes is again dependent on the experimental design that defines the ratios across yeast groups. Due to the compositional nature of TMT quantification (see p. 85), a group with very high ratios compared to the other two groups (here 1:8 and 2:8) will limit the detectability of the smaller ratio (here 1:2). Hence, it is not surprising that O'Connell *et al.* who mixed yeast peptides in ratios of 1:2:3 (instead of 1:2:8) into human background achieved a much higher sensitivity for detecting significant regulations in groups with the expected fold change of 2 (96 % versus 78 % in data presented here). Nevertheless, basic differences of MS2 and MS3-based acquisition can be deduced irrespective of the exact experimental design. Surprisingly, comparisons of the two different quantification approaches in the literature have so far been only based on the degree of recovery of known regulations and the absolute number of detected, significant changes [250, 326] meanwhile neglecting false positive rates that can potentially result in misleading conclusion. In the here conducted analyses, false positives were increased upon MS2-based quantification, this time affecting groups with expectedly larger ratios. Ultimately, the decision on which measurement mode to use depends on the consideration of how much quality (accuracy, sensitivity, specificity) can be sacrificed to improve quantity (proteome coverage) and which advanced measurement modes are available (see also pp. 170).

Filtering of acquired spectra based on precursor intensity fraction (PIF; fraction of intensity in the isolation window that is derived from the targeted precursor) can be an alternative approach for improvement of quantification accuracy when advanced instruments for MS3 mode measurements are not available. A PIF cut-off can remove a fraction of potentially highly distorted peptides. However, TMT intensities in some of the remaining spectra may still be compressed because

the PIF has limited predictive value for the actually existent ratio compression (based on personal experience and observations of others [250, 326, 327]). This is likely caused by an underestimation of co-isolation from raw file spectra that are used for PIF estimation owing to a specific signal cut-off that is already applied by the MS instrument software [245, 267]. In addition, different fragmentation efficiencies and reporter ion yields for co-isolated compared to targeted peptides can result in a significant contribution to reporter intensities even of very low abundant background peptides [326, 327]. Attempts have been made to computationally correct for co-isolation based on PIF values [245, 326], but this only works under the assumption that the co-isolated peptide(s) are of constant abundance in all conditions. While these corrections may be expedient on protein level where many quantitative data are aggregated, they are likely too inaccurate for individual PSMs [245] due to the low correlation of PIF and ratio compression and thus are not applicable on the level of individual peptides.

Besides quantification accuracy, identification rates were generally lower for TMT-labelled as compared to label-free peptides. This has been observed by others already [325, 349, 350], and various reasons and solutions have been proposed for this phenomenon. First, although the labelling reaction with NHS-ester reactive reagents has a high selectivity towards primary amines, overlabelling on hydroxyl and imidazole groups has been reported to occur [351-356] and was also detected in TMT titration experiments using adequate search parameters. The use of standard search parameters for TMT-labelled samples, however, would prohibit identifying those overlabelled peptides and thus decrease identification rates. Interestingly, also pre-processing of spectra by removing ions related to labelling reagents has been proven to have benefits for identification scores and thus identification rates [349, 357] suggesting that isobaric tag-labelled peptides are inherently more difficult to identify by current search algorithms. Furthermore, upon TMT-labelling, a primary amine is replaced by a tertiary amine which increases the gas phase basicity [358] and leads to a slight increase in the average charge state [327, 350, 359]. In addition, it was hypothesized that isobaric tags change the mobility of protons along the peptide backbone [349]. Higher charge states and altered proton mobility may lead to disparate or more complex fragmentation patterns and those may not sufficiently be accounted for by standard search algorithms which are optimized for doubly charged peptides [360]. In line with these assumptions, a charge reduction through application of ammonia vapour during ionization has been shown to improve identification rates for isobaric tag-labelled peptides [350]. Together this indicates that identification rates and thus (phospho)proteome coverage could generally be improved by adjusting search algorithms specifically to isobaric tag-labelled peptides and optimizing fragmentation modes to reduce adverse effects of altered charge states and proton mobility.

An MS3 method optimization series suggested that MSA may be such a useful fragmentation technique for the identification of TMT-labelled, phosphorylated peptides. This is not surprising since the additional activation of the neutral loss, which is even more profound for TMT-labelled than for label-free phosphopeptides due to the changed proton mobility [359], has already been reported to improve scores and localization certainty for phosphopeptides [238]. However, it is a rather slow fragmentation process. Interestingly, HCD showed an even slightly better performance with regard to absolute number of identifications despite lower identification rates. It reduces the formation of neutral losses similar to MSA and additionally preserves the lower mass range yielding richer MS/MS spectra compared to CID [234]. However, this increase may mostly be explained by the fast beam-type activation allowing for the acquisition of many more spectra than the slower fragmentation methods. In line with Hogrebe *et al.* [250] who recently also published a comparison of different quantification approaches for phosphorylated peptides, OT

readout performed better for phosphopeptide identification irrespective of the fragmentation mode used, presumably based on a better mass accuracy which enhances scores. Consequently, MSA-OT appeared most advantageous for identification of TMT-labeled phosphopeptides. With regard to quantification, significantly more ratio compression was observed when the neutral loss generated after CID fragmentation was allowed for the SPS and subsequent quantification. This is in conflict with other reports [248-250], which stated improved TMT intensities with no compromise on quantification accuracy for neutral loss methods. Based on their results, Erickson *et al.* have even argued that ratio distortion is not caused by co-isolated peptides of identical charge and mass but instead by low-level background fragmentation [248]. The most profound difference of here presented data to these studies was the abundance of peptides used for the evaluation: While here the aim was to determine ratio distortion for very low abundant peptides, the other laboratories assessed peptides that were of much higher abundance (in some cases even higher than the constant background). It can be reasoned that the increased ratio compression observed in the presented data resulted from co-isolated, non-changing phosphopeptides of the identical charge state as the targeted precursor which produced a neutral loss that was again of similar m/z as the targeted neutral loss and thus contributed to distorted intensities. The discrepancy with other studies suggests that this mechanism is, on average, not as relevant for higher abundant peptide species, although it remains unclear from their results whether the fraction of highly compressed peptides may have been considerably increased in these studies.

Other parameters that influenced quantitative performance were isolation windows and the number of SPS precursors. The beneficial influence of smaller isolation windows on ratio compression has been under debate [276], but early analyses may have been biased by the limited capabilities of early quadrupoles to reproducibly narrow isolations of precursors. This should be improved on newer instrument generations. A more recent report shows a clear reduction of ratio distortion with smaller windows [338], which was corroborated by the presented data even for MS3-based quantifications albeit to a lesser extent. Moreover, the previously observed negative correlation of accuracy and precision mediated by higher TMT intensities was also detected [338]. The same relation was evident when the number of fragment precursors for SPS was scaled up. More notches increase the probability of including a fragment from a co-isolated peptide rendering a very small number of SPS notches advantageous for average accuracy. However, few notches, at the same time, increased the fraction of strongly distorted peptides likely resulting from cases were only fragments from co-isolated but not the targeted peptide were selected. As a compromise between robust and accurate TMT quantification, an MS2 isolation window of 0.7 m/z and 10 SPS notches were chosen for MS3 methods.

## 4.4 Characterizing critical parameters for TMT labelling

Several protocols in which the amount of TMT reagent recommended by the vendor for peptide labelling was reduced have recently been published [298, 356, 361-363] and such economically optimized labelling workflows have successfully been applied to address a variety of biological questions [364-366]. However, details on the quantities and concentrations of reactants vary widely in the published literature and, to the best of the knowledge of the author, no systematic evaluation of the influence of reducing TMT-to-peptide ratios on the overall labelling performance has been reported to date. The conducted series of TMT-labelling experiments using different reagent and peptide concentrations, quantities, and ratios allowed for a systematic assessment of the influence of these parameters on the labelling reaction. Smaller reaction volumes and, consequently, higher TMT and peptide concentrations were advantageous for labelling efficiency as the

law of mass action would demand. Further, the smaller the ratio of TMT to peptides, the more crucial was the concentration of reagent and peptide. This can be readily explained by the competing reactions of labelling of primary amines and hydrolysis of the TMT reagent in aqueous conditions, which contributes to less efficient labelling in less concentrated protein and TMT solutions. Therefore, for reagent-to-peptide ratios of 1:1 (wt/wt), it is advisable to employ TMT and peptide concentrations of 10 mM (3.4 µg/µL) and 2 g/L, respectively, to ensure efficient labelling. Importantly, peptide concentrations should be determined directly before TMT labelling (as done in the optimization experiments) because, from experience, 30 to 50 % of the initial protein quantity can be lost during digestion and subsequent desalting procedures and these losses may vary between sample types and laboratories.

Moreover, it needs to be stressed that careful handling of the TMT reagent is inevitable (as described in the manufacturer protocol) when working with low TMT-to-peptide ratios to avoid loss of active reagent as a result of hydrolysis caused by absorbed water from ambient air. This is of particular relevance when TMT leftovers need to be stored. In the author's experience, unused TMT reagent can readily be kept in anhydrous ACN at -20 °C or -80 °C for at least 3 months without any drop in labelling efficiencies. For long-term storage, TMT should be aliquoted in an inert atmosphere and stored dried down and under exclusion of water. This can easily be realized by performing the aliquoting procedure in a bin filled with argon, and aliquots can be kept under argon or with a desiccant. By this means, aliquoted TMT reagents have been stored for up to a year without any decline in labelling performance.

Although the optimized protocol, in principle, can be adapted to peptide quantities in the low microgram range by appropriately decreasing reaction volumes, handling very small volumes, particularly TMT reagent in 100 % ACN, is not very practical and can lead to inaccuracies. Consequently, it is advisable to increase the relative reaction volume for peptide quantities below 50 µg and compensate for the lower TMT and peptide concentrations (e.g. 5 mM TMT and 1 g/L peptide) by concomitantly increasing the TMT-to-peptide ratio (e.g. to 2:1). For peptide quantities below 10 µg, even higher reagent-to-peptide ratios are likely required [367]. Alternatively, it is conceivable that higher ACN concentrations may have a positive effect on labelling efficiency (due to lower reagent hydrolysis) particularly for less concentrated samples or low absolute sample quantities and would facilitate the use of the desired TMT-to-peptide ratio of 1:1. Although labelling experiments at different ACN concentrations were included during the optimization process, these always also involved variations of other parameters such as TMT or peptide concentration. Hence, a systematic assessment of the influence of ACN on the labelling reaction would require further experiments preferentially using small peptide quantities.

Small changes in the pH of the reaction buffer can also affect labelling efficiency and overlabelling. Typically, more alkaline pH values promote the inactivation of NHS-esters due to hydrolysis [351]. This is particularly relevant when the excess of the labelling reagent is limited. For example, a TMT-to-peptide ratio of 1:1 roughly corresponds to a 2.5 x molar excess of TMT reagent over the estimated molarity of primary amines in a perfectly digested human proteome. Therefore, the pH (and purity) of the peptide solution must be controlled properly to ensure a reproducible outcome. At the same time, pH values lower than the pKa values of the primary amines of lysine and peptide N-termini result in a higher degree of protonation at equilibrium, which hinders the reaction with TMT. Since labelling was performed at pH 8.5, this effect is illustrated by the higher fraction of non-labelled lysine residues (pKa ~10.5) compared to N-termini (pKa ~7.7) in all samples that show significant underlabelling in the single-shot analyses. However, also an opposite trend of N-termini being preferentially underlabelled at pH 8.5 using various TMT, peptide, and ACN

concentrations was observed, particularly in samples that show near complete labelling. This is also in line with reports from literature [194, 367]. This may be explained by the fact that the higher pKa value of the ε-amine also typically (depending on the solvent) corresponds to a higher nucleophilicity in the deprotonated state and consequently a higher reactivity towards TMT. Thereby, the labelling of lysine residues may be kinetically favoured over peptide N-termini under certain conditions. However, which reaction conditions determine preferential underlabelling of either the ε- or α-amine remains elusive and needs further investigation. Notably, it has been shown for pH values of up to 8.5 that the increase in reactivity of both primary amines exceeds the accelerated hydrolysis rate of NHS esters [354, 368] providing the basis for conducting the TMT labelling reaction at pH 8.5.

Considering that the pKa values of the side chains of lysine and tyrosine are similar [329], one would expect that the reactivity of tyrosine would also increase at elevated pH, rendering it more prone to react with TMT at more basic pH values. However, several studies investigating labelling of amino acids and peptides using NHS esters reported that the abundance of acylated tyrosine residues is enhanced only at a more acidic pH, whereas more alkaline pH values favour N-acylation [353, 354, 368]. This may be explained by the lower stability of tyrosine acylation in basic conditions that can be harnessed to reverse overlabelling by adding hydroxylamine and thus increasing the pH above 9 to quench the labelling reaction [352, 355, 368, 369]. The reversal of overlabelling by hydroxylamine may also account for the overall lower fraction of O-acylation at higher TMT-to-peptide ratios observed in the titration experiments using 12.5 to 200 µg peptides compared to the ones using 100 to 800 µg peptides. In the latter series of experiments, Tris buffer at pH 8 instead of hydroxylamine was used to quench the labelling reaction.

The high prevalence of TMT-labelled serine and threonine residues seems surprising considering their very high pKa values [330] which must result in a high degree of protonation of their hydroxyl groups at pH 8.5 and, therefore, a low susceptibility to react with TMT. Indeed, an early study investigating the reactivity of NHS esters towards amino acids could not detect serine and threonine derivatives at pH 7.4 [351]. In contrast, others have found that hydroxyl-containing amino acids in peptides are reactive towards NHS esters when histidine is in close proximity, notably in -2 or +2 position, to the labelled amino acid (H-X-[STY], [STY]-X-H) [352, 354, 355, 370]. This implies that pKa values can change drastically depending on the molecular environment of amino acids. Indeed, overlabelled peptides were also strongly enriched in histidine in presented data, and serine, threonine, or tyrosine residues were 3 to 11 times more likely to be identified in a TMT labelled state when they were part of the H-X-[STY] motif. The pKa of histidine is lower than that of N-termini [329] which would, in principle, promote histidine modification by TMT. In fact, this has been suggested to occur in solid-phase labelling protocols under slightly acidic conditions [356]. However, for the applied in-solution labelling protocol (performed at basic pH), the results of database searches allowing TMT as a variable modification on histidine provided no plausible evidence that histidine itself is prevalently labelled. This is in accordance with studies reporting a transient, very labile modification of histidine under neutral to alkaline conditions with spontaneous hydrolysis of the formed N-acylimidazole that has a half-life in the range of minutes [351, 353]. This may also explain the preferential overlabelling of histidine containing peptides via an increase in the local concentration of TMT by an initial derivatization of histidine and a subsequent reaction of the N-acylimidazole intermediate with proximal hydroxyl-containing amino acids [354]. Besides, histidine could also lead to an increase in the nucleophilicity of hydroxyl containing amino acids via hydrogen bonding between the side chains resulting in a higher reactivity towards TMT [370]. Of note, the enrichment of histidine in O-acylated peptide sequences likely also accounts

for the, on average, higher intensities observed for overlabelled peptides compared to correctly and underlabelled peptides. Irrespective of absolute abundance, histidine-containing peptides exhibit generally higher intensities than non-histidine containing peptides due to improved ionization mediated by its gas phase basicity.

Interestingly, a weak but consistent increase in spectra identification rates with decreasing TMT-to-peptide ratios could be noticed in all searches of titration experiments specifying TMT as variable modification on lysine and peptide N-termini. As already suggested by Böhm *et al.* [356], this increase in identification rates may be ascribed to a reduction in the fraction of overlabelled peptides which was simultaneously noticed at decreased TMT-to-peptide ratios. A similar observation was made by Miller *et al.* who detected a higher fraction of modified tyrosine residues with higher reagent-to-peptide ratios [352]. This may be a result of different rates of TMT hydrolysis and the reaction with primary amines versus hydroxyl groups. O-acylation has been found to proceed up to 20 times slower than N-acylation [352, 353] at least when no histidine was present in close proximity. Therefore, employing relatively low reagent quantities that can be fully consumed by reacting with all primary amines as well as accounting for some reagent hydrolysis would suppress O-acylation and thus reduce overlabelling. Although not investigated here, shortening the reaction time might further minimize overlabelling.

## 4.5 Conclusion

A range of different sample and experimental designs was employed to provide the basis for a well-justified decision on which approach is most suitable for reproducible and accurate quantification of individual peptides across several conditions. While label-free quantification expectedly showed the best accuracy and dynamic range, missing values substantially limited the ability to track the quantitative behaviour of individual peptides across different samples. Hence, they will ultimately restrict the inference of information about peptidoforms across conditions in a biological setup and may complicate or bias statistical approaches that often benefit from (or are only applicable in the case of) a complete quantitative matrix across samples. In contrast, TMT-based quantification featured an unmatched precision and a, for the most part, complete quantitative matrix within a single TMT-plex thereby improving overall proteome coverage. Together, this readily enabled the identification of peptide level changes across conditions even when acquiring TMT reporter ions in the much less accurate MS2 mode as long as replicates are available. In absence of replicates, however, MS3-based TMT quantification represented a compromise with improved dynamic range and accuracy but reduced proteome coverage compared to MS2 data. If time is available, this can be overcome by deeper sample fractionation and higher measurement costs. This, in turn, would also further reduce ratio distortion [244] which even in MS3 mode could not be overcome completely. In general, TMT approaches have the additional advantage of requiring less input material per condition. This can prove beneficial especially when sample amounts are limited or larger quantities are needed for enrichment of PTMs. The disadvantage of higher costs for sample preparation due to expensive isobaric labelling reagents was overcome by adjusting the labelling protocol. The optimized in-solution labelling procedure is cost-effective without any sacrifice in labelling efficiency or robustness and reduces the amount of required TMT reagent by 8-fold relative to the vendor's protocol. Thus, it represents a further improvement of previously published labelling workflows [298, 356, 361, 362]. In conclusion, the MS3-based quantification approach was considered most expedient for analyses of proteoforms with a bottom-up approach enabling robust, quantitative measurements of single peptides across multiple conditions.

*„The discussion is somewhat too long*
*but has the merit of dissecting carefully all critical points […].”*

**- Reviewer #3**

# IV TOWARDS PEPTIDOFORM-RESOLVED TURNOVER ESTIMATES

**1 Introduction and summary ..... 101**

**2 Experimental designs ..... 103**

**3 Results ..... 105**

3.1 Design of a pulsed SILAC-TMT experiment ..... 105

3.2 Benchmarking the pulsed SILAC-TMT approach ..... 107

3.3 Evaluation of determinants of protein turnover ..... 109

3.4 Turnover diversity of peptides ..... 114

**4 Discussion and conclusion ..... 118**

4.1 Technical and data analysis challenges for (pSILAC-TMT) turnover studies ..... 118

4.2 Biological implications of measured protein turnover ..... 119

4.3 Conclusion ..... 123

The following chapter is largely based on the publication "*Peptide level turnover measurements enable the study of proteoform dynamics*" [371] published in Molecular & Cellular Proteomics on May 1, 2018.

**Author contributions for** [371]

J.Z. developed the methodology, conceptualized the research, performed experiments, analysed the data, and wrote the manuscript; J.Z., set up the PRM assay with help from D.P.Z., and M.W.; J.Z. and D.P.Z. acquired MS data; C.M. developed the tool for curve fitting; P.S. performed structure prediction of proteins and was responsible for integration of protein turnover data into ProteomicsDB; J.Z., B.K., M.W., and D.P.Z. revised the manuscript.

**Supplementary material for** [371]

Additional supplementary tables are freely available for download at the publisher's website (https://www.mcponline.org/content/17/5/974/tab-figures-data). The MS proteomics raw data and complete MaxQuant search results have been deposited to the ProteomeXchange Consortium (http://www.proteomexchange.org/) via the PRIDE [324] partner repository with the data set identifier PXD012703.

# 1 Introduction and summary

Cellular proteins exist in a dynamic state in which they are continuously destructed and re-constructed [97], and the coordination of protein synthesis and degradation for controlling and adjusting protein abundance is a fundamental process in cellular homeostasis. In order to expand the understanding of the concerted action of these two processes, protein turnover has been studied for decades (reviewed in [372]). Traditional approaches have made use of a multitude of methodologies notably pulse-chase radiolabelling [157, 373], inhibition of protein synthesis [152], or tagging of endogenous proteins with fluorescent dyes [153, 154]. However, radioactive labelling only allows for the analysis of bulk protein turnover or of the stability of single proteins. Treatment of cells with translation inhibitors, on the other hand, disrupts cell homeostasis and half-lives determined in this way might not fully reflect the actual endogenous degradation process. Similarly, measuring the stability of fluorescently tagged and overexpressed proteins might not resemble physiological protein half-lives.

In recent years, advances in MS based technologies in conjunction with SILAC [167] have dramatically improved protein turnover measurements. Nowadays, dynamic SILAC experiments enable the parallel measurement of turnover characteristics of thousands of endogenous proteins expressed at physiological levels [113, 168-170, 305, 374, 375]. Yet, despite ongoing efforts, technical issues still exist and important cellular mechanisms affecting protein stability at the molecular level still often remain elusive. For example, comparison of different studies often show limited correlation of protein turnover rates and sometimes arrive at contrasting conclusions about which protein properties might affect half-lives [125, 153, 163, 170, 305, 375]. In addition, different cellular stabilities have been reported for the same protein depending on its localization [170], the cellular condition [375], protein interactions [376], or its post-translational modification state [163]. Moreover, differences in turnover rates have also been detected for splice variants of the same gene [162]. Together with the fact that more than 200 different types of protein modifications have been described [14] and that nearly all multi-exon genes have been shown to be alternatively spliced [29, 30], this demands studying turnover on the level of individual proteoforms, but such analyses have largely been neglected in the past.

The investigation of proteoform dynamics with bottom-up proteomics approaches is not trivial. First, a robust and accurate method for quantification of pulsed SILAC labelled peptides across multiple time points is needed because often only single peptide sequences distinguish non-canonical from canonical isoforms or modified from non-modified proteins. However, the standard dynamic SILAC approach suffers from substantial missing quantitative values across pulse time-points measured as separate samples. This issue is amplified when increasing the number of measured pulse time-points and matters a lot when analysing data at the peptide level. While protein level quantification can make use of averaging several peptide measurements thereby increasing the robustness of turnover estimation, every missing value may severely lower accuracy at the peptide level. As indicated by initial results of the comparison of MS1-based (label-free) and TMT-based quantification (see chapter III), multiplexing and subsequent quantification of all time-points in the same spectrum may overcome this issue. The general feasibility of combining pulsed SILAC with iTRAQ 4-plex labelling of peptides derived from different pulse time-points has previously been demonstrated in *Streptomyces coelicolor* [176] albeit with low proteome coverage. Subsequently, TMT labelling of pulsed SILAC samples has been proposed in a review by Hughes and Krijgsveld [377] and recently demonstrated by Welle *et al*. [302] who determined turnover

dynamics for 1,276 human proteins in a single MS3 based experiment. Yet, none of the above studies have specifically addressed turnover at the level of proteoforms.

Here, the merits of combining dynamic SILAC and TMT-labelling of ten pulse time-points for peptide-resolved analysis of protein turnover were evaluated. For this purpose, a robust normalization method for multiplexed turnover data was established. Moreover, a new approach to compute absolute protein copy numbers per cell from TMT data was developed. Pulsed SILAC-TMT and standard dynamic SILAC data showed high concordance and SILAC-TMT hyperplexing, additionally, enabled high proteome coverage (6,000 proteins) within reasonable time (two days) of LC-MS measurements. Systematic evaluation of replicates showed that robust single peptide level turnover measurements can be achieved if experiments are conducted carefully. Facilitated by the deep proteome coverage, several examples of post-transcriptional and post-translational processing leading to differential protein stabilities were identified. For example, N-terminally processed peptides exhibited both faster and slower turnover behaviour compared to other peptides of the same protein. In addition, the suspected proteolytic processing of the fusion protein FAU was substantiated by measuring vastly different stabilities of the cleavage products. Furthermore, differential peptide turnover suggested a previously unknown mechanism of activity regulation by post-translational destabilization of Cathepsin D as well as the DNA helicase BLM. Finally, the comprehensive dataset enabled a detailed re-evaluation of molecular determinants of proteome stability in steady-state cells and uncovered that oxidative stress may contribute to the high turnover of proteins in the respiratory chain complex I. To enable further research on the topic, protein stability data have been made available in ProteomicsDB [378].

## 2 Experimental designs

The following sections provide an overview of the experimental procedures and methods that are specific for the evaluation of the pulsed SILAC-TMT approach. Details on individual experimental steps, data processing and analyses are specified in chapter II (pp. 48, and pp. 52). The rationale of the experimental design is described in more detail in the results section (pp. 105).

For an initial assessment of the pulsed SILAC-TMT method (Figure IV-1A), K0R0 cells were switch to K8R10 medium and lysed directly before medium exchange (0 h time-point) and after 1, 3, 6, 10, 16, 24, 32, 40, and 50 h. After digestion, peptides were labelled using TMT, combined, separated into six bRP fractions, and analysed employing an MS2 or MS3 method for TMT quantification. In the main experiment (Figure IV-1B), four HeLa cell culture replicates and an MS measurement duplicate were performed to assess the reproducibility of the pulsed SILAC-TMT approach and enable statistical evaluation of differences in peptide turnover rates. Two cell batches were switched from K0R0 to K8R10 and two replicates were switched from K8R10 to K0R0 and lysed 0, 1, 3, 6, 10, 16, 24, 34 and 48 h after medium exchange. This time, an "infinite" time-point was included for which cells were grown in K8R10 (or K0R0) medium for ≥ 10 cell doublings and checked for >99.9 % label incorporation. These were seeded in K8R10 (or K0R0) medium concurrently to K0R0 (or K8R10) labelled cells and lysed at the same time as the cells of the 48 h time-point. In addition, for one replicate, fractional SILAC labelling of six time-points (1, 3, 6 ,10, 24, 48 h) was directly analysed on MS1 level omitting the TMT labelling step to enable a comparison of the two different quantification approaches (Figure IV-1B). To achieve similar measurement time, peptides from the six time-points were separated into four fractions, while TMT-labelled samples were separated into 24 fractions.

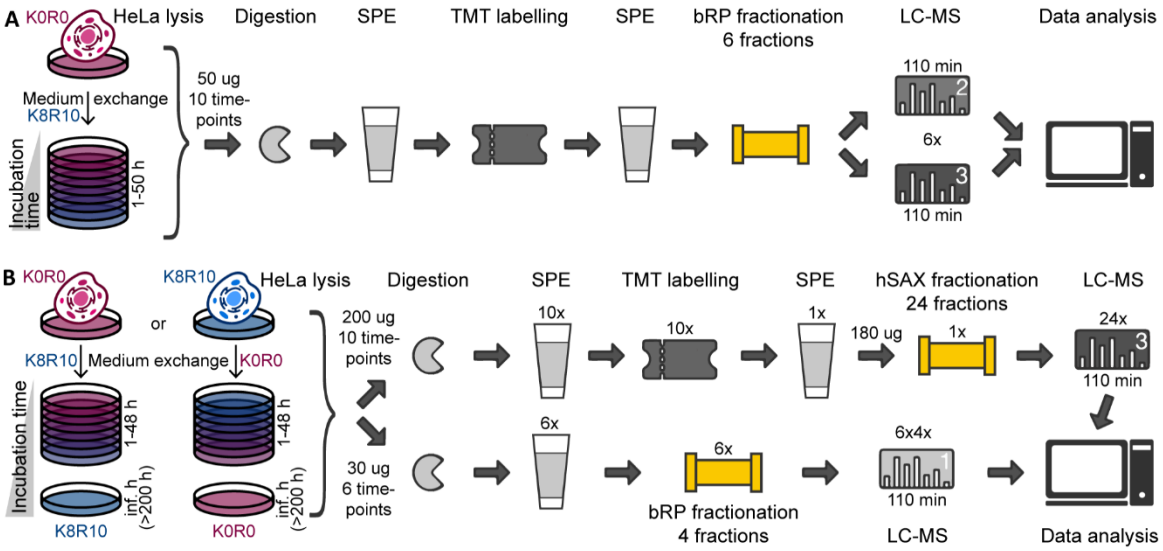



**Figure IV-1 | Experimental designs for the evaluation of the pulsed SILAC-TMT approach.** (*A) For the initial comparison of MS2 and MS3-based TMT quantification, peptides from ten different pulse-time points were labelled with TMT, fractionated using bRP STAGE tips, and measured in MS2 and MS3 mode. (B) For the main experiment (upper workflow), four pulsed SILAC replicates including label swaps and ten time-points were TMT-labelled, and pulsed SILAC peptides were subjected to hSAX fractionation. In parallel, 6 time-points of one replicate were processed omitting the TMT-labelling step and performing bRP fractionation for each time-point (lower workflow).*

All samples were measured on a Fusion Lumos Tribrid mass spectrometer injecting an amount corresponding to 1.2-1.5 µg protein digest, and peptides were separated using a 100 min linear

gradient from 4 to 32 % solvent B. For MS analysis of SILAC samples without TMT label, MS2 spectra were recorded in the Orbitrap at 15K resolution after HCD fragmentation (28 % NCE) using an isolation window of 1.6 m/z, an AGC target value of 1e5, a maxIT of 50 ms, and a fixed first mass of 100 m/z. In the MS2-based TMT method, the isolation window was set to 1.2 m/z, the AGC target value to 1.2e5, the maxIT to 100 ms, the NCE to 33 %, and the fixed first mass to 120 m/z. For both methods, cycle time and dynamic exclusion were set to 2 and 60 s, respectively. In the MS3-based TMT method, MS2 spectra for peptide identification were recorded in the ion trap in rapid scan mode upon fragmentation via CID (NCE of 35 %, activation Q of 0.25) and using an AGC target value of 2e4 and a maxIT of 100 ms (isolation window 0.7 m/z, dynamic exclusion of 90 s). MS3 spectra for TMT quantification were obtained in the Orbitrap at 60K resolution (scan range 100-1,000 m/z, charge dependent isolation window from 1.3 (2+) to 0.7 (5 - 6+) m/z, AGC of 1.2e5 charges, maxIT of 110 ms) following synchronous selection of the 10 most intense peptide fragments in the ion trap and fragmentation via HCD using a NCE of 55 %. Cycle time was set to 2 s.

Raw data were searched against the human Swiss-Prot database using MaxQuant v1.5.5.1. For pulsed SILAC samples without TMT label, Lys0/Arg0 and Lys8/Arg10 were specified as metabolic labels, whereas for pulsed SILAC-TMT samples, TMT10 was specified as label within a reporter ion MS3 experiment type and K8 and R10 were set as additional variable modifications. Andromeda score and delta score cut-offs for modified peptides were disabled. To obtain labelling rate constants, TMT data were normalized based on steady-state assumptions and equations following first-order kinetics were fitted to peptide data. After filtering for high quality curve fits (K: 0-5; B: 0-0.3; A: 0.67-1.5; $R^2 \geq 0.8$), data were condensed at the peptide and protein level and analysed regarding determinants of protein turnover and peptidoform-specific turnover behaviour. To determine protein and peptide half-lives (T1/2), cell doubling rates were estimated by fitting an exponential growth equation to cell numbers counted every 12 h and subtracted from labelling rates to obtain protein degradation rates (k). Half-life was then calculated as ln(2)/k (see pp. 52 for details).

For the investigation of the turnover of respiratory chain complex I proteins upon induction of oxidative stress (Figure IV-2), HeLa cells were treated with the complex I inhibitor rotenone and the complex I specific substrates glutamate and malate. After a 3 or 8 h pulse with heavy medium, ratios of newly synthesized to formerly existing proteins (heavy-to-light ratios) were evaluated in a PRM assay (pp. 54) and compared to control cells that were treated either with glutamate and malate or with DMSO. All cell culture conditions were evaluated as triplicates to enable statistical analysis of turnover differences.

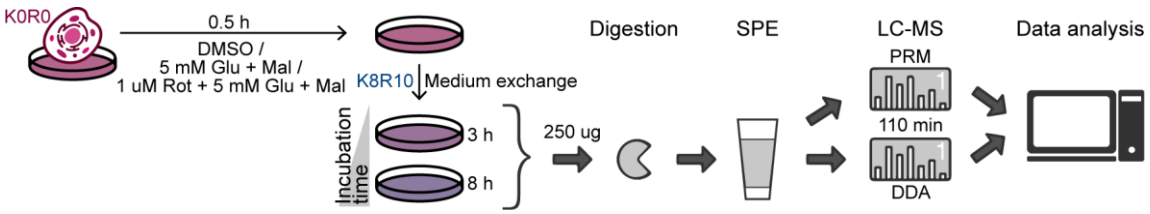



**Figure IV-2 | Workflow for the assessment of respiratory chain complex I turnover upon rotenone-induced oxidative stress.** *K0R0 cells were treated with complex I inhibitor rotenone and the complex I substrates glutamate and malate before medium was switched to K8R10. H/L ratios were analysed in a PRM assay, meanwhile using quantitative information of DDA runs for normalization.*

# 3 Results

## 3.1 Design of a pulsed SILAC-TMT experiment

The aim was to design an experimental workflow that facilitates the systematic determination of protein turnover measured as SILAC label incorporation or loss on a proteome-wide scale ideally providing proteoform resolution. This demanded (i) a deep proteome profiling method to ensure good protein and peptide coverage, (ii) an adequate number of pulse time-points enabling high accuracy of turnover estimation, (iii) a robust quantification method for single peptides across all time-points, and (iv) high reproducibility of (peptide) rate estimations. To meet these requirements, an approach combining TMT labelling of ten different pulsed SILAC time-points together with peptide fractionation by hSAX chromatography was employed (Figure IV-3). For calculation of meaningful peptide ratios for curve fitting, a maximum value corresponding to the total abundance of a particular peptide for both, increasing (synthesis) and decreasing (degradation) label, was required. This was achieved by allocating the first TMT channel to cells lysed directly before the pulse start (0 h) and reserving the last TMT channel for cells that were already completely labelled with the SILAC amino acids that were provided during the pulse (inf. h, in practice > 200 h, Figure IV-3). This experimental design also allowed for estimation of ratio compression in the outermost TMT channels resulting from co-isolation of oppositely labelled peptides, which can distort turnover rate estimations. In addition, this enabled the calculation of protein abundances as copies per cell from TMT intensities by utilizing the fraction of the MS1 intensity that was linked to the first or last TMT channel of peptides showing degradation or synthesis, respectively (see pp. 48). Intermittent time-points were chosen based on three premises: (i) Obtaining a high temporal resolution for early time-points to facilitate accurate quantification based on the assumed first order kinetics; (ii) Ensuring exponential growth of HeLa cells during the entire pulse period to maintain the steady-state assumption which required avoiding growth inhibition due to high cell densities at later time-points; (iii) Achieving a comparable MS1 intensity of SILAC pairs to increase the probability of fragmentation (given the DDA mode for MS analysis), and thus quantification of fractional SILAC labelling of both, the K0R0 and the K8R10 labelled peptides. Based on these considerations, HeLa cells were lysed 1, 3, 6, 10, 16, 24, 34, and 48 h after medium exchange (Figure IV-3). It should be noted that optimal time-points may differ for different cell lines depending on the respective cell doubling rates.

In order to correct for mixing errors of TMT-labelled digests, data were normalized based on the prerequisite that the sum of K0R0 and the K8R10 peptide intensities should be constant across different time-points (i.e. TMT channels), as equal protein amounts were employed for each time-point (Appendix Figure 0-7), for detailed information see pp. 48). Subsequent to normalization, curves for estimation of turnover rates, determined from the kinetics of SILAC label incorporation or loss, were fitted to quantitative data of all peptide evidence assuming exponential protein degradation (Figure IV-3). The normalization procedure improved the overall quality of the curve fitting as indicated by an overall shift of the $R^2$ distribution to higher values (Appendix Figure 0-7) In addition, the number of successful curve fits after filtering also increased (in total 210,704 before and 238,489 after normalization in all 4 cell culture replicates). A number of criteria were systematically evaluated in order to remove poor quality peptide curve fits, for example, due to low TMT intensities or high ratio compression. A graphical user interface is implemented in the R package “proturn” (https://github.com/mengchen18/proturn) for curve fitting visualization and assessment of filter criteria.

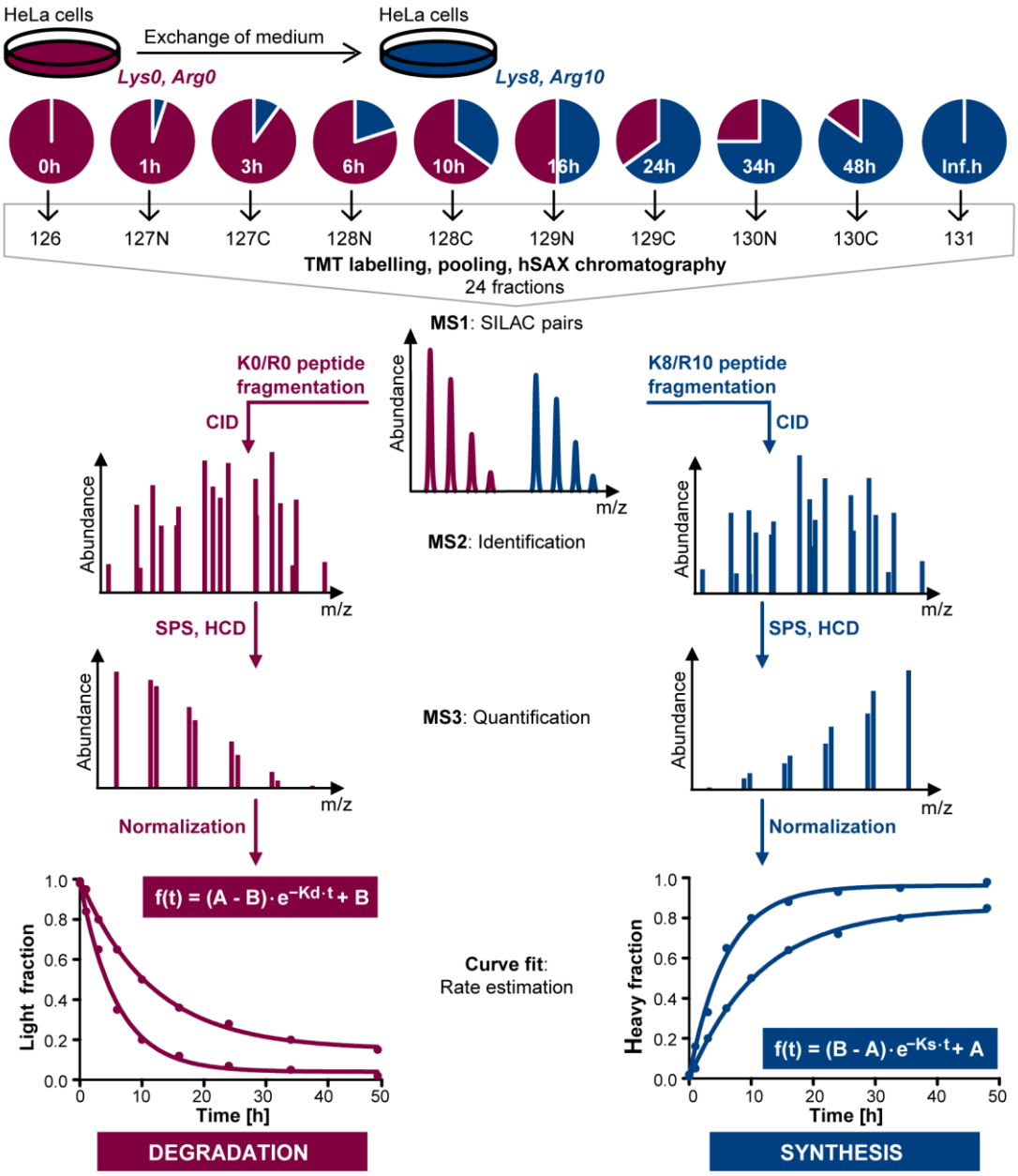



**Figure IV-3 | Schematic representation of the multiplexed pulsed SILAC-TMT strategy for estimation of protein synthesis and degradation employed in this study.** *Cells grown in K0R0 containing medium were pulsed labelled with medium supplemented with K8R10 and lysed after 10 different time-points (inf. h corresponds to ≥ 10 cell doublings). After digestion, peptides derived from different time-points were labelled with TMT, pooled, and fractionated using hSAX chromatography. Peptides were identified using MS2 spectra and quantified via MS3 scans. Decreasing and increasing labels represent protein degradation and synthesis. Assuming exponential protein degradation, one-phase decay and association functions were applied for estimation of the rates of K0R0 label decrease and K8R10 label increase (A: curve maximum; B: curve offset; K: turnover rate; see Chapter II, General Methods for a detailed explanation of the curve fitting).*

For calculation of half-life times (T1/2), labelling rates indicative of protein or peptide turnover were corrected for cell doubling to obtain rates of protein synthesis and degradation. To account for minor differences in growth behaviour and thus to improve accuracy of half-life time calculations, cell doubling rates were determined for each cell culture replicate separately using the identical cell batch and applying the same conditions (e.g. cell medium exchange) as for the corresponding, simultaneously conducted pulse experiment. Cell doubling times for the four cell culture replicates ranged from 26.3 h to 30.9 hours (Appendix Figure 0-8).

## 3.2 Benchmarking the pulsed SILAC-TMT approach

After developing an experimental design for a pulsed SILAC-TMT study, MS2- and MS3-based quantification were compared and the acquired data were benchmarked against the commonly used, non-multiplexed pulsed SILAC approach. Further, the reproducibility of turnover rate estimations on protein and peptide level was assessed.

**Comparison to other SILAC based protein turnover measurement approaches** – It is well known that isobaric tagging circumvents missing values occurring in data dependent MS1-based quantification approaches. However, quantification based on fragment ions of isobaric tags can suffer from ratio compression resulting from co-isolation of peptides featuring differing quantitative behaviour. Hence, it was of particular interest to assess and minimize such ratio compression as much as possible to avoid an adulteration of labelling rate estimations. Our experimental design enabled estimation of ratio distortion in either the first or the last TMT channel. Indeed, a median residual intensity of 28.0 % in the 0 h time-point of synthesis curves indicated that severe ratio compression was present for SILAC-TMT samples that were measured employing an MS2 readout (Appendix Figure 0-9 A). In contrast, the MS3-based quantification method using the same sample and fractionation reduced the median residual intensity to 1.8 %. To further minimize ratio distortion and concomitantly increase proteome coverage, a more extensive fractionation scheme was employed in following experiments (Figure IV-1). In this final setup, residual intensities in the outermost TMT channels were still detectable, but data filtering based on curve fitting parameters resulted in less than 10 % ratio compression for more than 80 % of all utilized peptide evidence (Figure IV-4 A).

In order to address if our TMT-multiplexed, pulsed SILAC approach using MS3-based quantification provides results similar to the standard dynamic SILAC workflow, aliquots of identical lysates from pulsed HeLa cells were processed in either way. Resulting samples were analysed expending equal amounts of LC-MS measurement time, using the same function for curve fitting, and applying identical filtering criteria after curve fitting. Hence, TMT-labelled and pooled samples derived from 10 pulse time-points (measured in 24 fractions, Figure IV-3) were compared to 6 non-tagged pulsed SILAC samples (1, 3, 6, 10, 24, 48 h, each measured in 4 fractions). MS3-based ratios of labels across time-points and derived turnover rates were in good agreement with those calculated from the classical MS1-based pulsed SILAC method as indicated by an overall correlation of R=0.70 (Figure IV-4 B) and exemplified by the virtually identical labelling curves of the protein STAT3 (Figure IV-4 C). Importantly, rates determined by either of both approaches also correlated as well with already published HeLa protein rates as these literature data correlated among different laboratories (R=0.51-0.54, Figure 0-9 B). However, for 41 % of filtered synthesis and degradation curve fits in the MS1-based approach, intensities were not detectable for all 6 time-points. A head-to-head comparison of MS1 and MS3 measurements disclosed that these missing intensity values across SILAC isotope pairs or time-points can considerably decrease accuracy of rate estimations as exemplified by the protein STAT6 (Figure IV-4 D) resulting in half-lives of 31.3 and 23.6 hours for the pSILAC and pSILAC-TMT approach, respectively. More generally, it became apparent that the precision of MS3-based quantification of reporter ions devoid of missing values enhanced the overall goodness of the curve fits in the multiplexing strategy to a median $R^2$ of 0.98 compared to 0.94 in the MS1-based approach (Figure IV-4 E). As a result, turnover dynamics could be determined for 83 % of proteins identified with the multiplexing strategy, whereas only 58 % of protein identifications in the MS1-based quantification approach passed filter criteria after curve fitting. Also facilitated by a deeper fractionation, the pulsed SILAC-TMT strategy yielded 6,035

proteins with quantified turnover compared to 3,600 proteins in the classical pulsed SILAC approach using the same amount of measurement time yet covering ten instead of six pulse timepoints (Figure IV-4 F).

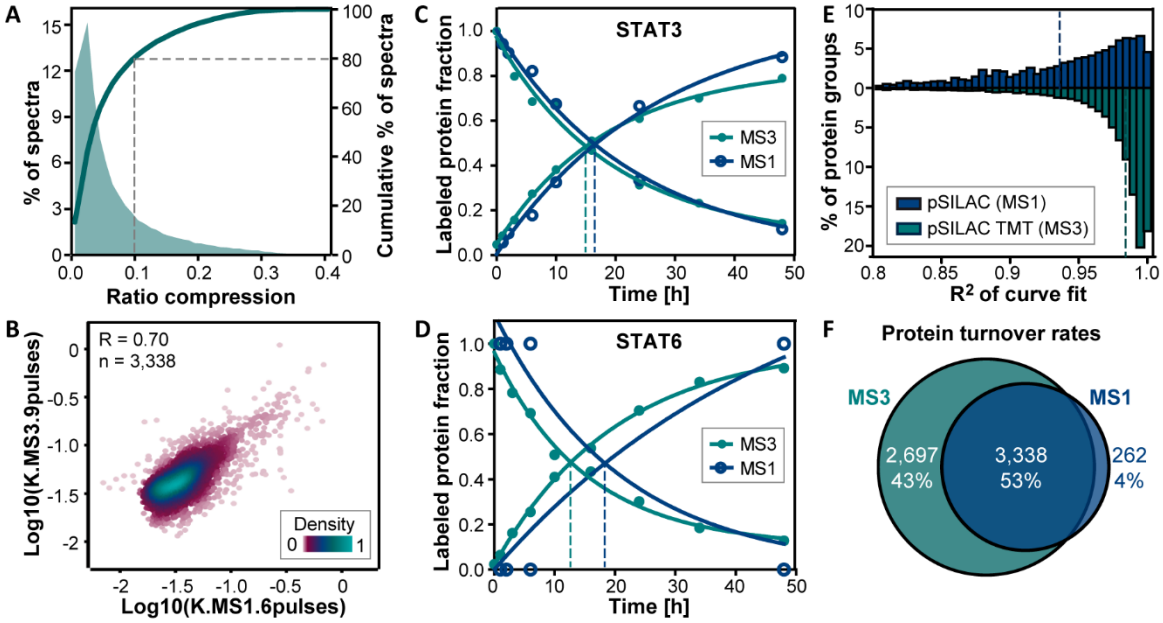



**Figure IV-4 | Comparison of MS1 (pulsed SILAC) and MS3 (pulsed SILAC-TMT) based quantification.** *(A) The fraction of MS3 spectra as a function of the detected ratio compression (from measuring the residual intensities in outermost TMT channels) illustrates that >80 % of all fitted and filtered spectra showed less than 10 % residual intensities. (B) Correlation analysis of log transformed labelling rates shows good agreement between the MS1 and MS3-based quantification approaches. (C) Labelling characteristics measured for the protein STAT3 either using the MS1 or MS3 strategy yielded consistent data. (D) Fractional labelling determined for the protein STAT6 in which MS1 data points were missing (one SILAC isotope pair signal missing for 1, 3, 6, and 48 h time points and no data for the 10 and 24 h time points) led to substantial differences in curve fits between MS1 and MS3 data. (E) Distributions of coefficients of determination ($R^2$) of curve fits display consistently higher values for the MS3 compared to the MS1 approach (dotted lines: medians). (F) Comparison of the number of proteins with determined turnover parameters shows a higher number for the MS3 strategy.*

**Reproducibility of estimations of protein and peptide turnover rates** – Next, the reproducibility of the pulsed SILAC-TMT approach was assessed. To do so, four pulsed SILAC experiments using different HeLa cell batches were performed, two of which were subjected to a SILAC label swap. In addition, fractions of one replicate were measured twice providing a technical MS replicate. After data processing and filtering, turnover rates were computed for on average 5,957 protein groups per cell culture replicate (Figure IV-5 A). For 71-76 % of all proteins, information on both protein label increase and decrease was available providing an internal duplicate measurement of protein turnover rates for each sample in a steady-state system. In all four cell culture replicates combined, synthesis and/or degradation curves were obtained for 55,067 protein group unique peptides (59,586 peptides when also counting oxidized forms) assigned to 7,203 proteins (Figure IV-5 B) with a median sequence coverage of 17.4 %. In total, turnover rates were computed for more than 86 % of all identified proteins groups and for 6,083 proteins, rates from both label increase and decrease were available. Labelling rate pairs showed a median Pearson's correlation coefficient of 0.64 and a median CV of 15 % (Figure IV-5 C). When rates of label decrease and increase were treated separately, the technical MS duplicate and the cell culture quadruplicates exhibited a median R of 0.80 and 0.77 and a median CV of 8 % and 18 %, respectively (Figure IV-5 C), demonstrating good precision of the pulsed SILAC-TMT approach. Interestingly, turnover

rate determinations on peptide level were as reproducible as for proteins (median R of 0.84 and 0.72 and median CV of 8 % and 18 %, Appendix Figure 0-10 A). Notably, for technical replicates more than 82 % of both estimated protein and peptides rates showed a CV of less than 20 %. However, when correlating rates obtained from label increase and decrease within a sample, the level of concordance dropped in particular on peptide level (median R of 0.51 and CV of 0.16, Appendix Figure 0-10 A). This observation was likely attributable to residual ratio distortion that still affected some peptides. This adulterated turnover rate determination in opposite ways for label incorporation and loss, eventually more strongly deteriorating correlation analysis on peptide than on protein level. Likewise, a weak correlation (R of 0.36) of peptide labelling rates with CVs computed from increasing and decreasing curve pairs was identified suggesting that ratio compression somewhat more strongly affects rate determinations of high turnover peptides (Appendix Figure 0-10 B) as one might expect. However, it is important to point out that there was generally no correlation of turnover rates and replicate CVs (exemplified for peptide rates across technical replicates in Appendix Figure 0-10 C, R=0.04). This encouragingly implies the absence of an overall precision bias depending on the turnover rate, meaning that rate determination is reliable across the measured range of fast and slow turnover peptides or proteins (for examples see Figure IV-5 D, Appendix Figure 0-10 D).

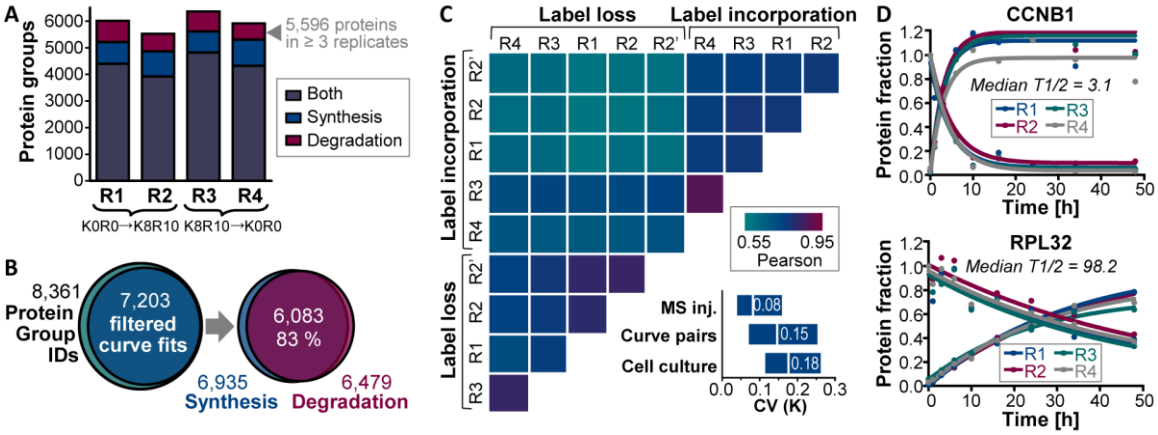



**Figure IV-5 | Reproducibility of protein turnover rate determination using pulsed SILAC-TMT.** *(A) Bar chart illustrates that turnover rates were determined for 5,528 to 6,367 protein groups per cell culture replicate (R1-R4). (B) In total, rates were obtained for 7,203 proteins, and for 83 % of this turnover information was available from both label increase and decrease. (C) Correlation matrix depicts color-coded Pearson's correlation coefficients for log transformed protein turnover rates determined from synthesis and degradation curves for cell culture (R1-R4) and MS injection (R2 and R2') replicates. The boxplots (25$^{th}$-50th-75$^{th}$ percentile) show the coefficients of variation of label incorporation and label loss rates across MS injections, curve pairs within a sample and cell culture replicates. (D) Examples of the reproducibility of turnover determination across cell culture replicates are displayed for the high turnover protein G2/mitotic-specific cyclin-B1 (CCNB1) and the stable 60S ribosomal protein L32 (RPL32).*

## 3.3 Evaluation of determinants of protein turnover

Estimated protein turnover rates spanned three orders of magnitude resulting in calculated half-lives ranging from minutes (exemplified by Serine/threonine-protein kinase SIK1) to thousands of hours (for Fatty acid desaturase 2 (FADS2), Figure IV-6 A). The median half-life of all proteins was 37.8 h. Apart from an expected slight underrepresentation of the membrane and the extracellular subproteomes, our set of proteins proved to be functionally representative for the entire human proteome (Appendix Figure 0-11 A). In addition, protein copy numbers per cell determined from the SILAC-TMT data were in good agreement with published data [306, 307] (R=0.72 and 0.85,

Appendix Figure 0-11 B) and covered several orders of magnitude substantiating the highly representative character of the data at hand (Appendix Figure 0-11 C). Hence, the current compilation of protein turnover data provided a good opportunity to investigate the influence of protein properties and functions on protein stability.

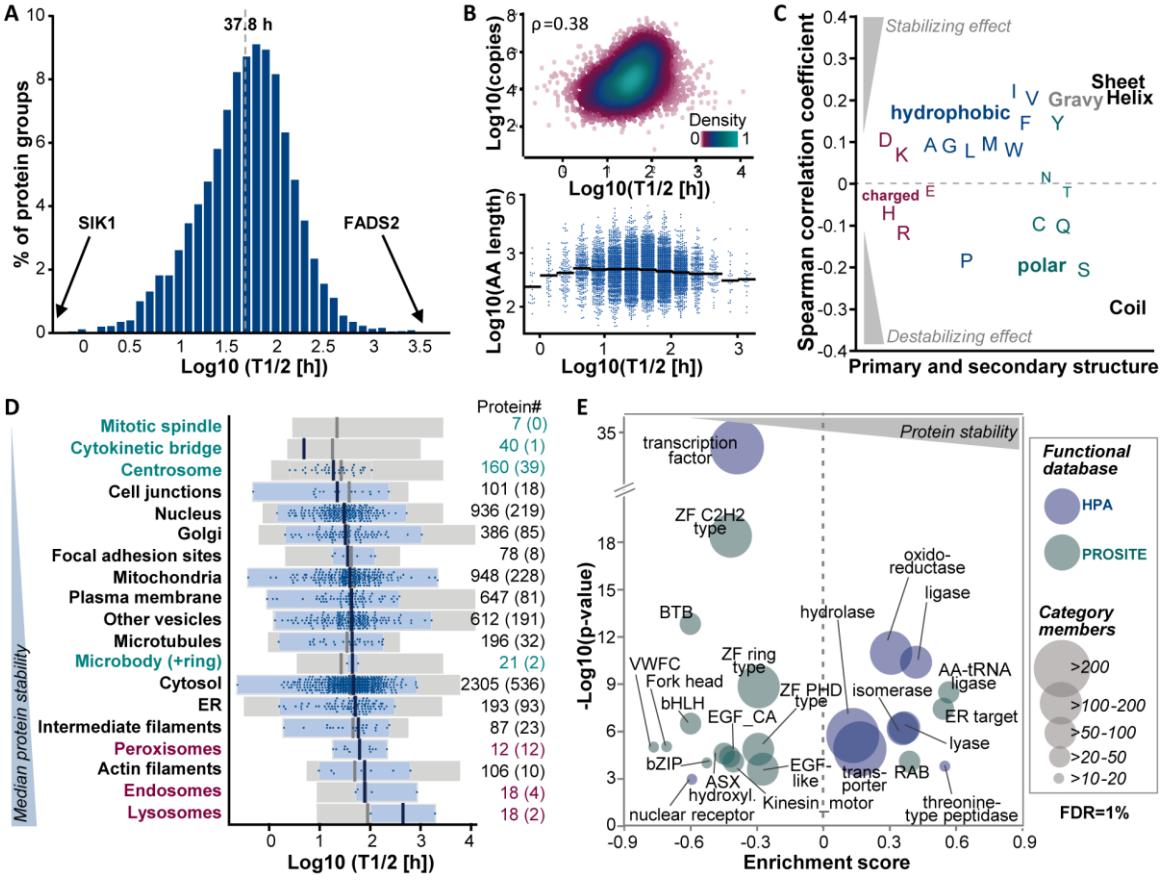



**Figure IV-6 | Analysis of protein half-lives relating to intrinsic protein properties and functions**. *(A) The distribution of protein half-lives in HeLa cells is displayed. The median protein half-life of all proteins is 37.8 hours. (B) Correlations of protein half-lives and copy numbers per cell (upper panel, ρ: Spearman rank correlation coefficient) and protein length (bottom panel) are shown. Solid black lines indicate the median length of all proteins per the half-life bin. (C) Spearman rank correlation coefficients are depicted for the correlation of protein half-lives and amino acid composition, amino acid properties, and protein secondary structure elements. (D) Floating bar charts illustrate the range of protein half-lives as a function of cellular localization (according to the Human Protein Atlas (HPA) and MitoCharta project). Proteins that are part of cell structures involved in cell division are shown in green, endo-, lyso- and peroxisome-associated proteins are shown in red. Grey boxes display all proteins associated with the respective subcellular location, blue boxes refer to proteins (blue dots) which were exclusively found in this particular cell compartment or structure. Numbers on the right indicate how many proteins are in each category and numbers in brackets refer to proteins with exclusive localization. (E) The scatter plot shows significantly enriched categories after a one-dimensional functional enrichment analysis (1 % FDR) using protein domain and family information provided by the PROSITE and HPA databases. The size of each circular shape indicates the number of proteins in each category.*

**Intrinsic determinants of protein stability** – Two factors that may affect protein turnover could be protein size and cellular abundance, as these co-determine the energy costs caused by re-synthesis of a certain protein species after its degradation in a steady-state system. Indeed, protein half-lives were positively (albeit not strongly) correlated with protein abundance (ρ=0.38, Figure IV-6 B). In contrast, and perhaps surprisingly, protein size did not show any consistent or global effect on protein stability (Figure IV-6 B). Other protein properties that potentially influence

turnover are the primary and secondary structure as well as the hydrophobicity of a protein. Overall, weak correlations of protein half-lives and amino acid content or predicted proportions of α-helix, β-sheet, and coil structures (ρ between -0.28 and 0.23) were observed. For example, a high percentage of hydrophobic amino acids and an ordered secondary structure were associated with longer half-lives, whereas polar amino acids and a disordered structure (often showing high proline content) seemed to rather destabilize proteins (Figure IV-6 C, Appendix D). It is noteworthy that all examined protein features correlated with each other to some extent. For instance, the more abundant a protein was, the smaller (ρ=-0.40) and the less polar (ρ=-0.30) and disordered (ρ=-0.22) it tended to be (Appendix Figure 0-11 E). In order to investigate whether susceptibility to aggregation might be associated with cellular protein turnover, protein half-lives were compared to corresponding melting points that have recently been reported for HeLa proteins [308]. However, no general dependency of cellular protein turnover on thermal stability could be determined (Appendix Figure 0-11 F).

The localization of proteins might also affect their stability. To examine a potential spatial regulation of turnover, proteins were grouped according to their subcellular location reported by the human protein atlas (HPA) [292] and the MitoCharta 2.0 [291] projects. Again, proteins in these categories spanned a wide range of stability even when only assessing proteins that were exclusively found at a single location (Figure IV-6 D). However, endo-, lyso-, and peroxisomal proteins (median T1/2 of 80.2 h) appeared to be more stable compared to the overall cellular proteome. Of note, the small number of data points limits the generalizability of this observation. Conversely, proteins that constitute members of mitotic cell structures (centrosome, mitotic spindle, cytokinetic bridge, and midbody, median T1/2 of 24.9 h) exhibited shorter median half-lives. This potentially reflects the need for rapid regulation of abundance during different phases of the cell cycle. In contrast, actin and intermediate filaments and proteins exclusively located in the endoplasmic reticulum (ER) were on average slightly more stable compared to nuclear and cell junction proteins (median T1/2 of 77.8, 58.7 and 50.5 versus 30.3 and 22.4 h). Other localizations did not show any considerable trend towards an overall stabilization or destabilization of associated proteins.

Next, the relation of protein half-lives to annotated functions was investigated using protein domain and family information provided by the PROSITE [288] and HPA [287] databases. A functional 1D enrichment analysis illustrated the significantly shorter half-lives of transcription factors containing zinc finger (ZF), fork head, basic helix-loop-helix (bHLH) and leucine zipper domains (bZIP), as well as nuclear receptors (Figure IV-6 E). Examples included members of the STAT (Signal transducer and activator of transcription) family, the transcriptional regulators MAX and MYC as well as retinoic acid and androgen receptors. In contrast, several families of enzymes, notably oxidoreductases, ligases, lyases, isomerases, and hydrolases were significantly overrepresented in more stable proteins (Figure IV-6 E). Interestingly, the aforementioned transcription factors and enzymes also clearly differed in the biochemical features assessed above. Enzymes did not only possess longer half-lives compared to transcription factors (55.4 h vs. 17.6 h), but were also more abundant (50,000 vs. 9,000 copies per cell) and more hydrophobic (44 % vs. 37 % hydrophobic amino acids) and exhibited much less disordered secondary structures (54 % vs. 84 % coil structure, Appendix Figure 0-11 G). Other distinctively more stable functional protein groups included cell and organelle membrane associated transporters and, interestingly, proteins with a C-terminal KDEL motif, which targets proteins to the ER. The latter indicates that proteins that permanently and exclusively reside in the ER lumen, like for instance protein disulphide isomerases, are indeed more stable as already suggested by the HPA subcellular location annotations (Figure IV-6 D). Furthermore, the Rab family of GTPases, which regulate vesicular trafficking, exhibited

significantly longer half-lives. On the contrary, kinesin-like proteins, which are involved in mitosis via the control of chromosome segregation, were enriched in high turnover proteins. Likewise, proteins bearing ASX hydroxyl and EGF like domains appeared to be rather short-lived. It is important to note that these domains often co-occur on extracellular proteins (e. g. fibrillins and fibulins), and the detected comparably short half-lives might therefore rather reflect the intracellular transit time before these secreted proteins are lost from the pool of analysed proteins than their actual stability. The same argument applies for proteins featuring von Willebrand factor type C (VWFC) repeats that include amongst others fibrillar collagens. Taken together, the collective turnover dataset facilitated the analysis of protein properties and functions affecting protein stability, but no universal protein immanent factors deterministically influencing half-lives were identified.

**Oxidative stress as a regulating factor for NADH dehydrogenase stability** – Besides protein intrinsic factors such as structure and function, protein half-lives might also be regulated by molecular interactions. To this end, the stability of CORUM complex members [379] was evaluated. Proteins reported to be part of a protein complex exhibited overall longer half-lives (median 51.9 h) compared to proteins which are not listed in the CORUM database (median 44.1 h) suggesting a stabilizing effect of protein interactions and complex formation. In particular, the proteasome and ribosome (and their precursors) were significantly enriched in more stable proteins (Figure IV-7 A). Interestingly, respiratory chain complex I (NADH dehydrogenase) members were the only proteins participating in complexes and the electron transport chain that showed overall significantly shorter half-lives (median 9.5 h, p-value=1.84e-11, Figure IV-7 A-B). Together with the Ubiquinol cytochrome C oxidoreductase (respiratory chain complex III), the NADH dehydrogenase is the main site of superoxide radical formation caused by electron leakage in the respiratory chain (Figure IV-7 A-B). Therefore, it was hypothesized that the high turnover of complex I proteins may be an adaptive mechanism to compensate oxidative stress by replacing damaged complex members and thus to maintain the functionality of the electron transport chain which is needed for energy generation by oxidative phosphorylation. Following this assumption, the turnover of NADH dehydrogenase proteins should be accelerated upon enhanced oxidative stress. In order to test this hypothesis, HeLa cells were treated with the complex I inhibitor rotenone in combination with glutamate and malate to increase oxidative stress specifically at complex I [380] (Figure IV-2). By inhibiting electron transfer from iron-sulphur centres to ubiquinone, rotenone treatment should lead to a backload of electrons that should be further amplified by the increased electron supply provided by the NADH dehydrogenase substrates glutamate and malate (Figure IV-7 C). The rotenone and control treatments (either DMSO or solely glutamate and malate) were followed by a 3 or 8 h pulse in K8/R10 SILAC medium. To overcome missing quantitative data across treatment conditions and replicates, a parallel-reaction-monitoring assay was developed to quantify heavy-to-light ratios, which reflected the fraction of newly synthesized to the total protein amount. In total, 43 peptides representing 27 complex I proteins were monitored in their K0R0 and K8R10 labelled states. For both pulse time-points, rotenone treated cells showed a clear shift towards higher heavy-to-light (H/L) ratios compared to control cells implying that the overall fraction of newly synthesized complex I members and thus their turnover increased (Figure IV-7 D, Appendix Figure 0-12). Proteins featuring peptides with a significantly different H/L ratio were mainly located at the so-called IF site where electrons are transferred from NADH to FMN (Flavin mononucleotide) and further passed down the chain of iron-sulphur centres (Figure IV-7 E). Taken together, these results indicate that the turnover of many respiratory chain complex I members is

accelerated upon the blockade of electron transfer to ubiquinone suggesting that a general regulation of their half-lives by oxidative stress may exist.

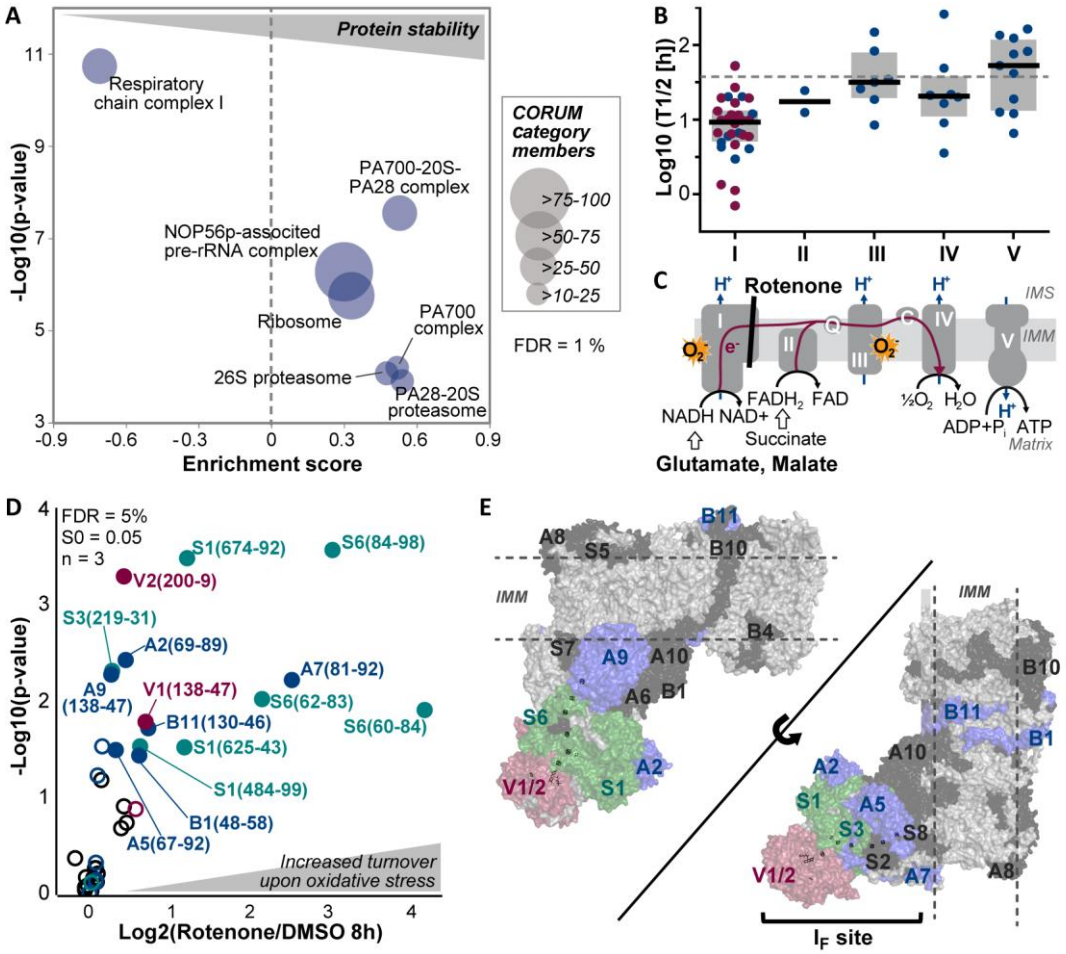
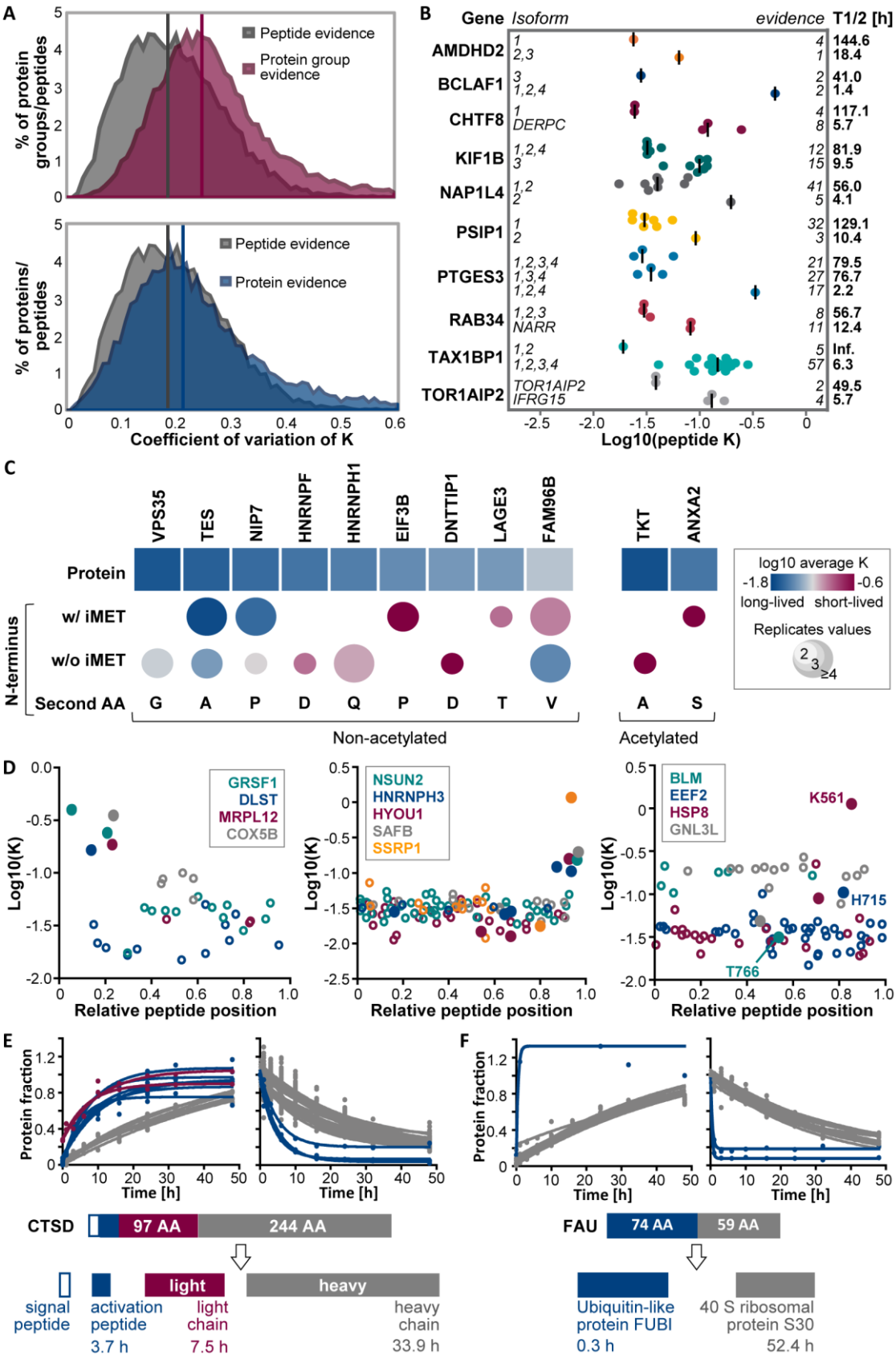



**Figure IV-7 | Analysis of turnover of respiratory chain complex I proteins in response to rotenone-induced, oxidative stress.** *(A) One-dimensional enrichment analysis (1 % FDR) using CORUM database annotations revealed that respiratory chain complex I (NADH dehydrogenase) was significantly enriched in high turnover proteins. The size of each circular shape indicates the number of proteins in each complex. (B) Scatter dot plots show protein half-lives of members of the different respiratory chain complexes. Black lines indicate the median half-life of proteins within each complex and proteins marked in red were followed up by rotenone and/or glutamate and malate treatment in the subsequent PRM assays. (C) The schematic representation of the different complexes of the respiratory chain (I-V) illustrates sites of metabolic reactions and superoxide production (IMM: Inner mitochondrial membrane; IMS: Intermembrane space; Q: Ubiquinone; C: Cytochrome C). (D) The Volcano plot shows the results of triplicate PRM assays monitoring the turnover of 22 members of the respiratory chain complex I in response to rotenone (1 uM), glutamate and malate (5 mM) treatment. Peptides exhibiting a significantly higher turnover upon treatment are illustrated by filled circles and labelled with the subunit, peptide start and end positions. Colours refer to particular proteins shown in panel E. (E) Crystal structures of respiratory chain complex I proteins are displayed. NADH dehydrogenase proteins with significantly increased turnover upon rotenone treatment are coloured in red, blue and green. Iron sulphur clusters are shown as sticks. Subunits coloured in black were detected in the PRM assay, but did not show a significant change in turnover upon treatment after an 8 h pulse. Subunits in light grey could not be robustly monitored in the PRM assay.*

## 3.4 Turnover diversity of peptides

Expecting a cell to be able to respond quickly to cellular stimuli such as rotenone induced, oxidative stress as shown above, inevitably leads to the hypothesis that protein turnover could be dynamically regulated by post-translational protein modifications. For the above example on the turnover of respiratory chain proteins, it was unfortunately not possible to test this hypothesis directly because information on methionine oxidized respiratory chain peptides and their non-oxidized counterparts was not available. In general, the comparison of oxidized and non-modified counterpart peptides did not show any global shift in turnover due to oxidation (Appendix Figure 0-13 A). This is not surprising assuming that protein and peptide oxidation largely occurs during sample processing and, hence, would not alter the measured, cellular turnover. Moreover, a Student's t-Test did not reveal any significant differences of individual oxidized and non-modified peptide pairs (FDR=5 %, S0=0.05). Yet, it was generally, and quite unexpectedly, observed that turnover rates determined from all spectrum evidences for a certain peptide sequence showed less variation (median CV of 18 %) than rates derived from all spectrum evidences for a protein group (median CV of 25 %, Figure IV-8 A upper panel). A protein group can contain peptides that originated from different expressed protein isoforms, if unique peptide(s) are identified solely for one of these isoforms and only shared peptides have been identified for the other isoform(s). Thus, the global difference in CV values of protein groups compared to peptides suggested that protein isoforms (that are unavoidably included in a protein group) might indeed often differ in their turnover behaviour. Given the availability of four cell culture replicates, statistical testing was feasible to prioritize protein groups consistently containing peptides with considerably varying stabilities. To do so, only peptides for which a turnover rate was determined at least 3 times (from synthesis or degradation curves or different replicates) and which belong to protein groups containing at least three of these peptides were included (25,313 peptides assigned to 3,130 protein groups). A two-sided, 5 % FDR corrected t-Test yielded 425 peptides from 305 protein groups for which turnover rates significantly differed (Appendix Figure 0-13 B). Amongst these, several

**Figure IV-8 | Analysis of peptidoform-resolved turnover.** *(A) Distributions of coefficients of variation of turnover rates across spectra are displayed for peptides and proteins. The upper panel shows the distribution of CVs of all spectra for the same protein group (irrespective of the number of proteins or protein isoforms in each group). The bottom panel depicts the CV distribution of all spectra for protein groups that only contain a single protein. Medians are indicated by vertical lines in the corresponding colour. (B) Scatter dot plots show the distributions of turnover rate constants for peptides of different protein isoforms. Peptides corresponding to a particular gene share the same colour. Median peptide rates across replicates are displayed as vertical lines. (C) Turnover rates are illustrated for different modified N-terminal peptides and corresponding proteins. Only N-termini and proteins with statistically significantly different rates are displayed (see also Appendix C). (D) Scatter plots show peptide turnover rates as a function of the relative position within the protein sequence. Zero denotes the protein N-terminus and 1 denotes the C-terminus. Peptides derived from the same protein share the same colour, and closed circles mark peptides that exhibited a significantly different turnover compared to the rate of the whole protein (see also Appendix C). The left panel illustrates examples for mitochondrial proteins in which the N-terminal transit peptides show a higher turnover than other peptides of the same protein. The middle panel depicts a similar analysis but for proteins with higher turnover of C-terminal peptides. The right panel shows examples for proteins in which one peptide displayed a strong difference in turnover compared to other peptides of the same protein which often but not always encompassed known modification sites. (E) Fractional peptide labelling is displayed for Cathepsin D (CTSD), a protein that is proteolytically processed into a signal peptide, an activation peptide (blue circles), a light chain (red circles) and a heavy chain (grey circles). (F) Fractional peptide labelling is depicted for the fusion protein FAU. Peptides representing the Ubiquitin-like protein FUBI are shown in blue, peptides from the 40S ribosomal protein S30 are shown in grey.*




protein groups containing different splice variants were identified. For example, the only peptide that was exclusively assigned to isoform 2 of Nucleosome Assembly Protein 1-like 4 (NAP1L4) was

much less stable (T1/2=4.1 h) compared to peptides occurring in both isoforms 1 and 2 (T1/2=56.0 h, Figure IV-8 B). It has to be pointed out that these differences might also be caused by a putative stabilizing modification, which happens to occur in the identified isoform 1 specific sequence stretch, hence making the unmodified counterpart peptide appear less stable compared to the other peptides of the protein. However, the occurrence of different half-lives for different isoforms was also observed in cases where isoforms were unambiguously identified via several unique peptides and which thus were assigned to different protein groups as exemplified by chromosome transmission fidelity protein 8 homolog (CHTF8, 5.7 vs. 117.1 h for isoform 1 vs. splice variant DERPC, Figure IV-8 B). Notably, most of these protein groups that separated isoforms featuring significantly different turnover also possessed considerably different primary sequences and thus physicochemical properties.

After the removal of all protein groups that contained more than one protein isoform, turnover rates of all peptide evidences belonging to a single protein still showed overall higher CVs (median CV of 22 %) than rates derived from all evidences for single peptides (median CV of 18 %, Figure IV-8 A, lower panel). This demonstrates that alternative splicing alone does not sufficiently explain the variation in turnover across peptides assigned to the same gene. In fact, many peptides located at the protein N-terminus were found to exhibit significantly different turnover rates. To explore this further, rates of differently modified N-terminal peptides of the same protein were compared among each other in addition to the comparison to the overall protein turnover rate. Even without enrichment of N-terminal peptides, rates were obtained for 343 N-terminal peptides from 306 proteins. About half of these peptides (53 %) neither contained the initiator Met residue nor were they acetylated. After filtering, differences were statistically evaluated for 287 proteins (Appendix Figure 0-13 C). Eleven N-termini significantly differed in their turnover rates, but the effect was not consistent for the type of modification (Figure IV-8 C). While, for instance, the N-terminal peptide without the initiator methionine of HNRNPH1 appeared to be turned over more quickly compared to the whole protein, it was the other way around for the mitotic spindle-associated MMXD complex subunit MIP18 (FAM96B).

Higher turnover rates for peptides located near the N-terminus were also identified for mitochondrial proteins like the G-rich sequence factor 1 (GRSF1, Figure IV-8 D, left panel). In fact, these peptides were part of or spanned cleavage sites of transit peptides, which target nuclear encoded proteins to mitochondria. This suggests that these localization signals are in general rapidly cleaved off and degraded leading to mature, more stable proteins. Likewise, pro-peptides often appeared to be less stable compared to mature proteins as exemplified by Prosaposin which is cleaved into four different Saposins, a signal peptide, and several pro-peptides (Appendix Figure 0-13 D). Another case where proteolytic processing led to products with different apparent cellular stability is illustrated by Cathepsin D (CTSD), a protease that consists of a light and a heavy chain which are encoded by the same gene and is post-translationally cleaved [381] (Figure IV-8 E). Interestingly, not only did the activation peptide of CTSD exhibit a higher turnover (T1/2=3.7 h), but also its light chain showed a shorter half-life (7.5 h) compared to the heavy chain (33.9 h) which, in turn, implies a much lower abundance of the light compared to the heavy chain in steady state. Moreover, to our knowledge, the first experimental evidence for the post-translational cleavage of the protein produced by the fusion gene FAU was provided by the data. This is demonstrated by the considerably different stabilities of peptides corresponding to the Ubiquitin-like protein FUBI and the 40S ribosomal protein S30 part of the fusion protein (T1/2=0.3h vs. 52.4 h, Figure IV-8 F).

Apart from these rather intuitively explicable examples, mechanistic explanations for discrepancies between peptide and corresponding protein turnover rates were often less apparent. A group of proteins, for instance, exhibited a distinctively higher turnover for the most C-terminal located peptides (examples are shown in Figure IV-8 D, middle panel). This might hint to a C-terminal modification that stabilizes these proteins and thus would lead to a seemingly shorter half-life of the unmodified C-terminus. In fact, many peptides showing significantly different turnover rates encompassed reported modification sites. Amongst others, for example the only peptide of elongation factor 2 (EEF2) showing a substantially higher turnover rate contained His715 which is believed to be the only histidine in eukaryotes that is converted into diphthamide [382] (Figure IV-8 D, right panel). Furthermore, the peptide showing by far the highest turnover rate for the heat shock cognate 71 kDA protein (HSPA8) included a lysine (position 561) that was demonstrated to be trimethylated by the methyltransferase METTL21A. This, in turn, suggests a stabilizing effect of the modified proteoform of HSPA8 in addition to the previously described chaperone activity modulating effect that enhanced α-Synuclein aggregation and, thus, may play a role in Parkinson disease [383] (Appendix Figure 0-13 E). The most stable peptide of the bloom syndrome protein (BLM) comprised Thr766 which has been shown to be phosphorylated by Cyclin dependent kinase 1 (CDK1) potentially regulating its helicase activity during mitosis [384] (Appendix Figure 0-13 F). In contrast to these examples, many peptides also featuring significantly different rates did not encompass any known modification site (exemplified by GNL3L in Figure IV-8 D, right panel). Still, all the cases described above clearly demonstrate differential turnover rates for different proteoforms suggesting an association of post-transcriptional and post-translational processing with protein stability.

# 4 Discussion and conclusion

## 4.1 Technical and data analysis challenges for (pSILAC-TMT) turnover studies

Enabled by advances in MS based proteomic technologies in general and the introduction of the SILAC technology in particular, several attempts have been made in recent years to investigate endogenous proteome turnover [113, 168-170, 305, 374]. Still, generating such data at high quality is far from trivial and many factors have to be carefully considered when planning and executing such experiments. At the technical level, the accuracy of the commonly applied standard dynamic SILAC method employing MS1-based quantification can be impaired by missing data across different pulse time-points. Considering the utilization of combined, quantitative information of multiple peptides, this shortcoming may be tolerable at the protein level, but, especially for single peptides, such rate estimations can be unreliable. Therefore, a method was established that combines the quantitative precision of TMT-10plex labelling with pulsed SILAC labelling of cells, and the merits of this approach were evaluated thoroughly. Direct comparisons illustrated the extent of the missing value issue in the MS1-based quantification method. Absent intensities across SILAC pairs and time-points reduced the number of successfully determined protein labelling rate constants and concomitantly decreased the quality of curve fits. This was the case even though missing values were already minimized by utilizing the automatic SILAC pair identification and the match-between-runs function implemented in the MaxQuant software [385] and even though less stringent criteria regarding missing values were applied for the curve fitting. While TMT labelling effectively overcomes the issue of missing values and thus facilitates determination of turnover rates at single peptide level, it is well known that quantification using isobaric tags suffers from ratio distortion caused by co-isolated peptides. This drawback can have massive consequences for rate determinations, in particular if a K0R0 and a K8R10 peptide are co-fragmented since they would show the exact opposite abundance behaviour. Indeed, when applying MS2-based quantification of reporter ions, severe ratio compression was detected often rendering determined labelling rates plain invalid. This was also underlined by the fact that, after correction of these turnover rates for cell doubling, almost 50 % of all quantified proteins had negative half-lives that typically result from either an underestimation of labelling rates or an overestimation of cell doubling rates. By contrast, using a more extensive fractionation scheme and a MS3-based TMT quantification strategy, ratio distortion was minimized and negative values were obtained for less than 3 % of all proteins which is a very small fraction compared to other published turnover studies. Moreover, turnover rates obtained from this experimental workflow were in good agreement with rates determined via the classical approach based on MS1 quantification. Overall, it must be concluded that an MS3-based measurement is required for proper estimation of protein turnover when using a pulsed SILAC-TMT format.

In contrast to the standard pulsed SILAC approach, the SILAC-TMT hyperplexing strategy provided a duplicate measurement of turnover rates for all cases where both peptides of a SILAC pair were fragmented. This helped to assess the quality of rate estimations offering an additional level of confidence within the same experiment. In fact, a comparison of rates calculated from label increase and decay exposed those, mainly high turnover peptides for which residual TMT ratio compression affected determined rates. Although not investigated in this study, this internal duplicate may be particularly helpful for the evaluation of turnover rates in non-steady-state, dynamic or disturbed systems such as cell differentiation or upon cell treatments. The data normalization and curve fitting functions used were based on the assumption that the average abundance of all proteins does not change during the course of the experiment (steady state) and, therefore, that the

synthesis rate of a protein equals its degradation rate. For this reason, cell seeding densities and culture conditions were carefully elaborated to assure that HeLa cells were no longer in lag phase or not yet entering the stationary growth phase during all SILAC pulses. Still, cells were cultured under somewhat non-continuous conditions as nutrients were depleted and metabolites accumulated over the course of the experiment. These factors can potentially cause cells to respond with a change in abundance of certain proteins. Consequently, it cannot be precluded the possibility that the applied curve fitting algorithms might not perfectly describe the detected labelling behaviour for all proteins.

In light of the above, an even greater challenge may be posed when aiming to establish an appropriately adopted model for labelling kinetics under non-steady-state circumstances, which would account for changes in protein abundance during an experiment. In addition, one would typically want to compare different conditions eventually obtaining absolute protein stabilities. However, when proliferating cell models are used, measured labelling kinetics are not only dependent on protein degradation and re-synthesis, but are also critically influenced by sheer cell doubling. This complicates comparability since cell-doubling rates are likely to be different between cell treatments and may even be changing during the course of a single pulse experiment under non-steady-state conditions. This emphasizes the need for an accurate method to measure temporally resolved labelling kinetics and for the correct determination of cell doubling times, which must be conducted under the same culture conditions using identical cell batches.

Steady-state assumptions and the consequently applied exponential synthesis and degradation models also presume that the probability of a protein being degraded stays constant over its lifetime. This alone might not hold true for all proteins. In fact, newly synthesized proteins have previously been reported to show shorter half-lives before they enter a second, more stable state [157]. Recently, McShane *et al.* found that about 10 % of all proteins detected in their study show such a non-linear degradation behaviour [376]. Many of these proteins were members of protein complexes that were produced in super-stoichiometric amounts substantiating the assumption of protein stabilization due to complex formation, which has been termed cooperative stability [386]. Besides, co-translational ubiquitination and rapid degradation of misfolded proteins immediately after synthesis have also been demonstrated [387, 388] suggesting that a biphasic degradation behaviour could also be a result of cellular quality control mechanisms. For evaluation of non-linear degradation kinetics, an initial discrimination of newly synthesized and aged proteins would be needed, for example via a combination of different pulse and chase time-points [157] or a second metabolic label [376]. As most other published turnover surveys, the presented work does not provide such resolution. Instead, the data describes an average behaviour of the different states of a protein likely dominated by the turnover characteristics of the most abundant one.

## 4.2 Biological implications of measured protein turnover

The systematic evaluation of replicates, followed by the assessment of the reproducibility for turnover estimations attested a reliable precision of the SILAC-TMT approach on the protein as well as the peptide level. Moreover, the four cell culture replicates did not only allow for testing of statistically significant differences, but should also increase the robustness of combined turnover information from replicates, thereby providing a unprecedentedly comprehensive and high-quality dataset on endogenous protein turnover. Still, it has to be noted that very quickly turned over proteins, particularly those that are completely turned over within 1 h, will be missing or underrepresented in the dataset simply due to the choice of time-points and curve fitting constraints. Nevertheless, the representative proteome coverage and high quality of the data allowed

for re-assessment of potential determinants of protein stability that are inherent to the protein itself.

A positive correlation between protein abundance and turnover, but a negative correlation between protein abundance and length was observed which corroborates findings by others [113, 170, 283]. Making highly abundant proteins stable and small may offer a route for cells to avoid excessive energy consumption considering the costs for re-synthesis of degraded proteins. It should be noted though that for high abundant proteins, even slow degradation rates will lead to degradation of a large number of molecules. For the 60S ribosomal protein L18, for example, only 0.13 % of all molecules were measured to be degraded per hour (corrected for cell doubling). However, at a cellular abundance of, on average, more than 1 Mio copies per cell, this still results in the degradation and subsequent re-synthesis of about one RPL18 molecule every 3 seconds illustrating the energy efforts for the maintenance of the default turnover of very stable, but highly abundant proteins. Protein size (length) has also been reported to be associated with protein stability based on measurements using fluorescently tagged proteins in mammalian cells and yeast [125, 153]. While perhaps intuitive, as making longer proteins should contribute to energy consumption during translation, others subsequently argued that this is not the case for endogenous proteins [170, 305] and the here presented data confirms that there is no global correlation of protein length with protein stability.

However, a weak correlation of protein half-lives with features of primary and secondary protein structure was detected. Polar amino acids, proline (which is known to disrupt ordered structures), and random coils were associated with short-lived proteins, the latter confirming earlier reports [125, 305] and being in line with the requirement of an disordered structure for proteasome binding and subsequent degradation [108]. In addition, overall sequence and structure differences appeared to explain, at least in part, the significantly different stabilities of functionally distinct proteins exemplified by transcription factors and enzymes. It has been known for a long time that the hydrophobic effect drives protein folding thereby reducing the surface area of proteins and the solvent accessibility of hydrophobic amino acids and leading to more ordered structures [389, 390]. Conversely, a more polar and random protein structure with relatively larger surface areas could possibly lead to a higher accessibility for modifications and interactions with other proteins, which could potentially induce protein degradation. This assumption would be in line with the hypothesis of cooperative protein stability since surfaces of proteins in complexes are also less solvent exposed [386, 391]. Indeed, it has even been hypothesized that disordered proteins with larger surface areas tend to engage in more promiscuous interactions, and are also more likely to have pathological effects when overexpressed [392]. This higher dosage sensitivity of disordered proteins would provide a conceptual explanation for the observed inverse correlation of the content of random coils with protein abundance. Overall, the observed higher turnover rate of rather disordered proteins could be a regulatory mechanism that protects cells from toxic protein aggregates.

In terms of single amino acids, serine had the strongest association with protein stability. One might speculate that its destabilizing effect might in part be related to its involvement in the formation of phosphodegrons, amino acid motifs that are recognized by E3 ligases which ubiquitinylate and thus mark respective proteins for degradation [132]. In contrast, charged amino acids did not appear to have any consistent effect on endogenous protein turnover in our data. This is somewhat contradictory to results of a fluorescence-based genomic tagging study, which found

glutamate, aspartate, lysine and arginine to be enriched in stable proteins [153]. This again suggests that protein stabilities derived from genomic tagging approaches investigating overexpressed proteins might not be readily transferable to endogenous protein half-lives.

In general, the transferability of turnover data obtained from different biological systems and experiments remains a subject of debate. A high conservation of protein stability has been claimed for human and murine proteins [169], across closely related rodent species [393] and for different primary cell types of the hematopoietic lineage [114] when using identical methodologies and conducting the comparative study in the same laboratory. However, discrepancies in turnover datasets appear generally more profound across different laboratories (Appendix Figure 0-9 B and according to [163, 170]). This may to some extent be of technical nature and, for example, be related to different measurement strategies, detection of different peptides for the same protein which may display distinct turnover rates, data quality or differing analysis approaches further underscoring the challenge to reproducibly determine protein turnover. However, turnover rates published from different laboratories but for the same cell line were found to correlate better among each other than across different cell lines which indicates some cell line-specific component affecting turnover. This may not be surprising considering the reported relationships of protein abundance, localization, or interactions with protein stability. Although features such as amino acid content and structure will not differ for the same protein between various cell lines, expression patterns, main splice variants, and complex partners may vary considerably and thus influence measured protein stability. Moreover, it has to be noted that the HeLa cells used in this study are highly aneuploid. It has already been shown that proteins derived from amplified gene regions often feature a higher degradation rate [376], which may provide an additional explanation for the observed discrepancies across cell lines. In addition, the here explored dynamic state of the proteome in proliferating cells may not resemble protein stability in whole organisms where most cells reside in a non-dividing state. Nevertheless, our turnover rates correlated well with data derived from arrested HeLa cells [169]. Overall, while some comparative cross-species or cross-cell type studies have been published recently [114, 393], the degree of conservation of protein stability still needs to be explored more systematically also including different laboratories.

Another factor that has been described to influence turnover rates is the localization of proteins. In the present study, only minor differences in half-lives of proteins assigned to different cell locations were observed. However, it has to be noted that potentially different turnover rates for the same protein species localizing to various cell compartments cannot be distinguished in the presented work since whole cell extracts were analysed. A spatially resolved study would provide further insights, especially considering recent observations that proteins frequently localize to multiple cell compartments [292]. Indeed, after subcellular fractionation, Boisvert *et al*., detected differing stabilities of the same proteins depending on their localization [170]. In particular, they detected complex subunits to exhibit longer half-lives after complex assembly which has recently been further supported by McShane *et al.* [376].

As mentioned earlier, the present study does not provide the resolution to discriminate between free and assembled complex units. However, a generally longer half-life of proteins that are part of complexes was detected supporting the overall notion of a stabilizing effect of protein interactions. The only significant exception to this rule was the NADH dehydrogenase, complex I of the electron transport chain. Despite of their high abundance, members of this complex exhibited rather short half-lives. It was suspected that this might be related to damage inflicted by oxidative stress, which was corroborated by experiments performed under rotenone-induced, oxidative

stress conditions. Within the scope of this work, only NADH dehydrogenase proteins were investigated, thus further experiments are required to clarify whether the turnover of other respiratory chain proteins can also be regulated by other forms of oxidative stress. This may be of particular interest for members of complex III, as this is the second site of superoxide formation in the electron transport chain. Surprisingly, not all assayed subunits of respiratory chain complex I showed the same significant increase in turnover upon rotenone treatment indicating that single subunits mainly located around the iron-sulphur centres were substituted in the complex. Although a selective exchange of single subunits has been described before in other complexes [394, 395], the processes and mechanisms by which this occurs remain elusive. Notably, half-lives of proteins within some complexes (also including the respiratory chain complexes) showed major variations which might further support the principle of a selective degradation of single complex subunits. However, another explanation could be that measured turnover rates represent an average of the degradation behaviour of free and assembled subunits. As already stated, these two states might possess different stabilities and, in addition, not exhibit the same proportions for all complex members. Indeed, for NADH dehydrogenase, this presumption was substantiated by an observed negative correlation of copies and half-lives ($\rho$=-0.34) attesting a higher turnover for those subunits which must feature a bigger fraction of the free protein state.

Not only proteins within the same complex varied in their stabilities, but also even peptides assigned to the same protein group differed more in their turnover rates than what could be simply explained by technical variation. It has been stated before that isoforms and differentially modified proteins can exhibit different stabilities [162, 163, 169], but due to the restricted quantification accuracy at the peptide level, these kinds of analyses have hitherto largely been limited to comparisons of turnover dynamics of groups of peptides or the simple comparison of proteins included in a modification database to those not registered in this database. As demonstrated within this work, using the pulsed SILAC-TMT multiplexing approach facilitates an evaluation of turnover rates at the level of single peptides. Among the peptides with significantly different turnover times, several were constituents of distinct splice variants potentially representing isoform specific protein turnover. For most of these examples, physicochemical properties of annotated splice variants considerably differed, which further reinforces the notion that structure, hydrophobicity and abundance play a fundamental role in regulating protein stability. How this is controlled at a molecular level, however, remains largely elusive.

As noted above, certain sequence motifs termed degrons have been found to serve as recognitions signals for E3 ubiquitin ligases and are therefore connected to protein stability [107, 119, 396]. In addition to the aforementioned phosphodegrons, it has been demonstrated that the identity of N-terminal residues following the initiator methionine and N-terminal processing are associated with different protein stabilities [163, 397]. Indeed, proteins possessing an N-terminal alanine were significantly enriched in long-lived proteins, whereas lysine and glutamate appeared to have a rather destabilizing effect (1D enrichment analysis at 1 % FDR employing protein half-life data). However, the overall effect was small (enrichment scores between -0.14 and 0.09) indicating a high variation in the half-lives of proteins featuring the same N-terminal residue. Regarding N-terminal processing, Gawron *et al.* also detected a generally higher stability for peptides that retained the initiator Met residue (iMet) before valine and proline residues compared to those without it [163]. Consistent with this, the proline containing N-terminus of 60S ribosome subunit biogenesis protein NIP7 was detected to be less stable after iMet cleavage in our data. Conversely, the eukaryotic translation initiation factor 3 subunit B (EIF3B) protein and the mitotic spindle-associated MMXD complex subunit MIP18 (FAM96B), which contain an N-terminal Pro and Val

residue respectively, displayed the opposite behaviour with iMet containing peptides featuring a much higher turnover. Overall, this indicates that effects of N-degrons do not necessarily need to be consistent across proteins and that additional factors like accessibility [118] may be involved in N-terminal regulation of protein stability.

Importantly, with regard to potentially modified peptides, it needs to be emphasized that the disappearance of SILAC labelled peptides can be caused either by actual degradation of the proteoform related to this detected peptide or by the peptide entering another (modification) state. In other words, a higher turnover rate of iMet containing peptides could also be related to the rate of N-terminal proteolytic processing. The same applies to other irreversible post-translational modifications and in general to all cases where the detected peptide or protein is lost from the pool of analysed species as it is the case, for example, for secreted proteins. Accordingly, it is conceivable that the high turnover rate of the peptide encompassing His715 in EEF2 might also reflect the rate of diphthamide modification [382]. Similarly, the turnover of peptides comprising cleavage sites in CTSD and PSAP, may as well illustrate the proteolytic process itself. In contrast, peptides that are completely contained in a cleavage product should represent its actual stability as observed for mitochondrial transit peptides, the Ubiquitin-like part of FAU and also the light chain of CTSD. Two different products of CTSD which are likely to be related to the full length protein and the product resulting from the cleavage of the secretion signal peptide have previously been shown to differ in their turnover rates [162]. However, it has never been demonstrated that the amino terminal light chain of CTSD is less stable than the carboxyl-terminal heavy chain. Since both chains are associated via hydrophobic interactions to form the active site of CTSD [398], this potentially hints to a hitherto unknown control mechanism of CTSD activity by regulation of the abundance of the light chain via its higher turnover. Consequently, a stabilization of the light chain alone would rapidly increase the abundance of active CTSD.

A similar principle could underlie the regulation of BLM activity. This helicase has been proposed to be involved in DNA double strand repair [399] and described to be phosphorylated at Thr766 by the cell cycle regulating kinase CDK1 [384] which in turn has been demonstrated to be degraded upon genotoxic stress [400]. Together with the observation of a much higher stability of the Thr766 non-phosphorylated state, this suggests that BLM half-life can be increased via diminished phosphorylation on Thr766 as a result of reduced CDK1 activity following DNA damage (Appendix Figure 0-13 E). This post-translationally regulated stabilization would provide a rapid means to enhance helicase activity for DNA repair. This example illustrates that a high “default” protein turnover, which at first sight may appear to be disadvantageous from an energy efficiency point of view, can enable a cell to respond more flexibly and rapidly to altered cellular conditions without the need to induce transcription and translation. With respect to the further identification of differentially turned over proteoforms, it generally needs to be considered that, in bottom-up proteomics, every peptide that is analysed may include molecules derived from various proteoforms. As a result, when comparing unmodified (potential modification counterpart) peptides to the whole protein, only modifications that either very potently alter protein stability and/or exhibit an overall high occupancy can be indirectly identified as turnover regulating PTMs. In order to overcome the limitation of proteoform-shared peptides, a direct measurement of modified peptides and a comparison to their unmodified counterparts would prove beneficial.

## 4.3 Conclusion

In summary, this pulsed SILAC-TMT proof-of-concept study provides one of the most comprehensive, high quality turnover datasets in a single cell line to date with high temporal resolution of

endogenously expressed, untagged proteins in a steady-state cell system. Observed protein degradation rates spanned more than four orders of magnitude demonstrating that protein turnover must be a highly regulated process. Although multiple protein features were associated with turnover on the protein level, correlations were often only weak, which diminishes their overall predictive value and underscores the intricacies of the underlying, regulatory processes. This is illustrated by an even higher variation of degradation rates at the peptide level (more than five orders of magnitude) demonstrating that post-transcriptional and post-translational processing plays an essential role in the dynamic regulation of protein stability and thus revealing an underappreciated dimension in the functional control of life.

*„The beginning of knowledge is the discovery of something we do not understand."*

**- Frank Herbert**

# V DECIPHERING MODIFICATION-SPECIFIC PROTEIN TURNOVER

**1 Introduction and summary ..... 131**

**2 Experimental Procedures ..... 133**

**3 Results ..... 135**

3.1 Two approaches for the investigation of PTM-specific turnover ..... 135

3.2 Characterization of peptidoforms with differential turnover ..... 138

3.3 Functional perspective on modification-specific turnover ..... 145

**4 Discussion and conclusion ..... 151**

4.1 Robust determination of modification-specific protein turnover ..... 151

4.2 Uncovering unknown global and protein-specific regulatory mechanisms ..... 153

4.3 Conclusion ..... 156

# 1 Introduction and summary

After protein turnover in steady state was discovered in the late 1930s [97], the dynamic nature of cellular proteins has been studied intensively [372], and its significance for maintaining cellular protein homeostasis has been widely acknowledged [84, 99, 401]. More recently, the utilization of isotope tracing methodologies in conjunction with MS-based readouts has enabled investigation of turnover of endogenous proteins on a proteome-wide scale [168]. Employing pulsed SILAC approaches, half-lives of cellular proteins have been shown to range between minutes and more than 1,000 hours and vary considerably between different cell types and tissues [114, 402, 403] demonstrating the high degree of regulation imposed on protein synthesis and degradation processes. For several proteins, PTMs have been implicated in protein stability suggesting that the differential turnover observed in different biological systems may be post-translationally controlled. Most prominently, poly-Ubiquitin chains linked via lysine at position 48 typically mark proteins for proteasomal degradation [76]. Further, phosphorylated motifs that serve as recognition elements for E3 Ubiquitin ligases can induce ubiquitination and subsequent protein degradation. Such phosphodegrons have been identified to be responsible for periodic degradation of cell cycle regulators and enable progression of cell proliferation [122, 132]. In contrast to this positive PTM crosstalk, a competitive interplay between stabilizing acetylation and destabilizing ubiquitination has been observed for several transcription factors including p53, MYC, and SMAD7 [94, 404-409]. However, acetylation has also been shown to result in degradation of other transcriptional regulators such as HIF1α and GATA-1 [410, 411] illustrating the diversity and complexity of underlying regulatory mechanisms. Importantly, for most proteins, modification types and sites governing protein degradation remain elusive demonstrating the need for a global interrogation of modification-specific protein turnover.

Advances in the analyses of PTMs with MS-based proteomics technologies now facilitate the quantification of thousands of phosphorylation, ubiquitination, and acetylation sites within a single sample [194, 207, 215, 297] thereby theoretically allowing for the proteome-wide investigation of modification-regulated protein stability. However, only a limited number of studies have examined the impact of PTMs on protein degradation [162, 412, 413]. They either compared protein half-lives that were computed including or excluding all available phosphopeptides [162], inferred degrons from responses of co-occurring phosphorylation and ubiquitination sites upon proteasome inhibition [412], or relied on the integration of protein turnover data with counts of previously observed modifications sites [413]. While these investigations have their own merit, none of these studies directly assessed the steady-state turnover of individual modified peptidoforms limiting their informative value regarding the impact of individual modifications sites.

In an attempt to examine phosphorylation-, acetylation-, and ubiquitination-specific turnover globally for the first time, two strategies were employed. First, the pulsed SILAC-TMT approach, which has proven useful for time-course analyses of peptide turnover (see Chapter IV), was extended to measurements of the phosphoproteome and acetylome. Via adjustments of the data processing procedure including a computational removal of ratio compression and abundance corrections, pSILAC-TMT data quality was improved further. Second, phosphorylated, acetylated, and di-glycine peptides were analysed for four individual SILAC pulse time-points to obtain their relative turnover. The combined datasets yielded turnover information for >120,000 peptides carrying >30,000 modification sites and demonstrated that lysine acetylation predominantly decreased turnover of peptides, while ubiquitination mostly accelerated peptidoform turnover and

phosphorylation produced both effects. Statistical evaluation revealed an unexpectedly high degree of differentially turned-over peptidoforms within metabolic enzymes and protein complexes like the proteasome and ribosome indicating an under-appreciated level of post-translational regulation of their activities. Modification-specific analyses further identified new candidates for active degrons and suggested an additional layer of regulation by modifications that inhibit degron activity. An integration with structural information and drug treatment responses facilitated the identification of two distinct groups within acetylation sites that resulted in decreased peptide turnover. The first group was located within ordered secondary structures and partly antagonized destabilizing ubiquitination. The second group of acetylation sites was followed by proline and their slower turnover appeared to reflect an imbalance in modification kinetics caused by the lack of eraser enzyme activity instead of altered protein stability. Notably, di-glycine counterparts of slower turning-over acetylation sites frequently displayed a similar turnover as the non-modified site, suggesting that alternative mechanisms than a blocking of ubiquitination might slow down the turnover of many acetylated peptidoforms. Finally, the majority of modification sites with differential turnover lacked functional annotations illustrating the great potential of modification-specific turnover analyses for the identification of sites with potentially functional relevance.

## 2 Experimental Procedures

The following sections provide an overview of the experimental design and data processing strategies that are specific for the assessment of modification-regulated turnover. Details on individual experimental steps, data processing and analyses are specified in chapter II (see pp. 37).

Initially, enrichment procedures for acetylated and ubiquitin-remnant peptides using respective PTMScan® Kits (Cell Signaling Technology) needed to be implemented in the laboratory. Amongst other things, differing peptide quantities (1, 2, 3, 5, 10 mg), antibody bead amounts (5, 10, 20 µl settled beads), wash protocols (1x IAP buffer + 2x water, or 2x IAP buffer + 3x PBS), and elution steps (eluting up to 4x) were evaluated on the basis of the manufacturer's and other published protocols [207, 210]. Based on this multi-step optimization procedure (Appendix Figure 0-14, Figure ***0-15***), a protocol was established in which 10 µl settled beads and 2 to 5 mg of protein digest was used, beads were washed once/twice (acetyl/ubiquitin-remnant) with IAP buffer followed by two washes with PBS, and enriched peptides were eluted twice.

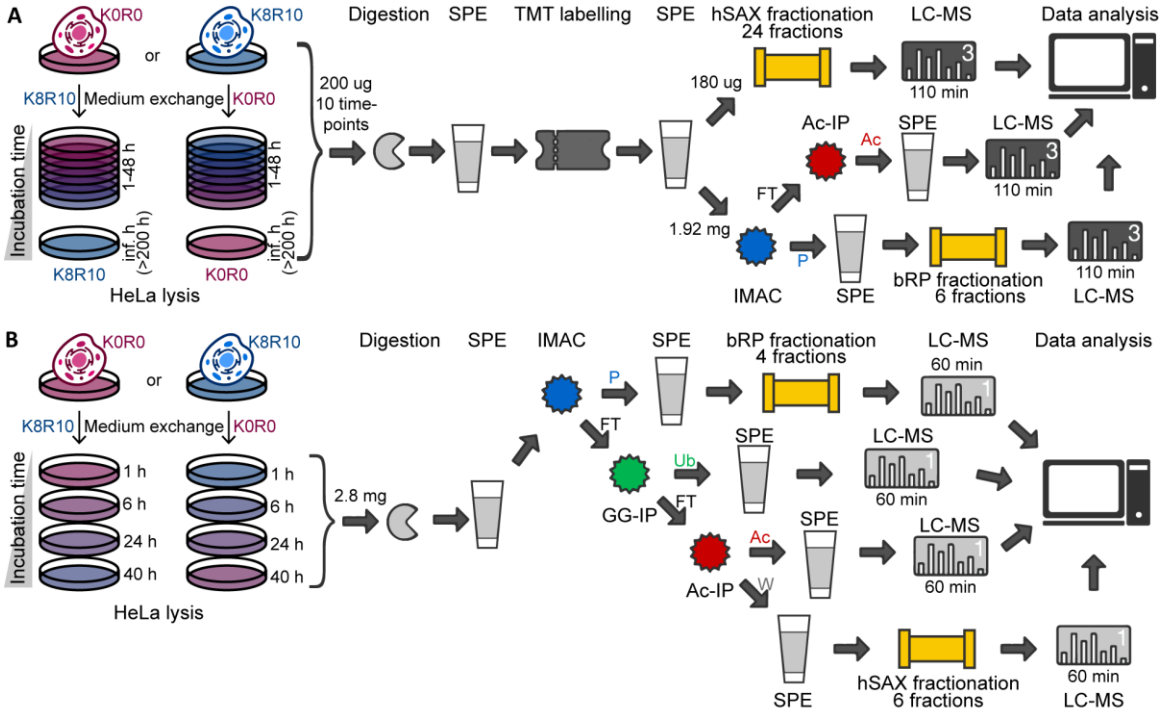



**Figure V-1 | Experimental designs of the two strategies for the assessment of modification-regulated protein turnover.** *For both experimental workflows, 4 replicates including two label-swaps were prepared. (A) For time-course data, pulsed SILAC-TMT experiments were conducted as described in chapter IV. An aliquot of labelled peptides was fractionated using hSAX chromatography to obtain non-modified peptidoforms. Residual peptides were subjected to serial phosphorylation and acetylation enrichments. (B) For single pulse data, individual pulses of 4 time-points were performed. Phosphorylated (P), di-glycine (GG, Ub-remnant), and acetylated (Ac) peptides were enriched sequentially. The whole proteome (W, flow-through containing non-modified peptides) and phosphoproteome were fractionated using bRP STAGE tips.*

To determine the turnover of modified peptides, four replicates of a time-course pSILAC-TMT experiment were performed exactly as described in chapter IV (pp. 105). This time, however, an enrichment of phosphorylated and acetylated peptides was included after TMT labelling (Figure V-1 A), and phosphorylated peptides were fractionated into 6 bRP tip fractions. In an alternative approach, 4 HeLa cell culture replicates of 4 single SILAC pulses (1, 6, 24, 40 h) were prepared meanwhile switching two replicates from K0R0 to K8R10 and the other two replicates from K8R10 to K0R0 (Figure V-1 B). Following protein extraction and digestion, a serial enrichment of phospho,

acetyl, and ubiquitin-remnant moieties was conducted. The flow-through containing the non-modified peptidoforms was fractionated via bRP STAGE tips into 6 fractions, whereas the phosphoproteome was separated into 4 fractions. No TMT labelling was performed.

Pulsed SILAC-TMT samples were measured, searched, and processed as described in chapter IV (pp. 103) with following modifications and additions: One-third of the phosphoproteome and half of the acetyl-peptide enrichment were each injected twice for MS measurement. MS2 spectra of phosphopeptides were acquired after HCD fragmentation using an NCE of 33 % and a cycle time of 2 s, whereas acetyl-peptides were fragmented with CID at 35 % NCE in a top10 method. Raw data were searched using MaxQuant v1.6.0.16 adding phosphorylation on serine, threonine, or tyrosine as variable modification for the phosphoproteomes and acetylation on K0 or K8 for acetylomes. Mass tolerance for ion trap spectra was set to 0.4 Da. The KRAB filter was adjusted (K: 0-5; $R^2 \geq 0.7$; A: 0.7-1.4; B: -0.15-0.25) to improve correlations and CVs of turnover rates across replicates and label incorporation and loss curves (Appendix Figure 0-16 A). Additionally, ratio compression was computationally reduced before normalization, and afterwards peptide intensities were corrected for abundance changes (see pp. 48, and pp. 135 for details).

Single pulse time-points were measured on a Q-Exactive HF-X mass spectrometer injecting amounts corresponding to 1 µg of whole proteome digest, one-fifth of the phosphopeptide enrichment, and half of the acetyl and di-glycine IPs. The latter were injected twice. Peptides were separated using a 50 min linear gradient from 4 to 32 % solvent B, except for phospho-peptides which were separated in a two-step gradient from 2 to 15 to 27 % solvent B. Up to 25, 15, and 12 precursors were selected for HCD fragmentation (26 % NCE) for non-modified, phosphorylated, and acetylated/di-glycine peptides, respectively. MS2 spectra were recorded at 15K resolution using an isolation window of 1.3 m/z, an AGC target value of 1e5 (2e5 for phosphopeptides), and a maxIT of 100 ms (22 for non-modified peptides). Raw data of single pulse samples were searched against the human Swiss-Prot database using MaxQuant v1.6.0.16. Lys0/Arg0 and Lys8/Arg10 were specified as metabolic labels, and phosphorylation on serine, threonine, and tyrosine, or acetylation or di-glycine on lysine were allowed as variable modifications for corresponding subproteome enrichments. Ratios of new-to-old (N/O) peptides and proteins, and corresponding turnover rates were computed as described in chapter II (pp. 52).

For counterpart analyses, non-modified peptides were mapped to modified sites on site not peptide sequence level for all PTMs, since acetylation and ubiquitination of lysine residues leads to missed cleavages. Hence, a peptide was defined as counterpart as long as it contained the unmodified amino acid that was identified with a modification in another peptide. Modifications were not filtered for localization probability but always automatically assigned to the most likely site. Student's t-tests were conducted separately for different datasets using turnover rates for pSILAC-TMT data and N/O ratios for single SILAC pulse experiments. Afterwards, significant hits were combined for functional enrichment and motif analyses as described in the results section (pp. 138).

# 3 Results

## 3.1 Two approaches for the investigation of PTM-specific turnover

For determination of the turnover of modified peptidoforms, two different approaches were pursued. For both, only acetylated, phosphorylated, ubiquitinated (i.e. di-glycine containing) and non-modified sequences were counted as distinct peptides (peptidoforms), whereas quantitative information of oxidized peptide sequences was integrated with non-oxidized versions.

**Pulsed SILAC-TMT approach and refinement of data processing** – In the first approach, the previously established pulsed SILAC-TMT method (see chapter III, Figure IV-3) was combined with an enrichment of phosphorylated and acetylated peptides to obtain their time-resolved label incorporation and loss behaviour. After data normalization, curve fitting and filtering, a disproportional discard of curves for modified peptidoforms was observed. While 53 % of evidence entries for non-modified peptides passed the KRAB-filter, only 17 % (ac-K) to 49 % (ph-S) of evidence entries mapping to modified peptides resulted in successful curve fitting (Figure V-2 A). It was assumed that this was at least partly caused by the, on average, higher ratio compression particularly for acetylated and tyrosine-phosphorylated peptides. Only 55, 48, and 45 % of evidence entries for N-terminal acetylated, lysine acetylated, and tyrosine-phosphorylated peptides, respectively, featured a ratio compression below 10 % compared to 62 % of entries for non-modified peptides (Figure V-2 B). To correct for this bias, ratio compression was computationally tackled assuming that it originated from co-isolation and fragmentation of peptides with respective opposing quantitative characteristics. More precisely, for each peptide that exhibited a label loss (or incorporation), the average label incorporation (or loss) curve was normalized to the apparent co-isolated intensity in the last (or first) TMT channel and then subtracted from the individual peptide intensities (Appendix Figure 0-16 B). In addition, it was investigated to which extent the premise of a constant abundance of peptides during the pulse time-course was fulfilled. To this end, abundances of peptides that were identified in both labelling states were assessed by harnessing the sum of their fractional MS1 intensities for each time-point. Although an ANOVA at 1 % FDR did not identify any peptides that changed significantly in abundance across time-points, partly large fold changes (FCs) were observed with 41 % of peptides showing at least a 2-fold difference between any two of the time-points. Especially modified peptidoforms were affected by these variations, for instance, 64 to 66 % of serine-phosphorylated and lysine-acetylated peptidoforms displayed at least a 2-fold change (Figure V-2 C). Since fluctuations in abundance can alter estimated turnover rates, an adjustment for such abundance variations was performed to improve accordance between label loss and incorporation curves (Appendix Figure 0-16 C). In brief, the knowledge of a relative abundance increase (or decrease) was utilized to adjust ratios, to which turnover curves were fitted, accordingly downwards (or upwards). Importantly, curves were corrected based on median abundance changes across replicates, and only turnover curves for which quantitative peptide information on both labelling states was available could be adjusted in this way (41 % of peptides).

After these refinements in processing of pulsed SILAC-TMT data, the number of successful curve fits increased by 17 %, most considerably for evidence entries identifying acetylated peptides (46 % increase, Figure V-2 D). Moreover, the correlations of turnover rates across replicates improved from an average Pearson correlation coefficient of 0.65 to 0.71 (Figure V-2 E), which was mainly a result of enhanced correlations of rates derived from label incorporation and loss curves (improvement from R=0.59 to 0.67, Appendix Figure 0-16 A). While the increase in the overall

number of distinct peptides for which a turnover rate could be determined was due to the *in silico* removal of ratio compression, abundance corrections mainly enhanced the correlation across replicates thereby improving the robustness of rate estimations (Figure V-2 F). In the end, nearly 82,000 peptidoforms (17,418 modified ones) assigned to 8,430 protein groups were available in the pulsed SILAC-TMT dataset for the assessment of modification-specific turnover.

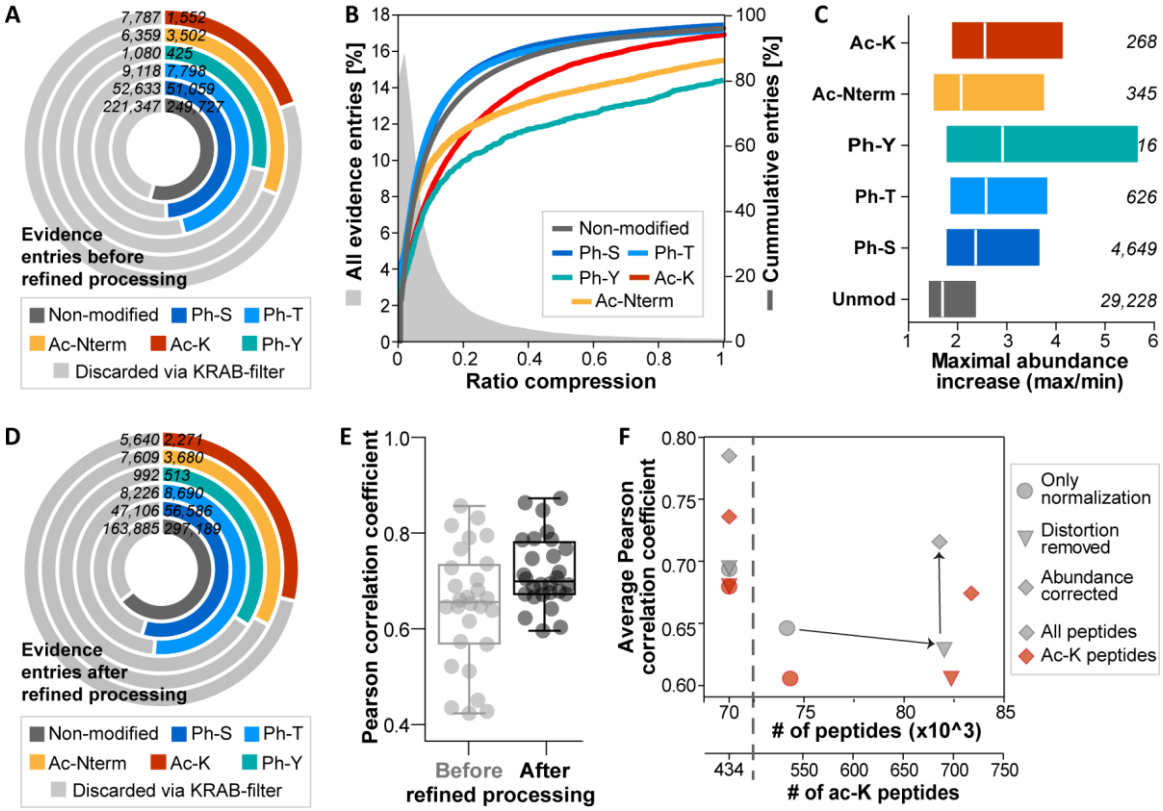



**Figure V-2 | Refinement of pulsed SILAC-TMT data processing.** *(A) The doughnut plot illustrates the fraction of evidence entries identifying peptides with various modifications and passing the KRAB-filter before data processing was revised (see also Appendix Figure 0-16). Numbers of evidence entries are indicated. (B) Ratio compression of unfiltered evidence entries was estimated based on intensities in the outermost channels that are supposed to be empty and is displayed for the all evidence entries (area) and separately for those matching to different modified and non-modified peptidoforms (lines). (C) Boxplots (25$^{th}$-50$^{th}$-75$^{th}$ percentile) show the maximal fold increase of peptidoform intensities (average of replicates) across the course of the pulse time-series. Numbers of included peptides are indicated on the right. (D) Same as in (A) is illustrated but after refinement of data processing including computational removal of ratio compression and abundance correction. (E) Pearson correlation coefficients of turnover rates obtained from label incorporation and loss curves of the four cell culture replicates are plotted before and after refined data processing. (F) Average Pearson correlation coefficients across replicates and the number of peptides for which a turnover rate could be determined are indicated for different refinement steps of the data processing procedure. The correlation coefficients on the left are obtained for peptides shared between different processing steps.*

**Single SILAC pulse time-points** – In the second approach, phosphorylated, ubiquitin-remnant-containing, and acetylated peptides were enriched from individual SILAC pulse experiments covering four time-points (1, 6, 24, and 40 h) with no additional TMT-labelling step. Instead of performing curve fitting to estimate absolute turnover rates, determined ratios of newly synthesized to older peptides (N/O ratios) were utilized as a proxy for the relative turnover of modified and non-modified peptidoforms within a certain pulse time-point. As expected, different pulses yielded globally shifted N/O ratios that were larger for later time-points (median of 0.1, 0.2, 1.1, and 1.8 for 1, 6, 24, and 40 h, Figure V-3 A). The number of peptidoforms quantified in both

labelling states was significantly lower for the 1 h pulse (22,891) compared to the other three pulse time-points (average of ~59,400) due to the overall small fraction of label incorporation and the challenging determination of very large ratios (Figure V-3 B). Different pulse time-points also tended to quantify more distinct peptidoforms than replicates of the identical time-point. Overall, N/O ratios of replicates from the same pulse correlated well with few exceptions (Figure V-3 C). Noteworthy, replicates that started with cell culture medium containing the same isotopically labelled amino acids featured a superior correlation of N/O peptide ratios (average R of 0.80) compared to replicates that started with different cell culture media (average R of 0.53, Figure V-3 C). This was especially apparent for the 1 h time-point where the poor correlation of label swap pairs (average R of 0.32) suggested that quantification of large ratios was considerably less reproducible and thus less accurate than if light and heavy intensities were more similar as it was the case for other pulse time-points. Interestingly, samples in which the light version of peptides showed on average higher intensities than the heavy versions consistently displayed the highest correlation of N/O ratios (K0R0-to-K8R10 switch for 1 and 6 h pulse, K8R10-to-K0R0 switch for 24 and 40 h pulse). The reason for this phenomenon remains ambiguous. In total, single pulse experiments provided N/O ratios for about 79,500 peptides (24,560 modified ones) mapping to 7,259 protein groups in any of the four time-points.

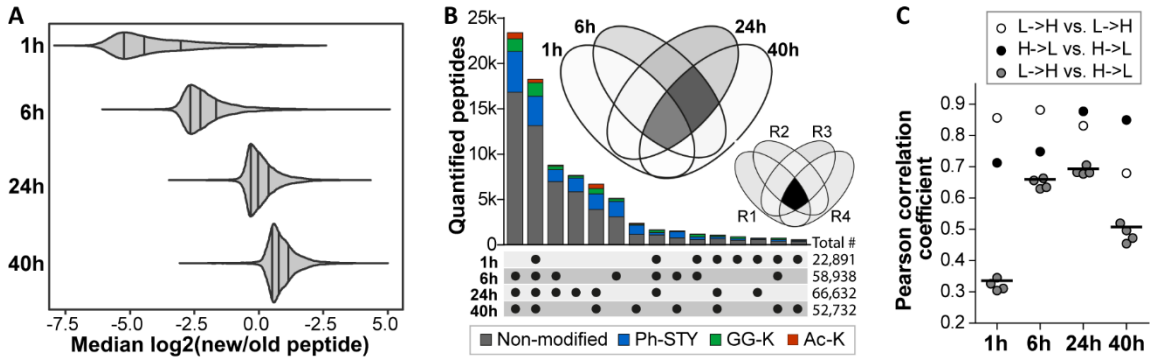



**Figure V-3 | Characteristics of the single time-point pulse dataset.** *(A) Violin plots illustrate distributions of N/O peptide ratios for increasing time-points. (B) The bar chart and the major Venn diagram display the number and overlaps of distinct peptidoforms with an N/O ratio for the different pulse time-points. Oxidized peptide versions were not counted as distinct. The shading of the Venn diagram reflects the degree of overlap across datasets. The smaller Venn diagram shows the overlaps across replicates of the 24 h pulse time-point as a comparison. (C) Pearson correlation coefficients of N/O ratios are plotted for different pulse time-points indicating whether they were derived from the comparison of replicates that start the pulse with the identical light (L = K0R0) or heavy (H = K8R10) cell culture medium (white and black circles) or with different cell culture media (grey circles, label swap correlations).*

**Congruence of both approaches** – When combining results from both experimental strategies, information of turnover was available for more than 121,000 peptidoforms assigned to more than 10,000 protein groups, and one third of peptides was quantified by both approaches (Figure V-4 A). This included 34,116 modified peptides covering 30,499 modification sites of which on fifth was shared between the pSILAC-TMT and the single pulse datasets (Figure V-4 B). In an attempt to integrate all data, turnover rates were computed for single SILAC pulse experiments using N/O ratios as described in Chapter II (pp. 52). Amino acid recycling estimated from pSILAC-TMT data (Appendix Figure 0-17 A) was factored in for rate computations since it improved coefficients of variation across pulse replicates especially for 1 h time-point (median of 0.55 vs. 0.30 without accounting for recycling). As already suggested from correlation analyses, the CV distribution for the 1 h pulse data was nevertheless broader as the ones of other pulse time-points (75$^{th}$ percentile of 0.37 vs. 0.14-0.17, Figure V-4 C). Remarkably, CVs increased further when turnover rates of all

single pulses were combined (median of 0.27 vs. 0.09-0.10), even if the 1 h pulse was omitted (median of 0.19). This indicated that N/O ratios from different pulses generally yielded somewhat diverging turnover rates, which was indeed corroborated by globally shifted rate distributions (0.22, 0.046, 0.033, 0.028 for 1 to 40 h, Appendix Figure 0-17 B). Similarly, turnover rate distributions of combined single SILAC pulses were slightly narrower and shifted towards smaller turnover rates in comparison to pulsed SILAC-TMT data (median of 0.034 vs. 0.038, Figure V-4 D), and a combination of both approaches somewhat increased CVs (Figure V-4 C). Hence, it was decided to analyse the datasets separately using N/O ratios for single SILAC pulses and turnover rates for pulsed SILAC-TMT data.

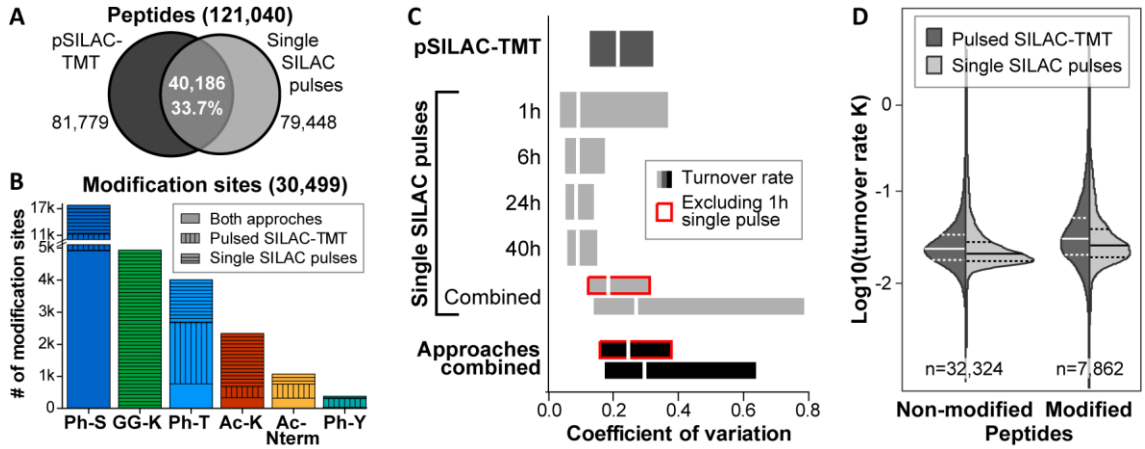
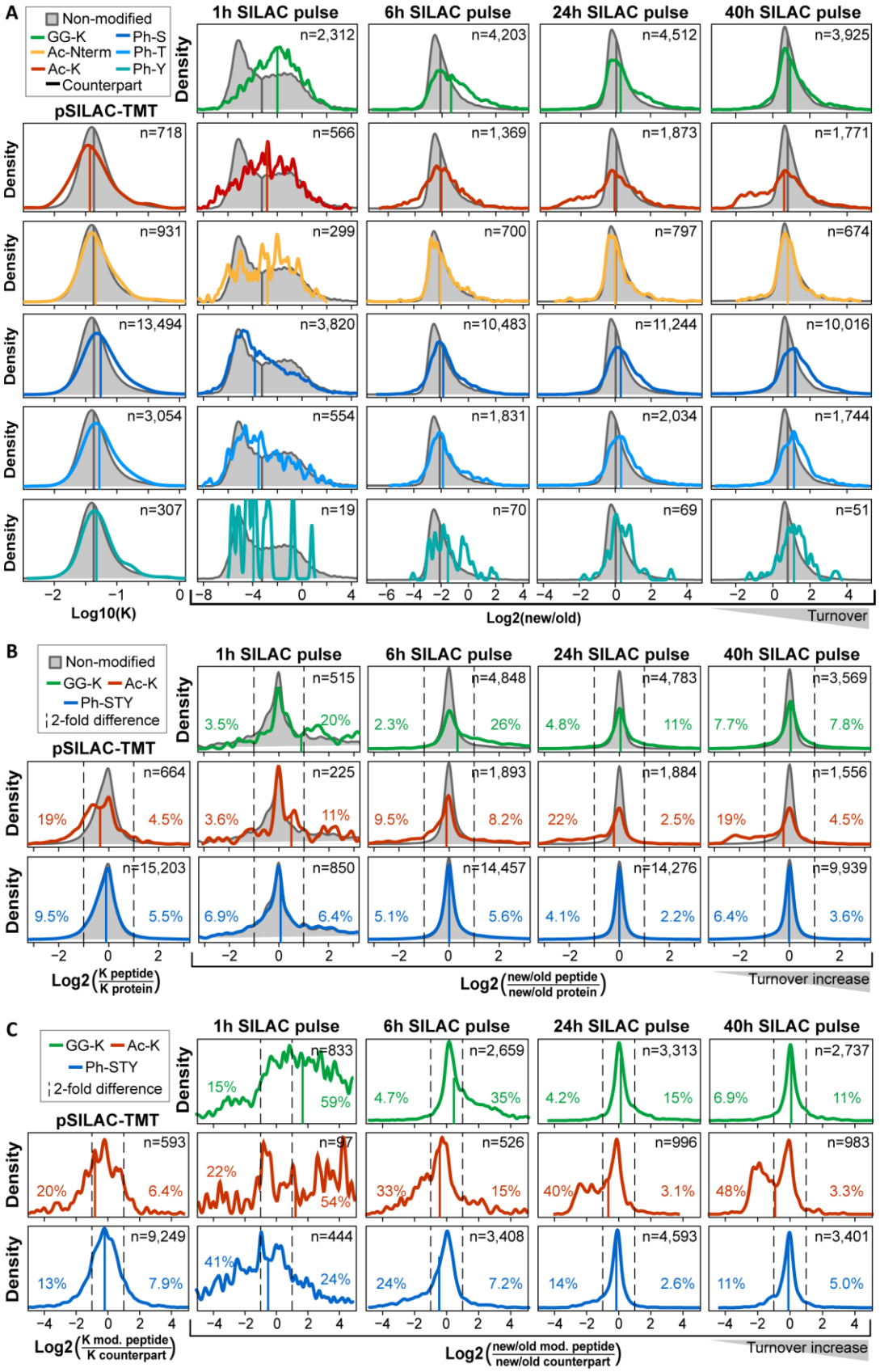



**Figure V-4 | Integration of data from the two different pulsed SILAC approaches.** *(A) Venn diagram and bar chart illustrate the number and overlaps of peptides (ignoring oxidized versions) and modification sites for which turnover behaviour could be quantified. (B) Boxplots (25$^{th}$-50$^{th}$-75$^{th}$ percentile) display distributions of coefficients of variation for N/O peptide ratios and estimated turnover rates for single datasets and combined data. (C) Violin plots reveal global differences in turnover rates between all modified (including acetylated, ubiquitinated, and phosphorylated) and non-modified peptides and the two different pSILAC approaches. Only peptides that were shared between the two different methods were included.*

## 3.2 Characterization of peptidoforms with differential turnover

A global comparison of post-translationally modified and non-modified peptides revealed a generally broader distribution of turnover rates for modified peptidoforms with medians shifted towards higher turnover indicating an overall decreased stability (e.g. 0.048 vs. 0.038 for pSILAC-TMT data, Figure V-4 D). To investigate modification-specific turnover in more detail, turnover rate and N/O ratio distributions of peptides with different modification types were examined separately (Figure V-5 A). Further, turnover of modified peptides was set in relation to their corresponding proteins (Figure V-5 B) or non-modified counterpart peptides (Figure V-5 C) to reveal protein-specific, stabilizing or destabilizing effects of post-translational modifications.

**Figure V-5 | Global differences in turnover of modified peptidoforms.** *Medians of cell culture replicates are plotted. (A) Turnover rate and N/O ratio distributions are shown for modified peptides in comparison to non-modified, non-counterpart peptides. Counterpart peptides include all peptides that encompass an unmodified amino acid that was found in a modified state in any of the acquired datasets. Vertical, coloured lines mark the medians of respective distributions. (B) Ratios of turnover rates or N/O fractions of modified peptides to their corresponding protein are displayed. For computation of protein turnover rates, at least 3 peptides were required including all peptides irrespective of their modification state. (C) Ratios of turnover rates or N/O fractions of modified peptides to their non-modified counterpart peptide are plotted. Coloured numbers indicate the fraction of peptides that featured a 2-fold slower or faster turnover compared to the protein or counterpart peptide.*




**Global differences in the turnover of modified peptidoforms** – The observed, global difference of modified and non-modified peptides was mainly driven by the large fraction of phosphorylated peptides. They featured, on average, a higher turnover as non-modified peptides

irrespective of the modified amino acid (STY) and for both experimental approaches (Figure V-5 A). Strikingly, this clear shift vanished when the turnover ratios of phosphorylated peptides to their corresponding proteins or counterpart peptides were examined (Figure V-5 B, C). This suggests that phosphorylation was predominantly detected on high-turnover proteins. While ratio distributions had medians close to zero, they exhibited a wide range, which indicated that still many phosphorylation events could alter measured protein turnover. Especially the comparison to non-modified counterpart peptides implied that the majority of such regulating phosphosites increased turnover (62-84 % of peptides that showed a ≥ 2-fold turnover difference, Figure V-5 C).

Comparable to phosphorylation, ubiquitin-remnant peptides displayed a global shift towards higher turnover in comparison to non-modified peptides (Figure V-5 A). This destabilizing effect had been expected based on the well-established involvement of ubiquitination in proteasomal protein degradation. However, the observed shift decreased with increasing pulse time and was not apparent anymore for the 40 h time-point. A large fraction of 20 to 60 % of di-glycine peptides showed an at least doubled turnover compared to the corresponding protein or counterpart peptide in the 1 and 6 h time-point, whereas this fraction dropped to 7.8 to 15 % for the 24 and 40 h pulses (Figure V-5 B, C). After longer pulse times, up to 7.7 % of di-glycine peptides even exhibited an at least 2-fold decrease in turnover compared to their protein supporting the notion that the functional relevance of ubiquitination transcends its role as degradation signal.

N-terminal acetylation resulted in only minor differences in overall turnover compared to non-modified peptides. For lysine acetylation, however, a globally decelerated turnover compared to non-modified peptides was observed, which was especially apparent in pSILAC-TMT data and after longer single SILAC pulses (Figure V-5 A). A comparison of the turnover of acetyl-lysine peptides and their assigned proteins or non-modified counterparts equally demonstrated their vastly slower turnover with almost half of them showing more than 2-fold decreased turnover in the 40 h pulse samples (Figure V-5 B, C). Noteworthy, also lysine acetylated peptides with more than a doubled turnover were detected, but mostly at a much lower frequency (2.5-15 %) compared to ≥ 2-fold stabilizing acetyl-lysine peptides (9.5-48 %). Only for the comparison to counterpart peptides in the 1 h pulse samples, these higher-turnover peptides accounted for the majority of acetyl-peptides (54 %). Noteworthy, this may also be an artefact of the partially poor quantitative performance observed for 1 h pulse replicates (Figure V-3 C, Figure ***V-4*** C). Therefore, statistical test were performed as a next step to account for replicate variance and prioritize modification sites that affect turnover with high confidence.

**Identification of peptides with significantly different turnover** – For each of the 5 datasets, significance of turnover differences between peptides and their corresponding proteins or counterparts was assessed separately. To this end, at least two turnover rates or N/O ratios were required in each group for any statistical test. Further, only peptides that mapped to proteins with at least three quantified peptides were considered in the comparisons to proteins. Hence, 64-77 % of all quantified peptides were included for the peptide-to-protein tests except for the 1 h SILAC pulse where only 5 % matched all required criteria (Appendix Figure 0-18 A). Naturally, proportions of quantified peptides used for the comparison of modified and counterpart peptides were smaller (3-7 % of all peptides, corresponding to 21-30 % of modified peptidoforms, Appendix Figure 0-18 A). All possible peptide pairs were allowed in the analyses resulting in several pairwise comparison for modified peptides for which more than one non-modified counterpart peptide was quantified (applied to 17-28 % of tested modified peptides). Statistical tests lead to differing fractions of significant hits ranging from 0 (for peptide-to-protein tests in the 1 h and 40 h pulse samples) to a quarter of performed pairwise comparisons (for modified peptide-to-counterpart

test in the 6 h replicates, Appendix Figure 0-18 B). The pulsed SILAC-TMT dataset yielded the highest number of peptides with significant different turnover to their assigned protein (6,697), whereas the 6 h SILAC pulse identified most modified peptides with significant differences in stability compared to their counterpart peptides (913). Collectively, 9,940 peptides assigned to 4,874 proteins scored in at least one of the statistical tests, which corresponded to ~11 % of tested peptides.

**Enriched protein functions among significant peptidoforms** – To set observed turnover differences in context with known functional annotations of peptidoforms and corresponding proteins, significant hits from all analyses were combined and subjected to Fisher's exact tests. To this end, turnover of a peptide was defined as significantly slower or faster when at least one of the statistical tests identified it to be significantly slower or faster. Peptides that were found in conflicting groups (faster and slower) in different datasets were discarded before enrichment analysis (50 for the peptide-to-protein and 18 for the modified peptide-to-counterpart comparison, corresponding to <1 % of significant hits, Appendix Figure 0-18 C).

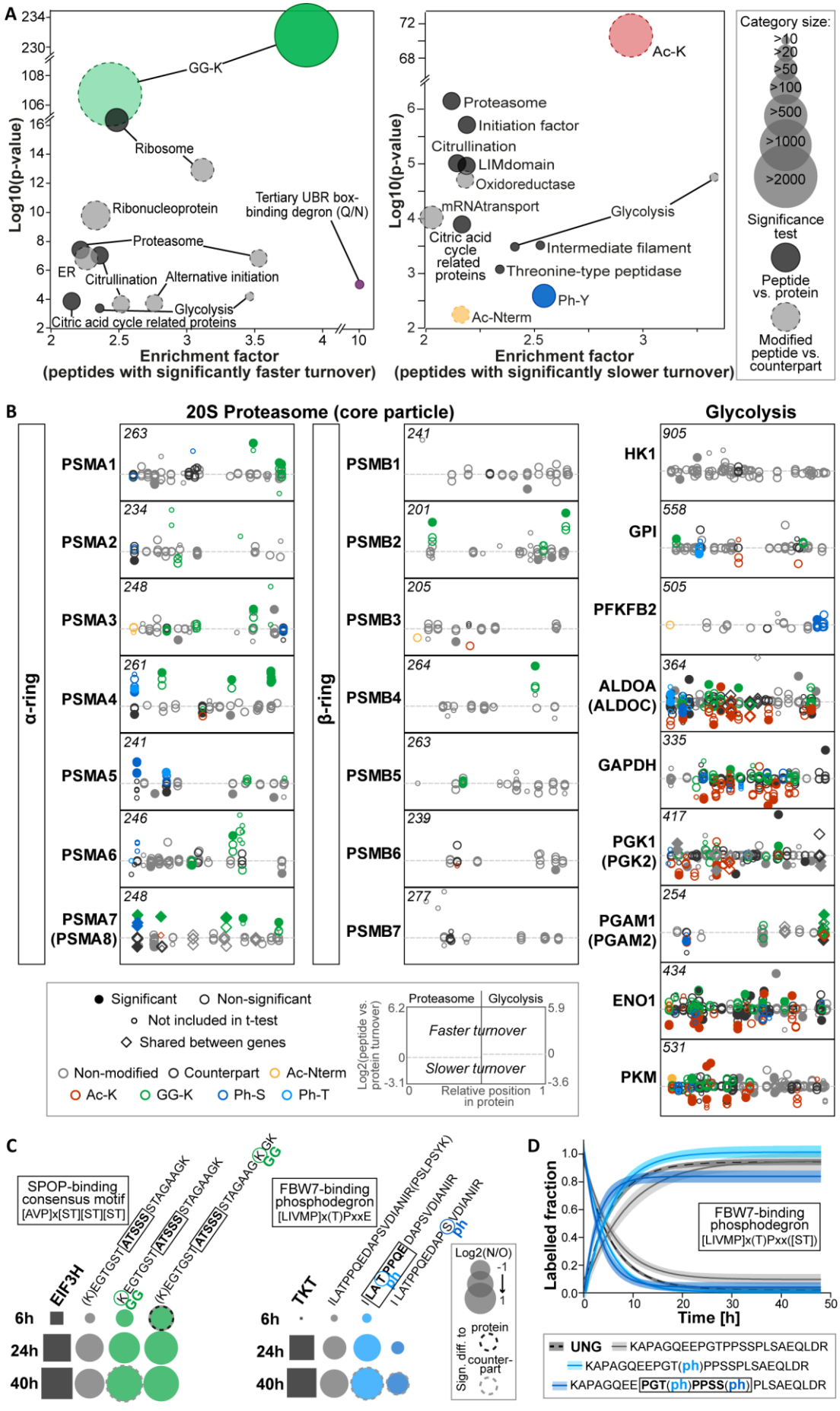
Fisher's exact tests using UniProt keywords, KEGG pathways, and HPA protein classes identified multiple categories that were enriched within peptides with both, significantly slower and faster turnover (Figure V-6 A). These included most notably annotations of large multi-protein complexes such as the proteasome, the spliceosome, and the nuclear pore complex, and of metabolic pathways like glycolysis and the citrate cycle (enrichment factors of >2.1, >1.6, >1.8, >2.4 and >2.1 for peptides with significantly different turnover compared to their protein). A more detailed examination of α-, β-, and regulatory subunits of the 26S proteasome revealed 153 of 895 quantified peptides with significantly different turnover corresponding to ~20 % of the 783 proteasomal peptides subjected to statistical testing. The majority of comparably faster turned-over peptides was post-translationally modified (72 of 98, mainly ubiquitinated), while most stabilized peptides accounted for non-modified peptidoforms (54 of 64). Regulated peptides were predominantly detected within subunits of the α-ring (PSMDA1-8, Figure V-6 B) and the base of the proteasome (ATPase subunits PSMC1-6 and non-ATPase subunits PSMD1-2, Appendix Figure 0-19). The high proportion of differentially turned-over and typically modified peptidoforms may hint at a hitherto under-appreciated, high degree of post-translational regulation of proteasome activity. Likewise, >21 % of statistically tested peptides mapping to proteins associated with the UniProt Keyword 'Glycolysis' exhibited a significantly different turnover to the protein or counterpart peptide. Approximately half of these regulated peptides carried a modification and 2/3$^{rd}$ featured a slower turnover. Notably, Fructose-bisphosphate Aldolase A (ALDOA), Phosphoglycerate Kinase 1 (PGK1), α-Enolase (ENO1), Glycerinaldehyde-3-phosphate Dehydrogenase (GAPDH), and Pyruvate Kinase (PKM) yielded the highest faction of modified and differentially turned-over peptides (Figure V-6 B) insinuating that their stability and activity may be strongly controlled by PTMs. Contrarily, only one phosphorylated peptide with significantly higher turnover was identified for the rate-limiting enzyme of glycolysis, 6-Phosphofructo-2-kinase (PFKFB2, pS466, log2 FC of 0.7), whose activity is regulated by pH and multiple allosteric activators and inhibitors (e.g. AMP/ATP, acetyl-CoA, and its product fructose-1,2-bisphosphate). Surprisingly, none of the modification sites on peptides mapping to the proteasomal subunits or glycolytic enzymes and featuring a significantly different turnover was annotated with a known effect in the PTM database PhosphoSitePlus. The differential turnover properties of these modified peptides yet suggest some regulatory relevance of the corresponding PTM sites. Their absent functional annotations demonstrate the knowledge gap between identification of modification sites and the characterization of their role in cellular processes.

Noteworthy, the UniProt keyword citrullination was also enriched among stabilized and destabilized peptidoforms (Figure V-6 A), which may hint at a general role of this modification in turnover regulation. To further investigate this hypothesis, the 40 proteins in this category containing for example histones, ribosomal proteins and transcription factors were searched for differentially turnover of non-modified peptides that encompass or (in case of trypsin cleavage) succeed potentially deiminated arginine residues annotated in UniProt. However, only 2 of 16 of such peptides showed a decreased turnover and at the same time did not carry any other modification that likely may impact turnover (peptides following potentially citrullinated TRIM28 residue 472 and CCDC86 residue 342 with log2 FCs of -0.7 and -1.4 in pSILAC-TMT data). Hence, the enrichment of this category was not directly mediated by citrullination itself, and a convincing global association between this modification and turnover needs yet to be established with further experiments.

Additional protein categories that were solely overrepresented among peptides with slower turnover included 'intermediate filament' and 'LIM domain', which both largely comprised structural proteins. For both protein classes, protein-protein interactions play an important role for protein function and may be regulated via modifications. Intermediate filaments form coiled-coil dimers and LIM domain-containing proteins can bind other proteins via their two contiguous zinc finger domains. Indeed, several phosphopeptides containing sites that have been linked to a molecular association function or intracellular localization were identified among stabilizing peptidoforms within these protein classes. As an example, Lamin A peptides spanning pS390, pS392, or both phosphorylation sites exhibited a significantly decreased turnover compared to the whole protein and counterpart peptide in the pSILAC-TMT and 6 h single pulse datasets, respectively (log2 FCs of -0.7 to -1.2 and -1.1 to -1.2). Phosphorylation on these residues has been described to induce disassembly of the nuclear lamina [414, 415]. It may be counterintuitive that phosphorylated monomers exhibit a slower turnover than non-phosphorylated dimers, but this finding could also indicate additional functionalities of pS390 and pS392 in Lamin A (see also discussion pp. 153).

Another category emerging from Fisher's exact tests contained translation initiation factors which showed a 2.2- and 1.9-fold enrichment among peptides with slower and higher turnover, respectively (Figure V-6 A). Roughly 15 % (163) of peptides mapping to 41 initiation factors and included in t-tests featured significantly different turnover characteristics and one third of these were modified peptidoforms (mostly ubiquitinated for destabilized and phosphorylated for stabilized peptides). Interestingly, several regulated, modified and non-modified peptides carried potential degron motifs that are listed in the ELM database [311] but have so far not been associated with

**Figure V-6 | Traits and functional annotations of assigned proteins for peptides with significantly different turnover.** *(A) Plots illustrate results of Fisher's exact tests at 1 % Benjamini-Hochberg FDR using peptides that were identified to feature significantly faster (left side) or slower (right side) turnover. The respective background was always defined by all other peptides included for t-test analyses as shown in Appendix* Figure 0-18 *B. Functional categories tested for significant enrichment included protein annotations (grey circles) from UniProt, KEGG, and HPA, and peptide traits (coloured circles) like modification state, localization within the protein, and degron sequences. Only more than 2-fold enriched categories are displayed. (B) The position of peptides within their corresponding proteins is plotted against the turnover difference to their proteins for 20S proteasome subunits and glycolytic enzymes that were enriched in both, significantly slower and faster degrading peptides. If peptides match to more than one gene, additional genes are indicated in brackets. The length of the major isoform is displayed in italics. (C) Average N/O ratios from replicates of single pulse time-points are illustrated for the Eukaryotic Translation Initiation Factor 3 subunit H (EIF3H) and Transketolase (TKT), and their modified and non-modified peptidoforms that comprise degron motifs. (D) Turnover curves are shown for Uracil-DNA Glycosylase (UNG) and its modified and non-modified peptidoforms that span a phosphodegron motif. Shaded areas indicate 95 % confidence intervals obtained from fitting to respective evidence entries from pSILAC-TMT quadruplicates.*


A
Log10(p-value)
Enrichment factor
(peptides with significantly faster turnover)
GG-K
Ribosome
Ribonucleoprotein
Tertiary UBR box-binding degron (Q/N)
Proteasome
ER
Citrullination
Alternative initiation
Glycolysis
Citric acid cycle related proteins
Enrichment factor
(peptides with significantly slower turnover)
Ac-K
Proteasome
Initiation factor
Citrullination
LIMdomain
Oxidoreductase
mRNAtransport
Glycolysis
Citric acid cycle related proteins
Intermediate filament
Threonine-type peptidase
Ph-Y
Ac-Nterm
Category size:
>10
>20
>50
>100
>500
>1000
>2000
Significance test
Peptide vs. protein
Modified peptide vs. counterpart
B
20S Proteasome (core particle)
Glycolysis
α-ring
β-ring
PSMA1
PSMA2
PSMA3
PSMA4
PSMA5
PSMA6
PSMA7 (PSMA8)
PSMB1
PSMB2
PSMB3
PSMB4
PSMB5
PSMB6
PSMB7
HK1
GPI
PFKFB2
ALDOA (ALDOC)
GAPDH
PGK1 (PGK2)
PGAM1 (PGAM2)
ENO1
PKM
Significant
Non-significant
Not included in t-test
Shared between genes
Non-modified
Counterpart
Ac-Nterm
Ac-K
GG-K
Ph-S
Ph-T
Log2(peptide vs. protein turnover)
Proteasome
Glycolysis
Faster turnover
Slower turnover
Relative position in protein
C
SPOP-binding consensus motif
[AVP]x[ST][ST][ST]
EIF3H
(K)EGTGST[ATSSS]STAGAAGK
6h
24h
40h
FBW7-binding phosphodegron
[LIVMP]x(T)PxxE
TKT
ILATPPQEDAPSVDIANIR(PSLPSYK)
[LATPPQE]DAPSVDIANIR
ILATPPQEDAPSVDIANIR
Log2(N/O)
Sign. diff. to
protein
counter-part
D
Labelled fraction
Time [h]
FBW7-binding phosphodegron
[LIVMP]x(T)Pxx([ST])
UNG
KAPAGQEEPGTPPSSPLSAEQLDR
KAPAGQEEPGT(ph)PPSSPLSAEQLDR
KAPAGQEE[PGT(ph)PPSS(ph)]PLSAEQLDR

the respective translation initiation factors. The Eukaryotic Translation Initiation Factor 3 subunit H (EIF3H), for instance, contains an S/T-rich, SPOP (speckle-type POZ protein)-binding consensus motif, and three ubiquitin-remnant-containing peptidoforms spanning this degron showed increased turnover (log2 FC of 0.5 to 1.4 to protein and 1.0 to 1.9 to counterpart, Figure V-6 C). In contrast, the non-modified versions exhibited a comparably slower turnover (e.g. log2 FC of -0.7 and -0.8 in pSILAC-TMT and 40 h pulse data). This may indicate that a high fraction of expressed EIF3H is by default targeted for degradation by the SPOP/Cullin-3 E3 ligase complex. Likewise, another non-modified peptide of EIF2A with a similar SPOP-binding motif (NTVLATWQ***PYTTS***K) featured a significantly slower turnover in the pSILAC-TMT dataset (log2 FC of -0.9). However, no ubiquitinated peptidoform, which would corroborate the notion of a functional SPOP/Cullin-3 degron, was detected. Further, EIF3B carries a potential, tertiary destabilizing N-degron (glutamine in second position [117]), and its unprocessed N-terminal peptide identified in pSILAC-TMT experiments showed a significantly higher turnover compared to the protein (log2 FC of 2.5). Contrarily, the N-terminal acetylated version detected in 24 h and 40 h pulse data showed no significant turnover difference to the protein. This is in line with a rapid and irreversible modification of EIF3B's N-terminus, either by N-terminal acetylation to protect the protein from degradation or by cleavage of the initiator methionine and further processing (i.e. deamidation and arginylation) of the glutamine residue to facilitate ultimately binding of N-recognins via their UBR-box. This results in EIF3B ubiquitination and subsequent degradation.

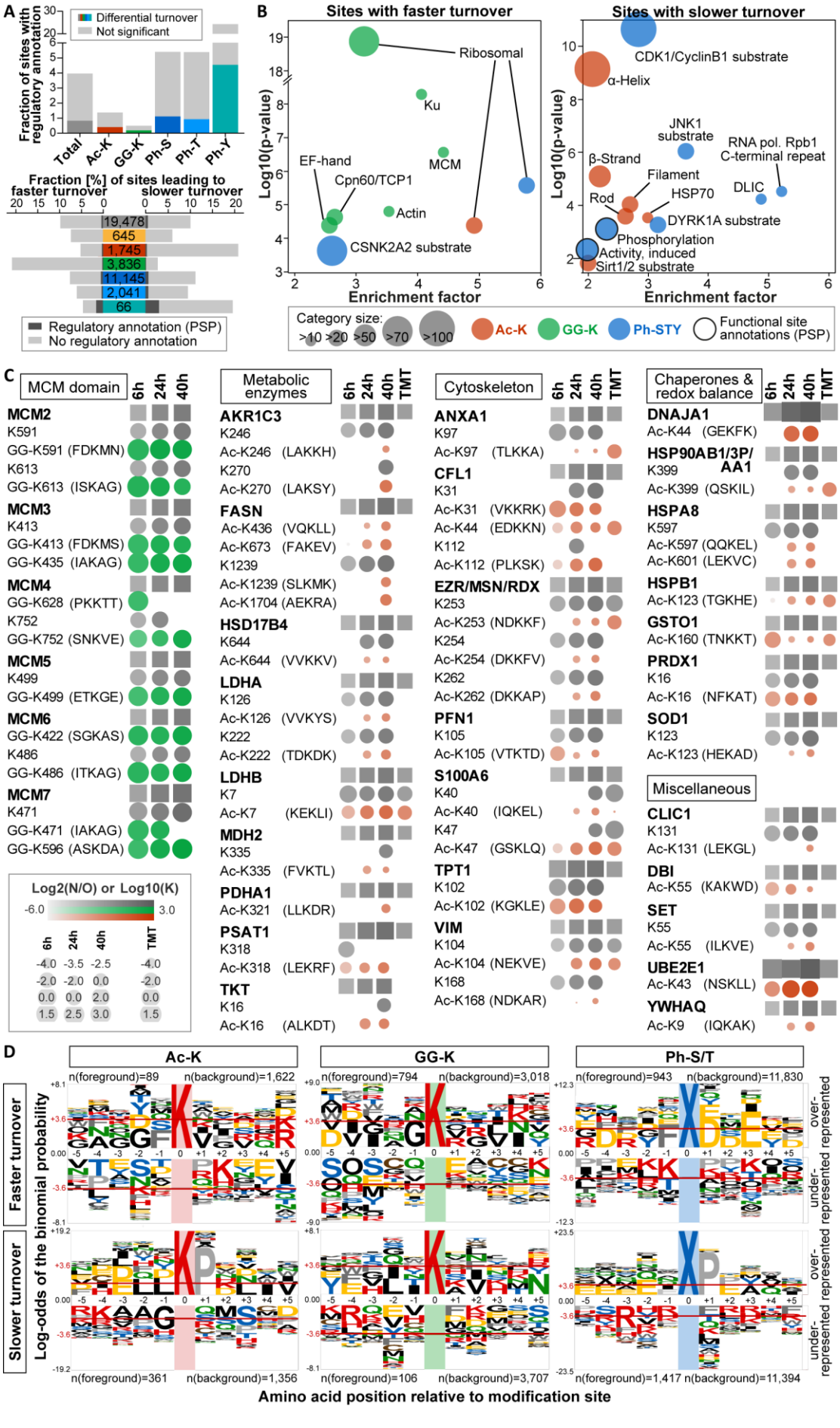
**Enriched peptide traits among significant peptidoforms** – Besides protein annotations, peptide characteristics like modification status, localization within the protein, and potential degron sequences were analysed for enrichment among peptides with significantly faster or slower turnover. UBR-box degrons with tertiary destabilizing glutamine or asparagine residues generally were enriched by a factor of 10 within destabilized peptides (Figure V-6 A). In addition to EIF3B, this was driven by Filamin C and Ubiquitin itself, which also featured faster turned-over N-terminal peptides with and without the initiator methionine (e.g. log2 FC of 3.4 and 2.2 in 24 h pulse samples). Of note, no other degron category was overrepresented among destabilized peptidoforms. Nevertheless, 180 peptides assigned to 154 protein groups and comprising a potential degron sequence exhibited significantly faster turnover compared to the protein or counterpart peptide. Among those, a doubly phosphorylated sequence of Uracil-DNA Glycosylase (UNG) was identified. The peptide with phosphorylated threonine 60 and serine 64 mapped to the consensus sequence of a FBW7 (F-box and WD repeat domain-containing 7)-binding phosphodegron and was significantly less stable compared to the non-modified counterpart peptide and the protein in pSILAC-TMT data (log2 FC of 1.3 and 1.0, Figure V-6 D). This observation substantiates a previous report that suggested that the ph-T60-ph-S64 pair functions as a phosphodegron in UNG [416]. The mono-phosphorylated peptidoform which only contained ph-T60 also exhibited a higher turnover (log2 FC of 0.6 and 0.3), but the difference in turnover rates was not significant. In addition to UNG, Transketolase (TKT), an enzyme of the pentose phosphate metabolic pathway, was found to carry a conditional FBW7-binding phosphodegron, which has so far not been characterised. The ph-T287-containing peptide showed a significantly decreased stability compared to non-modified counterpart peptides in the 40 h pulse experiment (average log2 FC of 0.8, Figure V-6 C). Interestingly, another peptide with a phosphorylated serine (pS295), which was located 8 residues downstream of the threonine that is phosphorylated to form the phosphodegron, was significantly stabilized (average log2 FC of -1.1). This may indicate an inhibitory role of this modification site towards the formation and activity of the phosphodegron. Regarding modification status, Ubiquitin-remnant peptides were clearly enriched within peptides with higher turnover, whereas lysine

acetylated, N-terminal acetylated, and tyrosine-phosphorylated peptidoforms were overrepresented among stabilized ones (Figure V-6 A). N-terminal acetylated and N- and C- terminal peptides, in general, also scored in Fisher's exact test using peptides with significantly higher turnover, and other phosphorylations and modification counterpart peptides were similarly enriched in higher- and slower-turnover peptidoforms, albeit overall to a lesser extent (enrichment factors of 1.2 to 1.5 in peptide-to-protein comparisons). Altogether, this illustrates a remarkably high degree of modification-specific turnover.

## 3.3 Functional perspective on modification-specific turnover

Modification-specific turnover was investigated in more detail in site- and motif-centric analyses. Importantly, in the following, site turnover will be referred to as the measured turnover of peptides carrying the respective modification and not the kinetics of the modification process itself. Analogous to the peptidoform enrichments, significant hits from all Student's t-tests were combined and sites were determined as influencing turnover significantly when any peptide encompassing this site (irrespective of other modifications in the peptide) was found to exhibit significantly different turnover in any of the statistical tests. This time, 120 of 19,600 statistically tested sites were identified on faster and slower turned-over peptides, mostly based on their assignment to different peptide sequences with a varying number of other modification sites (e.g. cases where mono- and di-phosphorylation on peptides showed opposing effects on turnover). These sites were discarded for further analyses resulting in 19,478 modified sites of which 13 % (for N-terminal acetylation) to 26 % (for ac-K) were identified to exhibit significantly different turnover compared to the non-modified site (Figure V-7 A). Only 4 % (773) of all sites were annotated to feature a regulatory function in the PSP database (for GG-/ac-K sites only 0.5 %/1.4 %). Noteworthy, this fraction increased within sites with significantly different turnover depending on the modification type by 11 % (phosphorylation) to 57 % (ubiquitination). Further, a fifth of all regulatory sites exhibited a differential turnover behaviour. Nevertheless, more than 96 % (3,708) of significant sites lacked any regulatory function in PSP illustrating the potential to spot hitherto unknown, functionally relevant modification sites conducting turnover analyses on peptidoform-level.

**Enriched sequence and site features among significant modifications** – PSP annotation together with UniProt sequence features, PFAM domains and data on writer and eraser enzymes [208, 309, 310] were utilized for site-centric enrichment analyses, which were performed for ac-K, GG-K, and ph-STY sites separately. Interestingly, modifications within ribosomal domains were over-represented among sites with higher turnover for all three modifications suggesting that PTMs may generally interfere with ribosomal function (also already indicated in the peptide-centric enrichment analyses, Figure V-6 A, Figure V-7 B). More precisely, 77 of 153 modified sites within ribosomal domains resulted in an accelerated protein turnover (9 ac-K, 10 ph-ST, and 58 GG-K sites), while only 4 modifications exhibited a stabilizing effect (RPL7A_acK150, RPL10A_acK91, RS11_acK45, MRTO4_pS80). For none of them functions were annotated in PSP. Likewise, no functions were listed for higher turnover di-glycine sites within additionally enriched PFAM categories. These included MCM domains of DNA replication licensing factors (MCM proteins) which control the initiation of DNA replication [417]. The high fraction of destabilizing Ubiquitin-remnants within the functional domain of all identified MCM proteins (11 of 12 detected GG-K sites, Figure V-7 C) indicates that ubiquitination of this C-terminal domain may be a common mechanism for MCM protein degradation. Destabilizing effects were also observed for most di-glycine sites on Ku domains of X-ray repair cross-complementing protein 5 and 6 (XRCC5/6; 16 of 19 GG-K sites, Appendix Figure 0-20), which form a DNA-binding heterodimer and are involved in


A
Differential turnover
Not significant
Fraction of sites with regulatory annotation
Total
Ac-K
GG-K
Ph-S
Ph-T
Ph-Y
Fraction [%] of sites leading to faster turnover
slower turnover
19,478
645
1,745
3,836
11,145
2,041
66
Regulatory annotation (PSP)
No regulatory annotation
B
Sites with faster turnover
Sites with slower turnover
Log10(p-value)
Enrichment factor
Ribosomal
Ku
MCM
EF-hand
Cpn60/TCP1
Actin
CSNK2A2 substrate
CDK1/CyclinB1 substrate
α-Helix
JNK1 substrate
RNA pol. Rpb1 C-terminal repeat
β-Strand
Filament
HSP70
DLIC
Rod
DYRK1A substrate
Phosphorylation Activity, induced Sirt1/2 substrate
Category size:
>10 >20 >50 >70 >100
Ac-K
GG-K
Ph-STY
Functional site annotations (PSP)
C
MCM domain
6h 24h 40h
MCM2
K591
GG-K591 (FDKMN)
K613
GG-K613 (ISKAG)
MCM3
K413
GG-K413 (FDKMS)
GG-K435 (IAKAG)
MCM4
GG-K628 (PKKTT)
K752
GG-K752 (SNKVE)
MCM5
K499
GG-K499 (ETKGE)
MCM6
GG-K422 (SGKAS)
K486
GG-K486 (ITKAG)
MCM7
K471
GG-K471 (IAKAG)
GG-K596 (ASKDA)
Log2(N/O) or Log10(K)
-6.0
3.0
6h 24h 40h TMT
-4.0 -3.5 -2.5 -4.0
-2.0 -2.0 0.0 -2.0
0.0 0.0 2.0 0.0
1.5 2.5 3.0 1.5
Metabolic enzymes
6h 24h 40h TMT
AKR1C3
K246
Ac-K246 (LAKKH)
K270
Ac-K270 (LAKSY)
FASN
Ac-K436 (VQKLL)
Ac-K673 (FAKEV)
K1239
Ac-K1239 (SLKMK)
Ac-K1704 (AEKRA)
HSD17B4
K644
Ac-K644 (VVKKV)
LDHA
K126
Ac-K126 (VVKYS)
K222
Ac-K222 (TDKDK)
LDHB
K7
Ac-K7 (KEKLI)
MDH2
K335
Ac-K335 (FVKTL)
PDHA1
Ac-K321 (LLKDR)
PSAT1
K318
Ac-K318 (LEKRF)
TKT
K16
Ac-K16 (ALKDT)
Cytoskeleton
6h 24h 40h TMT
ANXA1
K97
Ac-K97 (TLKKA)
CFL1
K31
Ac-K31 (VKKRK)
Ac-K44 (EDKKN)
K112
Ac-K112 (PLKSK)
EZR/MSN/RDX
K253
Ac-K253 (NDKKF)
K254
Ac-K254 (DKKFV)
K262
Ac-K262 (DKKAP)
PFN1
K105
Ac-K105 (VTKTD)
S100A6
K40
Ac-K40 (IQKEL)
K47
Ac-K47 (GSKLQ)
TPT1
K102
Ac-K102 (KGKLE)
VIM
K104
Ac-K104 (NEKVE)
K168
Ac-K168 (NDKAR)
Chaperones & redox balance
6h 24h 40h TMT
DNAJA1
Ac-K44 (GEKFK)
HSP90AB1/3P/AA1
K399
Ac-K399 (QSKIL)
HSPA8
K597
Ac-K597 (QQKEL)
Ac-K601 (LEKVC)
HSPB1
Ac-K123 (TGKHE)
GSTO1
Ac-K160 (TNKKT)
PRDX1
K16
Ac-K16 (NFKAT)
SOD1
K123
Ac-K123 (HEKAD)
Miscellaneous
CLIC1
K131
Ac-K131 (LEKGL)
DBI
Ac-K55 (KAKWD)
SET
K55
Ac-K55 (ILKVE)
UBE2E1
Ac-K43 (NSKLL)
YWHAQ
Ac-K9 (IQKAK)
D
Ac-K
GG-K
Ph-S/T
n(foreground)=89
n(background)=1,622
n(foreground)=794
n(background)=3,018
n(foreground)=943
n(background)=11,830
Faster turnover
Slower turnover
Log-odds of the binomial probability
over-represented
under-represented
n(foreground)=361
n(background)=1,356
n(foreground)=106
n(background)=3,707
n(foreground)=1,417
n(background)=11,394
Amino acid position relative to modification site

DNA repair pathways [418]. The same applied to many di-glycine sites within chaperonin domains of the protein-folding assistants included in the T-complex protein 1 family (17 of 31 GG-sites on CCT2-5/6A/7-8 in Cpn60/TCP1 category, Figure V-7 B). Similarly, di-glycine sites on Actin and EF hand domains within cytoskeleton proteins such as Myosins and S100 proteins often displayed an increased turnover (17 of 32 GG-K sites).

**Table V-1 | Motifs enriched among PTMs leading to significantly different peptidoform turnover.** *Consecutive motif enrichment analyses were performed using motifX* [319] (fg: foreground; bg: background).

| MOTIF | SCORE | FOLD ENRICHMENT | MATCHES IN FG | MATCHES IN BG |
|---|---|---|---|---|
| **Faster turnover** | | | | |
| *.[ST]$^{ph}$xxE.* | 6.7 | 1.5 | 158 (17 %) | 1,303 (11 %) |
| *.GK$^{GG}$.* | 3.7 | 1.5 | 87 (11 %) | 223 (7 %) |
| **Slower turnover** | | | | |
| *.Ex[ST]$^{ph}$P.* | 22.9 | 2.4 | 75 (5 %) | 253 (2 %) |
| *.[ST]$^{ph}$PE.* | 15.9 | 1.8 | 92 (7 %) | 429 (4 %) |
| *.[ST]$^{ph}$P.* | 9.2 | 1.2 | 662 (53 %) | 4,745 (44 %) |
| *.K$^{ac}$P.* | 14.2 | 3.8 | 47 (13 %) | 46 (3 %) |
| *.ExxK$^{ac}$.* | 5.2 | 2.1 | 42 (13 %) | 84 (6 %) |

In contrast, acetylation on structural proteins was widely associated with decreased turnover as exemplified by the over-representation of Rod domains among stabilizing ac-K sites (Figure V-7 B). They are composed of α-helices which shape coiled-coiled dimers as part of filamentous proteins such as Vimentin, Lamin A, and Keratins. As already suggested in the peptide-centric analyses, this indicates that turnover analyses may identify modifications that mediate potentially stabilizing protein-protein interactions. More general, 86 of 197 ac-K sites within α-helices and β-strands featured a slower turnover resulting in the enrichment of these structural features, and many of them were located within cytoskeleton proteins (Figure V-7 C). As a second protein class, a multitude of metabolic enzymes including several involved in glycolysis were identified to carry ac-K sites within ordered protein structures that led to a vastly decreased turnover (Figure V-7 C, Appendix Figure 0-20). Further, such sites were also detected on several proteins involved in DNA synthesis, mRNA processing, translation, and protein folding. Importantly, for some of these sites increased acetylation was observed upon KDAC inhibitor treatment using SAHA, CUDC-101, and Romidepsin (e.g. UBE2E1_acK43, FASN_acK1704, ENO1_acK80, PKM_acK433, Figure 0-21 A). Others have detected similar upregulations on additional sites (e.g. MDH2_acK335, VIM_acK104) using the Sirtuin inhibitors Sirtinol and Tenovin [310], and identified several ac-K sites to be regulated by the KATs CREB-binding protein (CBP) and p300 (e.g. UBE2E1_acK43, EEF1B2_acK185) [208]. This implies that at least a fraction of these modifications is subject to enzymatic control. Altogether, this may disclose

**Figure V-7 | Functional annotations and motifs of PTMs resulting in significantly different turnover.** *(A) Bar charts present fractions of sites with regulatory annotations in the PhosphoSitePlus database (upper panel) and residing on peptides with significantly different turnover (lower panel). (B) Enriched sequence features and site functions are shown for modifications on peptides with significantly faster (left panel) or slower turnover (right panel). Functional categories in enrichment analyses (Fisher's exact tests at 5 % Benjamini-Hochberg FDR) comprised kinase-substrate relations [309], site functions from PSP, and UniProt sequence and site features. Only at least 2-fold enriched categories with more than 10 members are displayed. (C) Average turnover is illustrated for modification sites (circles) within MCM domains (GG-sites) and α-helices or β-strands (ac-sites), and their corresponding proteins (squares). All peptides mapping to a modified or non-modified site were included for calculation of average N/O ratios or turnover rates K. Amino acids surrounding the modification site are indicated in brackets. (D) Probability logos for 11 amino acid long sequence windows were plotted using pLogo [322] and are illustrated for serine/threonine (X) phosphorylation, lysine acetylation, and ubiquitination that significantly influenced protein turnover (red horizontal line: p=0.05).*

what appears to be a prevalent mechanism by which acetylation within ordered structures decreases protein turnover.

Other enriched categories among sites with decelerated protein turnover included chaperones of the Heat shock protein 70 family (Figure V-7 B) which featured slow-turnover ac-K sites also in non-helical protein parts (e.g. HSP8_acK507/524/531/ 589). Further, ph-ST sites mapping to positions 2, 4, 5, and 7 in C-terminal heptapeptide repeats of the DNA-directed RNA polymerase II subunit RPB1 (POLR2A, Y[ST]P[STQ][ST]P[SRTEVKGN]) were enriched hinting at some relevance of C-terminal phosphorylation for POLR2A function. Likewise, turnover of Dynein light intermediate chains decreased with C-terminal phosphorylation (e.g. DLIC1_pS510/512/513/515/ 516) indicated by the over-representation of the DLIC PFAM category. The same applied to regulatory site categories 'phosphorylation' and 'activity, induced', which were enriched within ph-ST sites with slower turnover despite the scarcity of functional annotations.

**Enriched enzyme-substrate relations and PTM motifs among significant modifications** – Regarding protein kinase substrates (determined *in vitro* [309]), target sites of Casein Kinase II subunit α (CSNK2A2/CK2a2) were over-represented within ph-ST sites accelerating protein turnover. On the contrary, ph-ST sites mapping to substrates of G2/M-phase Cyclin dependent Kinase (CDK1/Cdc2), c-Jun N-terminal Kinase (MAPK8/JNK1), and Dual Specificity Tyrosine-phosphorylation-regulated Kinase 1A (DYRK1A) were enriched among stabilizing ph-ST sites (Figure V-7 B). Their preferential substrate motifs were also reflected by the results of motif enrichment analyses (Figure V-7 D, Table V-1). For faster turnover ph-ST sites, acidic amino acids were significantly overrepresented in +1 and +3 position, and slower turnover sites exhibited mainly an enrichment of proline in +1 position. Motif analysis of ac-K sites with decreased turnover equally revealed a proline in +1 position. This finding was surprising in light of concurrently enriched α-helices and β-strands since proline generally breaks these ordered secondary structures due to its rigid structure. Interestingly, ac-K sites followed by a proline were indeed depleted within slower turnover modifications sites in α-helices and β-strands compared to those that were not located within such ordered structures (2 % vs. 16 %). Hence, there are two major groups among acetylation sites that show a decreased turnover: One is located within ordered structures, and one followed by proline accounting for 23 and 12 % of slower-turnover ac-K sites, respectively.

To investigate whether certain writer or eraser enzymes are connected to ac-K sites with significantly different turnover, published data from drug treatment and transfection studies was utilized (19 KDAC inhibitors in HeLa cells [310], one CBP/p300 inhibitor in Kasumi-1 cells and p300 transfection in 293FT cells [208]). Although Sirtuin 1 and 2 inhibitor-responsive sites were slightly enriched among slower turned-over ac-K sites (based on 10 observations only), other acetylation sites that have been found to be regulated by class I, II and IV KDACs and CBP/p300 were depleted 2- to 5-fold. However, they were not enriched within faster turnover ac-K sites (Figure V-7 B). Additionally, proline was found to be significantly under-represented in +1 position for sites induced upon KDAC inhibitor treatment (only 4 of 302 sites) but not among CBP/p300 regulated sites (28 of 674, Appendix Figure 0-21 B). This suggests that the second class of slow turnover ac-K sites may be a result of the inability to remove acetylation enzymatically from lysine residues that are followed by a proline. This would inevitably lead to an accumulation of such modification over the lifetime of a protein and their higher probability to occur on 'older' proteins would present itself as decreased turnover of the modified proteoform.

**Counterplay of acetylation and ubiquitination** – In contrast to ac-K sites with slower turnover, no consistent patterns were identified for di-glycine sites with significantly different turnover

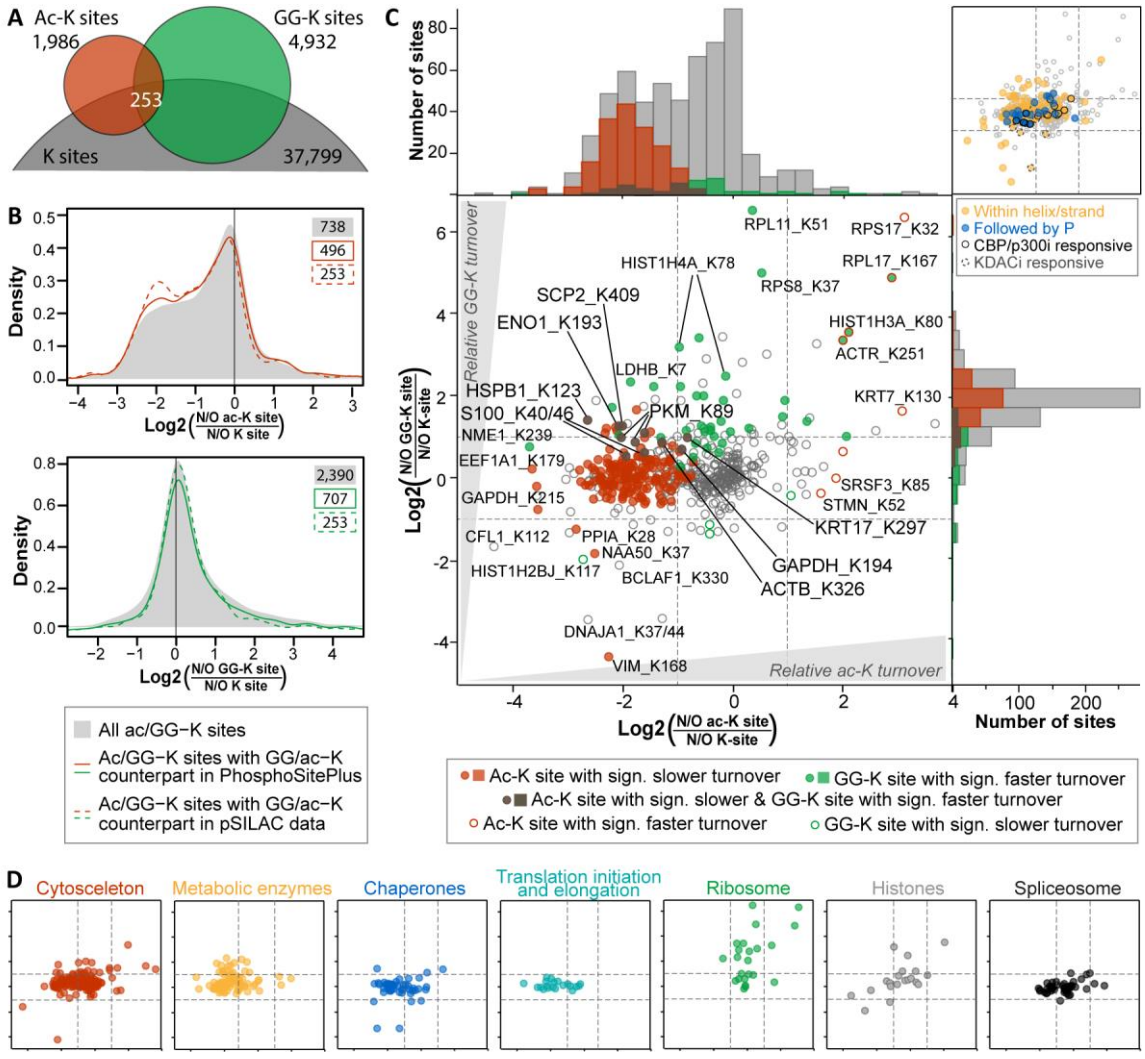



**Figure V-8 | Turnover counterplay of lysine modifications.** *(A) The overlap of modified and non-modified lysine residues is displayed for single pulse time-point data excluding the 1 h time-point. (B) Ratios of average N/O fractions of ac-K (upper panel) or GG-K sites (lower panel) to their non-modified counterpart site were computed including all peptides mapping to respective sites. Ratio distributions are shown for all identified modification sites, sites that list an ac-/GG-K pair in the PhosphoSitePlus database, and ac-/GG-K pairs actually identified in 6, 24, and 40 h single pulse data. Number of sites mapping to the three categories are indicated. (C) Relative turnover differences of ac-K and GG-K sites to non-modified lysine residues are illustrated and sites with significant differences are marked. Note that the same site can be displayed multiple times if it was detected in different single time-point pulse experiments, and that 22-26 % of plotted GG-K and ac-K data points could not be tested for significant differences due to missing replicate quantifications. Dotted lines indicate a 2-fold turnover difference of the modified compared to the unmodified site. The upper left panel identifies sites that are located within ordered secondary structures, followed by a proline, or regulated upon KDAC or KAT inhibitor (KDACi/KATi) treatment. (D) Scatter plots display the same data as in (C) but discriminating between members of different protein classes.*

except for a weak over-representation of glycine in -1 position for Ubiquitin-remnant sites with increased turnover (Figure V-7 D, Table V-1). The lack of congruence between motifs of slower turnover ac-K and faster turnover GG-K sites may give the impression that there is no strong counteracting effect on protein stability for acetylation and ubiquitination. However, this observation may also simply be biased by the preferential identification of different subpopulations of lysine sites within the two PTM subproteomes. Therefore, effects of different lysine modifications were assessed for sites that were identified in both modification states and as unmodified version. To

preclude artefacts from combining different pulse datasets, only sites that were detected in all three states within the same single pulse time-point were considered resulting in 253 lysine sites available for analysis (Figure V-8 A). Despite being small, this subset of ac-K and GG-K sites appeared to be a reasonably representative selection of sites indicated by a comparison to listed ac-K/GG-K pairs within the PSP database (Figure V-8 B). Interestingly, the proportion of acetylated sites shifted towards a decreased turnover was much larger than the fraction of Ubiquitin-remnant sites featuring an increased turnover. Only 9 sites were identified with significantly higher turnover in the di-glycine state and, at the same time, significantly slower turnover as acetylated version (note that a quarter of site pairs could not be tested for significant turnover differences due to missing replicate quantifications, Figure V-8 C). More than half of these oppositely regulated sites were located within α-helices or β-strands (HSPB1_K123, ENO1_K193, PKM_K89, S100A6_K40/47), while none was followed by a proline. Generally, ac-K sites with decreased turnover and followed by proline did typically not exhibit an accelerated turnover upon ubiquitination within the tested data subset, but many of the ac/GG-K pairs with counteracting effects were located within ordered, secondary structures. Noteworthy, similar differences were also observed among several protein classes (Figure V-8 D). A marked fraction lysine sites mapping to metabolic enzymes displayed decreased ac-K, but increased GG-K turnover. For translation initiation and elongation factors, however, ac-K sites often featured a slower turnover compared to the unmodified site, but ubiquitination did not clearly increase peptidoform turnover. In line with previous observations, modification of lysines within ribosomes frequently led to an accelerated turnover in particular for ubiquitination but in some cases also for acetylation. The direct comparison of turnover for acetylated, ubiquitin-remnant and unmodified sites suggests that, for a fraction of lysine sites, acetylation and ubiquitination indeed have opposing effects on protein stability. However, often a decreased turnover of ac-K sites was observed without any impact of the di-glycine counterpart modification on turnover implying that these lysine modifications do not function necessarily and globally in a competitive manner.

# 4 Discussion and conclusion

## 4.1 Robust determination of modification-specific protein turnover

For the study of modification-specific peptidoform turnover, two approaches were employed. Using the pSILAC-TMT method, label incorporation and loss are measured in a time-dependent manner and turnover rates of peptidoforms are estimated via curve fitting. In contrast, N/O ratios obtained in single pulse experiments provide the relative pace of turnover within a certain pulse time-point. Hence, no curve fitting and filtering is necessary which clearly simplifies overall data processing. However, the observable turnover range within a single time-point is more limited since relatively high or low turnover peptides more likely escape analysis due to missing quantitative information for one of the two SILAC channels. For pSILAC-TMT time-course experiments, extreme turnover behaviour will only affect the likelihood of fragmenting and obtaining a curve for only the newly synthesized (very high turnover) or only the 'old' SILAC labelled peptide species (very low turnover). While those peptidoforms with considerably faster or slower turnover may be the most interesting ones, they cannot be rescued for quantitative assessment in single pulse data in a reliable way because data imputation could strongly influence and bias results [328, 343] (see also pp. 87). The limitation of such missing quantitative data was particularly apparent for the earliest and latest single time-points, which showed overall decreased numbers of peptides quantified in both labelling states. Notably, this also entails the quantification of more distinct selections of peptides across different time-points than within replicates of the same time-point. This is important to mention since the subset of peptides identified for a protein or protein group will affect the overall assessment of protein turnover. Similarly, this can also impede the identification of peptidoforms with utterly differing turnover within the same pulse time-point because, the greater their difference, the less likely they are quantified in both labelling states. This may also explain the comparably small number of acetyl-K and GG-K peptides available for the investigation of their interplay within the same pulse time-point (during IP optimizations more than 1,700 lysine sites with both PTMs were identified). In earlier time-points, more ubiquitin-remnant peptides were quantified, while later time-points showed increasing quantifications of ac-K peptidoforms. It would have been possible to combine results from different time-points and as well include ac-K data from pSILAC-TMT experiments to increase the overlap between acetylated and di-glycine modified lysine residues, but it was decided to consider only sites within the same dataset to avoid potential artefacts based on the partly observed systematic shifts between turnover datasets. Likewise, 1 h pulse data were excluded due to the alarming poor correlations across label swap experiments, which were caused by distinct sub-populations of peptides that deteriorate the otherwise excellent concordance of remaining peptides. While this was most notable for 1 h pulse data, the other pulse time-points showed similar, albeit less profound biases. Quantitative discordance between label-swap experiments has already been reported in cross-linking experiments as a result of large peptide mass, low intensity, and heavily overlapping isotope clusters of light and heavy signals [419]. In addition, quantification accuracy may be impaired to some extent by isotope impurities of SILAC amino acids, but this is unlikely to produce such strong effects. Further systematic analyses are required to clarify the chief cause for erroneous quantification in single pulse SILAC experiments.

In contrast to single time-points pulses, the time-course characteristic of pSILAC-TMT data required more sophisticated data processing and curve fitting. Noteworthy, curve fitting and filtering constitutes an additional level of quality control. The prior knowledge about which labelled peptide should exhibit label incorporation and which one should resemble label loss, and their

automatic removal if they do not show the expected quantitative trend, can further decrease false identifications and poor quantifications. However, peptides that do not adhere well enough to the assumptions underlying the curve-fitting algorithm despite representing valid identifications are equally removed and thereby observable turnover behaviour is limited to *a priori* suppositions. The high proportion of evidence entries that did not pass the KRAB filter criteria suggested that this might have been a prevalent issue. To improve overall data quality and fraction of successful curve fits, two major adjustments in data processing were carried out compared to the initial establishment of the pSILAC-TMT approach (see Chapter IV). First, co-isolated, potentially distorting intensities from peptides with opposing characteristics (label loss vs. incorporation) were removed *in silico*. Several approaches have already been described for computational reduction of co-isolated intensities, but they rely on either spike-in standards [420] or the estimation of the precursor intensity fraction within the isolation window [245, 326], and the assumption of a constant, co-isolated background. The later prerequisite renders such approaches inapplicable to pSILAC-TMT data since co-isolated peptides in such experiments always feature decreasing or increasing intensities. Hence, instead of removing a constant background, average label incorporation and loss curves were subtracted from curves with apparent ratio compression (i.e. showing residual intensities in the necessarily empty channel). Of note, co-isolation of peptides with the same behaviour (e.g. both indicating label loss) would not result in a measurable ratio compression and thus cannot be corrected. In such cases, however, quantification accuracy would anyway only be influenced to a minor degree. As expected, removal of co-isolated intensities increased the number of successful curve fits especially for low abundant, modified peptides (ac-K and ph-Y) that suffered most from ratio distortion. In a second step, peptide intensities were corrected for abundance changes during the time-course of the experiment to improve accordance with steady-state assumptions underlying the curve-fitting algorithm (see pp. 52). Importantly, such corrections must be assessed carefully to minimize potential processing artefacts. For example, if only one entry for each curve is available, label incorporation and loss curves would be inevitably forced to yield identical turnover rates after correction. This would artificially improve correlations for comparisons of label loss and incorporation. Therefore, it was decided to require at least four curves per peptide and calculate correction factors globally (not separately for each replicate) thereby correcting primarily abundance changes that were reproducible across replicates. Consequently, only a fraction of peptides was subjected to such intensity adjustments, but the overall correlation across replicates still increased considerably.

Despite these improvements, the number of successful curve fits for acetylated peptides in the pSILAC-TMT approach was still comparably low. This was also caused by the drop in ID rates compared to single pulse experiments, which is commonly observed for TMT data [325, 349, 350, 355-357] (see also pp. 89). The acetylome coverage in turnover measurements can generally be improved in the future using larger input amounts (10-20 mg) and fractionation prior to acetyl-peptide enrichments [207]. With the optimized TMT protocol described above (pp. 77), it would still be feasible to perform such experiments in a pSILAC-TMT-like fashion and label large input amounts at reasonable costs. Labelling before fractionation would also minimize technical variation and maintain relative quantification accuracy across conditions. Alternatively, it is also conceivable to label peptides only after immunopurification to reduce overall reagent amounts and costs. This is also the only way to study ubiquitination with the pulsed SILAC-TMT approach. Ubiquitin-remnant peptides cannot be enriched after TMT labelling since the N-terminus of the di-glycine is masked by TMT. A TMT setup for turnover measurements of di-glycine remnant peptides would hold great promise since it would facilitate the detection of very high-turnover modification

sites compared to single time-point pulses (see above) especially when pulse time-points are adjusted accordingly. The feasibility of TMT labelling of di-glycine enriched peptides has been shown recently [421] and suggests that such workflows can readily be employed for time-dependent turnover measurements with increased coverage of acetylated and di-glycine peptides.

Utilization of Ubiquitin linkage-specific antibodies [422] in combination with turnover measurements could further help to deconvolute effects of different poly-ubiquitin chains and yield interesting insights in their possibly protein- and site-specific functions. This level of resolution is not provided in the data presented here, and it is possible that the N/O ratios measured in different time-points are indeed derived from a mix of several mono- and poly-ubiquitin chains. Modification enrichments on protein level could also initially separate different proteoforms (ubiquitinated and non-ubiquitinated ones) and enable an assessment of PTM crosstalk after enrichment of additional modifications (e.g. phosphorylation) in these different sub-populations [412]. In addition, PTM crosstalk can be investigated if several modifications within the same PTM subproteome are allowed in the database search (e.g. ubiquitination in the phosphoproteome). However, it is important to note that such strategies require careful and critical evaluation since they will exponentially enlarge the combinatorial search space that is already comparably large for pSILAC-TMT data (due to variable K8 and R10 modifications). Thus, the probability of high scoring random matches and false positive identifications of modified peptides will be increased as well [259, 423].

## 4.2 Uncovering unknown global and protein-specific regulatory mechanisms

The combination of the pSILAC-TMT and single pulse time-point approaches detected a multitude of peptidoforms and PTMs that featured a differential turnover compared to the whole protein or counterpart peptides. The data provided additional evidence for findings that have been described previously such as the overrepresentation of degron sequences on protein termini [119, 413], the degron activity of a doubly phosphorylated motif on UNG [416], and the destabilizing nature of glutamine in second position [117]. Further, hints on numerous, hitherto unknown, cellular regulations were implied by the data, of which only a limited number of examples could be presented here.

Generally, one could have expected to observe an enrichment of known (phospho)degron motifs among peptides with increased turnover. However, only the small category of tertiary N-degrons was overrepresented. The incapacity to identify destabilizing motifs predominantly on sequences with accelerated turnover likely illustrates the conditional nature of degrons and the negligence of information about relevant cellular context when analysing merely peptides. To induce degradation much more than only the short E3 ligase binding motif is required. The degron needs to be exposed to enable actual interaction with the E3 ligase, which also needs to be present in the same cellular compartment. Moreover, an Ubiquitin acceptor site must be available, and a loosely folded, degradation initiating region is required for interaction with the proteasome [108]. However, for annotation of degrons, the exclusive information utilized was the linear degron sequence. Additionally, degrons can often be switched on and off by modifications adding another layer of complex control. Potential degron motifs, quite unexpectedly, were frequently even identified on peptides with decreased turnover, and often those included additionally modifications within or around the degron sequence (e.g. phosphorylation was observed on 2/3$^{rd}$ of seemingly stabilized peptides containing a SPOP-binding degron). This suggests that these modifications might inhibit specific binding of E3 ligases to non-modified, constitutively active degrons and thus prevent degradation, a mechanism that has been described before [132]. Despite the challenges

for detection of degrons in a global approach, several candidates for hitherto unknown, potentially active degron motifs were identified when data were examined on a case-by-case level, including a SPOP-binding motif on EIF3H and a FBW7-binding phosphodegron on TKT.

Noteworthy, identification of modification-specific turnover of degron sequences (and other peptidoform in general) was greatly facilitated when modified and non-modified peptides were compared directly. Naturally, this analysis should yield larger differences than the comparison of the turnover of a modified peptide to the whole protein since the latter will partially include peptides originating from the proteoform(s) that contained the modified peptide thereby attenuating the effect size. This becomes especially relevant for PTMs with high occupancies. It is worth mentioning that the influence of very high stoichiometry modifications is generally more difficult to determine with the employed experimental setup because, in such cases, the turnover of the corresponding protein will largely resemble the turnover of the modified peptide. At the same time, the non-modified peptidoform has a high probability of escaping detection due to its low abundance, which in turn precludes a comparison of modified and non-modified counterpart peptides.

Of all examined modifications, lysine acetylation showed the most profound, global turnover changes and yielded the highest fraction of sites with significantly altered turnover. So far protein stability-regulating acetyl-modifications have been described mainly for transcription factors [409]. As suggested by the large group of acetylation sites located within α-helices and β-sheets and exhibiting a decreased turnover, such a regulatory mechanism might be much more common than currently anticipated and, for instance, also largely relevant for cytoskeleton proteins, metabolic enzymes, and chaperones. For a fraction of these ac-K sites, destabilizing di-glycine countersites were identified supporting the notion of a direct competitive crosstalk between acetylation and ubiquitination by blocking ε-amino groups of lysine residues [409]. Yet, there was also a large proportion of apparently stabilizing ac-K sites, for which the ubiquitin counterpart displayed a similar turnover as the non-modified site, suggesting alternative mechanisms might slow down the turnover of the acetylated peptidoform. These could include crosstalk to more distant sites within the same protein, altered protein conformations, active stabilization of secondary structures, or regulation of protein interactions. As an example, acetylation of PDHA1 on K321, which resulted in decreased turnover, has been reported to mediate interactions to PDK1 thereby regulating the activity of the mitochondrial pyruvate dehydrogenase complex [424]. It remains to be determined which of these mechanisms explain slower turnover for individual proteins, but evidence so far suggests that multiple processes may lead to the observed deceleration in turnover.

The fact that acetylation occupancies are typically very low [205] makes it at least unlikely that ac-K modifications within secondary structures are indispensable for protein folding and structural maintenance in a global level. Further, this raises the question how relevant the turnover-increasing effect may be in a cellular context if only a tiny fraction of respective proteins exists in an acetylated state. While such mechanisms may not play a significant role under normal conditions, they could become crucial in disease states when modification stoichiometries may change considerably. Decreased activity of the deacetylating enzyme Sirtuin 3, for example, has been associated to the survival-promoting initiation of the Warburg effect in cancer cells involving increased abundances of PDHA1_acK321, PGK1, and LDHA (for which several stabilizing ac-K sites were identified) [424, 425]. Further, acetylation of Tau protein on different sites has been linked with promotion as well as prevention of its aggregation, which is a marker for Alzheimer's disease and other neurological pathologies [426, 427]. Likewise, drug treatments can markedly change abundances of acetylated proteoforms (e.g. 20 x increase on PKM_acK433 upon CUDC-101 treatment).

In conjunction with the huge turnover changes that have been measured frequently for acetylation (e.g. 6 x slower turnover of PKM_acK433), such alterations in the modification landscape could dramatically alter functional states of target proteins. Even if effect sizes of turnover changes are comparably small, the impact on protein stability could be significant if they affect proteins with very high default turnover (such as the Ubiquitin-conjugating enzyme UBE2A1). These examples illustrate that measured turnover differences may be highly relevant to gain mechanistic insights into disease states and drug mode of actions even for sites that feature a low occupancy in steady state.

The observation of widespread non-enzymatic acetylation in certain cellular contexts [72, 73, 428] begs the question whether observed differentially, turned-over ac-K sites are subject to enzymatic control and thus can potentially be manipulated by drug interventions. It has been suggested that non-enzymatic modifications predominantly take place within clusters of basic residues on mitochondrial proteins favoured by elevated ac-CoA concentrations and a slightly higher pH in mitochondria compared to other cell compartments [73, 429]. Of note, the minority (11 %) of ac-K sites with significantly different turnover mapped to proteins associated with mitochondria (according to MitoCharta [291]). Further, motif enrichment analyses indicated that basic residues around the modification site are rather underrepresented, and there was evidence that numerous ac-K sites with decreased turnover are regulated by KATs [208]. Together this suggests that most ac-K sites with differential turnover are under control of writer enzymes. Nevertheless, it cannot be excluded that they are at least partly subject to non-enzymatic acetylation, as well.

Besides acetylation within ordered secondary structures, ac-K sites followed by a proline were identified as a second group of acetylation with slower turnover. Integration of turnover data with acetylome profiles after HDAC inhibitor treatments suggested that these modifications might be difficult to remove enzymatically, possibly due to the strong structural constraints implicated by proline. Thus, acetylation on KP motifs likely accumulates over the lifetime of respective proteins. Importantly, such events would be in discordance with the steady-state assumption that is the basis for interpreting measured peptide turnover as proxy for proteoform lifetime. This requires modifications to occur reversibly and be observable on recently synthesized and long-existing proteins with the same likelihood. Consequently, the observed decrease in turnover most likely does not imply an increase in protein stability upon acetylation (though, conversely, this cannot be excluded), but rather reflects precisely the imbalance of the acetylation writing and erasing processes (i.e. the dynamics of the modification process). The high fraction of slower-turnover ac-KP sites suggests that the inability to remove such modifications may be prevalent and calls for an investigation of potential functional consequences of this disturbance of the writer-eraser equilibrium. Noteworthy, the discordance with steady-state assumptions could also explain in part the low success rate of acetylated peptides for passing curve-fitting criteria. If ac-KP acetylation accumulates on proteins with already very slow turnover, observed label incorporation and loss of acetyl-peptides easily can be much slower than one would expect from cell doubling alone and such shallow “curves” may yield poor curve fits and be filtered out.

Interestingly, proline in +1 position was also enriched among ph-ST sites with decreased turnover. Little is known about phosphatase motif specificities, but it can be speculated that a fraction of these slower turning-over sites is also a result of accumulation of phosphorylation due to a lack of phosphatase activity towards the structurally very distinct S/TP sites. As an example, S390 and S392 in Lamin A are followed by proline and featured slower turnover upon phosphorylation. An induction of these modification sites has been reported to lead to the disassembly of the nuclear lamina, which is crucial during mitosis [414] and for the release of viral nucleocapsids [415]. It

appears unlikely that Lamin A in monomeric state is more stable than in its assembled form since protein-protein (complex-like) interactions typically lead to a stabilization of proteins [376, 386]. This renders the hypothesis of a missing phosphatase activity appealing. This would lead to the observed decrease in turnover and imply that LMNA_pS290/292 is irrecoverable and that Lamin A re-synthesis is required for functional recovery of the nuclear lamina.

Some of the examples described above illustrate that a detected difference in turnover among peptidoforms/modifications does not necessarily indicate a direct protein stabilizing or destabilizing effect, but could be a result of a multitude of cellular regulations (or the lack thereof, see also p. 123). Logically, however, it becomes clear that differential turnover of modified peptidoforms must be based on some form of a potentially interesting and relevant, cellular process. This consideration is intriguing especially because the majority of modified peptidoforms with significantly different turnover comprised PTMs with unknown function. Fructose-bisphosphate Aldolase A, for instance, was heavily modified and numerous modified peptides featured a differential turnover, which is astonishing considering that not a single modification site is annotated with a function in the PSP database [294]. Likewise, the proteasome comprised an extraordinary high fraction of differentially turned-over peptidoforms, but PSP lists a function for only six modifications sites (exclusively phosphorylations) and none of them was identified here. This applies similarly to the ribosome, which featured numerous modified peptides with increased turnover. Interestingly, two studies reported profound differences in turnover of the proteasome across several primary, hematopoietic cell types [114] and a poor correlation of protein and mRNA levels across HeLa cell lines specifically for ribosomal proteins [403]. The data suggest that this may be caused by varying levels of PTMs that tightly control protein abundance and activity.

## 4.3 Conclusion

Here, turnover of thousands of phosphorylated, acetylated and ubiquitin-remnant peptides was analysed for the first time, and a large body of differentially turned-over modified peptidoforms was identified. The possible mechanisms underlying turnover differences are diverse in nature rendering the interpretation of the data non-trivial. However, sites with differential turnover are believed to represent a rich resource of PTMs with supposedly regulatory function. The detection of such modification sites could be in particular helpful for acetylation and ubiquitination sites, which are chronically underrepresented among annotated regulatory sites. Certainly, a discrimination of ubiquitin-linkage types in the future would prove useful to advance the understanding of their functional distinctions. Moreover, effects of modifications on turnover may be cell type and context dependent demanding investigations on modification-specific turnover in several different cell systems. Finally, the coverage on lysine PTMs can be readily improved employing larger input amounts, which should enable conclusion also about low occupancy modifications on lower abundant proteins. Importantly, an integration of such turnover data with results of PTM-based, drug treatment studies can lay the foundation to prioritize functionally relevant sites within drug-responsive PTMs. The complexity of cellular processes makes the inference of the type of function from turnover measurements alone difficult, but drug-responsive sites with significantly different turnover represent prime candidates for further molecular investigations of their function. By narrowing down the number of relevant prospects, mechanistic studies on the mode of action or resistance of drugs can be greatly facilitated, demonstrating the tremendous potential of modification-specific turnover analyses.

*„Je mehr ich weiß, desto mehr erkenne ich, dass ich nichts weiß."*

**- Albert Einstein (adapted from Plato)**

# VI GENERAL DISCUSSION AND FUTURE PERSPECTIVES

**1 Protein turnover as new dimension in drug research.....163**

**2 The proteoform challenge.....166**

**3 New proteomic technologies on the rise.....170**

**4 The future of proteomics: Higher, faster, further?.....172**

# 1 Protein turnover as new dimension in drug research

The dimensions of protein turnover reflect the cellular proteostasis state, which is critically controlled by complex cellular machineries such as ribosomes, chaperones, and proteasomes. Since it became increasingly apparent in recent years that many diseases and ageing are hallmarked by deregulations of the proteostasis network [84, 135], proteomic turnover studies are gaining attention in the drug research field. Further, numerous drugs that target important proteostasis players have been developed to exploit the connection of proteome imbalance and diseases like cancer therapeutically. As maybe most prominent examples of proteostasis drugs, proteasome inhibitors (e.g. Bortezomib, Carfilzomib, Ixazomib) and HSP90 inhibitors (Geldanamycin, Tanespimycin) have proven effective, for instance, for treatment of multiple myeloma [146-148]. In addition, molecules targeting other members of the UPS [430] have been approved or are currently under (pre)clinical investigation for cancer treatment, including E3 enzyme modulators (e.g. Nutlins [431], Lenalidomide [432]), DUB inhibitors (e.g. P5091 [433]), and proteasome substrate recognition inhibitors (e.g. Ubistatin A/B [434]). Owing to their target molecules, some of these drugs affect large subsets of the proteome broadening their applicability for various indications but potentially also increasing the risk for toxic side effects [435]. As more targeted approach, proteolysis targeting chimeras (PROTACs) are emerging for selective protein degradation (reviewed in [436]). They consist of two moieties, one small molecule inhibitor binding to the protein of interest and one ligand recruiting an E3 ubiquitin ligase, which then ubiquitinates the target protein and marks it for degradation. Although PROTACs typically feature excellent target selectivity, some have shown off-targets effects [436], illustrating the importance of a global evaluation of PROTAC effects. Moreover, the molecular basis of sensitivity, resistance, and toxic side effects oftentimes is not well understood for proteostasis-regulating agents, further emphasizing the need for unbiased methods to investigate proteostasis mechanisms and drug responses on a molecular and proteome-wide level.

**Considerations for analyses of non-steady-state turnover** – Studying changes in protein turnover upon drug treatment can broaden our understanding of molecular actions of drugs and of the principles underlying the regulation of proteostasis. However, the analysis of drug perturbation experiments is complicated by inherent violations of steady-state assumptions, such as the potential discordance of synthesis and degradation rates. Consequently, first-order equations may not describe labelling kinetics properly. Curve-filtering procedures applied to time-resolved labelling data, like those obtained with the pSILAC-TMT approach, potentially will remove exactly the most interesting candidates (for examples see HSP90 inhibitor treatment in [375]). Alternatively to time-dependent analyses, single time-point measurements and more straightforward SILAC ratio comparisons can be performed [437]. This comes with the downside of losing the time resolution unless the analysis is carried out for several time-points separately. Within this thesis, major differences in computed turnover rates have been observed for different time-points already under assumed steady-state conditions (see Chapter V). This demonstrates that multiple time-point measurements may be advisable to obtain a more comprehensive picture of ongoing regulations. As already outlined earlier, drug perturbations are further typically accompanied by changes in cell proliferation, which can significantly alter measured turnover rates. Hence, a correction across different drug treatment experiments is indispensable. Importantly, adjustments need to be undertaken with great care since small variations (e.g. measurement uncertainties in the cell-doubling rate) can have a big impact on corrected degradation rates and protein half-lives in a comparative setup. It could also be expedient to correct for changes in cell doubling behaviour via

mean or median normalization approaches (i.e. equalizing medians of SILAC ratio or turnover rate distributions for different cellular conditions) [375]. It is worth mentioning though that, from personal experience, overall shapes of ratio or rate distributions can change dramatically for drug treatments that have extensive effects on proteostasis and thus cell doubling (e.g. proteasome inhibitors). This can impede meaningful normalization and interpretation of data on the level of single proteins. Generally, the requirement for a correction of cell doubling differences could be circumvented by utilizing cell systems that reside in a non-dividing state like primary, differentiated, or growth-arrested cells [114, 169, 302]. However, this entails the disadvantage of typically slow label incorporation, and large ratios are difficult to measure (accurately) in classical pulsed SILAC experiments. Therefore, advanced experimental designs have been used to study turnover in non-dividing cells, for example by starting the pulse with a 50 % labelled cell line [169]. Noteworthy, pSILAC-TMT measurements could prove beneficial in such setups as well, due to the high precision of TMT and the less biased identification and quantification of high and slow turnover proteins. On the other side, perturbation experiments pose another challenge specifically for pSILAC-TMT experiments since data normalization for mixing errors across channels is built on the assumption that the overall abundance of most proteins does not change during the course of the experiment. This may not be true for investigations of drugs with very broad effects. However, it might be possible to omit this normalization step with minor impact on data quality, if mixing errors are minimized by careful evaluation of peptide content before pooling of TMT labelled samples.

Besides proper data processing, another crucial step for turnover studies in perturbed systems is the interpretation of measured turnover changes. Although incorporation and loss curves commonly are referred to as synthesis and degradation curves, it is important to understand that they cannot measure synthesis and degradation independently. Determined label incorporation and loss reflect the net turnover, which means a change in either degradation or synthesis will affect dynamics for both labels. More precisely, if the degradation rate decreases while the synthesis rate stays the same, the absolute amount of proteins with the new label will still increase because overall less proteins including the newly synthesized ones will be degraded (based on an identical probability of degradation for pre-existing and newly synthesized proteins). As a result, one will observe a slower decrease (i.e. a relative increase) in proteins carrying the old label and a faster increase in proteins with the new label. This can be misinterpreted easily as induced protein synthesis. Of note, an induction of synthesis without any change in the degradation rate actually will result in a similar trend because proportionally less of the pre-existing proteins will be degraded when more proteins with a new label are available for degradation. The same rules apply to an induction of degradation or a reduction in synthesis, which will both lead to interdependent slower increases in proteins with the new label, and a faster decrease of proteins with the pre-existing label. Naturally, the extent and dynamics of respective changes will indeed very much dependent on the actual process that changed. Yet, it might still be difficult to model dynamics accurately in a time-dependent manner, as cellular protein degradation and synthesis are likely to change dynamically and in an interrelated manner over time after drug treatment.

**Elucidating molecular actions of (proteostasis) drugs** – Theoretical considerations left aside, conclusions about whether synthesis or degradation change can still be drawn if data are utilized adequately. To this end, information about a SILAC label ratio and a protein abundance is required for both conditions to decipher doubtlessly whether and how synthesis or degradation have changed upon perturbation. This may become clearer with an example using ratios of newly synthesized to pre-existing proteins (in this thesis named N/O ratios, e.g. as determined in [113]) and

total protein levels (N+O). If N/O ratios are elevated upon perturbation, this can reflect an increase in protein synthesis (more N is synthesized) or in degradation (proportionally more O is degraded). Taking additionally the actual total protein expression (N+O) into account can reveal which process is affected. An increase, for instance, will identify synthesis as the altered process. Notably, the same interpretation is possible from absolute abundances of newly synthesized proteins and ratios of newly synthesized to total proteins (in this thesis named abundance-corrected label incorporation ratios, e.g. similar to [170]). Multiple other combinations of ratio and abundance information can be exploited, but their interpretation can change depending on the variables used (e.g. abundance corrected ratios are different from ratios of SILAC label abundance to initial protein abundance as determined in label incorporation curves of the pSILAC-TMT experiment). This demonstrates the importance of a careful assessment of all information that is provided within different experimental setups. Recently, Savitski *et al.* designed a method for evaluation of PROTAC targets and off-targets that employs TMT-labelling of pSILAC samples but combines different treatments and label swaps instead of pulse times within one TMT-plex [437]. This approach provides all necessary information to deduce the type of proteostasis regulation upon drug treatment and can be adjusted flexibly to include several treatment time-points or drug concentrations. Hence, it certainly will be very useful for a wide range of research questions in the future.

Evidently, comparative turnover studies in non-steady-state systems can be quite challenging and can lead to misinterpretation if they are not conducted thoroughly. Furthermore, a proper and informative assessment of drugs that can cause extensive proteome remodelling and induce cell death very quickly may even be impractical within a pulsed SILAC setup that requires a certain duration of label incorporation. Therefore, alternative approaches are needed for proteome-wide evaluation of proteostasis drugs' mode of actions. One such strategy is provided by the measurement of PTM changes upon drug perturbation. This can yield insights into immediate responses upon very short treatment times and thus potentially disclose molecular mechanisms that drive drug efficacy as well as resistance mechanisms. If such experiments are performed in a concentration-dependent manner, another level of important information is added, namely the potency with which certain effects are induced. Typically, however, PTM readouts upon drug perturbation are only conducted after single-dose treatments, occasionally adding a time dimension (e.g. see for Bortezomib [421, 438]). Noteworthy, time- and concentration-dependent PTM response analyses using Bortezomib and Carfilzomib (and concentration-dependent treatments with ~20 additional drugs and drug combination) have been carried out already during this PhD work, but results could not be covered within this thesis due to time and space constraints. An integration of these large datasets of drug-responsive PTMs with modifications sites that feature differential turnover in the here presented modification-specific turnover atlas is of utmost interest. This could considerably aid interpretation of drug perturbation experiments and identify altered PTMs that are of particular, functional relevance. As outlined in Chapter V, differential turnover does not necessarily reflect a cellular stabilization or destabilization of the corresponding protein itself, but it should generally spot sites that regulate or are regulated by notable cellular mechanisms. As an example, a change in the occupancy of ac-KP sites that were observed to decrease turnover can indicate the enhanced or reduced activity of a corresponding writer enzyme. Efforts to conduct and systematically integrate such drug concentration-dependent, proteome-wide screens of PTM changes are ongoing and hold great promise for the elucidation of mode of actions of all kinds of drugs as well as of principles underlying cellular signalling.

**Identifying molecular vulnerabilities of diseases** – In addition to the investigation of drug effects, turnover measurements can also uncover differences in the steady-state dynamics across

cellular systems [114, 171, 302]. Since this default turnover is an important determinant of a cell's flexibility to react rapidly to changed environments and mirrors its overall proteostasis state, such comparative analyses may provide clues on potential molecular vulnerabilities that are specific to certain disease states. As an example, an overall very high turnover may mark diseases that are particularly vulnerable towards proteasome inhibition [84]. Further, disease-relevant proteins that exhibit a very slow turnover could be prime targets for PROTAC drugs [436]. Similarly, global turnover comparisons may also reveal causes for drug resistance. Proteins that are turned over very rapidly in a specific disease, for instance, would be poor targets for covalently binding molecules [439], which therefore likely would not be effective in this disease. For this certain class of drugs, turnover profiles are of general interest since different stabilities of targets and off-targets can influence their effective, cellular selectivity and efficacy. To elucidate relationships between turnover and drug sensitivity/resistance in a systematic way and potentially identify molecular turnover markers, one could acquire turnover profiles, at best resolved for modified peptidoforms, for a large panel of cell lines and correlate them with drug response data (according to similar approaches using protein abundance profiles [440, 441]). Of note, such comparisons would again require a careful evaluation how (or if at all) measured turnover rates should be corrected for overall differences in cell doubling.

Future efforts should be directed towards the identification of particularly fast turnover proteins, though these might be difficult to study not least due to their presumed low abundance. However, these proteins might be of particular interest considering that short-lived proteins are enriched in molecules that regulate primary cell functions (like transcription factors). Further, quickly turned-over proteins bear a high chance of being effectively regulated in abundance via post-translational stabilization, for instance, resulting from drug interventions. Finally, the time may have come for more systematic, comparative turnover studies, be it across cell line panels or upon drug perturbations, and simultaneously acquiring information about modification-specific turnover in that process could also help to elucidate fundamental regulatory mechanisms of cellular proteostasis.

## 2 The proteoform challenge

The expression of the same gene can have diverse functional consequences depending on post-transcriptional alterations [29, 33, 85-88] and post-translational modifications of the amino acid sequence [89, 91, 93]. This demonstrates the importance of protein analyses for the understanding of the molecular processes of life. Yet, several challenges have to be faced in proteomics in particular for proteoform analyses.

**Interrogations on protein versus peptide level** – As outlined in the introduction, it can be difficult to identify different splice variants unambiguously with bottom-up proteomic approaches due to their often largely conserved sequence stretches and the protein inference problem [264]. Consequently, peptides mapping to differing subsets of isoforms frequently end up in the same protein group. Despite reasonable proteome coverage, this was also commonly the case for the turnover dataset presented here, and differential turnover was not only detected across distinct protein groups containing different splice forms, but occasionally also across peptides within a single protein group, which were assigned to varying splice variants. This illustrates the merits of studying bottom-up proteomes on peptide instead of protein group level for identification of isoform-specific regulations. Of note, this also applies to differential protein expression surveys where diverging abundance ratios of different peptides may suggest some interesting, underlying biology [264].

Additionally, the information on co-occurrence of functionally interacting PTMs within the same protein [14, 96] is lost upon protein digestion if those PTMs are not in close proximity and covered by the same peptide. Noteworthy, cleverly combined PTM enrichment methods on protein and peptide level can tackle this issue and enable an investigation of PTM crosstalk on more distant sites of a protein [412]. Alternatively, top-down approaches allow for the analysis of whole proteins preserving the information on proteoform sequence and modifications and thus simplifying quantitative comparisons [442]. However, they often have difficulties to identify and localize PTMs precisely [443], suffer more severely from sensitivity limitations than peptide-centric approaches, and are not yet capable of disentangling complex proteoform mixtures efficiently [179]. Generally, several challenges in the analysis of proteoforms and their functional interpretation largely apply to both approaches.

**Detection and identification challenges** – Unlike genomics and transcriptomics, proteomic methodologies cannot benefit from *in vitro* amplification of analyte molecules, thus excellent analytical sensitivity is crucial for the analyses of proteins and peptides. In addition, a comprehensive detection is complicated by the much higher dynamic range of the expression of proteins compared to transcripts [113]. Especially modified proteoforms can be of low abundance but yet of high functional relevance. This is exemplified by the well-known sub-stoichiometric nature of tyrosine phosphorylation even within the phospho-subproteome, which most likely reflects its tight regulation within signalling networks and otherwise low structural importance [63, 64]. Hence, it is not surprising that only a small number of phospho-tyrosine peptides has been identified in the PTM turnover datasets. With enrichment approaches specific for phospho-tyrosine sites [203, 204] their coverage can be enhanced in future studies. In general, analytical sensitivity and dynamic range of proteomic analyses are being improved perpetually by advances in sample preparation workflows [189, 207, 444], instrumentation [445, 446], and computational data processing [255, 344].

Global proteoform analyses are further typically based on protein sequence databases for identification of isoforms. However, such reference proteomes often still lack annotations in particular for less common or individual events such as rare splice variants, alternative TIS, SAPs or somatic cancer mutations. For example, publications report ~100,000 minor human splice forms in medium to high abundance [30], but standard databases commonly only include ~50,000 [447]. This can be overcome partly with the use of repositories tailored to specialized research questions (e.g. TISdb [448], dbSNP [449], COSMIC [450]), or sample-specific databases that can be generated from genomic or transcriptomic data. Such proteogenomic approaches have been applied successfully to reduce the number of potentially translated splice variants for peptide mapping to the sequence database and thus increased the number of protein groups containing only a single isoform [188]. They also verified expression of genomic variants on protein level albeit with rather low success rate especially after manual curation [188, 274]. It remains to be clarified whether this reflects cellular control mechanisms preventing high expression of potentially dysfunctional proteins or whether this is mainly a result of technical constraints like incomplete ion series in MS2 spectra and increased FDRs for searches including large databases and highly homologous peptide sequences [451].

Similar FDR issues arise when too many variable modifications are allowed for database searches due to an exponential enlargement of the combinatorial search space and a concomitantly increased likelihood of random matches [259, 423]. It is conceivable that this to some extent also applied to the PTM data of pulsed SILAC-TMT experiments since 5 to 6 variable modifications were included (e.g. K8, R10, ac-K, ac-K8, ox-M, ac-Nterm for acetylomes), and score and delta score cut-

offs for modified peptides were lowered to default values for non-modified sequences to avoid penalizing heavy labelled peptides. Subsequent data processing and filtering likely reduced false identifications (see also Chapter IV and V), and global insights will not be distorted by slightly higher FDRs, but single observations still can be wrong. More generally, this applies to all probability-based searches and pinpoints that individual findings that are highlighted when reporting results require careful examination of the identification and quantification quality on a case-by-case basis. This can be of particular relevance when it comes to localization of PTMs, which is oftentimes accounted for by applying additional filters for localization probabilities on site level (e.g. >0.75 for so-called class I sites [263]). Noteworthy, in the presented work, identifications of modified peptides were not filtered for localization probabilities. Likewise, for integration of data on site level, always only the most likely site within an identified peptide was included irrespective of its localization probability. Second (and third etc.) most likely sites were not considered. This practice was based on the following considerations: (i) For global, peptide-centric analyses, the simple information that a modification is present was sufficient and the exact localization was initially negligible; (ii) Acetylation and Ubiquitin-remnants typically anyway showed overall very high localization probabilities (average close to 1) due to the limited number of lysine residues within tryptic peptides; (iii) Filtering for localization probability would have led to an underrepresentation of phosphopeptides with (multiple) neighbouring serine and threonine residues since precise localization is particularly challenging within such sequences. One could even argue that, from a biological perspective, it may be of minor importance where exactly a phosphorylation occurs within a serine and threonine-rich sequence patch. Still, this strategy inevitably will result in an increased number of false site localizations compared to more stringent filtering approaches. Synthetic peptide pools (employed for dynamic range analyses in Chapter III), for instance, showed 3 to 14 % false localizations for MS3-TMT samples without filtering, which decreased to 2 to 11 % after filtering for a localization probability higher than 0.75 (6-16 % vs. 4-10 % for label free quantification). Importantly, all examples for individual modification sites that were pointed out to alter turnover (Chapter V) were examined for localization probabilities of all evidence entries included for their quantification. Localization probabilities were typically higher than 0.90 with the exception of the phosphodegron in UNG for which localization probabilities of 0.53 to 0.76 were observed for serine 64 (due to the neighbouring serine 63).

In recent years, so-called mass-tolerant, open, or blind database searches have gained attention [255, 452, 453]. Here, MS1 mass tolerances are increased and thus matching of peptides with unknown modifications is allowed within several 100 Da of mass difference. Thereby, the identification of unexpected and novel modifications reported as delta masses is facilitated. Open searches can provide hints on general differences in the modification landscape across tissues [454] and be very useful to assess potential artefacts of different sample processing protocols [455]. However, the likelihood of random matches within these searches again is increased limiting their broad applicability. It is important to point out that false identifications can also emerge from the omission of modifications that are present in samples in significant quantities, most notably if very similar modifications are allowed in the search at the same time. This relation forms the basis of the on-going debate as to whether the identifications of non-canonical phosphorylations on histidine, aspartate, glutamate, lysine, arginine, and cysteine using mass spectrometry are genuine [65]. The relevance of such false positive hits was illustrated in overlabelling searches during optimization of the TMT protocol (Chapter III). A high number of TMT-labelled histidine residues was identified (within S/T/Y peptides) when TMT-labelled serine, threonine and tyrosine residues were excluded from the list of variable modifications. It becomes clear that employed

search parameters and sequence databases are crucial for the quality of identification and subsequent quantification of MS data. The completeness and the size of the search space have to be balanced carefully and adapted appropriately to the specific sample type.

**Missing functional links** – Databases searches typically yield large amounts of identifications for proteins and modifications that have to be put into meaningful biological context. To aid interpretation of proteomic data, numerous knowledge repositories provide information on protein functions, structures, localizations, and interactions, and on functions, effects, and writers of modification sites [43, 285, 286, 289-294]. While such resources are tremendously helpful, they are naturally incomplete despite all efforts to assemble as many curated information from experiments and published literature as possible. This becomes especially apparent for PTM databases such as PhosphoSitePlus [294], which is often utilized to annotate regulatory functions of identified protein modifications. Typically, MS experiments detect many modification sites that are not listed in the database and even those that are included rarely provide an assigned regulatory function (e.g. <10 % of PTMs on human proteins in PhosphoSitePlus). In addition, such databases are inherently biased in their information content towards certain well-studied proteins thereby complicating the discovery of hitherto unknown regulatory mechanisms. Even though not all PTMs may necessarily feature a specific biological function at all, it is rather unlikely that most modification events are just "noise in the system" [456]. Yet, unravelling regulatory site functions can be laborious, time-consuming, and sometimes unsuccessful, illustrating the need for methods that can spot and prioritize functionally relevant modification sites in a time- and cost-efficient manner. Measuring the turnover of modified peptidoforms could be one such approach. An integration of sites regulating peptide turnover and sites changing, for instance, in certain disease states or upon drug treatment could greatly facilitate the decision on which site alterations may be worthwhile to follow up on. Increasing the depth of the modification-specific protein turnover analysis and expanding it to other cell systems and modifications could considerably supplement the here presented modification-regulated turnover atlas and further facilitate identification of functionally relevant modification sites.

During exploitation of turnover data, it was also attempted to integrate information on writer and eraser relationships with detected modification sites. However, similar to regulatory functions, annotations of (de)modifying enzymes are sparse in data repositories. Ultimately, three recently published datasets were utilized, one measuring the acetylome response after treatment with 19 KDAC inhibitors [310], the second assessing alteration in the acetylome after knockout or selective inhibition of CBP/p300 [208], and the third screening for kinase substrates *in vitro* [309]. In the latter approach, HeLa lysate was dephosphorylated using alkaline phosphatase, heat-treated to inactivate the added enzyme, and then supplemented with one of 385 purified kinases. Consequently, the information about substrate structure and intracellular localization is lost, which can limit the identification of biologically meaningful kinases substrates. On the other hand, drug treatment approaches take place in the cellular context, but the correct inference of substrates can be complicated by a limited selectivity of utilized drugs or secondary effects caused by downstream signalling especially for longer treatment times (here 16 h). Indeed, many HDAC inhibitors are assumed to feature poor selectivity [310] and even many allegedly very selective drugs have been found to be remarkably unselective [457]. Hence, all these approaches suffer from some inherent biases, but they still represent an enormous step forward for the elucidation of enzyme-substrate relations, and especially the HDAC inhibitor screen proved useful for the interpretation of the decreased turnover observed for ac-KP sites.

Altogether, this illustrates that the analysis of proteoforms with proteomics and, above all, their functionalization pose major analytical challenges. Despite much progress made in recent years, further technical improvements are expected to accelerate the speed at which new biological knowledge is gained in the future.

## 3 New proteomic technologies on the rise

Several challenges in proteoform analyses, such as the unambiguous identification of single isoforms or the robust quantification of individual (modified) peptides across multiple conditions, can be tackled partly by enhancing the depth, reproducibility, consistency, and specificity of bottom-up proteomic measurements. In this context, several (new) approaches for data acquisition and subsequent data analyses have shown enormous capabilities to broaden and deepen overall peptide and thus proteoform coverage.

**Alternative strategies for data acquisition** – An acquisition regime that is specifically aiming at alleviating missing quantitative values across label-free samples, is represented by data independent acquisition (DIA). The idea of DIA has been established 15 years ago [458], but due to improved instrumentation and computational strategies it is gaining attention recently. In contrast to DDA, DIA methods record a comprehensive map of all fragment ions over retention time without the selection of certain MS1 precursors and independent of the sample content. This is typically achieved by repeated cycling through several consecutive and comparably wide isolation and fragmentation windows (usually >10 m/z) which together completely cover a desired MS1 mass range [459]. While DIA can theoretically overcome the inconsistent sampling of DDA, it entails complex data analysis workflows that have to disentangle the highly chimeric spectra produced by co-fragmentation of 10s to 100s of peptide precursors. Further, the analysis of DIA data commonly relies on spectral libraries containing information on previously acquired fragment spectra and retention times of peptides that are expected to be present in the sample [459]. Generation of appropriate decoy spectral libraries considering all peptide query parameters is far from being trivial and it is under debate whether current approaches [460-462] underestimate true FDRs. Nevertheless, recent studies yielded excellent proteome coverage with >10,000 proteins in a 6 h single shot [462] illustrating the great potential of DIA analyses for comprehensive proteome interrogation in the future.

The BoxCar acquisition method has achieved similar coverage within 100 min gradient time and increased the number of peptides consistently quantified across ten replicates by >40 % compared to standard DDA [463]. It successfully combines the principles of gas-phase fractionation [464] and match-between-runs [255, 465]. The method essentially relies on compiling an accurate m/z and retention time map with high dynamic range and transferring peptide identifications from a library that was previously acquired using deeply fractionated samples. More precisely, the entire MS1 mass range is sub-divided (‘fractionated’) into multiple narrow segments (‘boxes’) that are adjusted for the overall m/z density and recorded in subsequent MS1 scans. Hence, the total number of collected ions per MS1 scan is distributed among a smaller number of precursors thereby increasing the signal-to-noise and dynamic range. Meanwhile overall longer MS1 acquisition times are balanced by reducing the number of MS2 scans. This makes the identification performance highly dependent on the project-specific peptide library and a reproducible chromatography to ensure a high quality of the feature matching procedure. Noteworthy, currently no control for false matches is implemented. Yet, the BoxCar method demonstrates how the combination of

different concepts can considerably improve proteome coverage within single shot analyses and might be especially useful when only limited sample amounts are available.

Other acquisition strategies that have shown improvements for peptide and protein coverage aim at enhancing identification rates and utilizing MS measurement time more efficiently via smart acquisition schemes. Most of these approaches are facilitated by decision tree-based MS methods that allow flexible settings for instrument parameters such as injection times and fragmentation modes depending on features of peptide precursors. This is exemplified by the CHOPIN (CHarge Ordered Parallel Ion aNalysis) method, which can readily be implemented on Lumos instruments. It routes precursors for fragment spectra acquisition to either the Orbitrap or the ion trap depending on their charge state and intensity [466]. Thereby, the advantages of both mass analysers are exploited in an optimal fashion. The concept of intelligent data acquisition is taken one step further by methods that employ real-time database searches to aid a decision on fragmentation of precursors. Recently, such approaches have demonstrated clear benefits for MS3-based measurements of TMT-labelled peptides [467]. Based on the identification of an MS2 spectrum, real-time decisions are taken on whether an MS3 scan should be triggered (only if the peptide was identified) and which fragments should be selected for the SPS (only fragments belonging to the target peptide and still carrying a TMT). Thus, needless and time-consuming MS3 scans are avoided, and both the peptide coverage and the quantification accuracy are improved markedly, which renders this novel method particularly attractive for PTM-subproteomes of pSILAC-TMT samples.

Quantification accuracy of isobaric tag-labelled samples further has been increased by utilizing ion mobility devices [278, 279]. Those minimize co-isolation by separating peptides based on their collisional cross sections (size and shape) in addition to retention time and m/z. Generally, the concept of ion mobility separation, which had been developed already half a century ago [468], has recently experienced a renaissance for proteomic applications. The additional dimension of separation has shown merit also for label-free samples in certain settings [469, 470] and is now available on new instrument platforms [471, 472] enabling a broader application in the future.

**Advanced approaches for data processing** – Above outlined DIA and BoxCar methods illustrate that data acquisition and processing approaches are often closely linked. Yet, there are several strategies for sophisticated analyses that are suitable for different types of raw data irrespective of the acquisition schemes. An example can be provided by the prediction of fragment spectra including accurate intensity information that can enhance the confidence in a PSM compared to mere m/z information. However, the rules dictating intensity ratios for a given peptide sequence, charge state, fragmentation mode, and energy are in their entirety so complex that they cannot be grasped by manual exploration of spectra. Hence, it was only recently that the deep neural networks like Prosit, which was trained on large amounts of spectra from synthetic peptides [231], accomplished the goal to compile high-quality spectra and retention time information for any desired (non-modified) peptide, also those that were not present in the training dataset [473-475]. The *in silico* generated spectra have demonstrated usability for DIA analyses and drastically improved the discrimination of target from decoy hits in DDA approaches thereby partly overcoming the above mentioned limitations of very large search spaces [475]. Further expansions of such neural networks to post-translationally (and chemically) modified peptides can incredibly enhance the confidence with which PTMs can be identified and thus have enormous potential to reshape how analyses of acquired spectra will be performed in the future.

Similar to the spectra identification strategy by Prosit, several advanced approaches for quantification are useful independent of specialized data acquisition. The processing pipeline of the IonStar method [476], for instance, should be applicable to any label-free DDA data (or other MS1 based quantification methods). Via chromatographic alignment in the 3D space (retention time, m/z, intensity), sensitive feature detection, and subsequent stringent outlier filtering, IonStar has achieved remarkably better sensitivity, accuracy, and precision, and a much lower rate of missing values (<0.2 % across 100 samples) compared to other processing software for label-free data. Likewise, a data processing procedure for enhanced detection of large SILAC ratios has been reported recently [114]. It was able to better disentangle overlapping isotope peaks and thus considerably increased coverage and quality of turnover estimations for very long-lived proteins. Hence, this method also may be able to improve the depth and quantitative accuracy for N/O peptide ratios derived from single time-point pulses, in particular for the 1 h SILAC pulse (see Chapter V). It has yet to be examined how these algorithms behave regarding peptide-level quantification for peptidoform analyses (e.g. the outlier filtering in the IonStar workflow could remove potentially interesting peptides), but generally these examples illustrate the crucial contribution of sophisticated computational methods to the overall performance of proteomic workflows.

## 4 The future of proteomics: Higher, faster, further?

Owing to the ongoing rapid advances in technologies, MS-based proteomics has revolutionized the breadth and depth with which protein expression can be studied in diverse cellular systems. However, functional interpretation of the large amounts of data generated from proteomic profiling oftentimes remains difficult. A major future challenge for biomedical research, in general, will be posed by the systematic assignment of protein functions and dysfunctions to health and disease states [16]. The integration of results from different techniques and research questions (e.g. proteoform turnover and PTM responses upon drug treatment) could hold great potential to facilitate the elucidation of functional relations and underlying causes of disease. To this end, advanced computational methods will be indispensable to enable systematic data analysis and integration. Very recently, there have been astonishing developments in the area of computational proteomics especially regarding deep learning, and this field will very likely become a new important driver of future progress. Besides the mentioned artificial neuronal networks for spectra prediction, computational deep learning approaches are emerging for all kind of biomedical research including drug discovery and structural biology [477]. It is conceivable that advanced machine learning algorithms will soon also extensively aid and accelerate data interpretation on a functional level. Yet, the usefulness of such algorithms will considerably dependent on the quality of training data and the accessibility for non-computer scientist, which both necessitates a close collaboration of scientists with distinct expertise. Ultimately, computational approaches can be of incredible help for generation of hypothesis, but their validation still has to be carried out in a real biological setting.

Further advancements in the understanding of molecular mechanisms underlying different disease states on the level of proteomes could also enable a broader and more immediate utility of proteomics methodologies for clinical applications. While genomic approaches are increasingly involved in the clinical decision-making for cancer therapies, clinical cancer proteomics so far mostly has been a retrospective analysis of patient cohorts, which typically merely results in the classification of molecular subgroups in an attempt to identify therapeutic targets [274, 340]. However, the ultimate goal for clinical proteomics should be to provide meaningful guidance on

treatment decisions in real-time. Yet, many questions regarding time, costs, details of implementation, and most importantly benefit remain to be answered before proteomic approaches can enter the daily clinical routine and lead to recommendations for therapeutic interventions based on proteomic profiles of a patients' tumour material. Despite all these challenges, recent innovations in sample preparation, instrumentation, and computational proteomics justify cautious optimism that the application of proteomics for precision medicine is within reach.

# BIBLIOGRAPHY


1. Crick F: **Central Dogma of Molecular Biology**. *Nature* 1970, **227**(5258):561-563.
2. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H *et al*: **Mass-spectrometry-based draft of the human proteome**. *Nature* 2014, **509**:582-587.
3. Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S *et al*: **A draft map of the human proteome**. *Nature* 2014, **509**:575-581.
4. Dickson D: **Gene estimates rises as US and UK discuss freedom of access**. *Nature* 1999, **401**(311).
5. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al*: **Initial sequencing and analysis of the human genome**. *Nature* 2001, **409**(6822):860-921.
6. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA *et al*: **The Sequence of the Human Genome**. *Science* 2001, **291**(5507):1304-1351.
7. International Human Genome Sequencing C: **Finishing the euchromatic sequence of the human genome**. *Nature* 2004, **431**(7011):931-945.
8. Salzberg SL: **Open questions: How many genes do we have?** *BMC Biol* 2018, **16**(1):94-94.
9. Basrai MA, Hieter P, Boeke JD: **Small Open Reading Frames: Beautiful Needles in the Haystack**. *Genome Research* 1997, **7**(8):768-771.
10. Saghatelian A, Couso JP: **Discovery and characterization of smORF-encoded bioactive polypeptides**. *Nature Chemical Biology* 2015, **11**:909-916.
11. Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL, Humphery-Smith I: **Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium**. *ELECTROPHORESIS* 1995, **16**(1):1090-1094.
12. Licatalosi DD, Darnell RB: **RNA processing and its regulation: global insights into biological networks**. *Nature Reviews Genetics* 2010, **11**:75-87.
13. de Klerk E, 't Hoen PAC: **Alternative mRNA transcription, processing, and translation: insights from RNA sequencing**. *Trends in Genetics* 2015, **31**(3):128-139.
14. Deribe YL, Pawson T, Dikic I: **Post-translational modifications in signal integration**. *Nature Structural & Molecular Biology* 2010, **17**:666-672.
15. Lisitsa A, Moshkovskii S, Chernobrovkin A, Ponomarenko E, Archakov A: **Profiling proteoforms: promising follow-up of proteomics for biomarker discovery**. *Expert Review of Proteomics* 2014, **11**(1):121-129.
16. Aebersold R, Agar JN, Amster IJ, Baker MS, Bertozzi CR, Boja ES, Costello CE, Cravatt BF, Fenselau C, Garcia BA *et al*: **How many human proteoforms are there?** *Nature Chemical Biology* 2018, **14**:206-214.
17. Smith LM, Kelleher NL, Proteomics CfTD: **Proteoform: a single term describing protein complexity**. *Nature methods* 2013, **10**(3):186-187.
18. Tan H, Bao J, Zhou X: **Genome-wide mutational spectra analysis reveals significant cancer-specific heterogeneity**. *Scientific Reports* 2015, **5**:12566.
19. Wu J-R, Zeng R: **Molecular basis for population variation: From SNPs to SAPs**. *FEBS Letters* 2012, **586**(18):2841-2845.
20. Tonegawa S: **Somatic generation of antibody diversity**. *Nature* 1983, **302**(5909):575-581.
21. Cooper MD, Alder MN: **The Evolution of Adaptive Immune Systems**. *Cell* 2006, **124**(4):815-822.
22. Berget SM, Moore C, Sharp PA: **Spliced segments at the 5' terminus of adenovirus 2 late mRNA**. *Proceedings of the National Academy of Sciences* 1977, **74**(8):3171-3175.
23. Chow LT, Gelinas RE, Broker TR, Roberts RJ: **An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA**. *Cell* 1977, **12**(1):1-8.
24. Beadle GW, Tatum EL: **Genetic Control of Biochemical Reactions in Neurospora**. *Proceedings of the National Academy of Sciences* 1941, **27**(11):499-506.
25. Ingram VM: **Gene Mutations in Human Hemoglobin: the Chemical Difference Between Normal and Sickle Cell Hemoglobin**. *Nature* 1957, **180**(4581):326-328.
26. Gilbert W: **Why genes in pieces?** *Nature* 1978, **271**(5645):501-501.
27. Wahl MC, Will CL, Lührmann R: **The Spliceosome: Design Principles of a Dynamic RNP Machine**. *Cell* 2009, **136**(4):701-718.
28. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq**. *Nature Methods* 2008, **5**:621-628.

29. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB: **Alternative isoform regulation in human tissue transcriptomes**. *Nature* 2008, **456**:470-476.
30. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ: **Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing**. *Nature Genetics* 2008, **40**:1413-1415.
31. Pickrell JK, Pai AA, Gilad Y, Pritchard JK: **Noisy splicing drives mRNA isoform diversity in human cells**. *PLoS genetics* 2010, **6**(12):e1001236-e1001236.
32. Shabalina SA, Ogurtsov AY, Spiridonov NA, Koonin EV: **Evolution at protein ends: major contribution of alternative transcription initiation and termination to the transcriptome and proteome diversity in mammals**. *Nucl Acids Res* 2014, **42**(11):7132-7144.
33. Reyes A, Huber W: **Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues**. *Nucl Acids Res* 2018, **46**(2):582-592.
34. Benne R, Van Den Burg J, Brakenhoff JPJ, Sloof P, Van Boom JH, Tromp MC: **Major transcript of the frameshifted coxll gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA**. *Cell* 1986, **46**(6):819-826.
35. Gott JM, Emeson RB: **Functions and mechanisms of RNA editing**. *Annual Review of Genetics* 2000, **34**(1):499-531.
36. Rosenthal JJC: **The emerging role of RNA editing in plasticity**. *The Journal of Experimental Biology* 2015, **218**(12):1812-1821.
37. Picardi E, Manzari C, Mastropasqua F, Aiello I, D'Erchia AM, Pesole G: **Profiling RNA editing in human tissues: towards the inosinome Atlas**. *Scientific Reports* 2015, **5**:14941.
38. Picardi E, D'Erchia AM, Lo Giudice C, Pesole G: **REDIportal: a comprehensive database of A-to-I RNA editing events in humans**. *Nucl Acids Res* 2017, **45**(D1):D750-D757.
39. Kochetov AV: **Alternative translation start sites and hidden coding potential of eukaryotic mRNAs**. *BioEssays* 2008, **30**(7):683-691.
40. Rogozin IB, Kochetov AV, Kondrashov FA, Koonin EV, Milanesi L: **Presence of ATG triplets in 5' untranslated regions of eukaryotic cDNAs correlates with a 'weak' context of the start codon**. *Bioinformatics* 2001, **17**(10):890-900.
41. Lee S, Liu B, Lee S, Huang S-X, Shen B, Qian S-B: **Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution**. *Proceedings of the National Academy of Sciences* 2012, **109**(37):E2424-E2432.
42. Kozak M: **Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes**. *Cell* 1986, **44**(2):283-292.
43. The UniProt Consortium: **UniProt: the universal protein knowledgebase**. *Nucl Acids Res* 2016, **45**(D1):D158-D169.
44. Lipmann FA, Levene PA: **Serinephorphoric acid obtained on hydrolysis of vitellinic acid**. *Journal of Biological Chemistry* 1932, **98**(1):109-114.
45. Phillips DM: **The presence of acetyl groups of histones**. *Biochem J* 1963, **87**(2):258-263.
46. Burnett G, Kennedy EP: **The enzymatic phosphorylation of proteins**. *Journal of Biological Chemistry* 1954, **211**(2):969-980.
47. Kleff S, Andrulis ED, Anderson CW, Sternglanz R: **Identification of a Gene Encoding a Yeast Histone H4 Acetyltransferase**. *Journal of Biological Chemistry* 1995, **270**(42):24674-24677.
48. Walsh CT, Garneau-Tsodikova S, Gatto Jr. GJ: **Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications**. *Angewandte Chemie International Edition* 2005, **44**(45):7342-7372.
49. Varland S, Osberg C, Arnesen T: **N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects**. *Proteomics* 2015, **15**(14):2385-2401.
50. Bernd W, Guido K, Klaus S: **Functions of Propeptide Parts in Cysteine Proteases**. *Current Protein & Peptide Science* 2003, **4**(5):309-326.
51. Canaff L, Bennett HPJ, Hendy GN: **Peptide hormone precursor processing: getting sorted?** *Molecular and Cellular Endocrinology* 1999, **156**(1):1-6.
52. Paetzel M, Karla A, Strynadka NCJ, Dalbey RE: **Signal Peptidases**. *Chemical Reviews* 2002, **102**(12):4549-4580.
53. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S: **The Protein Kinase Complement of the Human Genome**. *Science* 2002, **298**(5600):1912-1934.
54. Alonso A, Sasin J, Bottini N, Friedberg I, Friedberg I, Osterman A, Godzik A, Hunter T, Dixon J, Mustelin T: **Protein Tyrosine Phosphatases in the Human Genome**. *Cell* 2004, **117**(6):699-711.
55. Shi Y: **Serine/Threonine Phosphatases: Mechanism through Structure**. *Cell* 2009, **139**(3):468-484.

56. Ubersax JA, Ferrell Jr JE: **Mechanisms of specificity in protein phosphorylation**. *Nat Rev Mol Cell Biol* 2007, **8**:530-541.
57. Roy J, Cyert MS: **Cracking the Phosphatase Code: Docking Interactions Determine Substrate Specificity**. *Science Signaling* 2009, **2**(100):re9.
58. Mandell DJ, Chorny I, Groban ES, Wong SE, Levine E, Rapp CS, Jacobson MP: **Strengths of Hydrogen Bonds Involving Phosphorylated Amino Acid Side Chains**. *Journal of the American Chemical Society* 2007, **129**(4):820-827.
59. Yaffe MB: **Phosphotyrosine-binding domains in signal transduction**. *Nat Rev Mol Cell Biol* 2002, **3**(3):177-186.
60. Reinhardt HC, Yaffe MB: **Phospho-Ser/Thr-binding domains: navigating the cell cycle and DNA damage response**. *Nat Rev Mol Cell Biol* 2013, **14**:563-580.
61. Olsen JV, Vermeulen M, Santamaria A, Kumar C, Miller ML, Jensen LJ, Gnad F, Cox J, Jensen TS, Nigg EA *et al*: **Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy During Mitosis**. *Science Signaling* 2010, **3**(104):ra3.
62. Batth TS, Papetti M, Pfeiffer A, Tollenaere MAX, Francavilla C, Olsen JV: **Large-Scale Phosphoproteomics Reveals Shp-2 Phosphatase-Dependent Regulators of Pdgf Receptor Signaling**. *Cell Reports* 2018, **22**(10):2784-2796.
63. Sharma K, D'Souza Rochelle CJ, Tyanova S, Schaab C, Wiśniewski Jacek R, Cox J, Mann M: **Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling**. *Cell Reports* 2014, **8**(5):1583-1594.
64. Hunter T: **Tyrosine phosphorylation: thirty years and counting**. *Current opinion in cell biology* 2009, **21**(2):140-146.
65. Hardman G, Perkins S, Brownridge PJ, Clarke CJ, Byrne DP, Campbell AE, Kalyuzhnyy A, Myall A, Eyers PA, Jones AR *et al*: **Strong anion exchange-mediated phosphoproteomics reveals extensive human non-canonical phosphorylation**. *The EMBO Journal*, **38**:e100847.
66. Steeg PS, Palmieri D, Ouatas T, Salerno M: **Histidine kinases and histidine phosphorylated proteins in mammalian cell biology, signal transduction and cancer**. *Cancer Letters* 2003, **190**(1):1-12.
67. Brown JL, Roberts WK: **Evidence that approximately eighty per cent of the soluble proteins from Ehrlich ascites cells are Nalpha-acetylated**. *Journal of Biological Chemistry* 1976, **251**(4):1009-1014.
68. Drazic A, Myklebust LM, Ree R, Arnesen T: **The world of protein acetylation**. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 2016, **1864**(10):1372-1401.
69. Allfrey VG, Faulkner R, Mirsky AE: **Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis**. *Proceedings of the National Academy of Sciences of the United States of America* 1964, **51**(5):786-794.
70. Narita T, Weinert BT, Choudhary C: **Functions and mechanisms of non-histone protein acetylation**. *Nat Rev Mol Cell Biol* 2019, **20**(3):156-174.
71. Paik WK, Pearson D, Lee HW, Kim S: **Nonenzymatic acetylation of histones with acetyl-CoA**. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* 1970, **213**(2):513-522.
72. Schwer B, Eckersdorff M, Li Y, Silva JC, Fermin D, Kurtev MV, Giallourakis C, Comb MJ, Alt FW, Lombard DB: **Calorie restriction alters mitochondrial protein acetylation**. *Aging Cell* 2009, **8**(5):604-606.
73. Wagner GR, Payne RM: **Widespread and Enzyme-independent Nε-Acetylation and Nε-Succinylation of Proteins in the Chemical Conditions of the Mitochondrial Matrix**. *Journal of Biological Chemistry* 2013, **288**(40):29036-29045.
74. Fujisawa T, Filippakopoulos P: **Functions of bromodomain-containing proteins and their roles in homeostasis and cancer**. *Nat Rev Mol Cell Biol* 2017, **18**:246-262.
75. Li Y, Wen H, Xi Y, Tanaka K, Wang H, Peng D, Ren Y, Jin Q, Dent Sharon YR, Li W *et al*: **AF9 YEATS Domain Links Histone Acetylation to DOT1L-Mediated H3K79 Methylation**. *Cell* 2014, **159**(3):558-571.
76. Komander D, Rape M: **The Ubiquitin Code**. *Annual Review of Biochemistry* 2012, **81**(1):203-229.
77. Deol KK, Lorenz S, Strieter ER: **Enzymatic Logic of Ubiquitin Chain Assembly**. *Frontiers in Physiology* 2019, **10**:835.
78. Dikic I, Wakatsuki S, Walters KJ: **Ubiquitin-binding domains — from structures to functions**. *Nat Rev Mol Cell Biol* 2009, **10**:659-671.
79. Komander D, Clague MJ, Urbé S: **Breaking the chains: structure and function of the deubiquitinases**. *Nat Rev Mol Cell Biol* 2009, **10**:550-563.

80. Ponomarenko EA, Poverennaya EV, Ilgisonis EV, Pyatnitskiy MA, Kopylov AT, Zgoda VG, Lisitsa AV, Archakov AI: **The Size of the Human Proteome: The Width and Depth**. *Int J Anal Chem* 2016, **2016**:7436849-7436849.
81. Phanstiel D, Brumbaugh J, Berggren WT, Conard K, Feng X, Levenstein ME, McAlister GC, Thomson JA, Coon JJ: **Mass spectrometry identifies and quantifies 74 unique histone H4 isoforms in differentiating human embryonic stem cells**. *Proceedings of the National Academy of Sciences* 2008, **105**(11):4093-4098.
82. Shyu A-B, Wilkinson MF, van Hoof A: **Messenger RNA regulation: to translate or to degrade**. *The EMBO Journal* 2008, **27**(3):471-481.
83. Wang F, Canadeo LA, Huibregtse JM: **Ubiquitination of newly synthesized proteins at the ribosome**. *Biochimie* 2015, **114**:127-133.
84. Harper JW, Bennett EJ: **Proteome complexity and the forces that drive proteome imbalance**. *Nature* 2016, **537**:328-338.
85. Yang X, Coulombe-Huntington J, Kang S, Sheynkman Gloria M, Hao T, Richardson A, Sun S, Yang F, Shen Yun A, Murray Ryan R *et al*: **Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing**. *Cell* 2016, **164**(4):805-817.
86. Kochetov AV, Sarai A, Rogozin IB, Shumny VK, Kolchanov NA: **The role of alternative translation start sites in the generation of human protein diversity**. *Molecular Genetics and Genomics* 2005, **273**(6):491-496.
87. Gandre S, Bercovich Z, Kahana C: **Mitochondrial localization of antizyme is determined by context-dependent alternative utilization of two AUG initiation codons**. *Mitochondrion* 2003, **2**(4):245-256.
88. Bab I, Smith E, Gavish H, Attar-Namdar M, Chorev M, Chen Y-C, Muhlrad A, Birnbaum MJ, Stein G, Frenkel B: **Biosynthesis of Osteogenic Growth Peptide via Alternative Translational Initiation at AUG85 of Histone H4 mRNA**. *Journal of Biological Chemistry* 1999, **274**(20):14474-14481.
89. Prabakaran S, Lippens G, Steen H, Gunawardena J: **Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding**. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 2012, **4**(6):565-583.
90. Huse M, Kuriyan J: **The Conformational Plasticity of Protein Kinases**. *Cell* 2002, **109**(3):275-282.
91. Hunter T: **Signaling—2000 and Beyond**. *Cell* 2000, **100**(1):113-127.
92. van der Horst A, de Vries-Smits AMM, Brenkman AB, van Triest MH, van den Broek N, Colland F, Maurice MM, Burgering BMT: **FOXO4 transcriptional activity is regulated by monoubiquitination and USP7/HAUSP**. *Nature Cell Biology* 2006, **8**(10):1064-1073.
93. Hafner A, Bulyk ML, Jambhekar A, Lahav G: **The multiple mechanisms that regulate p53 activity and cell fate**. *Nat Rev Mol Cell Biol* 2019, **20**(4):199-210.
94. Ito A, Kawaguchi Y, Lai C-H, Kovacs JJ, Higashimoto Y, Appella E, Yao T-P: **MDM2-HDAC1-mediated deacetylation of p53 is required for its degradation**. *The EMBO journal* 2002, **21**(22):6236-6245.
95. Gu W, Roeder RG: **Activation of p53 Sequence-Specific DNA Binding by Acetylation of the p53 C-Terminal Domain**. *Cell* 1997, **90**(4):595-606.
96. Venne AS, Kollipara L, Zahedi RP: **The next level of complexity: Crosstalk of posttranslational modifications**. *Proteomics* 2014, **14**(4-5):513-524.
97. Schoenheimer R, Clarke HT: **The Dynamic State of Body Constituents**. *Harvard University Press, Cambridge, Massachusetts, USA* 1942.
98. Richter K, Haslbeck M, Buchner J: **The Heat Shock Response: Life on the Verge of Death**. *Molecular Cell* 2010, **40**(2):253-266.
99. Buchberger A, Bukau B, Sommer T: **Protein Quality Control in the Cytosol and the Endoplasmic Reticulum: Brothers in Arms**. *Molecular Cell* 2010, **40**(2):238-252.
100. Hetz C, Papa FR: **The Unfolded Protein Response and Cell Fate Control**. *Molecular Cell* 2018, **69**(2):169-181.
101. Cohen-Kaplan V, Livneh I, Avni N, Cohen-Rosenzweig C, Ciechanover A: **The ubiquitin-proteasome system and autophagy: Coordinated and independent activities**. *The International Journal of Biochemistry & Cell Biology* 2016, **79**:403-418.
102. Collins GA, Goldberg AL: **The Logic of the 26S Proteasome**. *Cell* 2017, **169**(5):792-806.
103. Thibaudeau TA, Smith DM: **A Practical Review of Proteasome Pharmacology**. *Pharmacological Reviews* 2019, **71**(2):170-197.
104. Xu H, Ren D: **Lysosomal physiology**. *Annu Rev Physiol* 2015, **77**:57-80.
105. Park C, Cuervo AM: **Selective Autophagy: Talking with the UPS**. *Cell Biochemistry and Biophysics* 2013, **67**(1):3-13.

106. Peth A, Nathan JA, Goldberg AL: **The ATP Costs and Time Required to Degrade Ubiquitinated Proteins by the 26 S Proteasome**. *Journal of Biological Chemistry* 2013, **288**(40):29215-29222.
107. Ravid T, Hochstrasser M: **Diversity of degradation signals in the ubiquitin–proteasome system**. *Nat Rev Mol Cell Biol* 2008, **9**:679-689.
108. Peth A, Uchiki T, Goldberg AL: **ATP-Dependent Steps in the Binding of Ubiquitin Conjugates to the 26S Proteasome that Commit to Degradation**. *Molecular Cell* 2010, **40**(4):671-681.
109. Kenniston JA, Baker TA, Fernandez JM, Sauer RT: **Linkage between ATP Consumption and Mechanical Unfolding during the Protein Processing Reactions of an AAA+ Degradation Machine**. *Cell* 2003, **114**(4):511-520.
110. Kisselev AF, Akopian TN, Woo KM, Goldberg AL: **The Sizes of Peptides Generated from Protein by Mammalian 26 and 20 S Proteasomes: Implications for understanding the degradative mechanism and antigen presentation**. *Journal of Biological Chemistry* 1999, **274**(6):3363-3371.
111. Tsvetkov P, Reuven N, Prives C, Shaul Y: **Susceptibility of p53 Unstructured N Terminus to 20 S Proteasomal Degradation Programs the Stress Response**. *Journal of Biological Chemistry* 2009, **284**(39):26234-26242.
112. Kirkin V, McEwan DG, Novak I, Dikic I: **A Role for Ubiquitin in Selective Autophagy**. *Molecular Cell* 2009, **34**(3):259-269.
113. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M: **Global quantification of mammalian gene expression control**. *Nature* 2011, **473**(7347):337-342.
114. Mathieson T, Franken H, Kosinski J, Kurzawa N, Zinn N, Sweetman G, Poeckel D, Ratnu VS, Schramm M, Becher I *et al*: **Systematic analysis of protein turnover in primary cells**. *Nature Communications* 2018, **9**:689.
115. Varshavsky A: **Naming a targeting signal**. *Cell* 1991, **64**(1):13-15.
116. Bachmair A, Finley D, Varshavsky A: **In Vivo Half-Life of a Protein is a Function of its Amino-Terminal Residue**. *Science* 1986, **234**(4773):179-186.
117. Dougan DA, Micevski D, Truscott KN: **The N-end rule pathway: From recognition by N-recognins, to destruction by AAA+proteases**. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 2012, **1823**(1):83-91.
118. Shemorry A, Hwang C-S, Varshavsky A: **Control of Protein Quality and Stoichiometries by N-Terminal Acetylation and the N-End Rule Pathway**. *Molecular Cell* 2013, **50**(4):540-551.
119. Mészáros B, Kumar M, Gibson TJ, Uyar B, Dosztányi Z: **Degrons in cancer**. *Science Signaling* 2017, **10**(470):eaak9982.
120. Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ, Pavletich NP: **Structure of the MDM2 Oncoprotein Bound to the p53 Tumor Suppressor Transactivation Domain**. *Science* 1996, **274**(5289):948-953.
121. Sakaguchi K, Saito Si, Higashimoto Y, Roy S, Anderson CW, Appella E: **Damage-mediated Phosphorylation of Human p53 Threonine 18 through a Cascade Mediated by a Casein 1-like Kinase: Effect on MDM2 binding**. *Journal of Biological Chemistry* 2000, **275**(13):9278-9283.
122. Carrano AC, Eytan E, Hershko A, Pagano M: **SKP2 is required for ubiquitin-mediated degradation of the CDK inhibitor p27**. *Nature Cell Biology* 1999, **1**(4):193-199.
123. Hon W-C, Wilson MI, Harlos K, Claridge TDW, Schofield CJ, Pugh CW, Maxwell PH, Ratcliffe PJ, Stuart DI, Jones EY: **Structural basis for the recognition of hydroxyproline in HIF-1α by pVHL**. *Nature* 2002, **417**(6892):975-978.
124. van der Lee R, Lang B, Kruse K, Gsponer J, Sánchez de Groot N, Huynen Martijn A, Matouschek A, Fuxreiter M, Babu MM: **Intrinsically Disordered Segments Affect Protein Half-Life in the Cell and during Evolution**. *Cell Reports* 2014, **8**(6):1832-1844.
125. Tompa P, Prilusky J, Silman I, Sussman JL: **Structural disorder serves as a weak signal for intracellular protein degradation**. *Proteins* 2008, **71**(2):903-909.
126. Murata S, Minami Y, Minami M, Chiba T, Tanaka K: **CHIP is a chaperone-dependent E3 ligase that ubiquitylates unfolded protein**. *EMBO reports* 2001, **2**(12):1133-1138.
127. Juszkiewicz S, Hegde RS: **Quality Control of Orphaned Proteins**. *Molecular Cell* 2018, **71**(3):443-457.
128. Wang F, Durfee Larissa A, Huibregtse Jon M: **A Cotranslational Ubiquitination Pathway for Quality Control of Misfolded Proteins**. *Molecular Cell* 2013, **50**(3):368-378.
129. Smith MH, Ploegh HL, Weissman JS: **Road to Ruin: Targeting Proteins for Degradation in the Endoplasmic Reticulum**. *Science* 2011, **334**(6059):1086-1090.
130. Hartl FU, Bracher A, Hayer-Hartl M: **Molecular chaperones in protein folding and proteostasis**. *Nature* 2011, **475**(7356):324-332.

131. Hipp MS, Park S-H, Hartl FU: **Proteostasis impairment in protein-misfolding and -aggregation diseases**. *Trends in Cell Biology* 2014, **24**(9):506-514.
132. Holt LJ: **Regulatory modules: Coupling protein stability to phopshoregulation during cell division**. *FEBS Letters* 2012, **586**(17):2773-2777.
133. Dikic I: **Mechanisms controlling EGF receptor endocytosis and degradation**. *Biochemical Society Transactions* 2003, **31**(6):1178-1181.
134. Asano S, Fukuda Y, Beck F, Aufderheide A, Förster F, Danev R, Baumeister W: **A molecular census of 26S proteasomes in intact neurons**. *Science* 2015, **347**(6220):439-442.
135. Ross CA, Poirier MA: **Protein aggregation and neurodegenerative disease**. *Nature Medicine* 2004, **10**:S10-S17.
136. Ekholm-Reed S, Goldberg MS, Schlossmacher MG, Reed SI: **Parkin-dependent degradation of the F-box protein Fbw7β promotes neuronal survival in response to oxidative stress by stabilizing Mcl-1**. *Molecular and cellular biology* 2013, **33**(18):3627-3643.
137. Lee Y, Karuppagounder SS, Shin J-H, Lee Y-I, Ko HS, Swing D, Jiang H, Kang S-U, Lee BD, Kang HC *et al*: **Parthanatos mediates AIMP2-activated age-dependent dopaminergic neuronal loss**. *Nature Neuroscience* 2013, **16**:1392-1400.
138. Chan NC, Salazar AM, Pham AH, Sweredoski MJ, Kolawa NJ, Graham RLJ, Hess S, Chan DC: **Broad activation of the ubiquitin–proteasome system by Parkin is critical for mitophagy**. *Human Molecular Genetics* 2011, **20**(9):1726-1737.
139. Winklhofer KF: **Parkin and mitochondrial quality control: toward assembling the puzzle**. *Trends in Cell Biology* 2014, **24**(6):332-341.
140. Ben-Zvi A, Miller EA, Morimoto RI: **Collapse of proteostasis represents an early molecular event in Caenorhabditis elegans aging**. *Proceedings of the National Academy of Sciences* 2009, **106**(35):14914-14919.
141. Back SH, Kaufman RJ: **Endoplasmic Reticulum Stress and Type 2 Diabetes**. *Annual Review of Biochemistry* 2012, **81**(1):767-793.
142. Özcan U, Yilmaz E, Özcan L, Furuhashi M, Vaillancourt E, Smith RO, Görgün CZ, Hotamisligil GS: **Chemical Chaperones Reduce ER Stress and Restore Glucose Homeostasis in a Mouse Model of Type 2 Diabetes**. *Science* 2006, **313**(5790):1137-1140.
143. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW: **Cancer Genome Landscapes**. *Science* 2013, **339**(6127):1546-1558.
144. Hanahan D, Weinberg Robert A: **Hallmarks of Cancer: The Next Generation**. *Cell* 2011, **144**(5):646-674.
145. Tsvetkov P, Adler J, Myers N, Biran A, Reuven N, Shaul Y: **Oncogenic addiction to high 26S proteasome level**. *Cell Death & Disease* 2018, **9**(7):773.
146. Crawford LJ, Walker B, Irvine AE: **Proteasome inhibitors in cancer therapy**. *J Cell Commun Signal* 2011, **5**(2):101-110.
147. Neckers L, Workman P: **Hsp90 Molecular Chaperone Inhibitors: Are We There Yet?** *Clinical Cancer Research* 2012, **18**(1):64-76.
148. Manasanch EE, Orlowski RZ: **Proteasome inhibitors in cancer therapy**. *Nature Reviews Clinical Oncology* 2017, **14**:417-433.
149. Schoenheimer R, Rittenberg D, Foster GL, Keston AS, Ratner S: **The application of nitrogen isotope N15 for the study of protein metabolism**. *Science* 1938, **88**(2295):599-600.
150. Pratt JM, Petty J, Riba-Garcia I, Robertson DHL, Gaskell SJ, Oliver SG, Beynon RJ: **Dynamics of Protein Turnover, a Missing Dimension in Proteomics**. *Molecular & Cellular Proteomics* 2002, **1**(8):579-591.
151. Dieterich DC, Link AJ, Graumann J, Tirrell DA, Schuman EM: **Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT)**. *Proceedings of the National Academy of Sciences* 2006, **103**(25):9482-9487.
152. Belle A, Tanay A, Bitincka L, Shamir R, O'Shea EK: **Quantification of protein half-lives in the budding yeast proteome**. *Proceedings of the National Academy of Sciences* 2006, **103**(35):13004-13009.
153. Yen H-CS, Xu Q, Chou DM, Zhao Z, Elledge SJ: **Global Protein Stability Profiling in Mammalian Cells**. *Science* 2008, **322**(5903):918-923.
154. Eden E, Geva-Zatorsky N, Issaeva I, Cohen A, Dekel E, Danon T, Cohen L, Mayo A, Alon U: **Proteome Half-Life Dynamics in Living Human Cells**. *Science* 2011, **331**(6018):764-768.
155. Khmelinskii A, Keller PJ, Bartosik A, Meurer M, Barry JD, Mardin BR, Kaufmann A, Trautmann S, Wachsmuth M, Pereira G *et al*: **Tandem fluorescent protein timers for in vivo analysis of protein dynamics**. *Nature Biotechnology* 2012, **30**:708-714.

156. Gilbert BE: **Protein turnover during maturation of mouse brain tissue**. *The Journal of Cell Biology* 1972, **53**(1):143-147.
157. Wheatley DN, Giddings MR, Inglis MS: **Kinetics of degradation of 'short-' and 'long-lived' proteins in cultured mammalian cells**. *Cell Biology International Reports* 1980, **4**(12):1081-1090.
158. Mosteller RD, Goldstein RV, Nishimoto KR: **Metabolism of individual proteins in exponentially growing Escherichia coli**. *Journal of Biological Chemistry* 1980, **255**(6):2524-2532.
159. Larrabee KL, Phillips JO, Williams GJ, Larrabee AR: **The relative rates of protein synthesis and degradation in a growing culture of Escherichia coli**. *Journal of Biological Chemistry* 1980, **255**(9):4125-4130.
160. Gerner C, Vejda S, Gelbmann D, Bayer E, Gotzmann J, Schulte-Hermann R, Mikulits W: **Concomitant Determination of Absolute Values of Cellular Protein Amounts, Synthesis Rates, and Turnover Rates by Quantitative Proteome Profiling**. *Molecular & Cellular Proteomics* 2002, **1**(7):528-537.
161. Jaleel A, Nair KS: **Identification of multiple proteins whose synthetic rates are enhanced by high amino acid levels in rat hepatocytes**. *American Journal of Physiology-Endocrinology and Metabolism* 2004, **286**(6):E950-E957.
162. Ahmad Y, Boisvert F-M, Lundberg E, Uhlen M, Lamond AI: **Systematic Analysis of Protein Pools, Isoforms, and Modifications Affecting Turnover and Subcellular Localization**. *Molecular & Cellular Proteomics* 2012, **11**(3):M111.013680.
163. Gawron D, Ndah E, Gevaert K, Van Damme P: **Positional proteomics reveals differences in N-terminal proteoform stability**. 2016, **12**(2):858-858.
164. Zinck R, Cahill MA, Kracht M, Sachsenmaier C, Hipskind RA, Nordheim A: **Protein synthesis inhibitors reveal differential regulation of mitogen-activated protein kinase and stress-activated protein kinase pathways that converge on Elk-1**. *Molecular and Cellular Biology* 1995, **15**(9):4930-4938.
165. Alber AB, Paquet ER, Biserni M, Naef F, Suter DM: **Single Live Cell Monitoring of Protein Turnover Reveals Intercellular Variability and Cell-Cycle Dependence of Degradation Rates**. *Molecular Cell* 2018, **71**(6):1079-1091.e1079.
166. Khmelinskii A, Meurer M, Ho C-T, Besenbeck B, Füller J, Lemberg MK, Bukau B, Mogk A, Knop M: **Incomplete proteasomal degradation of green fluorescent proteins in the context of tandem fluorescent protein timers**. *Molecular Biology of the Cell* 2016, **27**(2):360-370.
167. Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M: **Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics**. *Molecular & Cellular Proteomics* 2002, **1**(5):376-386.
168. Selbach M, Schwanhäusser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N: **Widespread changes in protein synthesis induced by microRNAs**. *Nature* 2008, **455**(7209):58-63.
169. Cambridge SB, Gnad F, Nguyen C, Bermejo JL, Krüger M, Mann M: **Systems-wide Proteomic Analysis in Mammalian Cells Reveals Conserved, Functional Protein Turnover**. *Journal of Proteome Research* 2011, **10**(12):5275-5284.
170. Boisvert F-M, Ahmad Y, Gierliński M, Charrière F, Lamont D, Scott M, Barton G, Lamond AI: **A Quantitative Spatial Proteomics Analysis of Proteome Turnover in Human Cells**. *Molecular & Cellular Proteomics* 2012, **11**(3):M111.011429.
171. Ly T, Endo A, Brenes A, Gierlinski M, Afzal V, Pawellek A, Lamond AI: **Proteome-wide analysis of protein abundance and turnover remodelling during oncogenic transformation of human breast epithelial cells**. *Wellcome Open Res* 2018, **3**:51.
172. Ma Y, McClatchy DB, Barkallah S, Wood WW, Yates JR: **HILAQ: A Novel Strategy for Newly Synthesized Protein Quantification**. *Journal of Proteome Research* 2017, **16**(6):2213-2220.
173. Beatty KE, Xie F, Wang Q, Tirrell DA: **Selective Dye-Labeling of Newly Synthesized Proteins in Bacterial Cells**. *Journal of the American Chemical Society* 2005, **127**(41):14150-14151.
174. Bagert JD, Xie YJ, Sweredoski MJ, Qi Y, Hess S, Schuman EM, Tirrell DA: **Quantitative, Time-Resolved Proteomic Analysis by Combining Bioorthogonal Noncanonical Amino Acid Tagging and Pulsed Stable Isotope Labeling by Amino Acids in Cell Culture**. *Molecular & Cellular Proteomics* 2014, **13**(5):1352-1358.
175. Schwanhäusser B, Gossen M, Dittmar G, Selbach M: **Global analysis of cellular protein translation by pulsed SILAC**. *Proteomics* 2009, **9**(1):205-209.
176. Jayapal KP, Sui S, Philp RJ, Kok Y-J, Yap MGS, Griffin TJ, Hu W-S: **Multitagging Proteomic Strategy to Estimate Protein Turnover Rates in Dynamic Systems**. *Journal of Proteome Research* 2010, **9**(5):2087-2097.

177. Aebersold R, Mann M: **Mass-spectrometric exploration of proteome structure and function**. *Nature* 2016, **537**:347-355.
178. Catherman AD, Skinner OS, Kelleher NL: **Top Down proteomics: facts and perspectives**. *Biochemical and biophysical research communications* 2014, **445**(4):683-693.
179. Schaffer LV, Millikin RJ, Miller RM, Anderson LC, Fellers RT, Ge Y, Kelleher NL, LeDuc RD, Liu X, Payne SH *et al*: **Identification and Quantification of Proteoforms by Mass Spectrometry**. *Proteomics* 2019, **19**(10):1800361.
180. Smith LM, Kelleher NL: **Proteoforms as the next proteomics currency**. *Science* 2018, **359**(6380):1106-1107.
181. Rosenberger G, Liu Y, Röst HL, Ludwig C, Buil A, Bensimon A, Soste M, Spector TD, Dermitzakis ET, Collins BC *et al*: **Inference and quantification of peptidoforms in large sample cohorts by SWATH-MS**. *Nature Biotechnology* 2017, **35**:781-788.
182. Mallick P, Kuster B: **Proteomics: a pragmatic perspective**. *Nature Biotechnology* 2010, **28**(7):695-709.
183. Shehadul Islam M, Aryasomayajula A, Selvaganapathy PR: **A Review on Macroscale and Microscale Cell Lysis Methods**. *Micromachines* 2017, **8**(3):83.
184. Médard G, Pachl F, Ruprecht B, Klaeger S, Heinzlmeir S, Helm D, Qiao H, Ku X, Wilhelm M, Kuehne T *et al*: **Optimized Chemical Proteomics Assay for Kinase Inhibitor Profiling**. *Journal of Proteome Research* 2015, **14**(3):1574-1586.
185. Sechi S, Chait BT: **Modification of Cysteine Residues by Alkylation. A Tool in Peptide Mapping and Protein Identification**. *Anal Chem* 1998, **70**(24):5150-5158.
186. Olsen JV, Ong S-E, Mann M: **Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues**. *Molecular & Cellular Proteomics* 2004, **3**(6):608-614.
187. Giansanti P, Aye Thin T, van den Toorn H, Peng M, van Breukelen B, Heck Albert JR: **An Augmented Multiple-Protease-Based Human Phosphopeptide Atlas**. *Cell Reports* 2015, **11**(11):1834-1843.
188. Wang D, Eraslan B, Wieland T, Hallström B, Hopf T, Zolg DP, Zecha J, Asplund A, Li L-h, Meng C *et al*: **A deep proteome and transcriptome abundance atlas of 29 healthy human tissues**. *Molecular systems biology* 2019, **15**(2):e8503.
189. Bekker-Jensen DB, Kelstrup CD, Batth TS, Larsen SC, Haldrup C, Bramsen JB, Sørensen KD, Høyer S, Ørntoft TF, Andersen CL *et al*: **An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes**. *Cell Systems* 2017, **4**(6):587-599.e584.
190. Manadas B, Mendes VM, English J, Dunn MJ: **Peptide fractionation in proteomics approaches**. *Expert Review of Proteomics* 2010, **7**(5):655-663.
191. Ruprecht B, Wang D, Chiozzi RZ, Li L-H, Hahne H, Kuster B: **Hydrophilic Strong Anion Exchange (hSAX) Chromatography Enables Deep Fractionation of Tissue Proteomes**. In: *Proteomics: Methods and Protocols.* Edited by Comai L, Katz JE, Mallick P. New York, NY: Springer New York; 2017: 69-82.
192. Rappsilber J, Mann M, Ishihama Y: **Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips**. *Nature Protocols* 2007, **2**:1896-1906.
193. Ducret A, Van Oostveen I, Eng JK, Yates JR, 3rd, Aebersold R: **High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry**. *Protein science : a publication of the Protein Society* 1998, **7**(3):706-719.
194. Mertins P, Tang LC, Krug K, Clark DJ, Gritsenko MA, Chen L, Clauser KR, Clauss TR, Shah P, Gillette MA *et al*: **Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography–mass spectrometry**. *Nature Protocols* 2018, **13**(7):1632-1661.
195. Doll S, Burlingame AL: **Mass Spectrometry-Based Detection and Assignment of Protein Posttranslational Modifications**. *ACS Chemical Biology* 2015, **10**(1):63-71.
196. Altelaar AFM, Munoz J, Heck AJR: **Next-generation proteomics: towards an integrative view of proteome dynamics**. *Nature Reviews Genetics* 2012, **14**:35.
197. Macek B, Mann M, Olsen JV: **Global and Site-Specific Quantitative Phosphoproteomics: Principles and Applications**. *Annual Review of Pharmacology and Toxicology* 2009, **49**(1):199-221.
198. Ruprecht B, Koch H, Medard G, Mundt M, Kuster B, Lemeer S: **Comprehensive and reproducible phosphopeptide enrichment using iron immobilized metal ion affinity chromatography (Fe-IMAC) columns**. *Molecular & Cellular Proteomics* 2015, **14**(1):205-215.
199. Porath J, Carlsson JAN, Olsson I, Belfrage G: **Metal chelate affinity chromatography, a new approach to protein fractionation**. *Nature* 1975, **258**(5536):598-599.
200. Andersson L, Porath J: **Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography**. *Analytical Biochemistry* 1986, **154**(1):250-254.

201. Ruprecht B, Koch H, Domasinska P, Frejno M, Kuster B, Lemeer S: **Optimized Enrichment of Phosphoproteomes by Fe-IMAC Column Chromatography**. In: *Proteomics: Methods and Protocols.* Edited by Comai L, Katz JE, Mallick P. New York, NY: Springer New York; 2017: 47-60.
202. Humphrey SJ, Karayel O, James DE, Mann M: **High-throughput and high-sensitivity phosphoproteomics with the EasyPhos platform**. *Nature Protocols* 2018, **13**(9):1897-1916.
203. Abe Y, Nagano M, Tada A, Adachi J, Tomonaga T: **Deep Phosphotyrosine Proteomics by Optimization of Phosphotyrosine Enrichment and MS/MS Parameters**. *Journal of Proteome Research* 2017, **16**(2):1077-1086.
204. Bian Y, Li L, Dong M, Liu X, Kaneko T, Cheng K, Liu H, Voss C, Cao X, Wang Y *et al*: **Ultra-deep tyrosine phosphoproteomics enabled by a phosphotyrosine superbinder**. *Nature Chemical Biology* 2016, **12**:959-966.
205. Hansen BK, Gupta R, Baldus L, Lyon D, Narita T, Lammers M, Choudhary C, Weinert BT: **Analysis of human acetylation stoichiometry defines mechanistic constraints on protein regulation**. *Nature Communications* 2019, **10**(1):1055.
206. Shaw PG, Chaerkady R, Zhang Z, Davidson NE, Pandey A: **Monoclonal antibody cocktail as an enrichment tool for acetylome analysis**. *Anal Chem* 2011, **83**(10):3623-3626.
207. Svinkina T, Gu H, Silva JC, Mertins P, Qiao J, Fereshetian S, Jaffe JD, Kuhn E, Udeshi ND, Carr SA: **Deep, Quantitative Coverage of the Lysine Acetylome Using Novel Anti-acetyl-lysine Antibodies and an Optimized Proteomic Workflow**. *Molecular & Cellular Proteomics* 2015, **14**(9):2429-2440.
208. Weinert BT, Narita T, Satpathy S, Srinivasan B, Hansen BK, Schölz C, Hamilton WB, Zucconi BE, Wang WW, Liu WR *et al*: **Time-Resolved Analysis Reveals Rapid Dynamics and Broad Scope of the CBP/p300 Acetylome**. *Cell* 2018, **174**(1):231-244.e212.
209. Rogers LD, Foster LJ: **Phosphoproteomics—finally fulfilling the promise?** *Molecular BioSystems* 2009, **5**(10):1122-1129.
210. Udeshi ND, Mertins P, Svinkina T, Carr SA: **Large-scale identification of ubiquitination sites by mass spectrometry**. *Nature Protocols* 2013, **8**:1950-1960.
211. Peng J, Schwartz D, Elias JE, Thoreen CC, Cheng D, Marsischky G, Roelofs J, Finley D, Gygi SP: **A proteomics approach to understanding protein ubiquitination**. *Nature Biotechnology* 2003, **21**(8):921-926.
212. Danielsen JMR, Sylvestersen KB, Bekker-Jensen S, Szklarczyk D, Poulsen JW, Horn H, Jensen LJ, Mailand N, Nielsen ML: **Mass Spectrometric Analysis of Lysine Ubiquitylation Reveals Promiscuity at Site Level**. *Molecular & Cellular Proteomics* 2011, **10**(3):M110.003590.
213. Xu G, Paige JS, Jaffrey SR: **Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling**. *Nature Biotechnology* 2010, **28**(8):868-873.
214. Kim W, Bennett Eric J, Huttlin Edward L, Guo A, Li J, Possemato A, Sowa Mathew E, Rad R, Rush J, Comb Michael J *et al*: **Systematic and Quantitative Assessment of the Ubiquitin-Modified Proteome**. *Molecular Cell* 2011, **44**(2):325-340.
215. Udeshi ND, Svinkina T, Mertins P, Kuhn E, Mani DR, Qiao JW, Carr SA: **Refined Preparation and Use of Anti-diglycine Remnant (K-ε-GG) Antibody Enables Routine Quantification of 10,000s of Ubiquitination Sites in Single Proteomics Experiments**. *Molecular & Cellular Proteomics* 2013, **12**(3):825-831.
216. Savaryn JP, Toby TK, Kelleher NL: **A researcher's guide to mass spectrometry-based proteomics**. *Proteomics* 2016, **16**(18):2435-2443.
217. Makarov A, Scigelova M: **Coupling liquid chromatography to Orbitrap mass spectrometry**. *Journal of Chromatography A* 2010, **1217**(25):3938-3945.
218. Fenn J, Mann M, Meng C, Wong S, Whitehouse C: **Electrospray ionization for mass spectrometry of large biomolecules**. *Science* 1989, **246**(4926):64-71.
219. Kebarle P, Verkerk UH: **Electrospray: From ions in solution to ions in the gas phase, what we know now**. *Mass Spectrometry Reviews* 2009, **28**(6):898-917.
220. Haag AM: **Mass Analyzers and Mass Spectrometers**. In: *Modern Proteomics – Sample Preparation, Analysis and Practical Applications.* Edited by Mirzaei H, Carrasco M. Cham: Springer International Publishing; 2016: 157-169.
221. Senko MW, Remes PM, Canterbury JD, Mathur R, Song Q, Eliuk SM, Mullen C, Earley L, Hardman M, Blethrow JD *et al*: **Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates**. *Anal Chem* 2013, **85**(24):11710-11714.

222. McAlister GC, Nusinow DP, Jedrychowski MP, Wühr M, Huttlin EL, Erickson BK, Rad R, Haas W, Gygi SP: **MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes**. *Anal Chem* 2014, **86**(14):7150-7158.
223. Allen JS: **An Improved Electron Multiplier Particle Counter**. *Review of Scientific Instruments* 1947, **18**(10):739-749.
224. Olsen JV, Schwartz JC, Griep-Raming J, Nielsen ML, Damoc E, Denisov E, Lange O, Remes P, Taylor D, Splendore M *et al*: **A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed**. *Molecular & Cellular Proteomics* 2009, **8**(12):2759-2769.
225. Makarov A: **Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis**. *Anal Chem* 2000, **72**(6):1156-1162.
226. Makarov A, Denisov E, Lange O: **Performance evaluation of a high-field orbitrap mass analyzer**. *Journal of the American Society for Mass Spectrometry* 2009, **20**(8):1391-1396.
227. Zubarev RA, Makarov A: **Orbitrap Mass Spectrometry**. *Anal Chem* 2013, **85**(11):5288-5296.
228. Denisov E, Damoc E, Lange O, Makarov A: **Orbitrap mass spectrometry with resolving powers above 1,000,000**. *International Journal of Mass Spectrometry* 2012, **325-327**:80-85.
229. He F, Emmett MR, Håkansson K, Hendrickson CL, Marshall AG: **Theoretical and Experimental Prospects for Protein Identification Based Solely on Accurate Mass Measurement**. *Journal of Proteome Research* 2004, **3**(1):61-67.
230. Steen H, Mann M: **The abc's (and xyz's) of peptide sequencing**. *Nat Rev Mol Cell Biol* 2004, **5**(9):699-711.
231. Zolg DP, Wilhelm M, Schnatbaum K, Zerweck J, Knaute T, Delanghe B, Bailey DJ, Gessulat S, Ehrlich H-C, Weininger M *et al*: **Building ProteomeTools based on a complete synthetic human proteome**. *Nature Methods* 2017, **14**:259-262.
232. Cleland TP, DeHart CJ, Fellers RT, VanNispen AJ, Greer JB, LeDuc RD, Parker WR, Thomas PM, Kelleher NL, Brodbelt JS: **High-Throughput Analysis of Intact Human Proteins Using UVPD and HCD on an Orbitrap Mass Spectrometer**. *Journal of proteome research* 2017, **16**(5):2072-2079.
233. Hunt DF, Yates JR, 3rd, Shabanowitz J, Winston S, Hauer CR: **Protein sequencing by tandem mass spectrometry**. *Proceedings of the National Academy of Sciences of the United States of America* 1986, **83**(17):6233-6237.
234. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M: **Higher-energy C-trap dissociation for peptide modification analysis**. *Nature Methods* 2007, **4**:709-712.
235. Roepstorff P, Fohlman J: **Proposal for a common nomenclature for sequence ions in mass spectra of peptides**. *Biomed Mass Spectrom* 1984, **11**(11):601.
236. Martin DB, Eng JK, Nesvizhskii AI, Gemmill A, Aebersold R: **Investigation of neutral loss during collision-induced dissociation of peptide ions**. *Anal Chem* 2005, **77**(15):4870-4882.
237. Boersema PJ, Mohammed S, Heck AJR: **Phosphopeptide fragmentation and analysis by mass spectrometry**. *Journal of Mass Spectrometry* 2009, **44**(6):861-878.
238. Schroeder MJ, Shabanowitz J, Schwartz JC, Hunt DF, Coon JJ: **A Neutral Loss Activation Method for Improved Phosphopeptide Sequence Analysis by Quadrupole Ion Trap Mass Spectrometry**. *Anal Chem* 2004, **76**(13):3590-3598.
239. Peterson AC, Russell JD, Bailey DJ, Westphall MS, Coon JJ: **Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics**. *Molecular & Cellular Proteomics* 2012, **11**(11):1475-1488.
240. Bourmaud A, Gallien S, Domon B: **Parallel reaction monitoring using quadrupole-Orbitrap mass spectrometer: Principle and applications**. *Proteomics* 2016, **16**(15-16):2146-2159.
241. Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Hamon C: **Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS**. *Anal Chem* 2003, **75**(8):1895-1904.
242. Bantscheff M, Boesche M, Eberhard D, Matthieson T, Sweetman G, Kuster B: **Robust and Sensitive iTRAQ Quantification on an LTQ Orbitrap Mass Spectrometer**. *Molecular & Cellular Proteomics* 2008, **7**(9):1702-1713.
243. Ow SY, Salim M, Noirel J, Evans C, Rehman I, Wright PC: **iTRAQ Underestimation in Simple and Complex Mixtures: "The Good, the Bad and the Ugly"**. *Journal of Proteome Research* 2009, **8**(11):5347-5355.
244. Ow SY, Salim M, Noirel J, Evans C, Wright PC: **Minimising iTRAQ ratio compression through understanding LC-MS elution dependence and high-resolution HILIC fractionation**. *Proteomics* 2011, **11**(11):2341-2346.

245. Savitski MM, Mathieson T, Zinn N, Sweetman G, Doce C, Becher I, Pachl F, Kuster B, Bantscheff M: **Measuring and Managing Ratio Compression for Accurate iTRAQ/TMT Quantification**. *Journal of Proteome Research* 2013, **12**(8):3586-3598.
246. Ting L, Rad R, Gygi SP, Haas W: **MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics**. *Nature Methods* 2011, **8**:937-940.
247. McAlister GC, Huttlin EL, Haas W, Ting L, Jedrychowski MP, Rogers JC, Kuhn K, Pike I, Grothe RA, Blethrow JD *et al*: **Increasing the Multiplexing Capacity of TMTs Using Reporter Ion Isotopologues with Isobaric Masses**. *Anal Chem* 2012, **84**(17):7469-7478.
248. Erickson BK, Jedrychowski MP, McAlister GC, Everley RA, Kunz R, Gygi SP: **Evaluating Multiplexed Quantitative Phosphopeptide Analysis on a Hybrid Quadrupole Mass Filter/Linear Ion Trap/Orbitrap Mass Spectrometer**. *Anal Chem* 2015, **87**(2):1241-1249.
249. Jiang X, Bomgarden R, Brown J, Drew DL, Robitaille AM, Viner R, Huhmer AR: **Sensitive and Accurate Quantitation of Phosphopeptides Using TMT Isobaric Labeling Technique**. *Journal of Proteome Research* 2017, **16**(11):4244-4252.
250. Hogrebe A, von Stechow L, Bekker-Jensen DB, Weinert BT, Kelstrup CD, Olsen JV: **Benchmarking common quantification strategies for large-scale phosphoproteomics**. *Nature Communications* 2018, **9**(1):1045.
251. Muth T, Hartkopf F, Vaudel M, Renard BY: **A Potential Golden Age to Come—Current Tools, Recent Use Cases, and Future Avenues for De Novo Sequencing in Proteomics**. *Proteomics* 2018, **18**(18):1700150.
252. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS: **Probability-based protein identification by searching sequence databases using mass spectrometry data**. *ELECTROPHORESIS* 1999, **20**(18):3551-3567.
253. Frewen BE, Merrihew GE, Wu CC, Noble WS, MacCoss MJ: **Analysis of Peptide MS/MS Spectra from Large-Scale Proteomics Experiments Using Spectrum Libraries**. *Anal Chem* 2006, **78**(16):5678-5684.
254. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M: **Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment**. *Journal of Proteome Research* 2011, **10**(4):1794-1805.
255. Tyanova S, Temu T, Cox J: **The MaxQuant computational platform for mass spectrometry-based shotgun proteomics**. *Nature Protocols* 2016, **11**:2301.
256. Gentzel M, Köcher T, Ponnusamy S, Wilm M: **Preprocessing of tandem mass spectrometric data to support automatic protein identification**. *Proteomics* 2003, **3**(8):1597-1610.
257. Hoopmann MR, Chavez JD, Bruce JE: **SILACtor: Software To Enable Dynamic SILAC Studies**. 2011, **83**(22):8403-8410.
258. Savitski MM, Wilhelm M, Hahne H, Kuster B, Bantscheff M: **A Scalable Approach for Protein False Discovery Rate Estimation in Large Proteomic Data Sets**. *Molecular & Cellular Proteomics* 2015, **14**(9):2394-2404.
259. Nesvizhskii AI: **A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics**. *Journal of Proteomics* 2010, **73**(11):2092-2123.
260. Chalkley RJ, Clauser KR: **Modification site localization scoring: strategies and performance**. *Molecular & Cellular Proteomics* 2012, **11**(5):3-14.
261. Taus T, Köcher T, Pichler P, Paschke C, Schmidt A, Henrich C, Mechtler K: **Universal and Confident Phosphorylation Site Localization Using phosphoRS**. *Journal of Proteome Research* 2011, **10**(12):5354-5362.
262. Savitski MM, Lemeer S, Boesche M, Lang M, Mathieson T, Bantscheff M, Kuster B: **Confident Phosphorylation Site Localization Using the Mascot Delta Score**. *Molecular & Cellular Proteomics* 2011, **10**(2):M110.003830.
263. Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M: **Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks**. *Cell* 2006, **127**(3):635-648.
264. Nesvizhskii AI, Aebersold R: **Interpretation of Shotgun Proteomic Data**. *The Protein Inference Problem* 2005, **4**(10):1419-1440.
265. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ: **Skyline: an open source document editor for creating and analyzing targeted proteomics experiments**. *Bioinformatics* 2010, **26**(7):966-968.
266. Toprak UH, Gillet LC, Maiolica A, Navarro P, Leitner A, Aebersold R: **Conserved Peptide Fragmentation as a Benchmarking Tool for Mass Spectrometers and a Discriminating Feature for Targeted Proteomics**. *Molecular & Cellular Proteomics* 2014, **13**(8):2056-2071.

267. Bantscheff M, Lemeer S, Savitski MM, Kuster B: **Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present**. *Analytical and Bioanalytical Chemistry* 2012, **404**(4):939-965.
268. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG: **Comparison of Label-free Methods for Quantifying Human Proteins by Shotgun Proteomics**. *Molecular & Cellular Proteomics* 2005, **4**(10):1487-1502.
269. Chelius D, Bondarenko PV: **Quantitative Profiling of Proteins in Complex Mixtures Using Liquid Chromatography and Mass Spectrometry**. *Journal of Proteome Research* 2002, **1**(4):317-323.
270. Ankney JA, Muneer A, Chen X: **Relative and Absolute Quantitation in Mass Spectrometry–Based Proteomics**. *Annual Review of Analytical Chemistry* 2018, **11**(1):49-77.
271. Pappireddi N, Martin L, Wühr M: **A Review on Quantitative Multiplexed Proteomics**. *ChemBioChem* 2019, **20**(10):1210-1224.
272. Cox J, Mann M: **MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification**. *Nature Biotechnology* 2008, **26**:1367-1372.
273. Argentini A, Goeminne LJE, Verheggen K, Hulstaert N, Staes A, Clement L, Martens L: **moFF: a robust and automated approach to extract peptide ion intensities**. *Nature Methods* 2016, **13**:964-966.
274. Mertins P, Mani DR, Ruggles KV, Gillette MA, Clauser KR, Wang P, Wang X, Qiao JW, Cao S, Petralia F *et al*: **Proteogenomics connects somatic mutations to signalling in breast cancer**. *Nature* 2016, **534**:55-62.
275. Archer TC, Ehrenberger T, Mundt F, Gold MP, Krug K, Mah CK, Mahoney EL, Daniel CJ, LeNail A, Ramamoorthy D *et al*: **Proteomics, Post-translational Modifications, and Integrative Analyses Reveal Molecular Heterogeneity within Medulloblastoma Subgroups**. *Cancer Cell* 2018, **34**(3):396-410.e398.
276. Wenger CD, Lee MV, Hebert AS, McAlister GC, Phanstiel DH, Westphall MS, Coon JJ: **Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging**. *Nature Methods* 2011, **8**:933-935.
277. Savitski MM, Sweetman G, Askenazi M, Marto JA, Lang M, Zinn N, Bantscheff M: **Delayed Fragmentation and Optimized Isolation Width Settings for Improvement of Protein Identification and Accuracy of Isobaric Mass Tag Quantification on Orbitrap-Type Mass Spectrometers**. *Anal Chem* 2011, **83**(23):8959-8967.
278. Pfammatter S, Bonneil E, Thibault P: **Improvement of Quantitative Measurements in Multiplex Proteomics Using High-Field Asymmetric Waveform Spectrometry**. *Journal of Proteome Research* 2016, **15**(12):4653-4665.
279. Schweppe DK, Prasad S, Belford MW, Navarrete-Perea J, Bailey DJ, Huguet R, Jedrychowski MP, Rad R, McAlister G, Abbatiello SE *et al*: **Characterization and Optimization of Multiplexed Quantitative Analyses Using High-Field Asymmetric-Waveform Ion Mobility Mass Spectrometry**. *Anal Chem* 2019, **91**(6):4010-4016.
280. Wühr M, Haas W, McAlister GC, Peshkin L, Rad R, Kirschner MW, Gygi SP: **Accurate Multiplexed Proteomics at the MS2 Level Using the Complement Reporter Ion Cluster**. *Anal Chem* 2012, **84**(21):9214-9221.
281. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M: **Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ**. *Molecular & Cellular Proteomics* 2014, **13**(9):2513-2526.
282. Weisser H, Nahnsen S, Grossmann J, Nilse L, Quandt A, Brauer H, Sturm M, Kenar E, Kohlbacher O, Aebersold R *et al*: **An Automated Pipeline for High-Throughput Label-Free Quantitative Proteomics**. *Journal of Proteome Research* 2013, **12**(4):1628-1644.
283. Wisniewski JR, Hein MY, Cox J, Mann M: **A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards**. *Mol Cell Proteomics* 2014, **13**(12):3497-3506.
284. Oveland E, Muth T, Rapp E, Martens L, Berven FS, Barsnes H: **Viewing the proteome: How to visualize proteomics data?** *Proteomics* 2015, **15**(8):1341-1355.
285. The Gene Ontology Consortium: **Expansion of the Gene Ontology knowledgebase and resources**. *Nucl Acids Res* 2016, **45**(D1):D331-D338.
286. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K: **KEGG: new perspectives on genomes, pathways, diseases and drugs**. *Nucl Acids Res* 2016, **45**(D1):D353-D361.
287. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, Wester K, Hober S *et al*: **Towards a knowledge-based Human Protein Atlas**. *Nature Biotechnology* 2010, **28**:1248-1250.

288. Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I: **New and continuing developments at PROSITE**. *Nucl Acids Res* 2013, **41**(D1):D344-D347.
289. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P *et al*: **The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible**. *Nucl Acids Res* 2016, **45**(D1):D362-D368.
290. Giurgiu M, Reinhard J, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Ruepp A: **CORUM: the comprehensive resource of mammalian protein complexes—2019**. *Nucl Acids Res* 2018, **47**(D1):D559-D563.
291. Calvo SE, Clauser KR, Mootha VK: **MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins**. *Nucl Acids Res* 2016, **44**(D1):D1251-D1257.
292. Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, Alm T, Asplund A, Björk L, Breckels LM *et al*: **A subcellular map of the human proteome**. *Science* 2017, **356**(6340):eaal3321.
293. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The Protein Data Bank**. *Nucl Acids Res* 2000, **28**(1):235-242.
294. Hornbeck PV, Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, Latham V, Sullivan M: **PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse**. *Nucl Acids Res* 2012, **40**:D261-D270.
295. Li S, Shen D, Shao J, Crowder R, Liu W, Prat A, He X, Liu S, Hoog J, Lu C *et al*: **Endocrine-Therapy-Resistant ESR1 Variants Revealed by Genomic Characterization of Breast-Cancer-Derived Xenografts**. *Cell Reports* 2013, **4**(6):1116-1130.
296. Marx H, Lemeer S, Schliep JE, Matheron L, Mohammed S, Cox J, Mann M, Heck AJR, Kuster B: **A large synthetic peptide and phosphopeptide reference library for mass spectrometry–based proteomics**. *Nature Biotechnology* 2013, **31**:557-564.
297. Mertins P, Qiao JW, Patel J, Udeshi ND, Clauser KR, Mani DR, Burgess MW, Gillette MA, Jaffe JD, Carr SA: **Integrated proteomic analysis of post-translational modifications by serial enrichment**. *Nature Methods* 2013, **10**:634-637.
298. Ruprecht B, Zecha J, Zolg DP, Kuster B: **High pH Reversed-Phase Micro-Columns for Simple, Sensitive, and Efficient Fractionation of Proteome and (TMT labeled) Phosphoproteome Digests**. In: *Proteomics: Methods and Protocols.* Edited by Comai L, Katz JE, Mallick P. New York, NY: Springer New York; 2017: 83-98.
299. Huang K-L, Li S, Mertins P, Cao S, Gunawardena HP, Ruggles KV, Mani DR, Clauser KR, Tanioka M, Usary J *et al*: **Proteogenomic integration reveals therapeutic targets in breast cancer xenografts**. *Nature communications* 2017, **8**:14864-14864.
300. Plubell DL, Wilmarth PA, Zhao Y, Fenton AM, Minnier J, Reddy AP, Klimek J, Yang X, David LL, Pamir N: **Extended Multiplexing of Tandem Mass Tags (TMT) Labeling Reveals Age and High Fat Diet Specific Proteome Changes in Mouse Epididymal Adipose Tissue**. *Molecular & Cellular Proteomics* 2017, **16**(5):873-890.
301. Wisniewski JR, Ostasiewicz P, Dus K, Zielinska DF, Gnad F, Mann M: **Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma**. *Molecular systems biology* 2012, **8**:611.
302. Welle KA, Zhang T, Hyrohorenko JR, Shen S, Qu J, Ghaemmaghami S: **Time-resolved analysis of proteome dynamics by TMT-SILAC hyperplexing**. *Molecular & Cellular Proteomics* 2016, **15**(12):3551-3563.
303. Bates DM, Chamber JM: **Nonlinear models in S**. In: *Statistical models in S.* Edited by Chambers JM, Hastie TJ: Wadsworth & Brooks/Cole; 1992.
304. Gray KA, Yates B, Seal RL, Wright MW, Bruford EA: **Genenames.org: the HGNC resources in 2015**. *Nucl Acids Res* 2015, **43**(D1):D1079-D1085.
305. Doherty MK, Hammond DE, Clague MJ, Gaskell SJ, Beynon RJ: **Turnover of the Human Proteome: Determination of Protein Intracellular Stability by Dynamic SILAC**. *Journal of Proteome Research* 2009, **8**(1):104-112.
306. Zeiler M, Straube WL, Lundberg E, Uhlen M, Mann M: **A Protein Epitope Signature Tag (PrEST) Library Allows SILAC-based Absolute Quantification and Multiplexed Determination of Protein Copy Numbers in Cell Lines**. *Molecular & Cellular Proteomics*, **11**(3):O111.009613.
307. Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, Pääbo S, Mann M: **Deep proteome and transcriptome mapping of a human cancer cell line**. *Molecular systems biology* 2011, **7**.

308. Leuenberger P, Ganscha S, Kahraman A, Cappelletti V, Boersema PJ, von Mering C, Claassen M, Picotti P: **Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability**. *Science* 2017, **355**(6327):eaai7825.
309. Sugiyama N, Imamura H, Ishihama Y: **Large-scale Discovery of Substrates of the Human Kinome**. *Scientific Reports* 2019, **9**(1):10503.
310. Schölz C, Weinert BT, Wagner SA, Beli P, Miyake Y, Qi J, Jensen LJ, Streicher W, McCarthy AR, Westwood NJ *et al*: **Acetylation site specificities of lysine deacetylase inhibitors in human cells**. *Nature Biotechnology* 2015, **33**:415-423.
311. Gouw M, Michael S, Sámano-Sánchez H, Kumar M, Zeke A, Lang B, Bely B, Chemes LB, Davey NE, Deng Z *et al*: **The eukaryotic linear motif resource – 2018 update**. *Nucl Acids Res* 2017, **46**(D1):D428-D434.
312. Guharoy M, Bhowmick P, Sallam M, Tompa P: **Tripartite degrons confer diversity and specificity on regulated protein degradation in the ubiquitin-proteasome system**. *Nature Communications* 2016, **7**:10239.
313. Sormanni P, Camilloni C, Fariselli P, Vendruscolo M: **The s2D Method: Simultaneous Sequence-Based Prediction of the Statistical Populations of Ordered and Disordered Regions in Proteins**. *Journal of Molecular Biology* 2015, **427**:982-996.
314. Guo R, Zong S, Wu M, Gu J, Yang M: **Architecture of Human Mitochondrial Respiratory Megacomplex I2III2IV2**. *Cell* 2017, **170**(6):1247-1257.e1212.
315. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J: **The Perseus computational platform for comprehensive analysis of (prote)omics data**. *Nature Methods* 2016, **13**:731-740.
316. Cox J, Mann M: **1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data**. *BMC Bioinformatics* 2012, **13**(16):S12.
317. Tibshirani R, Chu G, Narasimhan B, Li J: **samr: SAM: Significance Analysis of Microarrays. R package version 2.0** 2011.
318. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response**. *Proceedings of the National Academy of Sciences* 2001, **98**(9):5116-5121.
319. Schwartz D, Gygi SP: **An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets**. *Nature Biotechnology* 2005, **23**(11):1391-1398.
320. Wagih O, Sugiyama N, Ishihama Y, Beltrao P: **Uncovering Phosphorylation-Based Specificities through Functional Interaction Networks**. *Molecular & Cellular Proteomics* 2016, **15**(1):236-245.
321. Chou MF, Schwartz D: **Biological Sequence Motif Discovery Using motif-x**. *Current Protocols in Bioinformatics* 2011, **35**(1):13.15.11-13.15.24.
322. O'Shea JP, Chou MF, Quader SA, Ryan JK, Church GM, Schwartz D: **pLogo: a probabilistic approach to visualizing sequence motifs**. *Nature Methods* 2013, **10**:1211-1212.
323. Zecha J, Satpathy S, Kanashova T, Avanessian SC, Kane MH, Clauser KR, Mertins P, Carr SA, Kuster B: **TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach**. *Molecular & Cellular Proteomics* 2019, **18**(7):1468-1478.
324. Vizcaíno JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T *et al*: **2016 update of the PRIDE database and its related tools**. *Nucl Acids Res* 2016, **44**(Database issue):D447-D456.
325. Li Z, Adams RM, Chourey K, Hurst GB, Hettich RL, Pan C: **Systematic Comparison of Label-Free, Metabolic Labeling, and Isobaric Chemical Labeling for Quantitative Proteomics on LTQ Orbitrap Velos**. *Journal of Proteome Research* 2012, **11**(3):1582-1590.
326. Mertins P, Udeshi ND, Clauser KR, Mani D, Patel J, Ong S-e, Jaffe JD, Carr SA: **iTRAQ Labeling is Superior to mTRAQ for Quantitative Global Proteomics and Phosphoproteomics**. *Molecular & Cellular Proteomics* 2012, **11**(6):M111.014423.
327. Altelaar AFM, Frese CK, Preisinger C, Hennrich ML, Schram AW, Timmers HTM, Heck AJR, Mohammed S: **Benchmarking stable isotope labeling based quantitative proteomics**. *Journal of Proteomics* 2013, **88**:14-26.
328. O'Connell JD, Paulo JA, O'Brien JJ, Gygi SP: **Proteome-Wide Evaluation of Two Common Protein Quantification Methods**. *Journal of Proteome Research* 2018, **17**(5):1934-1942.
329. Grimsley GR, Scholtz JM, Pace CN: **A summary of the measured pK values of the ionizable groups in folded proteins**. *Protein science : a publication of the Protein Society* 2009, **18**(1):247-251.
330. S W Englander, N W Downer a, Teitelbaum H: **Hydrogen Exchange**. *Annual Review of Biochemistry* 1972, **41**(1):903-924.

331. O'Brien JJ, O'Connell JD, Paulo JA, Thakurta S, Rose CM, Weekes MP, Huttlin EL, Gygi SP: **Compositional Proteomics: Effects of Spatial Constraints on Protein Quantification Utilizing Isobaric Tags**. *Journal of Proteome Research* 2018, **17**(1):590-599.
332. Makarov A, Denisov E, Lange O, Horning S: **Dynamic range of mass accuracy in LTQ orbitrap hybrid mass spectrometer**. *Journal of The American Society for Mass Spectrometry* 2006, **17**(7):977-982.
333. Kaufmann A, Walker S: **Comparison of linear intrascan and interscan dynamic ranges of Orbitrap and ion-mobility time-of-flight mass spectrometers**. *Rapid Communications in Mass Spectrometry* 2017, **31**(22):1915-1926.
334. Scigelova M, Hornshaw M, Giannakopulos A, Makarov A: **Fourier Transform Mass Spectrometry**. *Molecular & Cellular Proteomics* 2011, **10**(7):M111.009431.
335. Gorshkov MV, Fornelli L, Tsybin YO: **Observation of ion coalescence in Orbitrap Fourier transform mass spectrometry**. *Rapid Communications in Mass Spectrometry* 2012, **26**(15):1711-1717.
336. Boldin IA, Nikolaev EN: **Theory of peak coalescence in Fourier transform ion cyclotron resonance mass spectrometry**. *Rapid Communications in Mass Spectrometry* 2009, **23**(19):3213-3219.
337. Werner T, Sweetman G, Savitski MF, Mathieson T, Bantscheff M, Savitski MM: **Ion Coalescence of Neutron Encoded TMT 10-Plex Reporter Ions**. *Anal Chem* 2014, **86**(7):3594-3601.
338. Martinez-Val A, Garcia F, Ximénez-Embún P, Ibarz N, Zarzuela E, Ruppen I, Mohammed S, Munoz J: **On the Statistical Significance of Compressed Ratios in Isobaric Labeling: A Cross-Platform Comparison**. *Journal of Proteome Research* 2016, **15**(9):3029-3038.
339. Tebbe A, Klammer M, Sighart S, Schaab C, Daub H: **Systematic evaluation of label-free and super-SILAC quantification for proteome expression analysis**. *Rapid Communications in Mass Spectrometry* 2015, **29**(9):795-801.
340. Zhang H, Liu T, Zhang Z, Payne SH, Zhang B, McDermott JE, Zhou J-Y, Petyuk VA, Chen L, Ray D *et al*: **Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer**. *Cell* 2016, **166**(3):755-765.
341. Brenes A, Hukelmann JL, Bensaddek D, Lamond AI: **Multi-batch TMT reveals false positives, batch effects and missing values**. *Molecular & Cellular Proteomics* 2019, **18**(10):1967-1980.
342. Karpievitch YV, Dabney AR, Smith RD: **Normalization and missing value imputation for label-free LC-MS analysis**. *BMC Bioinformatics* 2012, **13**(Suppl 16):S5.
343. Webb-Robertson B-JM, Wiberg HK, Matzke MM, Brown JN, Wang J, McDermott JE, Smith RD, Rodland KD, Metz TO, Pounds JG *et al*: **Review, Evaluation, and Discussion of the Challenges of Missing Value Imputation for Mass Spectrometry-Based Label-Free Global Proteomics**. *Journal of Proteome Research* 2015, **14**(5):1993-2001.
344. Hebert AS, Thöing C, Riley NM, Kwiecien NW, Shiskova E, Huguet R, Cardasis HL, Kuehn A, Eliuk S, Zabrouskov V *et al*: **Improved Precursor Characterization for Data-Dependent Mass Spectrometry**. *Anal Chem* 2018, **90**(3):2333-2340.
345. Possemato AP, Paulo JA, Mulhern D, Guo A, Gygi SP, Beausoleil SA: **Multiplexed Phosphoproteomic Profiling Using Titanium Dioxide and Immunoaffinity Enrichments Reveals Complementary Phosphorylation Events**. *Journal of proteome research* 2017, **16**(4):1506-1514.
346. Baeumlisberger D, Arrey TN, Rietschel B, Rohmer M, Papasotiriou DG, Mueller B, Beckhaus T, Karas M: **Labeling elastase digests with TMT: Informational gain by identification of poorly detectable peptides with MALDI-TOF/TOF mass spectrometry**. *Proteomics* 2010, **10**(21):3905-3909.
347. Shen J, Pagala VR, Breuer AM, Peng J, Bin M, Wang X: **Spectral Library Search Improves Assignment of TMT Labeled MS/MS Spectra**. *Journal of proteome research* 2018, **17**(9):3325-3331.
348. Huang F-K, Zhang G, Lawlor K, Nazarian A, Philip J, Tempst P, Dephoure N, Neubert TA: **Deep Coverage of Global Protein Expression and Phosphorylation in Breast Tumor Cell Lines Using TMT 10-plex Isobaric Labeling**. *Journal of Proteome Research* 2017, **16**(3):1121-1132.
349. Pichler P, Köcher T, Holzmann J, Mazanek M, Taus T, Ammerer G, Mechtler K: **Peptide Labeling with Isobaric Tags Yields Higher Identification Rates Using iTRAQ 4-Plex Compared to TMT 6-Plex and iTRAQ 8-Plex on LTQ Orbitrap**. *Anal Chem* 2010, **82**(15):6549-6558.
350. Thingholm TE, Palmisano G, Kjeldsen F, Larsen MR: **Undesirable Charge-Enhancement of Isobaric Tagged Phosphopeptides Leads to Reduced Identification Efficiency**. *Journal of Proteome Research* 2010, **9**(8):4045-4052.
351. Anjaneyulu PSR, Staros JV: **Reactions of N-hydroxysulfosuccinimide active esters***. *International Journal of Peptide and Protein Research* 1987, **30**(1):117-124.

352. Miller BT, Collins TJ, Nagle GT, Kurosky A: **The occurrence of O-acylation during biotinylation of gonadotropin-releasing hormone and analogs. Evidence for a reactive serine**. *Journal of Biological Chemistry* 1992, **267**(8):5060-5069.
353. Leavell MD, Novak P, Behrens CR, Schoeniger JS, Kruppa GH: **Strategy for selective chemical cross-linking of tyrosine and lysine residues**. *Journal of the American Society for Mass Spectrometry* 2004, **15**(11):1604-1611.
354. Mädler S, Bich C, Touboul D, Zenobi R: **Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities**. *Journal of Mass Spectrometry* 2009, **44**(5):694-706.
355. Wiktorowicz JE, English RD, Wu Z, Kurosky A: **Model Studies on iTRAQ Modification of Peptides: Sequence-dependent Reaction Specificity**. *Journal of Proteome Research* 2012, **11**(3):1512-1520.
356. Böhm G, Prefot P, Jung S, Selzer S, Mitra V, Britton D, Kuhn K, Pike I, Thompson AH: **Low-pH Solid-Phase Amino Labeling of Complex Peptide Digests with TMTs Improves Peptide Identification Rates for Multiplexed Global Phosphopeptide Analysis**. *Journal of Proteome Research* 2015, **14**(6):2500-2510.
357. Sheng Q, Li R, Dai J, Li Q, Su Z, Guo Y, Li C, Shyr Y, Zeng R: **Preprocessing Significantly Improves the Peptide/Protein Identification Sensitivity of High-resolution Isobarically Labeled Tandem Mass Spectrometry Data**. *Molecular & Cellular Proteomics* 2015, **14**(2):405-417.
358. Brauman JI, Riveros JM, Blair LK: **Gas-phase basicities of amines**. *Journal of the American Chemical Society* 1971, **93**(16):3914-3916.
359. Everley RA, Huttlin EL, Erickson AR, Beausoleil SA, Gygi SP: **Neutral Loss Is a Very Common Occurrence in Phosphotyrosine-Containing Peptides Labeled with Isobaric Tags**. *Journal of Proteome Research* 2017, **16**(2):1069-1076.
360. Sadygov RG, Cociorva D, Yates JR: **Large-scale database searching using tandem mass spectra: Looking up the answer in the back of the book**. *Nature Methods* 2004, **1**(3):195-202.
361. Edwards A, Haas W: **Multiplexed Quantitative Proteomics for High-Throughput Comprehensive Proteome Comparisons of Human Cell Lines**. In: *Proteomics in Systems Biology: Methods and Protocols.* Edited by Reinders J. New York, NY: Springer New York; 2016: 1-13.
362. Navarrete-Perea J, Yu Q, Gygi SP, Paulo JA: **Streamlined Tandem Mass Tag (SL-TMT) Protocol: An Efficient Strategy for Quantitative (Phospho)proteome Profiling Using Tandem Mass Tag-Synchronous Precursor Selection-MS3**. *Journal of Proteome Research* 2018, **17**(6):2226-2236.
363. Stepanova E, Gygi SP, Paulo JA: **Filter-Based Protein Digestion (FPD): A Detergent-Free and Scaffold-Based Strategy for TMT Workflows**. *Journal of Proteome Research* 2018, **17**(3):1227-1234.
364. Koch H, Wilhelm M, Ruprecht B, Beck S, Frejno M, Klaeger S, Kuster B: **Phosphoproteome Profiling Reveals Molecular Mechanisms of Growth-Factor-Mediated Kinase Inhibitor Resistance in EGFR-Overexpressing Cancer Cells**. *Journal of Proteome Research* 2016, **15**(12):4490-4504.
365. Paulo JA, O'Connell JD, Everley RA, O'Brien J, Gygi MA, Gygi SP: **Quantitative mass spectrometry-based multiplexing compares the abundance of 5000 S. cerevisiae proteins across 10 carbon sources**. *Journal of Proteomics* 2016, **148**:85-93.
366. Mirzaei M, Gupta VB, Chick JM, Greco TM, Wu Y, Chitranshi N, Wall RV, Hone E, Deng L, Dheer Y *et al*: **Age-related neurodegenerative disease associated pathways identified in retinal and vitreous proteome from human glaucoma eyes**. *Scientific Reports* 2017, **7**(1):12685.
367. de Graaf EL, Pellegrini D, McDonnell LA: **Set of Novel Automated Quantitative Microproteomics Protocols for Small Sample Amounts and Its Application to Kidney Tissue Substructures**. *Journal of Proteome Research* 2016, **15**(12):4722-4730.
368. Yang W-C, Mirzaei H, Liu X, Regnier FE: **Enhancement of Amino Acid Detection and Quantification by Electrospray Ionization Mass Spectrometry**. *Anal Chem* 2006, **78**(13):4702-4708.
369. Regnier FE, Julka S: **Primary amine coding as a path to comparative proteomics**. *Proteomics* 2006, **6**(14):3968-3979.
370. Miller BT, Kurosky A: **Elevated Intrinsic Reactivity of Seryl Hydroxyl Groups within the Linear Peptide Triads His-Xaa-Ser or Ser-Xaa-His**. *Biochemical and Biophysical Research Communications* 1993, **196**(1):461-467.
371. Zecha J, Meng C, Zolg DP, Samaras P, Wilhelm M, Kuster B: **Peptide Level Turnover Measurements Enable the Study of Proteoform Dynamics**. *Molecular & Cellular Proteomics* 2018, **17**(5):974-992.
372. Hinkson IV, Elias JE: **The dynamic state of protein turnover: It's about time**. *Trends in Cell Biology* 2011, **21**(5):293-303.

373. Dittrich A, Siewert E, Schaper F: **Determination of Protein Turnover Rates in the JAK/STAT Pathway Using a Radioactive Pulse-Chase Approach**. In: *JAK-STAT Signalling: Methods and Protocols.* Edited by Nicholson SE, Nicola NA. Totowa, NJ: Humana Press; 2013: 69-80.
374. Kristensen AR, Gsponer J, Foster LJ: **Protein synthesis rate is the predominant regulator of protein expression during differentiation**. *Molecular systems biology* 2013, **9**(1):689.
375. Fierro-Monti I, Racle J, Hernandez C, Waridel P, Hatzimanikatis V, Quadroni M: **A Novel Pulse-Chase SILAC Strategy Measures Changes in Protein Decay and Synthesis Rates Induced by Perturbation of Proteostasis with an Hsp90 Inhibitor**. 2013, **8**(11):e80423.
376. McShane E, Sin C, Zauber H, Wells JN, Donnelly N, Wang X, Hou J, Chen W, Storchova Z, Marsh JA *et al*: **Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation**. *Cell* 2016, **167**(3):803-815.e821.
377. Hughes C, Krijgsveld J: **Developments in quantitative mass spectrometry for the analysis of proteome dynamics**. *Trends in Biotechnology* 2012, **30**(12):668-676.
378. Schmidt T, Samaras P, Frejno M, Gessulat S, Barnert M, Kienegger H, Krcmar H, Schlegl J, Ehrlich H-C, Aiche S *et al*: **ProteomicsDB**. *Nucl Acids Res* 2017:gkx1029-gkx1029.
379. Ruepp A, Waegele B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Mewes HW: **CORUM: the comprehensive resource of mammalian protein complexes—2009**. *Nucl Acids Res* 2010, **38**(suppl_1):D497-D501.
380. Madungwe NB, Zilberstein NF, Feng Y, Bopassa JC: **Critical role of mitochondrial ROS is dependent on their site of production on the electron transport chain in ischemic heart**. *American Journal of Cardiovascular Disease* 2016, **6**(3):93-108.
381. Takahiko K, Koichi H, Shinsei G, Tetsuya F, Shiro M, Tamotsu M, Akira M: **Proteolytic processing sites producing the mature form of human cathepsin D**. *International Journal of Biochemistry* 1992, **24**(9):1487-1491.
382. Mittal N, Subramanian G, Bütikofer P, Madhubala R: **Unique posttranslational modifications in eukaryotic translation factors and their roles in protozoan parasite viability and pathogenesis**. *Molecular and Biochemical Parasitology* 2013, **187**(1):21-31.
383. Jakobsson ME, Moen A, Bousset L, Egge-Jacobsen W, Kernstock S, Melki R, Falnes PØ: **Identification and Characterization of a Novel Human Methyltransferase Modulating Hsp70 Protein Function through Lysine Methylation**. *Journal of Biological Chemistry* 2013, **288**:27752-27763.
384. Bayart E, Dutertre S, Jaulin C, Guo R-B, Xi XG, Amor-Guéret M: **The Bloom Syndrome Helicase is a Substrate of the Mitotic Cdc2 Kinase**. *Cell Cycle* 2006, **5**(15):1681-1686.
385. Tyanova S, Mann M, Cox J: **MaxQuant for In-Depth Analysis of Large SILAC Datasets**. In: *Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC): Methods and Protocols.* Edited by Warscheid B. New York, NY: Springer New York; 2014: 351-364.
386. Buchler NE, Gerland U, Hwa T: **Nonlinear protein degradation and the function of genetic circuits**. *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(27):9559-9564.
387. Schubert U, Antón LC, Gibbs J, Norbury CC, Yewdell JW, Bennink JR: **Rapid degradation of a large fraction of newly synthesized proteins by proteasomes**. *Nature* 2000, **404**(6779):770-774.
388. Duttler S, Pechmann S, Frydman J: **Principles of Cotranslational Ubiquitination and Quality Control at the Ribosome**. *Molecular Cell* 2013, **50**(3):379-393.
389. Baldwin RL: **Temperature dependence of the hydrophobic interaction in protein folding**. *Proceedings of the National Academy of Sciences* 1986, **83**(21):8069-8072.
390. Chothia C: **Structural invariants in protein folding**. *Nature* 1975, **254**:304-308.
391. Asher G, Reuven N, Shaul Y: **20S proteasomes and protein degradation "by default"**. *BioEssays* 2006, **28**(8):844-849.
392. Vavouri T, Semple JI, Garcia-Verdugo R, Lehner B: **Intrinsic Protein Disorder and Interaction Promiscuity Are Widely Associated with Dosage Sensitivity**. *Cell* 2009, **138**(1):198-208.
393. Swovick K, Welle KA, Hryhorenko J, Seluanov A, Gorbunova V, Ghaemmaghami S: **Cross-species comparison of proteome turnover kinetics**. *Molecular & Cellular Proteomics* 2018, **17**(4):580-591.
394. Johnson ES, Gonda DK, Varshavsky A: **Cis-trans recognition and subunit-specific degradation of short-lived proteins**. *Nature* 1990, **346**(6281):287-291.
395. Biederer T, Volkwein C, Sommer T: **Degradation of subunits of the Sec61p complex, an integral component of the ER membrane, by the ubiquitin-proteasome pathway**. *The EMBO Journal* 1996, **15**(9):2069-2076.

396. Varshavsky A: **The N-end rule pathway and regulation by proteolysis**. *Protein Science* 2011, **20**(8):1298-1345.
397. Kim H-K, Kim R-R, Oh J-H, Cho H, Varshavsky A, Hwang C-S: **The N-Terminal Methionine of Cellular Proteins as a Degradation Signal**. *Cell* 2014, **156**(1–2):158-169.
398. Baldwin ET, Bhat TN, Gulnik S, Hosur MV, Sowder RC, Cachau RE, Collins J, Silva AM, Erickson JW: **Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design**. *Proceedings of the National Academy of Sciences of the United States of America* 1993, **90**(14):6796-6800.
399. Langland G, Elliott J, Li Y, Creaney J, Dixon K, Groden J: **The BLM Helicase Is Necessary for Normal DNA Double-Strand Break Repair**. *Cancer Research* 2002, **62**(10):2766-2770.
400. Yoon CH, Miah MA, Kim KP, Bae YS: **New Cdc2 Tyr 4 phosphorylation by dsRNA-activated protein kinase triggers Cdc2 polyubiquitination and G2 arrest under genotoxic stresses**. *EMBO reports* 2010, **11**(5):393-399.
401. Balch WE, Morimoto RI, Dillin A, Kelly JW: **Adapting Proteostasis for Disease Intervention**. *Science* 2008, **319**(5865):916-919.
402. Fornasiero EF, Mandad S, Wildhagen H, Alevra M, Rammner B, Keihani S, Opazo F, Urban I, Ischebeck T, Sakib MS *et al*: **Precisely measured protein lifetimes in the mouse brain reveal differences across tissues and subcellular fractions**. *Nature Communications* 2018, **9**(1):4230.
403. Liu Y, Mi Y, Mueller T, Kreibich S, Williams EG, Van Drogen A, Borel C, Frank M, Germain P-L, Bludau I *et al*: **Multi-omic measurements of heterogeneity in HeLa cells across laboratories**. *Nature Biotechnology* 2019, **37**(3):314-322.
404. Li M, Luo J, Brooks CL, Gu W: **Acetylation of p53 Inhibits Its Ubiquitination by Mdm2**. *Journal of Biological Chemistry* 2002, **277**(52):50607-50611.
405. Grönroos E, Hellman U, Heldin C-H, Ericsson J: **Control of Smad7 Stability by Competition between Acetylation and Ubiquitination**. *Molecular Cell* 2002, **10**(3):483-493.
406. Patel JH, Du Y, Ard PG, Phillips C, Carella B, Chen C-J, Rakowski C, Chatterjee C, Lieberman PM, Lane WS *et al*: **The c-MYC Oncoprotein Is a Substrate of the Acetyltransferases hGCN5/PCAF and TIP60**. *Molecular and Cellular Biology* 2004, **24**(24):10826-10834.
407. Jin Y-H, Jeon E-J, Li Q-L, Lee YH, Choi J-K, Kim W-J, Lee K-Y, Bae S-C: **Transforming Growth Factor-β Stimulates p300-dependent RUNX3 Acetylation, Which Inhibits Ubiquitination-mediated Degradation**. *Journal of Biological Chemistry* 2004, **279**(28):29409-29417.
408. Kang S-A, Na H, Kang H-J, Kim S-H, Lee M-H, Lee M-O: **Regulation of Nur77 protein turnover through acetylation and deacetylation induced by p300 and HDAC1**. *Biochemical Pharmacology* 2010, **80**(6):867-873.
409. Sadoul K, Boyault C, Pabion M, Khochbin S: **Regulation of protein turnover by acetyltransferases and deacetylases**. *Biochimie* 2008, **90**(2):306-312.
410. Jeong J-W, Bae M-K, Ahn M-Y, Kim S-H, Sohn T-K, Bae M-H, Yoo M-A, Song EJ, Lee K-J, Kim K-W: **Regulation and Destabilization of HIF-1α by ARD1-Mediated Acetylation**. *Cell* 2002, **111**(5):709-720.
411. Hernandez-Hernandez A, Ray P, Litos G, Ciro M, Ottolenghi S, Beug H, Boyes J: **Acetylation and MAPK phosphorylation cooperate to regulate the degradation of active GATA-1**. *The EMBO journal* 2006, **25**(14):3264-3274.
412. Swaney DL, Beltrao P, Starita L, Guo A, Rush J, Fields S, Krogan NJ, Villén J: **Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation**. *Nature methods* 2013, **10**(7):676-682.
413. Martin-Perez M, Villén J: **Determinants and Regulation of Protein Turnover in Yeast**. *Cell Systems* 2017, **5**(3):283-294.e285.
414. Heald R, McKeon F: **Mutations of phosphorylation sites in lamin A that prevent nuclear lamina disassembly in mitosis**. *Cell* 1990, **61**(4):579-589.
415. Lee C-P, Huang Y-H, Lin S-F, Chang Y, Chang Y-H, Takada K, Chen M-R: **Epstein-Barr Virus BGLF4 Kinase Induces Disassembly of the Nuclear Lamina To Facilitate Virion Production**. *Journal of Virology* 2008, **82**(23):11913-11926.
416. Hagen L, Kavli B, Sousa MML, Torseth K, Liabakk NB, Sundheim O, Peňa-Diaz J, Otterlei M, Hørning O, Jensen ON *et al*: **Cell cycle-specific UNG2 phosphorylations regulate protein turnover, activity and association with RPA**. *The EMBO Journal* 2008, **27**(1):51-61.
417. Koonin EV: **A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication**. *Nucl Acids Res* 1993, **21**(11):2541-2547.

418. Walker JR, Corpina RA, Goldberg J: **Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair**. *Nature* 2001, **412**(6847):607-614.
419. Chen ZA, Fischer L, Cox J, Rappsilber J: **Quantitative Cross-linking/Mass Spectrometry Using Isotope-labeled Cross-linkers and MaxQuant**. *Molecular & Cellular Proteomics* 2016, **15**(8):2769-2778.
420. Karp NA, Huber W, Sadowski PG, Charles PD, Hester SV, Lilley KS: **Addressing Accuracy and Precision Issues in iTRAQ Quantitation**. *Molecular & Cellular Proteomics* 2010, **9**(9):1885-1897.
421. Rose C, Isasa M, Ordureau A, Prado Miguel A, Beausoleil Sean A, Jedrychowski Mark P, Finley Daniel J, Harper JW, Gygi Steven P: **Highly Multiplexed Quantitative Mass Spectrometry Analysis of Ubiquitylomes**. *Cell Systems* 2016, **3**(4):395-403.e394.
422. Newton K, Matsumoto ML, Wertz IE, Kirkpatrick DS, Lill JR, Tan J, Dugger D, Gordon N, Sidhu SS, Fellouse FA *et al*: **Ubiquitin Chain Editing Revealed by Polyubiquitin Linkage-Specific Antibodies**. *Cell* 2008, **134**(4):668-678.
423. Na S, Paek E: **Software eyes for protein post-translational modifications**. *Mass Spectrometry Reviews* 2015, **34**(2):133-147.
424. Fan J, Shan C, Kang H-B, Elf S, Xie J, Tucker M, Gu T-L, Aguiar M, Lonning S, Chen H *et al*: **Tyr Phosphorylation of PDP1 Toggles Recruitment between ACAT1 and SIRT3 to Regulate the Pyruvate Dehydrogenase Complex**. *Molecular Cell* 2014, **53**(4):534-548.
425. Finley Lydia WS, Carracedo A, Lee J, Souza A, Egia A, Zhang J, Teruya-Feldstein J, Moreira Paula I, Cardoso Sandra M, Clish Clary B *et al*: **SIRT3 Opposes Reprogramming of Cancer Cell Metabolism through HIF1α Destabilization**. *Cancer Cell* 2011, **19**(3):416-428.
426. Min S-W, Cho S-H, Zhou Y, Schroeder S, Haroutunian V, Seeley WW, Huang EJ, Shen Y, Masliah E, Mukherjee C *et al*: **Acetylation of Tau Inhibits Its Degradation and Contributes to Tauopathy**. *Neuron* 2010, **67**(6):953-966.
427. Cook C, Carlomagno Y, Gendron TF, Dunmore J, Scheffel K, Stetler C, Davis M, Dickson D, Jarpe M, DeTure M *et al*: **Acetylation of the KXGS motifs in tau is a critical determinant in modulation of tau aggregation and clearance**. *Human Molecular Genetics* 2013, **23**(1):104-116.
428. Weinert BT, Moustafa T, Iesmantavicius V, Zechner R, Choudhary C: **Analysis of acetylation stoichiometry suggests that SIRT3 repairs nonenzymatic acetylation lesions**. *The EMBO Journal* 2015, **34**(21):2620-2632.
429. Baeza J, Smallegan MJ, Denu JM: **Site-Specific Reactivity of Nonenzymatic Lysine Acetylation**. *ACS Chemical Biology* 2015, **10**(1):122-128.
430. Huang X, Dixit VM: **Drugging the undruggables: exploring the ubiquitin system for drug development**. *Cell Research* 2016, **26**(4):484-498.
431. Vassilev LT, Vu BT, Graves B, Carvajal D, Podlaski F, Filipovic Z, Kong N, Kammlott U, Lukacs C, Klein C *et al*: **In Vivo Activation of the p53 Pathway by Small-Molecule Antagonists of MDM2**. *Science* 2004, **303**(5659):844-848.
432. Krönke J, Fink EC, Hollenbach PW, MacBeth KJ, Hurst SN, Udeshi ND, Chamberlain PP, Mani DR, Man HW, Gandhi AK *et al*: **Lenalidomide induces ubiquitination and degradation of CK1α in del(5q) MDS**. *Nature* 2015, **523**:183-188.
433. Chauhan D, Tian Z, Nicholson B, Kumar KGS, Zhou B, Carrasco R, McDermott Jeffrey L, Leach Craig A, Fulcinniti M, Kodrasov Matthew P *et al*: **A Small Molecule Inhibitor of Ubiquitin-Specific Protease-7 Induces Apoptosis in Multiple Myeloma Cells and Overcomes Bortezomib Resistance**. *Cancer Cell* 2012, **22**(3):345-358.
434. Verma R, Peters NR, D'Onofrio M, Tochtrop GP, Sakamoto KM, Varadan R, Zhang M, Coffino P, Fushman D, Deshaies RJ *et al*: **Ubistatins Inhibit Proteasome-Dependent Degradation by Binding the Ubiquitin Chain**. *Science* 2004, **306**(5693):117-120.
435. Dou QP, Jeffrey AZ: **Overview of Proteasome Inhibitor-Based Anti-cancer Therapies: Perspective on Bortezomib and Second Generation Proteasome Inhibitors versus Future Generation Inhibitors of Ubiquitin-Proteasome System**. *Current Cancer Drug Targets* 2014, **14**(6):517-536.
436. Toure M, Crews CM: **Small-Molecule PROTACS: New Approaches to Protein Degradation**. *Angewandte Chemie International Edition* 2016, **55**(6):1966-1973.
437. Savitski MM, Zinn N, Faelth-Savitski M, Poeckel D, Gade S, Becher I, Muelbaier M, Wagner AJ, Strohmer K, Werner T *et al*: **Multiplexed Proteome Dynamics Profiling Reveals Mechanisms Controlling Protein Homeostasis**. *Cell* 2018, **173**(1):260-274.e225.
438. Ge F, Xiao C-L, Bi L-J, Tao S-C, Xiong S, Yin X-F, Li L-P, Lu C-H, Jia H-T, He Q-Y: **Quantitative Phosphoproteomics of Proteasome Inhibition in Multiple Myeloma Cells**. *PLOS ONE* 2010, **5**(9):e13095.

439. Singh J, Petter RC, Baillie TA, Whitty A: **The resurgence of covalent drugs**. *Nature Reviews Drug Discovery* 2011, **10**(4):307-317.
440. Gholami A, Hahne H, Wu Z, Auer F, Meng C, Wilhelm M, Kuster B: **Global Proteome Analysis of the NCI-60 Cell Line Panel**. *Cell Reports* 2013, **4**(3):609-620.
441. Frejno M, Zenezini Chiozzi R, Wilhelm M, Koch H, Zheng R, Klaeger S, Ruprecht B, Meng C, Kramer K, Jarzab A *et al*: **Pharmacoproteomic characterisation of human colon and rectal cancer**. *Molecular systems biology* 2017, **13**(11):951.
442. Jungblut PR: **The proteomics quantification dilemma**. *Journal of Proteomics* 2014, **107**:98-102.
443. Smith LM, Thomas PM, Shortreed MR, Schaffer LV, Fellers RT, LeDuc RD, Tucholski T, Ge Y, Agar JN, Anderson LC *et al*: **A five-level classification system for proteoform identifications**. *Nature Methods* 2019, **16**:939-340.
444. Zhu Y, Clair G, Chrisler WB, Shen Y, Zhao R, Shukla AK, Moore RJ, Misra RS, Pryhuber GS, Smith RD *et al*: **Proteomic Analysis of Single Mammalian Cells Enabled by Microfluidic Nanodroplet Sample Preparation and Ultrasensitive NanoLC-MS**. *Angewandte Chemie International Edition* 2018, **57**(38):12370-12374.
445. Kelstrup CD, Bekker-Jensen DB, Arrey TN, Hogrebe A, Harder A, Olsen JV: **Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics**. *Journal of Proteome Research* 2018, **17**(1):727-738.
446. Levy MJ, Washburn MP, Florens L: **Probing the Sensitivity of the Orbitrap Lumos Mass Spectrometer Using a Standard Reference Protein in a Complex Background**. *Journal of Proteome Research* 2018, **17**(10):3586-3592.
447. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Girón CG *et al*: **Ensembl 2018**. *Nucl Acids Res* 2018, **46**(D1):D754-D761.
448. Wan J, Qian S-B: **TISdb: a database for alternative translation initiation in mammalian cells**. *Nucl Acids Res* 2013, **42**(D1):D845-D850.
449. Day INM: **dbSNP in the detail and copy number complexities**. *Human Mutation* 2010, **31**(1):2-4.
450. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L *et al*: **COSMIC: somatic cancer genetics at high-resolution**. *Nucl Acids Res* 2016, **45**(D1):D777-D783.
451. Nesvizhskii AI: **Proteogenomics: concepts, applications and computational strategies**. *Nature Methods* 2014, **11**:1114.
452. Chick JM, Kolippakkam D, Nusinow DP, Zhai B, Rad R, Huttlin EL, Gygi SP: **A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides**. *Nature Biotechnology* 2015, **33**:743.
453. Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI: **MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics**. *Nature Methods* 2017, **14**:513-520.
454. An Z, Zhai L, Ying W, Qian X, Gong F, Tan M, Fu Y: **PTMiner: Localization and Quality Control of Protein Modifications Detected in an Open Search and Its Application to Comprehensive Post-translational Modification Characterization in Human Proteome**. *Molecular & Cellular Proteomics* 2019, **18**(2):391-405.
455. Zhang Y, Muller M, Xu B, Yoshida Y, Horlacher O, Nikitin F, Garessus S, Magdeldin S, Kinoshita N, Fujinaka H *et al*: **Unrestricted modification search reveals lysine methylation as major modification induced by tissue formalin fixation and paraffin embedding**. *Proteomics* 2015, **15**(15):2568-2579.
456. Cohen P: **The origins of protein phosphorylation**. *Nature Cell Biology* 2002, **4**(5):E127-E130.
457. Klaeger S, Heinzlmeir S, Wilhelm M, Polzer H, Vick B, Koenig P-A, Reinecke M, Ruprecht B, Petzoldt S, Meng C *et al*: **The target landscape of clinical kinase drugs**. *Science* 2017, **358**(6367):eaan4368.
458. Venable JD, Dong M-Q, Wohlschlegel J, Dillin A, Yates JR: **Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra**. *Nature Methods* 2004, **1**(1):39-45.
459. Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R: **Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial**. *Molecular systems biology* 2018, **14**(8):e8126.
460. Lam H, Deutsch EW, Aebersold R: **Artificial Decoy Spectral Libraries for False Discovery Rate Estimation in Spectral Library Searching in Proteomics**. *Journal of Proteome Research* 2010, **9**(1):605-610.
461. Rosenberger G, Bludau I, Schmitt U, Heusel M, Hunter CL, Liu Y, MacCoss MJ, MacLean BX, Nesvizhskii AI, Pedrioli PGA *et al*: **Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses**. *Nature Methods* 2017, **14**:921-927.

462. Muntel J, Gandhi T, Verbeke L, Bernhardt OM, 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**. *Molecular Omics* 2019, **15**(5):348-360.
463. Meier F, Geyer PE, Virreira Winter S, Cox J, Mann M: **BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes**. *Nature Methods* 2018, **15**(6):440-448.
464. Spahr CS, Davis MT, McGinley MD, Robinson JH, Bures EJ, Beierle J, Mort J, Courchesne PL, Chen K, Wahl RC *et al*: **Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry I.Profiling an unfractionated tryptic digest**. *Proteomics* 2001, **1**(1):93-107.
465. Geyer Philipp E, Kulak Nils A, Pichler G, Holdt Lesca M, Teupser D, Mann M: **Plasma Proteome Profiling to Assess Human Health and Disease**. *Cell Systems* 2016, **2**(3):185-195.
466. Davis S, Charles PD, He L, Mowlds P, Kessler BM, Fischer R: **Expanding Proteome Coverage with CHarge Ordered Parallel Ion aNalysis (CHOPIN) Combined with Broad Specificity Proteolysis**. *Journal of proteome research* 2017, **16**(3):1288-1299.
467. Erickson BK, Mintseris J, Schweppe DK, Navarrete-Perea J, Erickson AR, Nusinow DP, Paulo JA, Gygi SP: **Active Instrument Engagement Combined with a Real-Time Database Search for Improved Performance of Sample Multiplexing Workflows**. *Journal of Proteome Research* 2019, **18**(3):1299-1306.
468. McDaniel EW, Martin DW, Barnes WS: **Drift Tube-Mass Spectrometer for Studies of Low-Energy Ion-Molecule Reactions**. *Review of Scientific Instruments* 1962, **33**(1):2-7.
469. Hebert AS, Prasad S, Belford MW, Bailey DJ, McAlister GC, Abbatiello SE, Huguet R, Wouters ER, Dunyach J-J, Brademan DR *et al*: **Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer**. *Anal Chem* 2018, **90**(15):9529-9537.
470. Spraggins J, Djambazova K, Rivera E, Migas L, Neumann E, Fuetterer A, Suetering J, Goedecke N, Ly A, Plas RVd *et al*: **High-Performance Molecular Imaging with MALDI Trapped Ion-Mobility Time-of-Flight (timsTOF) Mass Spectrometry.** *Anal Chem* 2019, **91**(22):14552-14560.
471. Meier F, Brunner A-D, Koch S, Koch H, Lubeck M, Krause M, Goedecke N, Decker J, Kosinski T, Park MA *et al*: **Online Parallel Accumulation–Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer**. *Molecular & Cellular Proteomics* 2018, **17**(12):2534-2545.
472. Pfammatter S, Bonneil E, McManus FP, Prasad S, Bailey DJ, Belford M, Dunyach J-J, Thibault P: **A Novel Differential Ion Mobility Device Expands the Depth of Proteome Coverage and the Sensitivity of Multiplex Proteomic Measurements**. *Molecular & Cellular Proteomics* 2018, **17**(10):2051-2067.
473. Zhou X-X, Zeng W-F, Chi H, Luo C, Liu C, Zhan J, He S-M, Zhang Z: **pDeep: Predicting MS/MS Spectra of Peptides with Deep Learning**. *Anal Chem* 2017, **89**(23):12690-12697.
474. Tiwary S, Levy R, Gutenbrunner P, Salinas Soto F, Palaniappan KK, Deming L, Berndl M, Brant A, Cimermancic P, Cox J: **High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis**. *Nature Methods* 2019, **16**(6):519-525.
475. Gessulat S, Schmidt T, Zolg DP, Samaras P, Schnatbaum K, Zerweck J, Knaute T, Rechenberger J, Delanghe B, Huhmer A *et al*: **Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning**. *Nature Methods* 2019, **16**(6):509-518.
476. Shen X, Shen S, Li J, Hu Q, Nie L, Tu C, Wang X, Poulsen DJ, Orsburn BC, Wang J *et al*: **IonStar enables high-precision, low-missing-data proteomics quantification in large biological cohorts**. *Proceedings of the National Academy of Sciences* 2018, **115**(21):E4767-E4776.
477. Mamoshina P, Vieira A, Putin E, Zhavoronkov A: **Applications of Deep Learning in Biomedicine**. *Molecular Pharmaceutics* 2016, **13**(5):1445-1454.

# LIST OF ABBREVIATIONS

| | |
|---|---|
| 2DE | Two-dimensional gel-electrophoresis |
| AC | Alternating current |
| ac-CoA | Acetyl coenzyme A |
| ACN | Acetonitrile |
| AGC | Automatic gain control |
| AHA | Azidohomoalanine |
| ANOVA | Analysis of variance |
| ATP | Adenosine triphosphate |
| BONCAT | Bio-orthogonal non-canonical amino acid tagging |
| bRP | High pH reversed-phase (basic reversed-phase) |
| CID | Collision-induced dissociation |
| CV | Coefficient of variation |
| DC | Direct current |
| DDA | Data dependent acquisition |
| DIA | Data independent acquisition |
| DNA | Deoxyribonucleic acid |
| DUB | Deubiquitinase |
| ER | Endoplasmic reticulum |
| ESI | Electrospray ionisation |
| FA | Formic acid |
| FBS | Fetal bovine serum |
| FC | Fold change |
| FDR | False discovery rate |
| H/L ratio | Ratio of K8R10 to K0R0 labelled peptide or protein (heavy-to-light ratio) |
| HCD | Higher-energy collisional dissociation |
| HPA | Human Protein Atlas |
| HPLC | High-performance liquid chromatography |
| hSAX | Hydrophilic strong anion exchange |
| HSP | Heat-shock protein |
| IAP | Immunoaffinity purification |
| IMAC | Immobilized metal ion affinity chromatography |
| iMet | Initiator methionine |
| IT | Ion trap |
| iTRAQ | Isobaric Tags for Relative and Absolute Quantification |
| K0 | Lysine $^{12}C_6{}^{14}N_2$ (lysine containing only light isotopes of carbon and nitrogen) |
| K8 | Lysine $^{13}C_6{}^{15}N_2$ (lysine containing 8 stable, heavy isotopes of carbon and nitrogen) |
| KAT | Lysine acetyl transferase |
| KDAC | Lysine deacetylase |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| LC | Liquid chromatography |

| | |
|---|---|
| maxIT | Maximum injection time |
| mRNA | Messenger ribonucleic acid |
| MS | Mass spectrometry |
| MSA | Multistage activation |
| m/z | Mass-to-charge ratio |
| N/O ratio | Ratio of newly synthesized to “old” protein or peptide (new/old ratio) |
| NCE | Normalized collision energy |
| NHS | N-hydroxysuccinimide |
| NL | Neutral loss |
| NTA | Nitrilotriacetic acid |
| OT | Orbitrap |
| PBS | Phosphate-buffered saline |
| PDX | Patient-derived xenograft |
| PIF | Precursor intensity fraction |
| PRM | Parallel reaction monitoring |
| PROTAC | Proteolysis targeting chimeras |
| PSM | Peptide spectrum match |
| PSP | PhosphoSitePlus |
| PTM | Post-translational modification |
| pSILAC | Pulsed stable isotopic labelling of amino acids in cell culture |
| R | Pearson’s correlation coefficient |
| R0 | Arg-$^{12}C_6{}^{14}N_4$ (arginine containing only light isotopes of carbon and nitrogen) |
| R10 | Arg-$^{13}C_6{}^{15}N_4$ (arginine containing 10 stable, heavy isotopes of carbon and nitrogen) |
| RF | Radio frequency |
| RNA | Ribonucleic acid |
| ROC | Receiver operator characteristic |
| RP | Reversed-phase |
| rpm | Rounds per minute |
| RT | Room temperature |
| SAP | Single-amino acid polymorphism |
| SAX | Strong anion exchange |
| SILAC | Stable isotopic labelling of amino acids in cell culture |
| SNP | Single-nucleotide polymorphism |
| SPE | Solid-phase extraction |
| SPS | Synchronous precursor selection |
| STAGEtip | Stop-and-go-extraction tip |
| TAP | Tandem affinity purification |
| TFA | Trifluoroacetic acid |
| TIS | Translation initiation site |
| TMT | Tandem mass tag |
| Ub | Ubiquitin |
| UPS | Ubiquitin-proteasome system |
| YPD | Yeast peptone dextrose |

# LIST OF FIGURES

**Figure I-1** | Sources of proteome complexity. ..... 4
**Figure I-2** | Frequencies of curated post-translational modifications and types of modified residues of human proteins from the Swiss-Prot database. ..... 5
**Figure I-3** | Writer-reader-eraser principle. ..... 8
**Figure I-4** | Impact of proteoform diversity on gene function. ..... 9
**Figure I-5** | The ubiquitin-proteasome system (UPS). ..... 12
**Figure I-6** | The cellular proteostasis network. ..... 14
**Figure I-7** | Methods to quantify protein turnover. ..... 17
**Figure I-8** | Challenges in analysis of proteoforms with bottom-up proteomics. ..... 19
**Figure I-9** | Typical bottom-up proteomics workflow. ..... 21
**Figure I-10** | Enrichment approaches for modified peptides. ..... 22
**Figure I-11** | The Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer. ..... 23
**Figure I-12** | Mass filter and analysers in the Fusion Lumos mass spectrometer. ..... 25
**Figure I-13** | Peptide fragmentation upon collision-type dissociation methods. ..... 27
**Figure I-14** | Isobaric tags and data acquisition in MS3 mode. ..... 28
**Figure I-15** | Peptide and protein identification. ..... 31
**Figure I-16** | Common quantification approaches in bottom-up proteomics. ..... 33
**Figure III-1** | Experimental designs for the comparison of label-free and TMT-based quantification. ..... 64
**Figure III-2** | Generation of an MS3 standard. ..... 65
**Figure III-3** | Workflow for the optimization of the TMT-labelling protocol. ..... 67
**Figure III-4** | Linear dynamic range of label-free and TMT-based phosphopeptide quantification examined with synthetic peptide spike-ins. ..... 70
**Figure III-5** | Missing values and localization probabilities for label-free and TMT-based phosphopeptide quantification in synthetic peptide spike-in experiments. ..... 71
**Figure III-6** | Comparison of label-free and MS3-TMT based phosphopeptide identification and quantification for time-dependent 17-AAG treatments. ..... 72
**Figure III-7** | Evaluation of TMT-based phosphopeptide quantification in MS2 and MS3 mode using a yeast-human species-mix sample. ..... 74
**Figure III-8** | Statistical recovery of differential phosphopeptide abundances in the yeast-human species-mix sample ..... 75
**Figure III-9** | Optimization of MS2 acquisition for MS3 mode measurements of phosphopeptides. ..... 76
**Figure III-10** | Peptide titration experiments using the vendor recommended (A-D) and a down-scaled (E-H) TMT-labelling protocol. ..... 78
**Figure III-11** | Selection of peptide titration experiments using smaller peptide quantities. ..... 80
**Figure III-12** | TMT titration experiments using the downscaled TMT labelling strategy across laboratories. ..... 81
**Figure III-13** | Benchmarking the optimized protocol for deep-scale (phospho)proteomic analysis. ..... 83
**Figure IV-1** | Experimental designs for the evaluation of the pulsed SILAC-TMT approach. ..... 103
**Figure IV-2** | Workflow for the assessment of respiratory chain complex I turnover upon rotenone-induced oxidative stress. ..... 104
**Figure IV-3** | Schematic representation of the multiplexed pulsed SILAC-TMT strategy for estimation of protein synthesis and degradation employed in this study. ..... 106
**Figure IV-4** | Comparison of MS1 (pulsed SILAC) and MS3 (pulsed SILAC-TMT) based quantification. ..... 108
**Figure IV-5** | Reproducibility of protein turnover rate determination using pulsed SILAC-TMT. ..... 109
**Figure IV-6** | Analysis of protein half-lives relating to intrinsic protein properties and functions. ..... 110
**Figure IV-7** | Analysis of turnover of respiratory chain complex I proteins in response to rotenone-induced, oxidative stress. ..... 113
**Figure IV-8** | Analysis of peptidoform-resolved turnover. ..... 114

**Figure V-1** | Experimental designs of the two strategies for the assessment of modification-regulated protein turnover. ..... 133
**Figure V-2** | Refinement of pulsed SILAC-TMT data processing. ..... 136
**Figure V-3** | Characteristics of the single time-point pulse dataset. ..... 137
**Figure V-4** | Integration of data from the two different pulsed SILAC approaches. ..... 138
**Figure V-5** | Global differences in turnover of modified peptidoforms. ..... 138
**Figure V-6** | Traits and functional annotations of assigned proteins for peptides with significantly different turnover. ..... 142
**Figure V-7** | Functional annotations and motifs of PTMs resulting in significantly different turnover. ..... 147
**Figure V-8** | Turnover counterplay of lysine modifications. ..... 149
**Figure 0-1** | Qualitative and quantitative bias in the label-free measured samples with the highest phosphopeptide spike-in. ..... XXXI
**Figure 0-2** | Comparison of label-free and MS3-TMT based phosphopeptide identification for time-dependent 17-AAG treatments. ..... XXXI
**Figure 0-3** | Comparison of TMT-based quantification of non-phosphopeptides in MS2 and MS3 mode using a yeast-human species-mix sample. ..... XXXII
**Figure 0-4** | Statistical recovery of differential non-phosphopeptide abundances in the yeast-human species mixed sample. ..... XXXII
**Figure 0-5** | Evaluation of MS2 isolation windows (A-C) and number of fragment precursors for SPS (D-E) on quantitative performance of MS3 mode measurements. ..... XXXIII
**Figure 0-6** | Peptide and TMT titration experiments using the vendor recommended (A) and downscaled (B, C) TMT labelling strategies. ..... XXXIII
**Figure 0-7** | Normalization procedure and filtering criteria for pulsed SILAC-TMT data. ..... XXXVI
**Figure 0-8** | Cell doubling of HeLa cells in different cell culture replicates. ..... XXXVI
**Figure 0-9** | Comparison of MS1, MS2 and MS3-based methods for protein turnover rate estimation. ..... XXXVII
**Figure 0-10** | Reproducibility of peptide rate determination by pulsed SILAC-TMT. ..... XXXVIII
**Figure 0-11** | Biophysical and functional determinants of cellular protein stability. ..... XXXIX
**Figure 0-12** | Impact of oxidative stress on the turnover of respiratory chain complex I proteins. ..... XL
**Figure 0-13** | Analysis of peptidoform-resolved protein turnover. ..... XLI
**Figure 0-14** | Optimization of immunoprecipitations for acetyl-peptides. ..... XLII
**Figure 0-15** | Optimization of immuno-affinitpurification of di-glycine remnant peptides. ..... XLIII
**Figure 0-16** | Refinement of pulsed SILAC-TMT data processing. ..... XLIV
**Figure 0-17** | Estimation of turnover rates from single time-point pulse data. ..... XLIV
**Figure 0-18** | Peptides included in significance and enrichment analyses. ..... XLV
**Figure 0-19** | Peptides with significantly different turnover in subunits of the 19S proteasome forming the regulatory particle of the proteasome. ..... XLVI
**Figure 0-20** | Ubiquitin-remnant and acetylation sites leading to significantly different turnover. ..... XLVII
**Figure 0-21** | Enzymatic regulated acetyl-lysine sites. ..... XLVIII

# LIST OF TABLES

**Table III-1** | Parameters altered during optimization of MS2 acquisition in MS3 mode measurements of phosphopeptides. ..... 66
**Table III-2** | Overview of labelling conditions of different experiment series for optimization of TMT labelling. ..... 68
**Table III-3** | Numbers of quantified, modified peptides sequences identified in MS2 and MS3 mode in the whole cell and phosphoproteome of the TMT standard and their species assignment. ..... 73
**Table III-4** | Theoretical amount of functional groups in a complete digest of 100 µg of a human proteome. ..... 77
**Table V-1** | Motifs enriched among PTMs leading to significantly different peptidoform turnover. ..... 147
**Table 0-1** | Overview of LC-MS parameters for measurement of TMT titration samples. ..... XXXIV
**Table 0-2** | Dynamic range constraints for TMT quantification due to its compositional nature..... XXXV

# APPENDIX

**1 A case for robust peptide quantification ..... XXXI**

**1.1** A compromise between accurate dynamic range and reproducible coverage ..... XXXI

1.2 A versatile TMT standard ..... XXXII

1.3 A robust and cost-efficient TMT-labelling protocol ..... XXXIII

1.4 Confining the dynamic range of peptide quantification ..... XXXV

**2 Towards proteoform-resolved analysis of turnover ..... XXXVI**

**2.1** Design of a pulsed SILAC-TMT experiment ..... XXXVI

2.2 Benchmarking the pulsed SILAC-TMT approach ..... XXXVII

2.3 Evaluation of determinants of turnover ..... XXXIX

2.4 Turnover diversity of peptidoforms ..... XLI

**3 Deciphering modification-regulated turnover ..... XLII**

**3.1** Optimization of acetyl and di-glycine remnant enrichments ..... XLII

3.2 Two approaches for the investigation of PTM-specific turnover ..... XLIV

3.3 Characterization of peptidoforms with differential turnover ..... XLV

3.4 Functional perspective on modification-specific turnover ..... XLVII

# 1 A case for robust peptide quantification

## 1.1 A compromise between accurate dynamic range and reproducible coverage

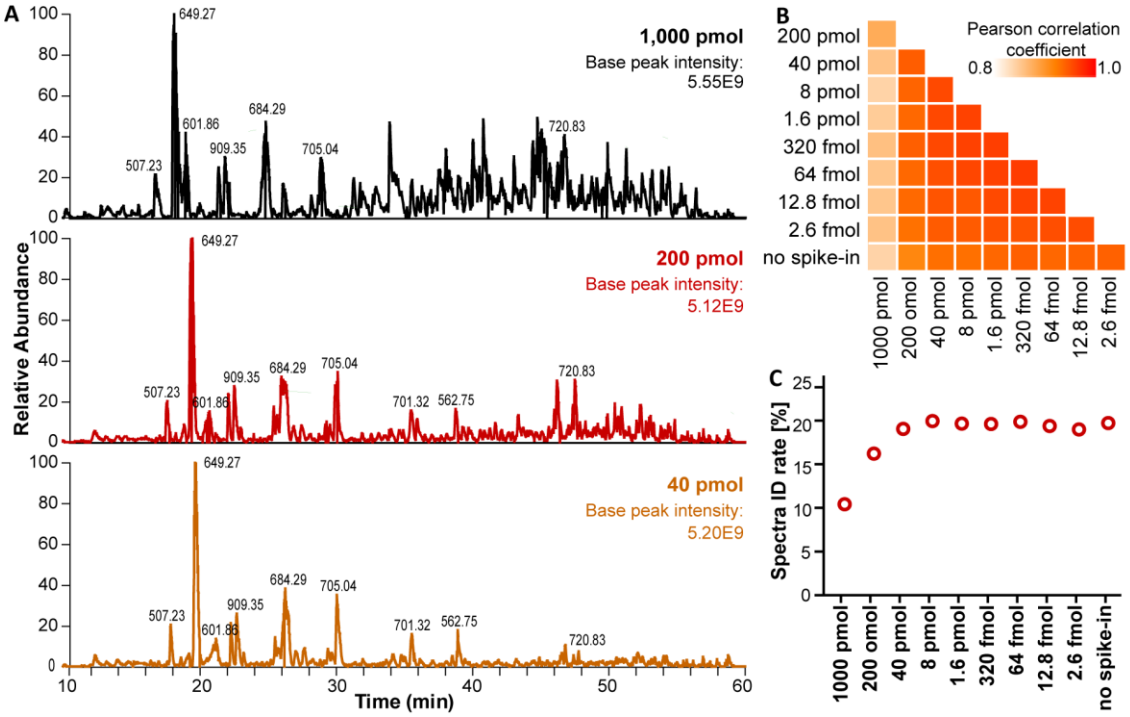



**Figure 0-1 | Qualitative and quantitative bias in the label-free measured samples with the highest phosphopeptide spike-in.** *(A) A retention time shift for early eluting peptides is visible in the sample with the highest quantity of spiked-in peptides (1,000 fmol estimated starting amount per peptide). (B) Pearson's correlation coefficients are displayed for the correlations of Hela peptides that are supposed to be present in constant amounts in label-free samples containing increasing amounts of synthetic phosphopeptides. (C) MS2 spectra identification* rates are shown for samples with different quantities of phosphopeptide spike-ins.

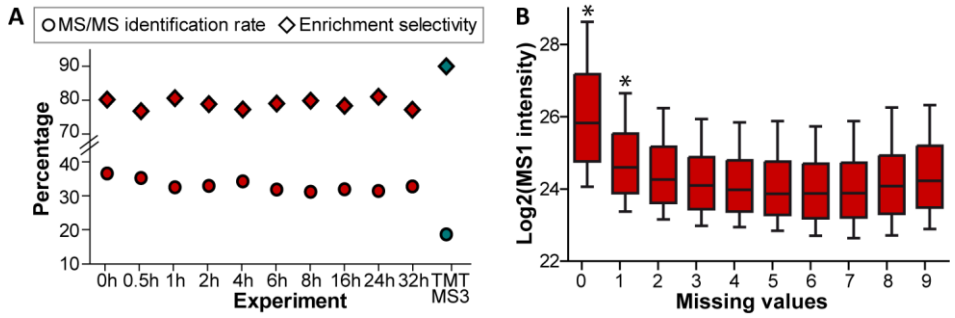



**Figure 0-2 | Comparison of label-free and MS3-TMT based phosphopeptide identification for time-dependent 17-AAG treatments.** *(A) Spectra identification rate and selectivity (fraction of phosphopeptides of all identified peptides) are shown for the two quantification approaches. (B) Boxplots illustrate intensities of peptides with increasing numbers of missing values in the label-free dataset. Asterisks mark distributions that are significantly different from all other distribution in a Kruskal-Wallis and Dunn's post hoc test (α=0.01; box: 25$^{th}$-50$^{th}$-75$^{th}$ percentile; whiskers: 10$^{th}$ and 90$^{th}$ percentile).*

## 1.2 A versatile TMT standard

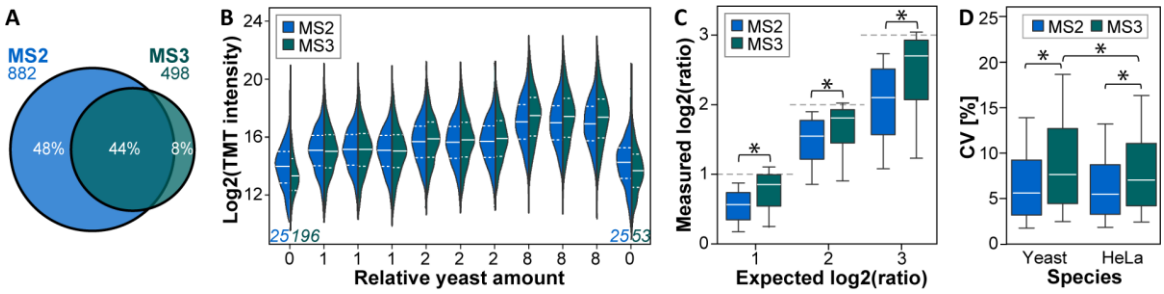



**Figure 0-3 | Comparison of TMT-based quantification of non-phosphopeptides in MS2 and MS3 mode using a yeast-human species-mix sample.** *(A) The overlap of identified yeast peptides is shown. (B) TMT intensity distributions of yeast peptides are displayed for different channels containing different amounts of yeast sample. Intensities detected in the first and the last channel arise from co-isolation of human peptides. The number of peptides with zero intensities in these channels is indicated. (C) Distributions of measured peptide intensity ratios (calculated from intensity averages of each group) are illustrated in comparison to the expected ratios for different yeast amounts. (D) Coefficients of variation (CV) are shown for HeLa and yeast peptides. Asterisks denote significantly different distributions (Mann-Whitney p-value < 0.0001).*

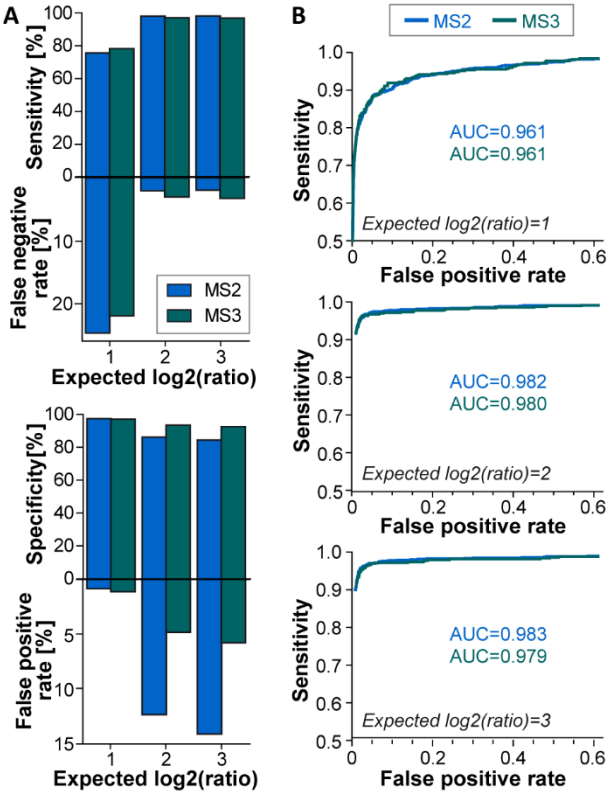



**Figure 0-4 | Statistical recovery of differential non-phosphopeptide abundances in the yeast-human species mixed sample.** *Sensitivity (significant yeast fraction) and specificity (non-significant human fraction) and corresponding false negative and positive rates are shown for different expected ratios and MS2 and MS3 mode measurements. Numbers are based on results of one-sided student's t-tests comparing groups of TMT channels that contain different yeast quantities and applying a 5% permutation based FDR and S0 cut-off. (B) ROC curves computed from corrected p-values (A) are displayed (AUC: area under the curve).*

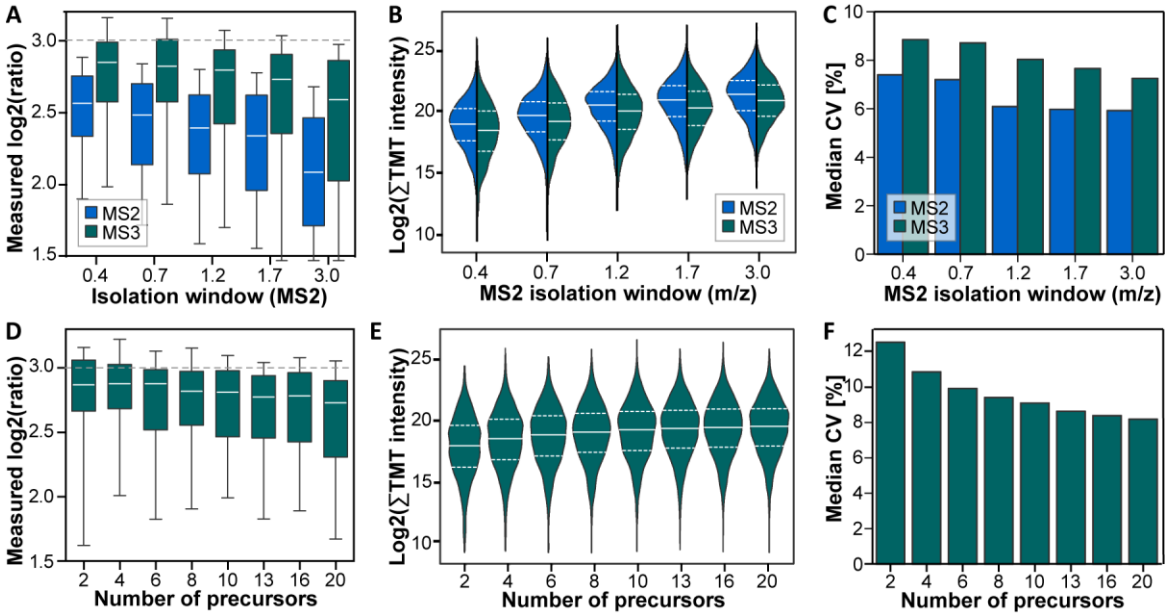



**Figure 0-5 | Evaluation of MS2 isolation windows (A-C) and number of fragment precursors for SPS (D-E) on quantitative performance of MS3 mode measurements.** *(A, D) The distributions of summed TMT intensities of all shared, identified peptides are displayed. (B, E) Bar charts indicate the median coefficients of variation (CVs). (C, F) Boxplots show the distributions of log ratios (calculated from intensity averages of each group) which would be expected to be 3. Only peptide shared across different methods were included.*

## 1.3 A robust and cost-efficient TMT-labelling protocol

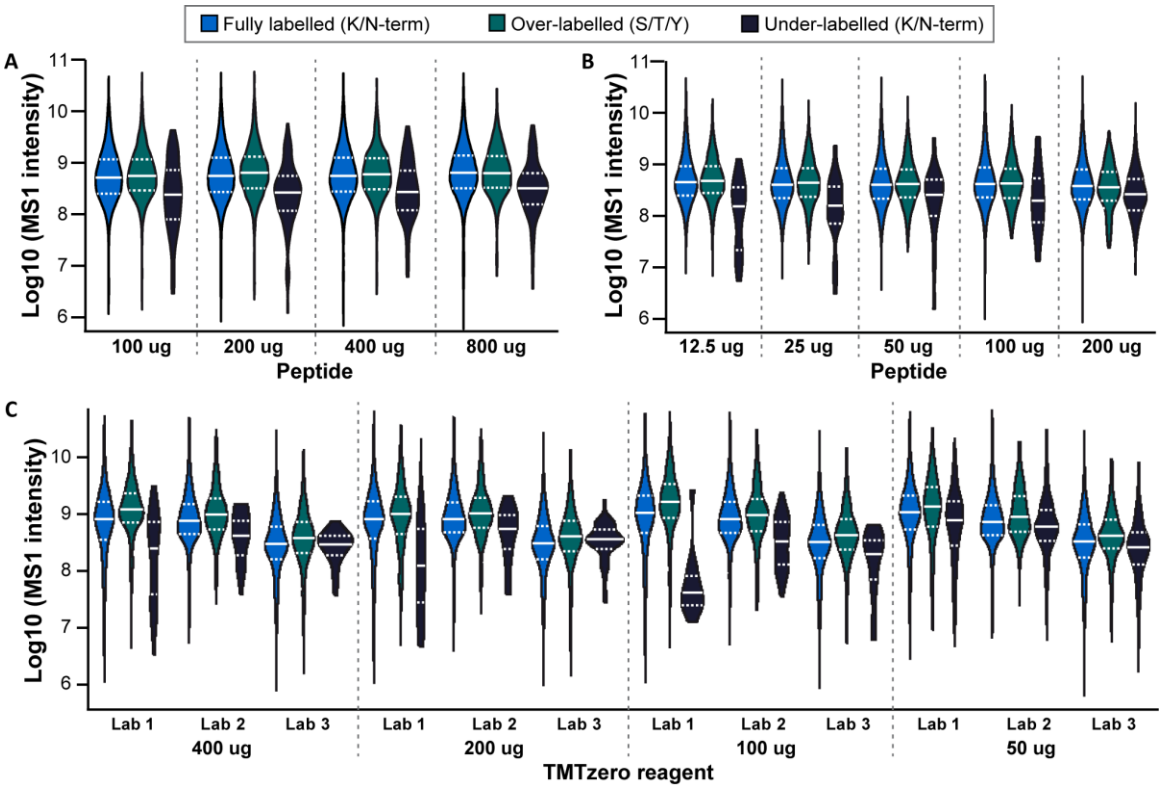



**Figure 0-6 | Peptide and TMT titration experiments using the vendor recommended (A) and downscaled (B, C) TMT labelling strategies.** *(A) MS1 intensities of fully, over- and underlabelled peptides are shown for the labelling experiment series illustrated in main Figure III-10 A. Intensities of fully labelled peptides were extracted from the underlabelling search. (B) Same as (A) but for the peptide titration experiments displayed in main figure Figure III-10 E. (C) Same as (A) but for the intra-lab, TMT titration experiments shown in main* Figure III-12 *A.*

**Table 0-1 | Overview of LC-MS parameters for measurement of TMT titration samples.** *The length of the effective gradient and method are specified in min, mass range, isolation window, and fixed first mass in m/z, and maxITs in ms. Material for trap (2 cm x 75 µm, 5 µm C18 resin, Reprosil PUR AQ) and analytical columns (1.9 or 3 µm C18 resin, Reprosil Gold) was purchased from Dr. Maisch GmbH (QE: Q Exactive).*

| | VENDOR / DOWN-SCALED 1 | DOWN-SCALED 2 | JURKAT/PDX | LOW IN-PUT | INTER-LAB 1 | INTER-LAB 2 | INTER-LAB | DEEP-SCALE |
|---|---|---|---|---|---|---|---|---|
| **LC system** | EASY-nLC 1200 | Ultimate 3000 RSLC-nano | EASY-nLC 1200 | EASY-nLC 1200 | EASY-nLC 1200 | EASY-nLC 1200 | Ultimate 3000 RSLC-nano | EASY-nLC 1200 |
| **Trap column** | ----- | Yes | ----- | ----- | ----- | ----- | Yes | ----- |
| **Analytical column** | 20 cm x 75 um, 1.9 um | 45 cm x 75 um, 3 um | 22 cm x 75 um, 1.9 um | 22 cm x 75 um, 1.9 um | 22 cm x 75 um, 1.9 um | 20 cm x 75 um, 1.9 um | 45 cm x 75 um, 3 um | 22 cm x 75 um, 1.9 um |
| **Flow rate (nL/min)** | 250 | 300 | 200 | 200 | 200 | 250 | 300 | 200 |
| **Solvent A** | 0.1 % FA, 3 % ACN | 0.1 % FA, 5 % DMSO | 0.1 % FA, 3 % ACN | 0.1 % FA, 3 % ACN | 0.1 % FA, 3 % ACN | 0.1 % FA, 3 % ACN | 0.1 % FA, 5 % DMSO | 0.1 % FA, 3 % ACN |
| **Solvent B** | 0.1 % FA, 90 % ACN | 0.1 % FA, 5 % DMSO in ACN | 0.1 % FA, 90 % ACN | 0.1 % FA, 90 % ACN | 0.1 % FA, 90 % ACN | 0.1 % FA, 90 % ACN | 0.1 % FA, 5 % DMSO in ACN | 0.1 % FA, 90 % ACN |
| **Gradient (B in A)** | 4-60 % | 8-34 % | 3-55 % | 3-55 % | 2-55 % | 5-60 % | 8-34 % | 6 to 30 % |
| **Effective gradient** | 97 | 20 | 20 | 20 | 20 | 17 | 20 | 84 |
| **Method length** | 110 | 30 | 30 | 30 | 30 | 30 | 30 | 110 |
| **MS system** | QE Plus | Lumos | QE Plus | QE Plus | QE HF-X | QE HF-X | QE HF-X | Lumos |
| **MS1 resolution** | 70k | 60k | 70k | 70k | 60k | 60k | 60k | 60k |
| **MS1 AGC** | 3.E+06 | 4.E+05 | 1.E+06 | 3.E+06 | 3.E+06 | 3.E+06 | 3.E+06 | 4.E+05 |
| **MS1 maxIT** | 50 | 10 | 5 | 5 | 10 | 10 | 10 | 50 |
| **MS1 mass range** | 350-2,000 | 360-1,500 | 300-2,000 | 300-1,800 | 300-2,000 | 300-2,000 | 300-2,000 | 350-1,800 |
| **MS2 resolution** | 70k | 15k | 35k | 15.5k | 30k | 30k | 30k | 50k |
| **MS2 AGC** | 5.E+04 | 5.E+04 | 5.E+04 | 5.E+04 | 5.E+04 | 5.E+04 | 5.E+04 | 6.E+04 |
| **MS2 maxIT** | 50 | 22 | 120 | 120 | 30 | 30 | 30 | 105 |
| **Precursor mode** | Top10 | Top20 | Top10 | Top10 | Top10 | Top10 | Top10 | 2 s |
| **Isolation window** | 0.7 | 0.7 | 2.5 | 2.5 | 0.7 | 0.7 | 0.7 | 0.7 |
| **Fixed first mass** | 120 | 100 | 100 | 100 | ---- | ---- | ---- | 110 |
| **NCE** | 32 | 31 | 31 | 30 | 31 | 32 | 31 | 36/38 |
| **Dynamic exclusion (s)** | 30 | 10 | 20 | 20 | 10 | 10 | 10 | 45 |

## 1.4 Confining the dynamic range of peptide quantification

**Table 0-2 | Dynamic range constraints for TMT quantification due to its compositional nature**. *Theoretical considerations reveal the influence of the experimental design on maximal detectable dynamic range of TMT quantification (encircled). The 5-fold dilution corresponds to the above conducted experiment including one empty TMT channel. Calculations are based on the employed AGC target value of 1e5 and the assumption that only TMT reporter ions are present, no co-isolation and ratio compression exists and ~100 charges are sufficient to induce a signal in the Orbitrap.*

| TMT CHANNEL | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | TOTAL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2-FOLD DILUTION | INTENSITY FRACTION | 0.50 | 0.25 | 0.13 | 0.063 | 0.031 | 0.016 | 8.0e-3 | 4.0e-3 | 2.0e-3 | **1** |
| | CHARGES | 5.0e4 | 2.5e4 | 1.3e4 | 6,300 | 3,100 | 1,600 | 800 | 400 | 200 | **1.0e5** |
| | DYNAMIC RANGE | -- | 2 | 4 | 8 | 16 | 32 | 64 | 130 | **260** | |
| 5-FOLD DILUTION | INTENSITY FRACTION | 0.80 | 0.16 | 0.032 | 6.4e-3 | 1.3e-3 | 2.6e-4 | 5.1e-5 | 1.0e-5 | 2.0e-6 | **1** |
| | CHARGES | 8.0e4 | 1.6e4 | 3.2e3 | 640 | 130 | 26 | 5 | 1 | -- | **1.0e5** |
| | DYNAMIC RANGE | -- | 5 | 25 | 130 | 630 | **3,100** | 1.6e4 | 7.8e4 | 3.9e5 | |
| 10-FOLD DILUTION | INTENSITY FRACTION | 0.90 | 0.090 | 9.0e-3 | 9.0e-4 | 9.0e-5 | 9.0e-6 | 9.0e-7 | 9.0e-8 | 9.0e-9 | **1** |
| | CHARGES | 9.0e4 | 9,000 | 900 | 90 | 9 | 1 | -- | -- | -- | **1.0e5** |
| | DYNAMIC RANGE | -- | 10 | 100 | **1,000** | 1e4 | 1e5 | 1e5 | 1e6 | 1e7 | |

# 2 Towards proteoform-resolved analysis of turnover

## 2.1 Design of a pulsed SILAC-TMT experiment

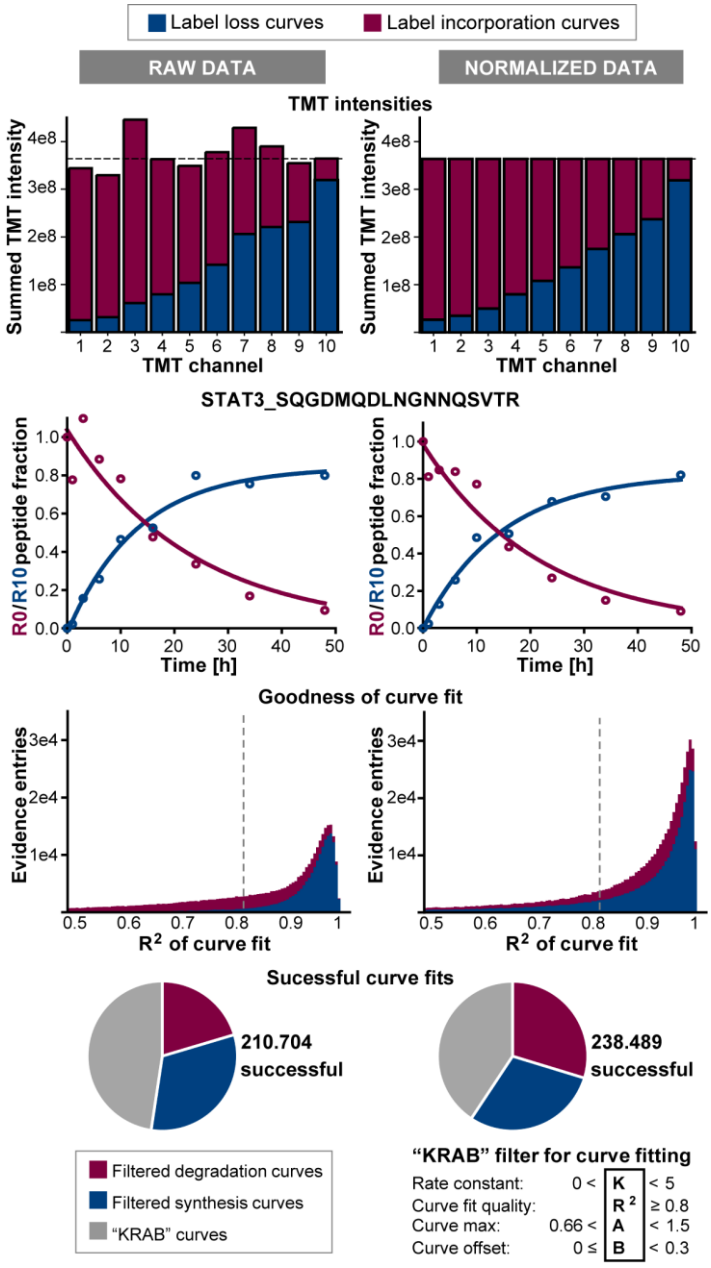



**Figure 0-7 | Normalization procedure and filtering criteria for pulsed SILAC-TMT data.** *In order to correct for potential mixing errors, TMT data were normalized based on the assumption that the total protein amount (light and heavy labelled protein) is equal across time-points. For this means, only peptides found in both labelling states were considered. This procedure decreased the variance of the data and improved the coefficient of determination ($R^2$) of the curve fit, ultimately increasing the number of peptide evidence entries passing the filter criteria after curve fitting for the rate K, the coefficient of determination $R^2$, the curve maximum A and the offset B ("KRAB"-filter).*

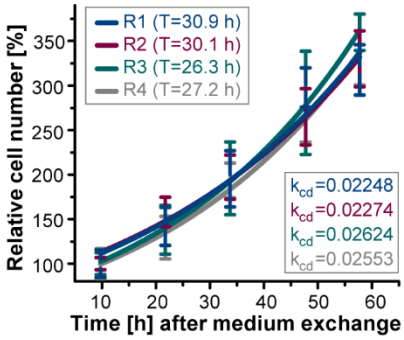



**Figure 0-8 | Cell doubling of HeLa cells in different cell culture replicates.** *For half-life calculations, rates of cell doubling were determined by cell counting for all four cell culture replicates separately in order to account for even minor differences in growth behaviour, for instance, due to variation in cell density before cell passage or initial seeding density.*

## 2.2 Benchmarking the pulsed SILAC-TMT approach

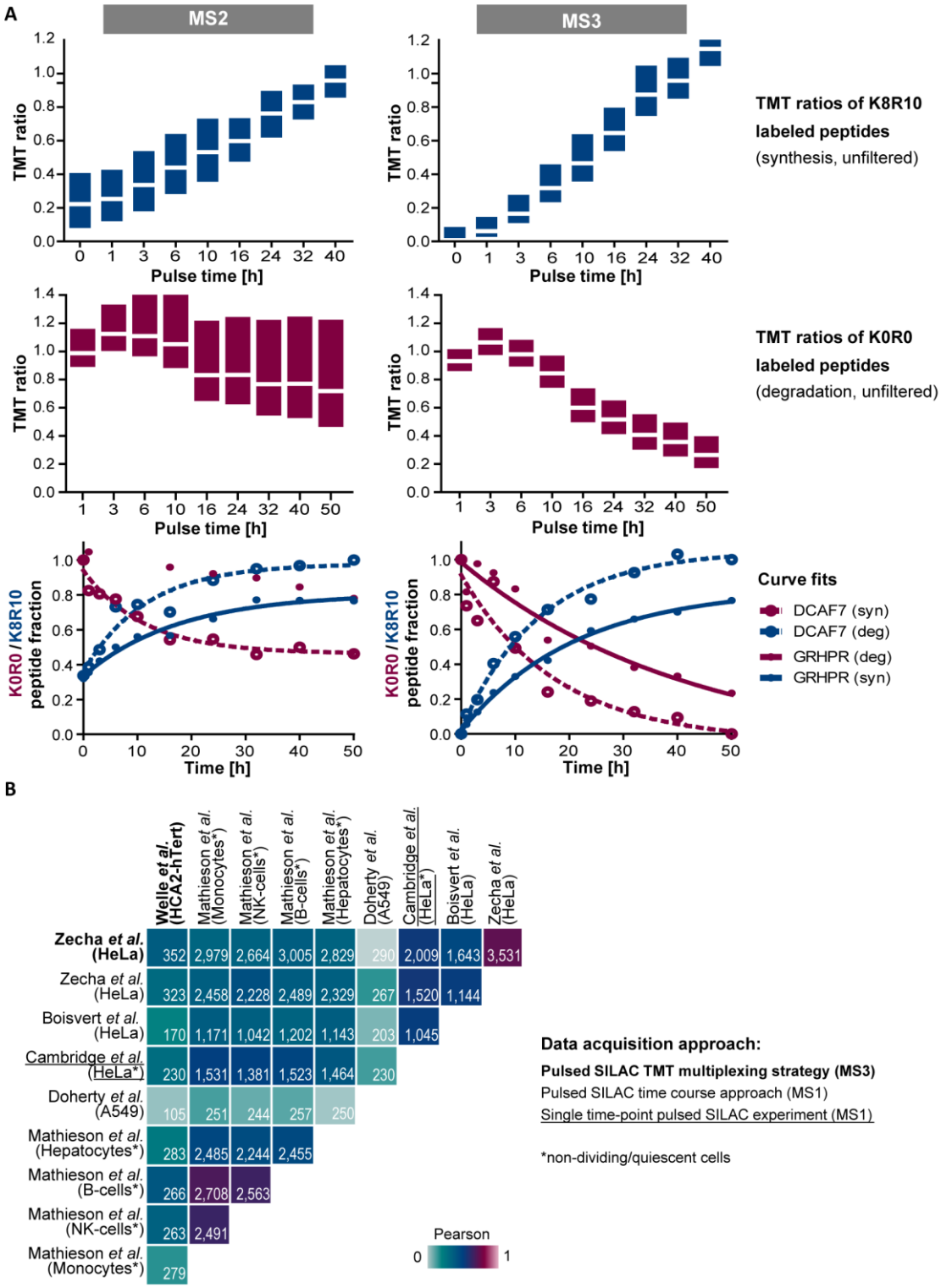



**Figure 0-9 | Comparison of MS1, MS2 and MS3-based methods for protein turnover rate estimation.** *(A) TMT-labelled pulsed SILAC lysates were fractionated into 6 fractions and measured using a MS2 and MS3-based method. Boxplots ($25^{th}$-$50^{th}$-$75^{th}$ percentile) of TMT intensity ratios show that MS2-based quantification suffered from ratio compression which severely distorted subsequent curve fittings and rate estimations. (B) Correlation matrix depicts color-coded Pearson's correlation coefficients for log transformed protein rates between the present work (Zecha et al.) and previously published datasets. Numbers inside cells indicate the number of proteins available for each correlation analysis.*

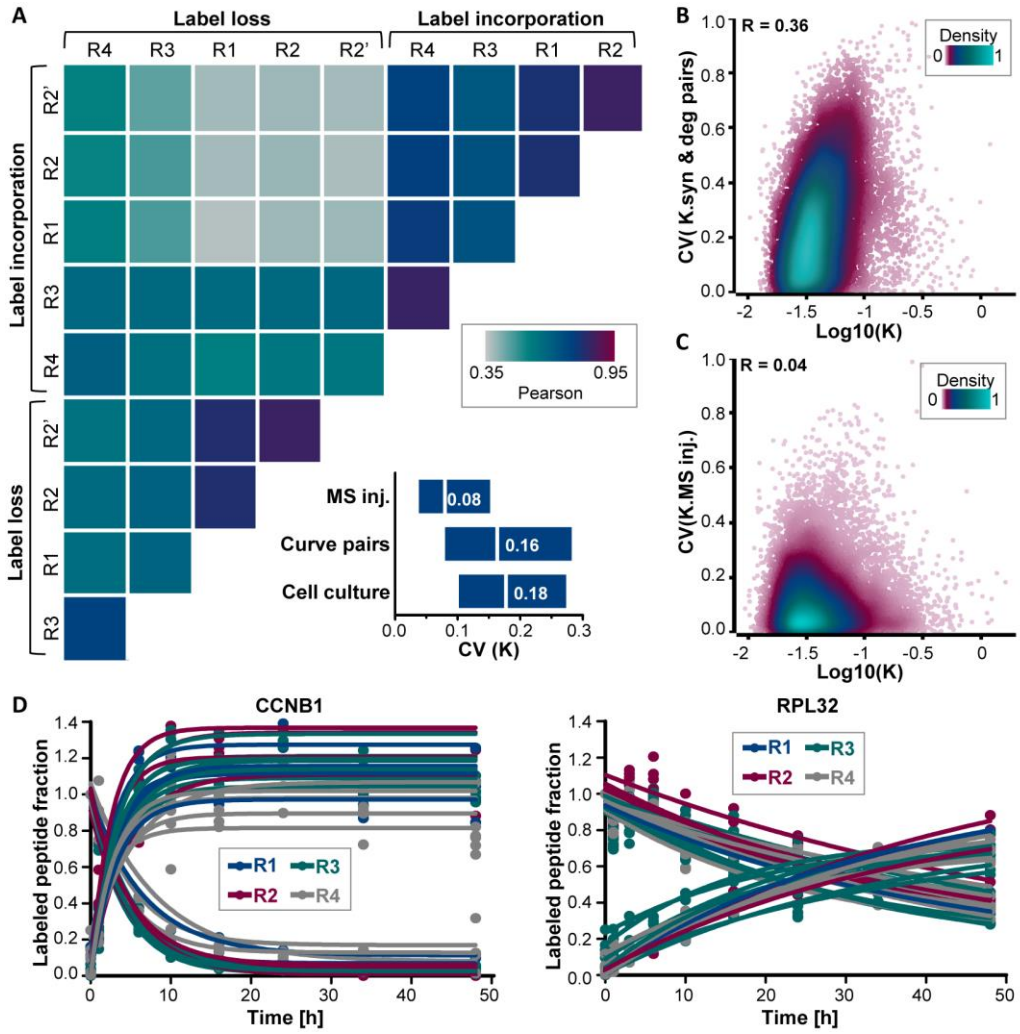



**Figure 0-10 | Reproducibility of peptide rate determination by pulsed SILAC-TMT.** *(A) Correlation matrix depicts color-coded Pearson's correlation coefficients for log transformed peptide turnover rates determined from synthesis and degradation curves for cell culture (R1-R4) and MS injection (R2 and R2') replicates. The boxplots ($25^{th}$-$50^{th}$-$75^{th}$ percentile) show the coefficients of variation of peptide turnover rates of label incorporation and label loss rates across MS injections, curves pairs within a sample and cell culture replicates. For some peptides, the precision of rate determination from increasing versus decreasing curves was compromised by residual ratio compression that deteriorated correlations. (B) Ratio distortion more likely affects curves of high turnover peptides (large values for K) as suggested by the weak positive correlation of rates and CVs computed from synthesis and degradation curve pairs (R: Pearson's correlation coefficient). (C) CVs across technical replicates did not correlate with rates indicating that synthesis and degradation behaviour of fast and slow turnover peptides can be identified with a comparably high precision. (D) Good agreement of peptide curve fits across cell culture replicates is exemplified for the fast turnover protein G2/mitotic-specific cyclin-B1 (CCNB1) and the very stable 60S ribosomal protein L32 (RPL32).*

## 2.3 Evaluation of determinants of turnover

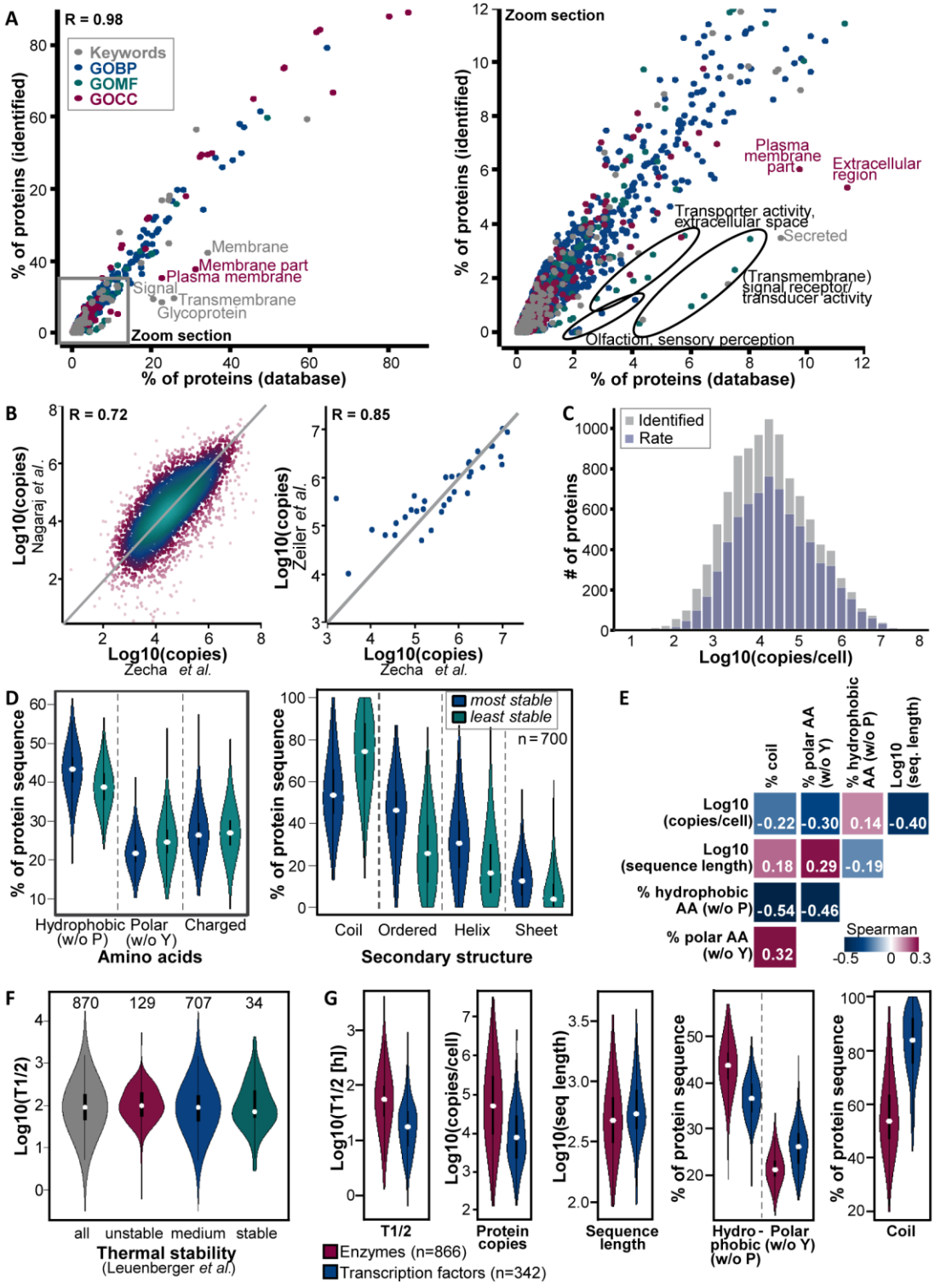



**Figure 0-11 | Biophysical and functional determinants of cellular protein stability.** *(A) The percentage of canonical protein isoforms in functional Gene Ontology (GO) and UniProt Keyword categories was highly correlated for all proteins from the database and the subset of proteins for which turnover rates were determined in this study (BP: biological process; MF: molecular function; CC: cellular compartment). Only membrane associated and extracellular proteins were underrepresented in the set of identified proteins. (B) Protein copies per cell, which were computed utilizing TMT and MS1 intensity information showed a strong correlation with data published by Nagaraj et al.*[307] *and Zeiler et al.* [306]. *(C) Copies of proteins for which*

*rates of turnover were determined spanned seven orders of magnitude. (D) Violin plots display that the 10 % most and least stable proteins (n=700) differed significantly in the proportion of hydrophobic and polar amino acids as well as in the fraction of disordered secondary structure. (E) Correlation matrix indicates Spearman rank correlation coefficients illustrating that protein features are interdependent. (F) No significant difference of cellular protein half-lives was identified between the three categories of thermal stability determined by Leuenberger et al.* [308]. *Numbers display the number of proteins in the respective category. (G) Enzymes, which were significantly enriched in stable proteins, differed in abundance, and primary and secondary structure from transcription factors that were on average much less stable.*

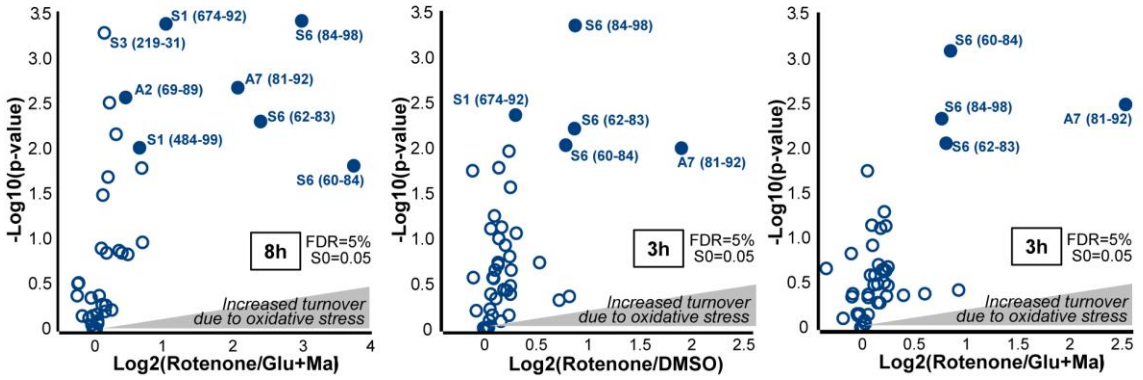



**Figure 0-12 | Influence of oxidative stress on the turnover of respiratory chain complex I proteins.** *Volcano plots illustrate that, after a 3 and 8h pulse with K8R10 medium, the heavy-to-light ratios of peptides belonging to NADH dehydrogenase proteins exhibited an overall shift towards higher ratios upon rotenone inhibition compared to the control treatments. This suggests an accelerated turnover due to increased oxidative stress. Significantly changing peptides are displayed by filled circles (two-sided t-test, n=3).*

## 2.4 Turnover diversity of peptidoforms

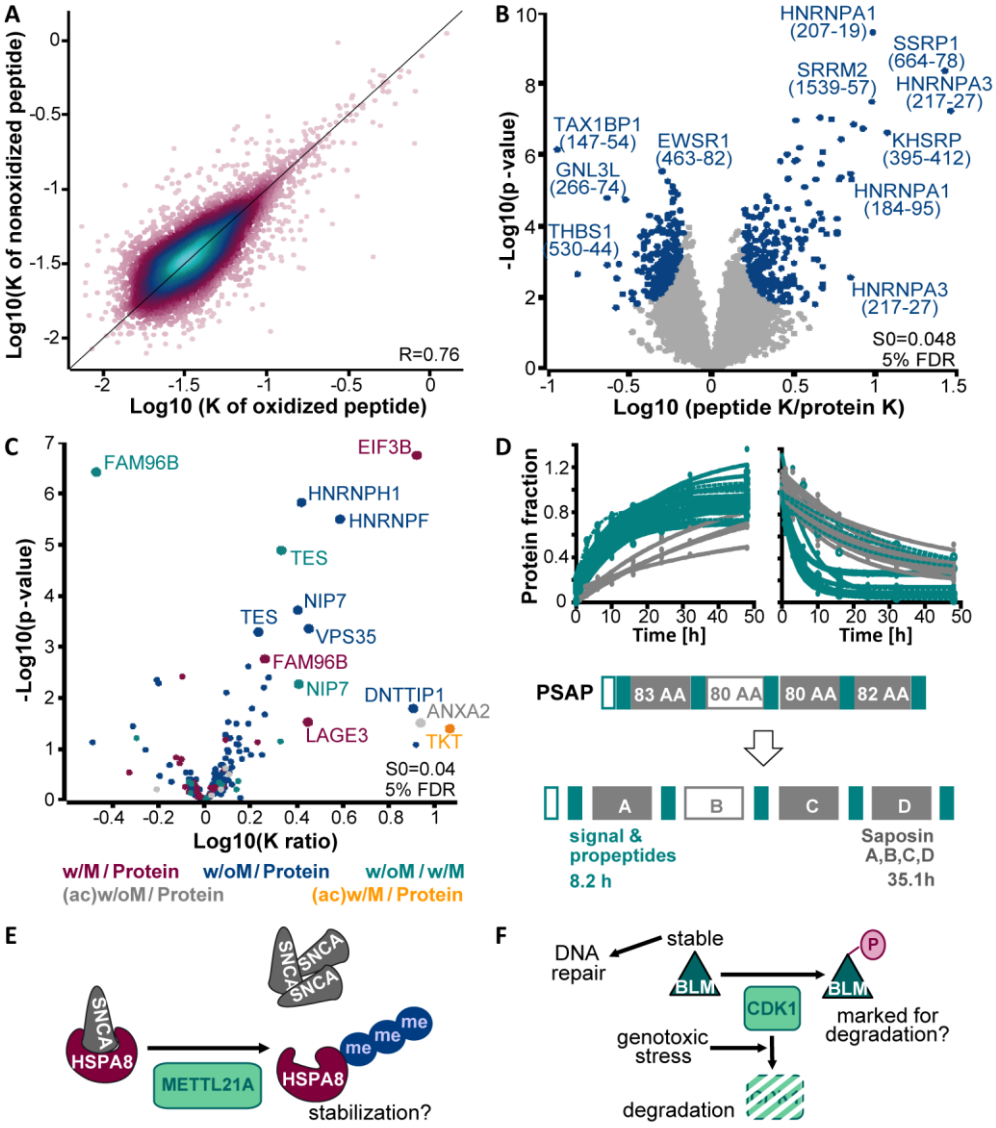



**Figure 0-13 | Analysis of peptidoform-resolved protein turnover**. *(A) Correlation analysis of log transformed labelling rates for oxidized peptides and their non-oxidized counterparts showed no global influence of detected oxidation on turnover. The analysis included peptide pairs from all 4 replicates (11,314 in total). (B) Turnover rates of all peptides were matched to corresponding protein rates and tested for significant differences in a two-sided t-test (S0=0.048, 5 % FDR). Peptides exhibiting significantly differing rates are coloured in blue. (C) N-terminal peptides rates were compared against each other and the associated protein rates in a two-sided t-test (S0=0.04, 5 % FDR). Significantly different pairs are annotated and also displayed in main* ***Figure IV-8 C****. (D) The propeptides of Prosaposin appeared to be much less stable than the mature saposins. Peptides encompassing the cleavage site of Saposin B and the following propeptide showed a slower turnover comparable to Saposin A, C and D, whereas peptides comprising the cleavage site of Saposin A and the successive propeptide showed a faster turnover matching that of propeptides (for Saposin B no peptides were detected, dotted lines indicate peptides that span cleavage sites). (E) The proposed stabilizing mechanism for the chaperone HSP8 via trimethylation of lysine 561 is displayed. This modification has also been described to lead to a reduced interaction with α-Synuclein (SNCA) and a subsequently increased aggregation* [383]*. (F) The connection between helicase BLM and CDK1 that likely destabilizes the helicase via a phosphorylation on threonine 766 is illustrated.*

# 3 Deciphering modification-regulated turnover

## 3.1 Optimization of acetyl and di-glycine remnant enrichments

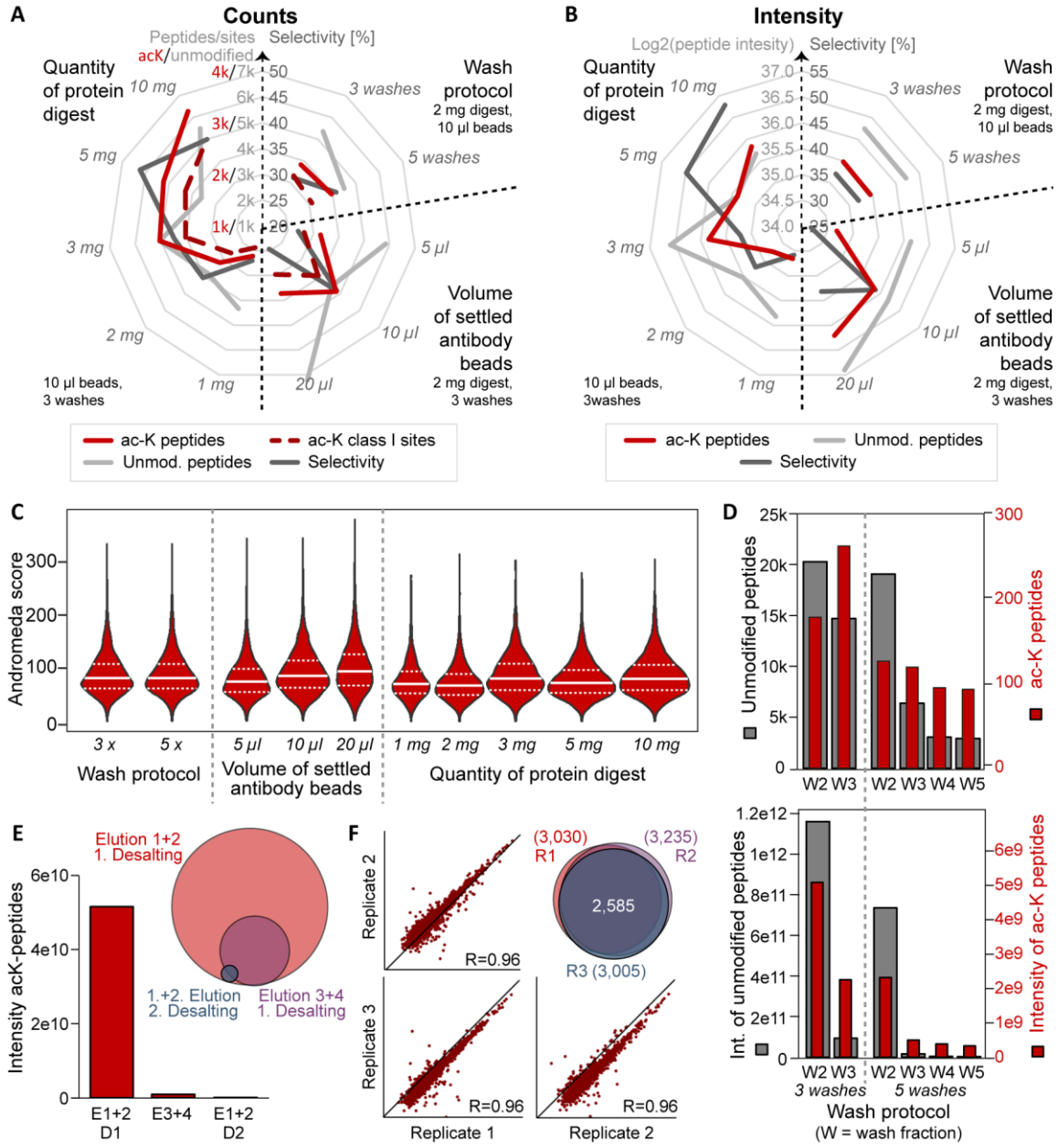



**Figure 0-14 | Optimization of immunoprecipitations for acetyl-peptides.** *Differing wash protocols (3 washes: 1x IAP buffer + 2x water: 5 washes: 2x IAP buffer + 3x PBS), antibody bead amounts, peptide quantities, and elution steps were tested using label-free HeLa digest in singlicates in independent experiment series. Match-between-runs was disabled for comparison in (A-C) but enabled for analysis in (D-E). (A) Numbers of acetylated (acK) and unmodified peptides, Class I acetyl-sites, and selectivity are plotted for indicated wash protocols, bead and peptide amounts. (B) Intensities of acetylated and unmodified peptides, and intensity-based selectivity are displayed for indicated wash protocols, bead and peptide amounts. (C) Score distributions of acetylated peptides are illustrated for indicated wash protocols, bead and peptide amounts. (D) Numbers and intensities of acetylated and unmodified peptides in the wash fractions of different wash protocols are shown (5 % injected into the MS). The first wash fraction was not measured. (E) Bound acetyl-peptides were eluted twice (E1+2) and then another two times (E3+E4). The desalting flow through of the first two elutions was loaded onto a second STAGE tip and desalted again (E1+2, D2). The intensity of acetylated peptides in different samples and their overlap is plotted. (F) Correlations (R: Pearson coefficient) and overlaps between 3 acetyl IP replicates are displayed.*

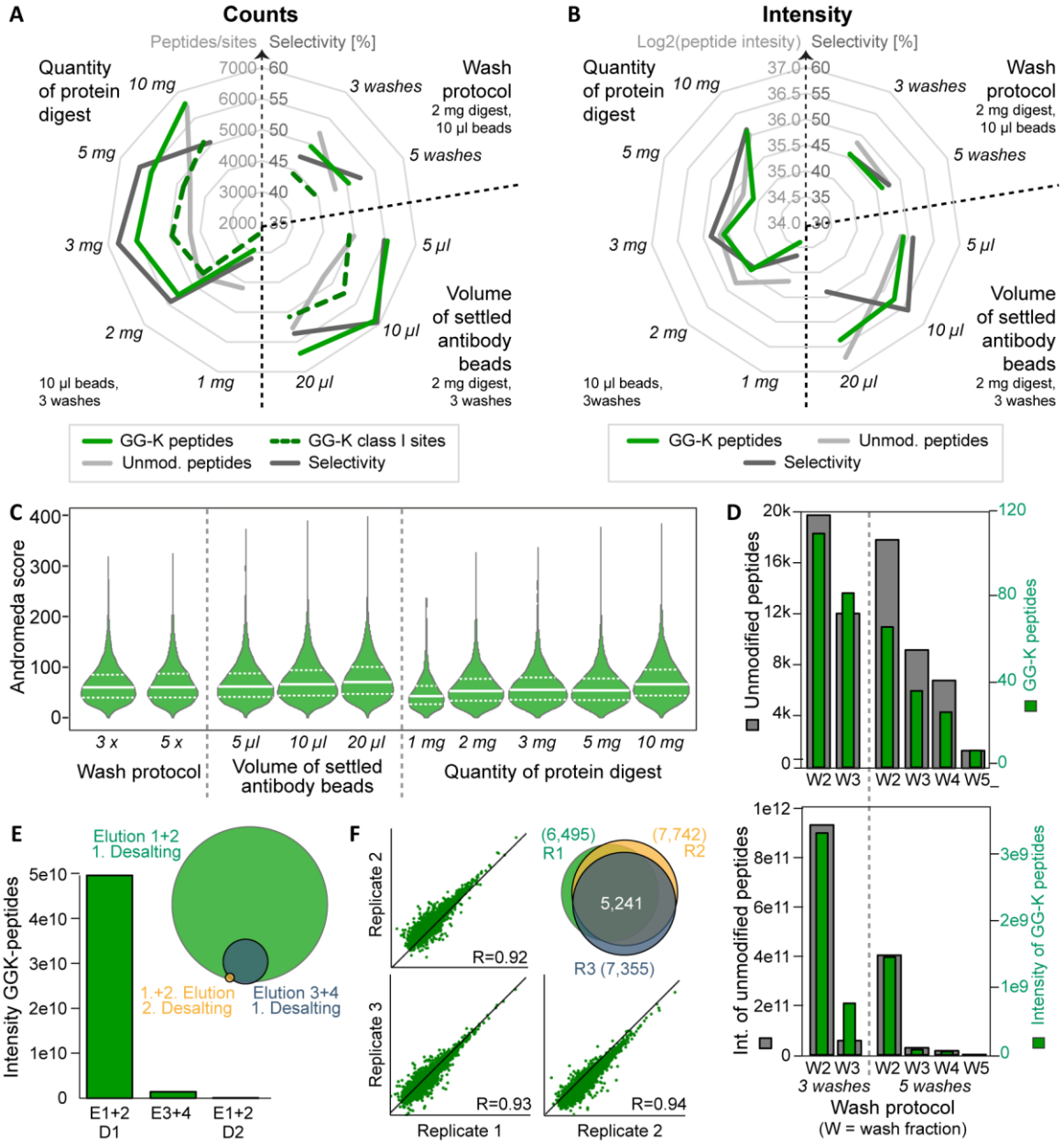



**Figure 0-15 | Optimization of immunoprecipitations of di-glycine remnant peptides**. *Differing wash protocols (3 washes: 1x IAP buffer + 2x water: 5 washes: 2x IAP buffer + 3x PBS), antibody bead amounts, peptide quantities, and elution steps were tested using label-free HeLa digest in singlicates in independent experiment series. Match-between-runs was disabled for comparison in (A-C) but enabled for analysis in (D-E). (A) Numbers of di-glycine remnant (GG-K) and unmodified peptides, Class I GG-K sites, and selectivity are plotted for indicated wash protocols, bead and peptide amounts. (B) Intensities of di-glycine remnant and unmodified peptides, and intensity-based selectivity are displayed for indicated wash protocols, bead and peptide amounts. (C) Score distributions of di-glycine remnant peptides are illustrated for indicated wash protocols, bead and peptide amounts. (D) Numbers and intensities of di-glycine remnant and unmodified peptides in the wash fractions of different wash protocols are shown (5 % injected into the MS). The first wash fraction was not measured. (E) Bound di-glycine peptides were eluted twice (E1+2) and then another two times (E3+E4). The desalting flow through of the first two elutions was loaded onto a second STAGE tip and desalted again (E1+2, D2). The intensity of di-glycine peptides in different samples and their overlap is plotted. (F) Correlations (R: Pearson coefficient) and overlaps between 3 di-glycine IP replicates are displayed.*

## 3.2 Two approaches for the investigation of PTM-specific turnover

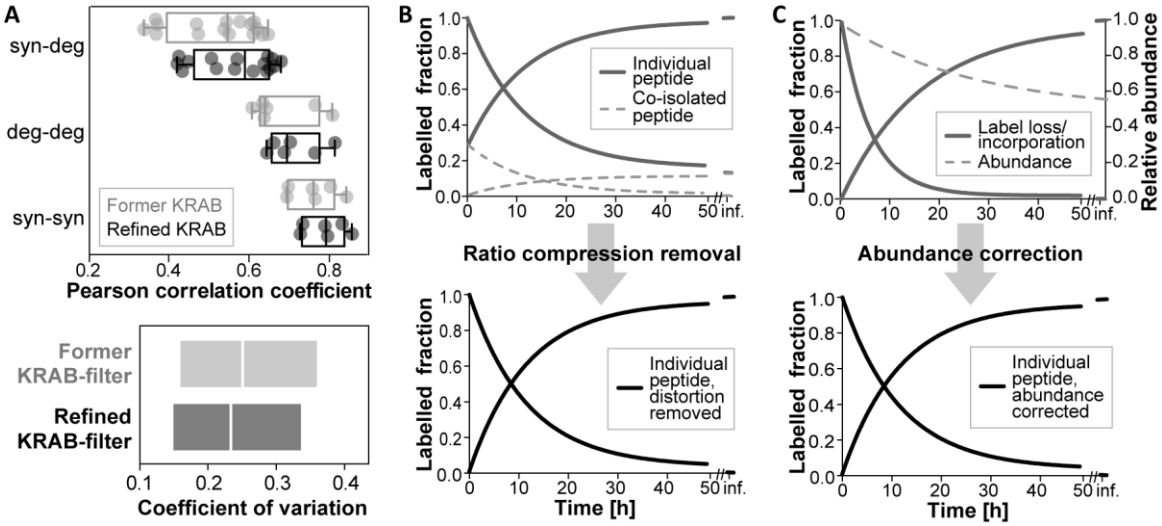



**Figure 0-16 | Refinement of pulsed SILAC-TMT data processing.** *(A) Pearson correlation coefficients and coefficients of variation are displayed for the former and refined KRAB-filter) that is applied to parameters of fitted curves (K: 0-5; B: -0-0.3; A: 0.67-1.5; $R^2$≥0.8 vs. K: 0-5; B: -0.15-0.25; A: 0.7-1.4; $R^2$≥0.7, see also Figure 0-7). (B) The principle of the computational removal of ratio compression is illustrated. It was based on the assumption that ratio compression of increasing/decreasing curves was caused by co-isolation of a peptide exhibiting the opposite, label loss/incorporation behaviour. To remove these co-isolated intensities, average label loss or incorporation curves were subtracted from individual peptide intensities. (C) The principle of the correction for abundance changes is shown. Instead of employing the channel indicating 100 % label incorporation as total peptide abundance (0 h or infinite h time-point) to calculate intensity ratios for each time-point, the actual peptide abundance at the respective time-point was derived from the sum of the heavy and light version of the peptide and used to compute respective ratios.*

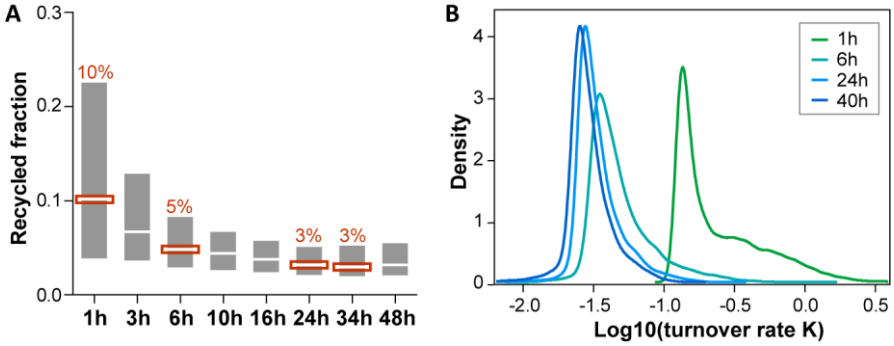



**Figure 0-17 | Estimation of turnover rates from single time-point pulse data.** *(A) Boxplots ($25^{th}$-$50^{th}$-$75^{th}$ percentile) illustrate the recycled fraction calculated from missed cleavage peptides that contained at least one isotopically heavy and light amino acid and were identified in the pulsed SILAC-TMT dataset (n=383). The median of the distributions at different time-points (indicated by a red box and number) was utilized to calculate turnover rates from single time-point pulse data. (B) The distributions of turnover rates obtained from different pulse time-points showed global shifts. Only peptides that were shared among all time-point were included for the analysis (n=18,661).*

## 3.3 Characterization of peptidoforms with differential turnover

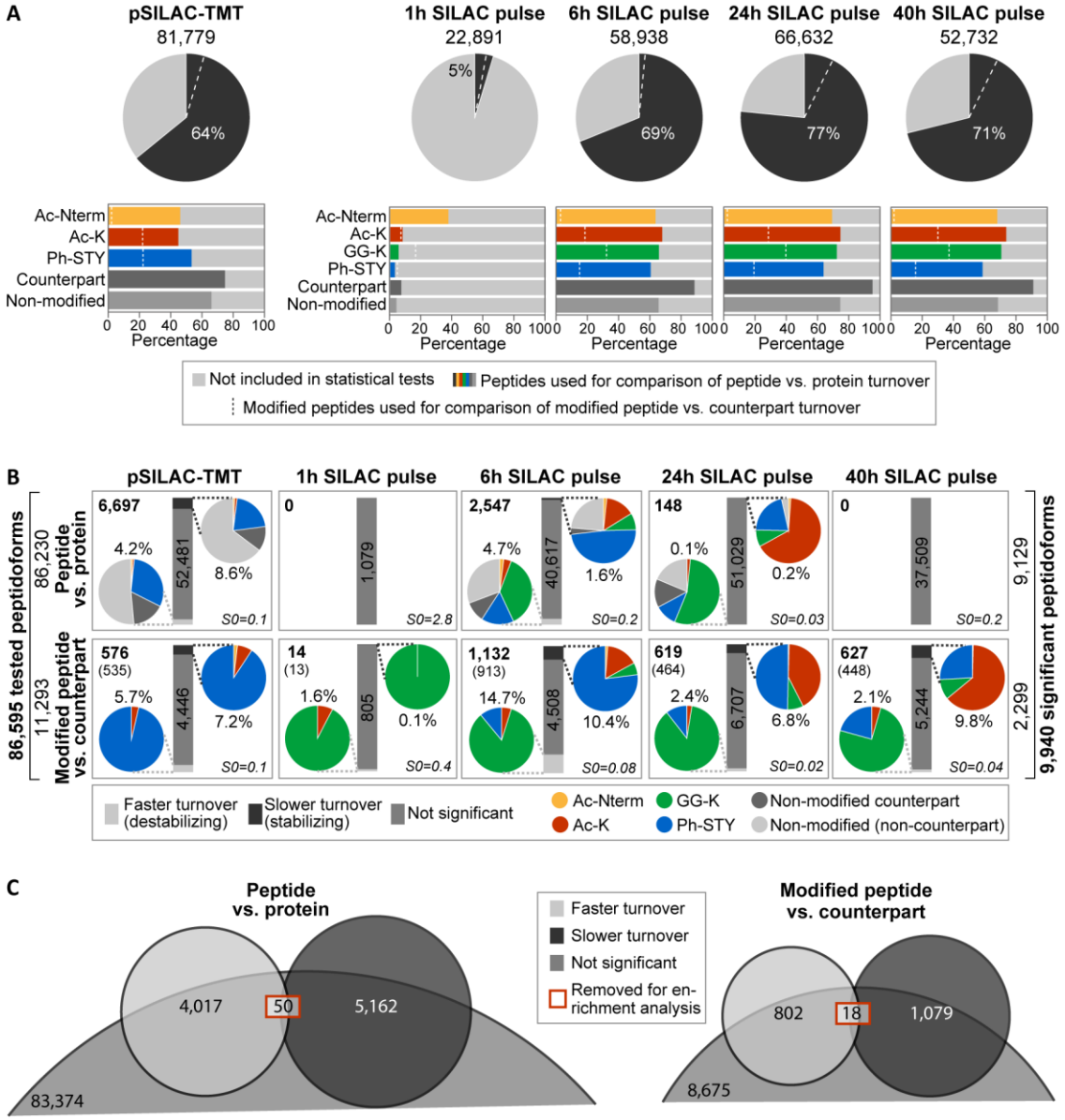



**Figure 0-18 | Peptides included in significance and enrichment analyses.** *(A) Pie diagrams illustrate the fraction of quantified peptides used to test statistical significance of turnover differences of (modified) peptides to their assigned proteins or counterpart peptides. Stacked bar charts show the same but discriminating between different modified peptides. Peptides not included for significance analyses did not fulfil criteria such as at least 2 quantifications per statistical group and at least 3 quantified peptides per protein. The total number of peptides for each dataset is indicated on top. (B) Significance of turnover differences between (modified) peptides and their proteins or counterpart peptides was tested for all datasets separately using two-sided Student's t-tests at 1 % permutation-based FDR (S0 shown in italics). Numbers of pairwise comparisons (vertical number), fractions of significant peptide hits (percentages), and their modification types (pie diagrams) are illustrated for each of the 10 statistical analyses. Note that for the modified peptide-counterpart comparison, modified peptides can be included in the analysis more than once if several non-modified counterpart peptides were quantified. The total number of significant hits for each test is indicated in bold font, and the numbers of significant, distinct modified peptides for the modified peptide-counterpart tests are displayed in brackets. The total numbers of tested and significant peptides are shown on the left and right side, respectively. (C) Venn diagrams display the number and overlap of peptides that were identified with significantly faster or slower turnover compared to their protein (left site) or counterpart peptide (right side) in all five datasets combined. Peptides that were found in conflicting groups (faster and slower, red boxes) in different datasets were discarded from enrichment analyses.*

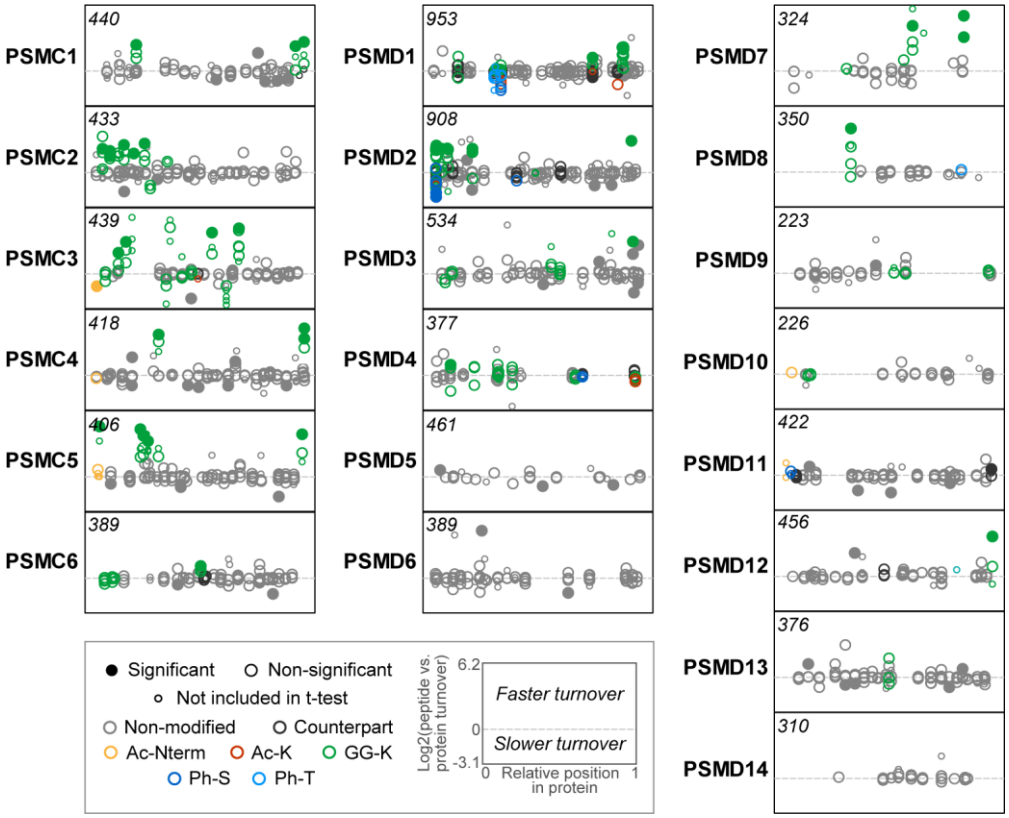



**Figure 0-19 | Peptides with significantly different turnover in subunits of the 19S proteasome forming the regulatory particle of the proteasome.** *The position of peptides within their corresponding 19S proteasome subunit is plotted against the turnover difference to their protein. Proteasomal proteins were enriched in both, significantly slower and faster degrading peptides. The length of the major isoform is displayed in italics.*

## 3.4 Functional perspective on modification-specific turnover

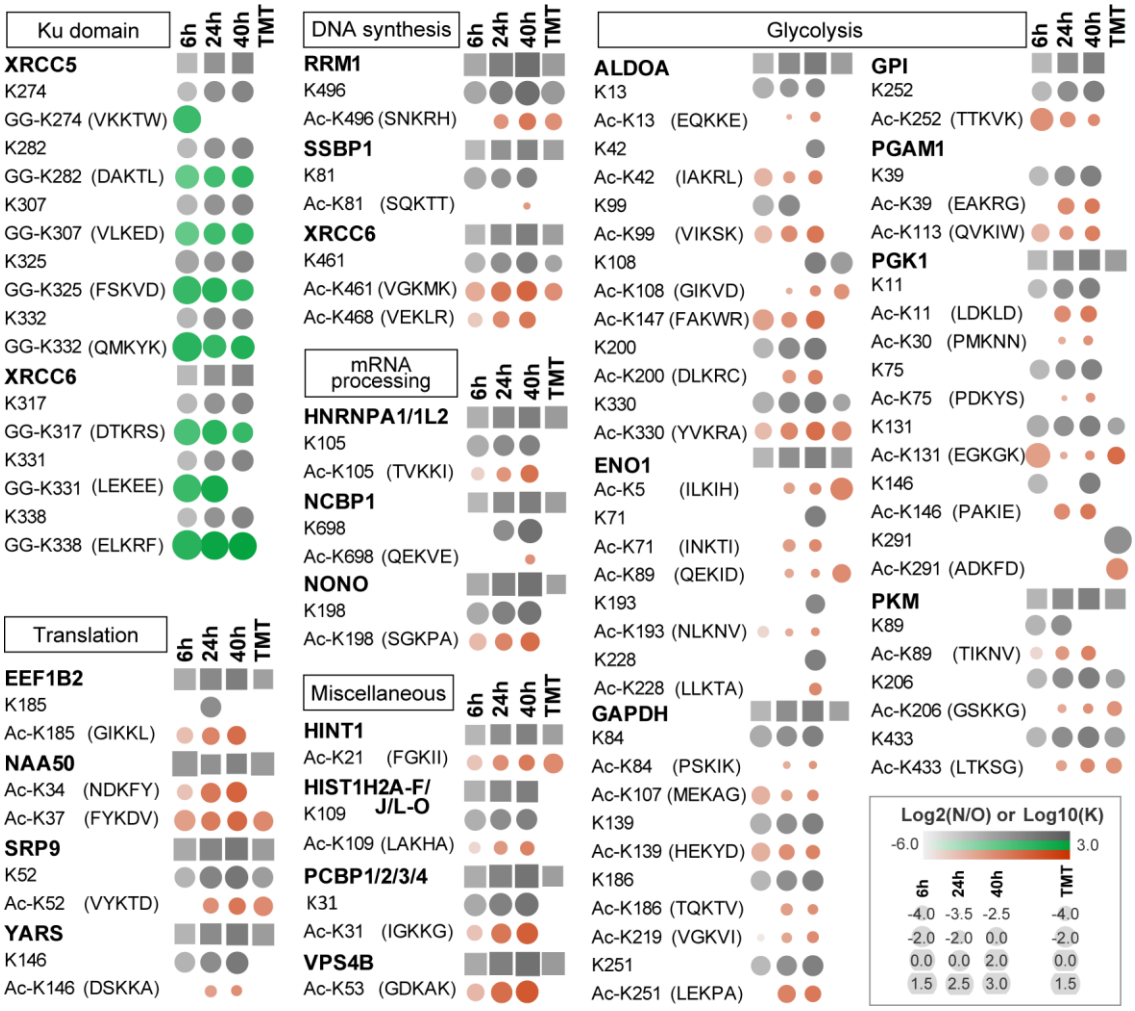



**Figure 0-20 | Ubiquitin-remnant and acetylation sites leading to significantly different turnover.** *Average turnover is illustrated for modification sites (circles) within Ku domains (GG-sites) and α-helices or β-strands (ac-sites), and their corresponding proteins (squares). All peptides mapping to a modified or non-modified site were included for calculation of average N/O ratios or turnover rates K. Amino acids surrounding the modification site are indicated in brackets.*

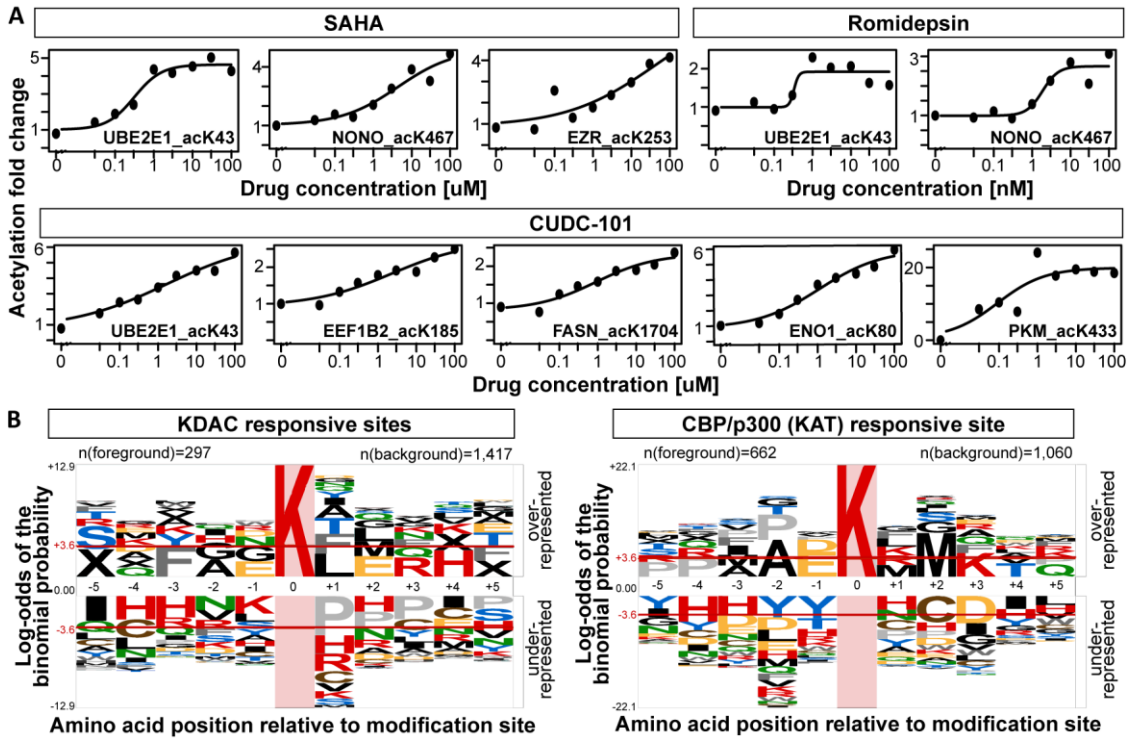



**Figure 0-21 | Enzymatic regulated acetyl-lysine sites.** *(A) Concentration-dependent increase of lysine acetylation upon HDAC inhibitor treatment of HeLa cells is displayed for ac-K sites that exhibited a decreased turnover (4 h for SAHA and CUDC-101, 16 h for Romidepsin). Samples were prepared equivalent to pSILAC-TMT experiments, but combining drug concentrations within a TMT-plex. (B) Probability logos for 11 amino acid long sequence windows of ac-K sites identified in the five turnover datasets were plotted using pLogo [322] and are illustrated for ac-K sites that showed an at least 2-fold increase upon KDAC inhibitor treatment (left panel), or more than a 2-fold regulation upon CBP/p300 inhibitor treatment or p300 transfection (right panel) in previously published data [208, 310] (red horizontal line: p=0.05).*

*“Glück ist das einzige, das sich verdoppelt, wenn man es teilt.”*

**- Albert Schweitzer**

*“If I have seen further, it is by standing on the shoulders of giants.”*

**- Isaac Newton**

# DANKSAGUNGEN

Geschafft! Endlich! Was war das doch für ein Reise... In den letzten Jahren habe ich viele Mitstreiter nach erfolgreichem Abschluss gehen, und noch viel mehr neue Kollegen kommen sehen, der gesamte Maschinenpark (inklusive der Kaffemaschine) wurde ausgetauscht und große, ambitionierte Projekte wurden neu ins Leben gerufen. Die Jahre sind viel schneller vergangen, als ich mir hätte vorstellen können, doch sie haben mich viel gelehrt. Ich habe Einiges erfahren über die Proteomik und Massenspektrometrie (q.e.d.), aber noch viel mehr Erkenntnisse über mich selbst gewonnen. Viele Menschen haben mich auf diesem Weg begleitet (in erster Linie alle KollegenInnen am Lehrstuhl) und waren Hilfe, Beistand und Ansporn zugleich. Ich kann hier nur einige wenige namentlich erwähnen.

Bernhard, mein Doktorvater, Danke dir für die Möglichkeiten, die du mir eröffnet hast, dein Vertrauen in meine Fähigkeiten, die wissenschaftlichen Freiheiten, die du mir eingeräumt hast, und deine Geduld, wenn es mal wieder länger gedauert hat. Danke, dass du mir das Gefühl gegeben hast, dass ich sagen durfte, was ich eben gedacht habe. Ich möchte dir auch ausdrücklich danken, dass du immer in erster Linie ‚Mensch' und nicht ‚Chef' warst. Damit machst du einen Unterschied. Ich rechne dir das alles sehr hoch an.

Daniel, den Anfang des Weges haben wir gemeinsam beschritten, dann hast du mich „abgehängt" (recht so!). Ohne dich wäre es nicht dasselbe gewesen. Du hast mit mir gelacht und mit mir geschimpft. Aber vor allem hast mich ertragen, als ich unerträglich war. Das werde ich dir nie vergessen. Danke!

Steffi, meine liebe Steffi, du warst in meinem Herzen von dem Moment an, als du mir einen Motivationsdrücker unmittelbar vor meiner Masterprüfung gegeben hast. Viele weitere sind gefolgt. Du hast meine PhD Zeit auf allen Ebenen so ungemein bereichert. Was hätte ich nur getan ohne dein ansteckendes Lachen! Danke, für dein immer offenes Ohr und deine unerschöpfliche Fähigeit, mich aufzumuntern.

Danke, Mathias, für dein wissenschaftliches Ohr und dein privates Ohr. Julia R., Lin, Fynn, Anton und Flo, ich danke euch für eure Zeit und Nerven, die ihr in eure Masterarbeiten und Forschungspraktika bei mir investiert habt. Es war lehrreich (auf so vielen Ebenen!) und eine ungemeine Freude mit euch zu arbeiten. Den Bond-Girls danke ich für all die Gespräche, ob wissenschaflich oder nicht, und die immer willkommene Versorgung mit Schokolade, Weihnachtsplätzchen und Obst. Mein ausdrücklicher Dank gilt auch Julia M. und Daniel dafür, dass sie sich die Zeit genommen haben, Teile meiner Arbeit Korrektur zu lesen.

Michaela, Antonio und Maria H., ich danke euch, dass ihr mir einen Rückzugsort geboten habt. Zeit mit euch zu verbringen, war wie ein Resetknopf zu drücken. Einfach nur befreiend!

Dan, Gloria und Jonas, euch danke ich dafür, dass ich auch Teil euer kleinen Familie bin. Durch eure Anwesenheit, ja durch eure bloße Existenz werde ich immer wieder daran erinnernt, was wirklich zählt. Bruderherz, ich kann hier nicht ausdrücken, wie wichtig du bist für mich und wie wichtig du auch in kritischen Phasen meines PhD warst. Das brauche ich auch nicht. Du weißt es.

Papa, Danke dir für die ‚Werkzeuge', die du mir gemeinsam mit Mama an die Hand gegeben hast. Du hast mir keine Vorschriften für mein Leben gemacht, mich meine ganz eigenen Entscheidungen treffen lassen. Du wolltest mir immer alles ermöglichen, dabei war Selbstständigkeit eines deiner obersten Erziehungsziele. Es hat sich ausgezahlt.

Ich danke auch Prof. Dr. Florian Bassermann und Prof. Dr. Matthias Selbach für die Bereitschaft, meine Arbeit zu begutachten, und Prof. Dr. Martin Klingenspor für die Übernahme des Prüfungsvorsitz.

Was bleibt noch zu sagen? Danke. Thank you. Merci. Xiè xiè nǐ. Grazie. Gracias. Spasibo. Efcharistó. Dziękuję. Dank u wel. Shnorhakalut'yun. An alle.

# LIST OF PUBLICATIONS

**Main publications of this thesis:**

[1] Zecha J, Satpathy S, Kanashova T, Avanessian SC, Kane MH, Clauser KR, Mertins P, Carr SA & Kuster B. **TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach.** *Molecular & Cellular Proteomics* 2019, **18**(7):1468-78. DOI: 10.1074/mcp.TIR119.001385

[2] Zecha J, Meng C, Zolg DP, Samaras P, Wilhelm M & Kuster B. **Peptide Level Turnover Measurements Enable the Study of Proteoform Dynamics.** *Molecular & Cellular Proteomics* 2018, **17**(5):974-92. DOI: 10.1074/mcp.RA118.000583

**Additional publications during PhD**: **equal contribution*

[3] Bian Y, Zheng R, Bayer FP, Wong C, Chang Y-C, Meng C, Zolg DP, Reinecke M, Zecha J, *et al.,* & Kuster B. **Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC–MS/MS.** *Nature Communications* 2020, **11**(1):1-12. DOI: 10.1038/s41467-019-13973-x

[4] Samaras P, Schmidt TK, Frejno M, Gessulat S, Reinecke M, Jarzab A, Zecha J, et al., Kuster B & Wilhelm M. **ProteomicsDB: A multi-omics and multi-organism resource for life science research.** *Nucleic Acid Research* 2020, **48**(D1):D1153-63. DOI: 10.1093/nar/gkz974

[5] Khakipoor S*, Giannaki M*, Theparambil SM, Zecha J, Kuster B, Heermann S, Deitmar JW & Roussa E. **Functional expression of electrogenic sodium bicarbonate cotransporter 1 (NBCe1) in mouse cortical astrocytes is dependent on S255-257 and regulated by mTOR.** *Glia* 2019, **67**(12):2264-78. DOI: 10.1002/glia.23682

[6] Wang, D*, Eraslan B*, Wieland T, Hallström BM, Hopf T, Zolg DP, Zecha J, et al. & Kuster B. **A deep proteome and transcriptome abundance atlas of 29 healthy human tissues.** *Molecular Systems Biology* 2019, **15**(2):e8503. DOI: 10.15252/msb.20188503

[7] Klaeger*, Heinzlmeir S*, Wilhelm M*, Polzer H, Vick B, Koenig PA, Reinecke M, Ruprecht B, Petzoldt S, Meng C, Zecha J, et.al. & Kuster B. **The target landscape of clinical kinase drugs.** *Science* 2017, **358**(6367):eaan4368. DOI: 10.1126/science.aan4368

[8] Ruprecht B, Zaal EA, Zecha J, Wu W, Berkers CR, Kuster B & Lemeer S. **Lapatinib resistance in breast cancer cells is accompanied by phosphorylation-mediated reprogramming of glycolysis.** *Cancer Research* 2017, **77**(8):1842-53. DOI: 10.1158/0008-5472.CAN-16-2976

[9] Ruprecht B, Zecha J, Zolg DP & Kuster B. **High pH reversed-phase micro-columns for simple, sensitive, and efficient fractionation of proteome and (TMT labeled) phosphoproteome digests**. in *Proteomics: Methods and Protocols* 2017, L. Comai, J.E. Katz, and P. Mallick, Editors. Springer New York, NY. 83-98, 2017. DOI: 10.1007/978-1-4939-6747-6_8

[10] Ruprecht B, Zecha J, Heinzlmeir S, Medard G, Lemeer S & Kuster B. **Evaluation of kinase activity profiling using chemical proteomics.** *ACS Chemical Biology* 2015, **10**(12):2743-52. DOI: 10.1021/acschembio.5b00616

**Manuscripts submitted:**

[11] Frejno M*, Meng C*, Ruprecht B*, Oellerich T, Scheich S, Kleigrewe K, Drecoll E, Samaras P, Hogrebe A, Helm D, Mergner J, Zecha J, *et al.*, & Kuster B. **ATLANTiC: Activity landscapes of tumor cell lines determine drug responses**.

[12] Jarzab A*, Kurzawa N*, Hopf T*, Moerch M, Zecha J, *et al.*, Hahne H, Savitski MM & Kuster B. **Meltome atlas – thermal proteome stability across the tree of life.**

*„Wer all seine Ziele erreicht, hat sie zu niedrig gewählt.“*

**- Herbert von Karajan**