University of Groningen

Chemical approaches for quantitative proteomics

Tian, Xiaobo

DOI:
[10.33612/diss.179003839](https://doi.org/10.33612/diss.179003839)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Tian, X. (2021). *Chemical approaches for quantitative proteomics*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.179003839>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chemical approaches for quantitative proteomics

Xiaobo Tian

2021

Chemical approaches for quantitative proteomics

PhD thesis

The research presented in this thesis was performed in the department of Analytical Biochemistry at the University of Groningen (RUG) and financially supported by the *China Scholarship Council* (CSC) with Grant Number 201706820013 and the Open Technology Programme of Toegepaste en Technische Wetenschappen (TTW) with project number 15230, which is financed by the Netherlands Organisation for Scientific Research (NWO).

Cover design: Xiaobo Tian

Layout: Xiaobo Tian

Printed by:

Copyright © Xiaobo Tian, Groningen, the Netherlands, 2021

All rights reserved. No part of this thesis may be reproduced in any form or by any means without prior permission of the author.

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof Dr. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Friday 24 September 2021 at 14:30 hours

by

Xiaobo Tian
born on 03 January 1991
in Sichuan, China

Supervisor

Prof. R.P.H. Bischoff

Co-supervisor

Dr. H. Permentier

Assessment Committee

Prof. F.J. Dekker

Prof. M. Wuhler

Prof. B. Thiede

*Dedicated to my beloved
wife
and all the members of my family*

Table of Contents	
Abbreviations	1
Chapter 1	3
Introduction to chemical isotope labeling-based quantitative proteomics and outline of the thesis	3
ABSTRACT	4
1 Introduction	4
2 MS1-based quantification	10
3 MS2-based quantification	19
4 Combined MS1 and MS2-based quantification	36
5 Isotope labeling-based quantification in DIA	37
6 Scope and outline of the thesis	38
References	42
Chapter 2	55
Selective maleylation-directed isobaric peptide termini labeling for accurate proteome quantification	55
ABSTRACT	56
1 Introduction	56
2 Experiments	59
3 Results and discussion	61
4 Conclusions	70
Acknowledgements	71
References	72
Supporting Information	76
Chapter 3	89
A collision-induced dissociation cleavable isobaric tag for peptide fragment-ion-based quantification in proteomics	89
ABSTRACT	90
1 Introduction	90
2 Experiments	92
3 Results and discussion	98
4 Conclusions	105
Acknowledgements	105
References	106
Supporting Information	109
Chapter 4	123
A versatile isobaric tag enables proteome quantification in data dependent and data independent acquisition mode	123
ABSTRACT	124
1 Introduction	124
2 Experiments	128
3 Results and discussion	130
4 Conclusions	141
Acknowledgements	141

References.....	143
Supporting Information.....	146
Chapter 5	161
The isotopic Ac-IP tag enables multiplexed proteome quantification in data-independent acquisition mode	161
ABSTRACT.....	162
1 Introduction.....	162
2 Experiments	164
3 Results and discussion.....	167
4 Conclusions.....	176
Acknowledgements.....	176
References.....	177
Supporting Information.....	180
Perspective and summary	199
Perspectief en samenvatting	205
Acknowledgments	211
List of publications	217

Abbreviations

Ac-AG	acetyl-alanine-glycine
Ac-IP	acetyl-isoleucine-proline
Ac-IPG	acetyl-isoleucine-proline-glycine
cICAT	cleavable isotope-coded affinity tag
CILAT	cleavable isobaric labeled affinity tag
CMTs	combinatorial isobaric mass tags
DDA	data-dependent acquisition
DIA	data-independent acquisition
DiART	deuterium isobaric amine reactive tag
DiLeu	N,N-dimethyl leucine
FAIMS	high-field asymmetric waveform ion mobility spectrometry
FT-ICR	Fourier transform ion cyclotron resonance
IBT	isobaric tags
ICAT	isotope-coded affinity tag
iTRAQ	isobaric tags for relative and absolute quantitation
IPTL	isobaric peptide termini labeling
LC-MS	liquid chromatography–mass spectrometry
MDHI	2-methoxy-4,5-dihydro-1H-imidazole
NCE	normalized collision energy
NHS	N-hydroxysuccinimide
RP-HPLC	reversed-phase high-performance liquid chromatography
SA	succinic anhydride
SILAC	stable isotope labeling by amino acids in cell culture
SMD-IPTL	selective maleylation-directed isobaric peptide termini labeling
SWATH-MS	sequential windowed acquisition of all theoretical fragment ion mass spectra
S/N	signal-to-noise
TMT	tandem mass tags

Chapter 1

Introduction to chemical isotope labeling-based quantitative proteomics and outline of the thesis

This chapter has been published in *Mass Spec. Rev.* 2021, 1-31

ABSTRACT

Advancements in liquid chromatography and mass spectrometry over the last decades have led to a significant development in mass spectrometry-based proteome quantification approaches. An increasingly attractive strategy is multiplex isotope labeling, which significantly improves the accuracy, precision and throughput of quantitative proteomics in the data-dependent acquisition (DDA) mode. Isotope labeling-based approaches can be classified into MS1-based and MS2-based quantification. At the beginning of this thesis, I give an overview of approaches based on chemical isotope labeling and discuss their principles, benefits and limitations with the goal to give insights into fundamental questions and to pave the way for the understanding of the following works. Besides, I discuss the current possibilities and limitations of multiplex isotope labeling approaches for the data-independent acquisition (DIA) mode, which is one of our main research branches.

1 Introduction

Proteomics aims at the comprehensive identification and quantification of all proteins in a biological system, to reveal the roles of proteins in biological, physiological and pathological processes. Proteomics data quality has been continuously improving in terms of accuracy and precision while throughput has increased leading to an ever-increasing flow of data. This is due to improvements of speed, sensitivity and resolution in liquid chromatography–mass spectrometry (LC-MS) as well as advances in data processing due to an increase in computational power and advanced algorithms. Proteomics is widely applied in many research areas, such as the exploration of biological mechanisms^{1,2}, the discovery of biomarkers³⁻⁵ and in drug design⁶. The most widely used strategy in LC-MS-based proteomics is bottom-up proteomics⁷, in which proteins are first proteolytically digested into peptides, which are usually separated by reversed-phase high-performance liquid chromatography (RP-HPLC). The masses of the peptides are determined at the MS1 level while masses of the corresponding fragment ions are obtained at the MS2 level after fragmentation in a collision cell. By

matching the precursor ions measured in MS1 and the fragment ions in MS2 with the theoretical masses generated from *in silico* digestion of the corresponding protein sequence, the peptides in the sample can be identified and this information is used to assign peptides to the corresponding proteins to arrive at their identification and quantification. A less widely used strategy is top-down proteomics, which stands for the direct analysis of intact proteins by mass spectrometry. For details about this approach, I refer readers to reviews by Kelleher and coworkers^{8,9}.

As proteomics research progressed, it has become increasingly clear that quantitative information is critical to relate proteomics data to actionable outcomes like the design of drugs or the development of biomarker-based diagnostic assays. Accurate quantification of the dynamically changing protein composition is the basis for understanding the functioning of biological systems. Based on whether isotope labeling is used or not, existing mass spectrometry-based proteome-wide quantitative methods can be classified into (I) label-free proteomics¹⁰⁻¹² and (II) label-based proteomic¹³⁻¹⁵, also named multiplexed quantitative proteomics. While the introduction part focuses on label-based proteomics and on chemical labeling in particular, I give a brief introduction of label-free proteomics in comparison.

1.1 Label-free proteomics

Representative features of label-free proteomics are discussed below and readers are referred to comprehensive reviews for further details¹⁶⁻¹⁸. In view of the fact that the label-free approach does not require derivatization with isotope labels and can be implemented on any type of mass spectrometer, it is widely used with good results due to improvements in instrument reproducibility and advanced data acquisition and data processing schemes¹⁹⁻²². However, label-free proteomics has inherent limitations in throughput, as it only analyzes one sample per LC-MS run, and in precision, as it does not correct for analytical variability. Relative quantitation is achieved, for example, by comparing the peak area of precursor ions at the MS1 level or by counting the number of MS2 spectra per peptide²³. It is therefore prone to be affected by retention time shifts,

changes of ionization efficiency and variations in sample loss during work-up from injection to injection. To avoid or alleviate these limitations, simultaneous analysis of multiple samples is a good remedy but it requires approaches that allow to discriminate between differentially labeled peptides containing the same sequence in the mass analyzer, while minimally affecting their physicochemical properties to minimize shifts in retention times or variable recoveries during sample work-up. Hence, stable isotope labeling is an ideal choice as isotopes have almost identical physicochemical properties but distinct masses.

1.2 Label-based proteomics

The technique of multiplexing in proteomics was first introduced in 1999, with the isotope-coded affinity tag (ICAT)^{13,24,25} and the ¹⁵N metabolic labeling method²⁶. After that, a series of other multiplexed quantification approaches have been published, including stable isotope labeling by amino acids in cell culture (SILAC)²⁷⁻²⁹, isobaric tags for relative and absolute quantitation (iTRAQ)^{14,30}, tandem mass tags (TMT)^{15,31} and isobaric peptide termini labeling (IPTL)³²⁻³⁷. All these approaches exploit different isotope labels to derivatize peptides or proteins that are derived from different samples. Most approaches involve derivatization of proteolytic peptides rather than intact proteins, because every proteolytic peptide contains at least one suitable labeling site, namely a primary amine group, thus permitting the labeling of all peptides in a sample using the same strategy. Since the differentially labeled peptides can be resolved by mass spectrometry and the labeled samples are pooled prior to LC-MS³⁸⁻⁴⁰, simultaneous analysis of multiple samples can be achieved, which circumvents signal variation due to, among others, variable sample loss during work-up and changing ionization efficiency between injections^{18,41,42}. Multiplexed approaches are expected to afford better quantification precision and have a higher throughput thus reducing the overall analysis time per sample. The commonly used heavy isotopes are ¹³C, ¹⁵N, ¹⁸O and ²H, with ²H being less popular due to the potential risk of varying LC retention times of the same peptide modified with a different number of deuterium atoms⁴³⁻⁴⁶,

which is caused by the differential interaction between hydrogen/deuterium and the reversed-phase stationary phase, since deuterium has a larger mass and smaller amplitude of vibrations than hydrogen^{47,48}. Hence, in spite of the fact that deuterium-based labels can extend the multiplexing capacity further, the majority of chemical labeling approaches utilizes ¹³C and ¹⁵N isotopes.

Figure 1. Schematic overview of MS1-based and MS2-based quantification.

A large number of multiplexed methods have been reported, which can be classified into sub-categories according to different criteria, as shown in **Figure 1**: (I) based on the stage at which peptides are quantified, they can be categorized into MS1-based quantification (e.g. ICAT^{13,24,25} and mTRAQ^{49,50}) and MS2-based quantification (e.g. TMT and iTRAQ); (II) based on whether the same peptides derived from different samples have the same mass after isotope labeling or not, they can be divided into isobaric labeling-based quantification (TMT, IPTL³² and EASI tag⁵¹), or isotopic labeling-based quantification (ICAT and m-pIDL⁵²); (III) based on the approach of incorporating the isotopes, they can be classified into chemical labeling-based quantification (e.g. ICAT and TMT), enzymatic labeling-based quantification (¹⁸O labeling^{53,54}) or metabolic labeling-based quantification (SILAC²⁷⁻²⁹); (IV) based on whether ultra-high resolution is required or not to distinguish mDa mass difference at the MS1 or MS2 level, they can be sub-categorized into tags not requiring ultra-high resolution (e.g. ICAT and 4-plex iTRAQ), tags requiring ultra-high resolution at the

MS1 level (DiPyrO and mdDiLeu) and tags requiring ultra-high resolution at the MS2 level (11-plex TMT and 21-plex DiLeu). It is notable that these classifications are based on different criteria (**Table 1**), which are not mutually exclusive. Although some tags, like 4-plex iTRAQ and DiLeu, do not have strictly the same precursor masses, they are classified as isobaric, since the mDa differences of precursor masses are not used at the MS1 level to discriminate between different labeling channels. Throughout this thesis, I will focus on chemical labeling-based quantitative proteomics approaches, in which proteolytic peptides are chemically labeled with distinct isotope labels and labeled samples are then pooled and analyzed in a single LC-MS run.

Table 1. Classification of chemical isotope labeling approaches based on different criteria.

	MS1 quantification	MS2 quantification	Isotopic labeling	Isobaric labeling	Ultra-high resolution at MS1	Ultra-high resolution at MS2
(c)ICAT	√		√			
Dimethylation	√		√			
ICPL	√		√			
mTRAQ	√		√			
mTMT	√		√			
DiPyrO	√		√		√	
mdDiLeu	√		√		√	
iTRAQ ^a		√		√		
TMT		√		√		√
DiLeu ^a		√		√		√
IBT		√		√		√
DiART		√		√		
IPTL		√		√		√
m-pIDL		√	√			√
Ac-IPG		√		√		
TMTc(+)		√		√		
EASI		√		√		
Ac-AG		√		√		

^a Although the iTRAQ and DiLeu series labels have mDa differences in precursor masses, they are classified as isobaric since the mDa differences are not revealed at the MS1 level in general use and they are already habitually referred to as isobaric.

1.3 Data acquisition strategies

Before discussing specific labeling strategies, it is necessary to introduce the main features of the two main MS2 spectra acquisition modes, namely data-dependent acquisition (DDA) and data-independent acquisition (DIA), as they are closely linked

to the advantages and disadvantages of multiplexed quantification approaches. In most multiplexed methods, collecting MS2 spectra is performed in DDA mode, in which a certain number of precursor ions is successively, but individually, selected for fragmentation during an LC-MS run in order of decreasing intensity. The resulting fragments are collected in discrete MS2 spectra thus providing a direct link between precursor ion and its fragments⁵⁵. DDA therefore has a bias toward high-abundance species and limits the detection of low-abundance peptides and ultimately proteins (**Figure 2**). The number of peaks that can be isolated for fragmentation increases with the scanning rate of the mass spectrometer, but there are usually more peaks in the MS1 spectra than the instrument can isolate in the timeframe of the analysis, usually the width of an LC peak, resulting in minor peaks being ignored. This problem increases with increasing complexity of the samples and is further related to the resolution of the chromatographic separation leading to the fact that highly complex samples must often be fractionated prior to the final LC-MS analysis, which increases the analysis time per sample. In addition, the stochastic nature of precursor ion selection in the DDA mode diminishes the run-to-run reproducibility often leading to the so-called ‘missing value problem’, which means that a given peptide may not be detected (fragmented) in every LC-MS run^{56,57}. This creates a problem for comparative data analysis. Various methods have been developed to arrive at a more comprehensive sampling of precursor ions in DDA^{58,59}. Dynamic exclusion is a commonly used strategy to prevent a peptide from triggering multiple MS2 events during the timeframe of an eluting chromatographic peak, as this dedicates time to a potentially already identified peptide while others escape analysis. Despite these limitations, DDA is still the most widely used data acquisition strategy in current discovery proteomics, because of well-developed data acquisition and data processing workflows that can be implemented on almost any MS instrument.

Figure 2. Difference in MS1 isolation windows for the DDA and DIA modes.

In contrast to DDA, isolation of precursors for fragmentation in DIA does not rely on peak intensities in MS1. Instead, all precursors present within a predefined isolation window will be fragmented and analyzed together, as shown in **Figure 2**. DIA largely addresses the problem of missing values⁶⁰, however, at the price that the direct link between precursor and fragment ions is lost. MS2 spectra of DIA are inherently more complex, since fragment ions derived from multiple, co-isolated precursor ions are present in the same MS2 spectrum, which makes analyzing DIA data more challenging than DDA data. Since the Sequential Windowed Acquisition of all Theoretical fragment ion Mass Spectra (SWATH-MS)⁶⁰ was proposed, DIA has made remarkable advances in recent years with the development of sophisticated data processing algorithms⁶⁰⁻⁶³. It is also notable that both the MS1 and MS2 data can be used for quantitation in DIA mode^{63,64}. While the application of stable isotope labeling is widely used in DDA mode, it is still in its infancy when it comes to DIA. At the end of this introduction, I will therefore highlight current developments, advantages and challenges of applying stable isotope labeling in DIA mode.

2 MS1-based quantification

MS1-based quantitative approaches are based on introducing distinct mass additions to peptides by isotopic labeling, so that the same peptides derived from different samples

have different masses. Quantification is achieved by comparing the peak areas or intensities of peptide ions for each labeling channel at the MS1 level^{13,65,66}. Subsequently, the quantitative information at the peptide level needs to be transferred to the protein level⁶⁷, which is a common challenge for DDA and DIA. Usually, a mass shift of at least 4 Da is incorporated into the isotopic tags to avoid overlap between the isotope envelopes of light- or heavy-labeled samples⁶⁸. The overlap resulting from smaller mass shifts (< 4 Da) renders relative quantification between differentially labeled peptide pairs/clusters more difficult since an additional deconvolution step is required. This is particularly challenging for deuterium containing labels, which induce a slight shift in retention time⁶⁹.

The distinct isotopes for MS1-based quantification can be incorporated into samples by three approaches: 1) chemical labeling (e.g. ICAT¹³ and mTMT⁷⁰) with synthetic tags targeted at specific reactive groups (amine or thiol); 2) enzymatic labeling^{53,54}, where proteins are enzymatically digested in the presence of ¹⁸O labeled water and two ¹⁸O atoms are incorporated into the newly formed C-terminal carboxyl groups of digested peptides^{71,72}; 3) metabolic labeling, such as SILAC²⁷⁻²⁹, where distinct isotope-labeled amino acids (e.g. Lys and Arg) are fed to cell cultures and metabolically incorporated into newly synthesized proteins. In the following discussion, I will focus on chemical labeling approaches. For in-depth discussions about enzymatic labeling-based and metabolic labeling-based quantification, I refer readers to other comprehensive reviews^{40,73}.

2.1 ICAT and cICAT tags

The first reported chemical isotope labeling tag is the Isotope-Coded Affinity Tag^{13,74,75} (ICAT) that consists of three functional parts, as shown in **Figure 3**: (1) a biotin moiety, which can be used to enrich the ICAT-labeled peptides from a digest mixture via (strept) avidin affinity chromatography; (2) an isotopic linker, which contains either eight ²H or ¹H atoms to differentially label peptides from different samples; (3) an iodoacetamide group, which specifically reacts with the sulfhydryl group at the side chain of cysteine

residues. Compared to label-free methods, the main improvement of ICAT was the ability to simultaneously quantify two samples in a single LC-MS run based on a subset of peptides containing at least one cysteine, which reduces sample complexity while preserving the identification coverage, since most proteins contain at least one cysteine residue despite the fact that cysteine is a rather rare amino acid⁷⁶. Although the merits of ICAT were clear, several limitations quickly came to light⁷⁷: First, the employed ^2H atoms in the linker alter the retention time in comparison to the ^1H -containing peptide. Second, fragmentation occurs at the biotin moiety, adversely impacting peptide identification due to unassignable fragment ions. Third, due the large mass of ICAT, labeled peptides may shift out of the optimum m/z range for detection, especially when more than one tag is incorporated in a peptide. In spite of these limitations, ICAT is widely considered as the pioneer of isotope labeling-based quantitative proteomics based on chemical labeling. Subsequently, a cleavable isotope-coded affinity tag (cICAT) was proposed⁷⁷⁻⁸¹, that allows removal of the biotin moiety prior to LC-MS at acidic pH thus reducing problems with peptide fragmentation. In addition, replacing ^2H with ^{13}C in the linker ensures the coelution of differentially labeled peptides during reversed-phase chromatography. However, the specificity of cICAT for cysteine remains a double-edged sword. It reduces the complexity of the proteome sample to allow a more detailed analysis of the subset of peptides containing cysteine, while missing all information about peptides containing no cysteine.

Figure 3. Structure design and isotope distribution of 2-plex ICAT.

2.2 Dimethylation

Dimethylation, a simple and cost-effective chemical labeling strategy, was first proposed in 2003⁶⁵. It is based on reductive amination chemistry, where the N-terminal amine group of a peptide/protein and the amine group at the side chain of Lys residues are reacted with formaldehyde, which is available as $^{12}\text{C}^1\text{H}_2\text{O}$, $^{13}\text{C}^1\text{H}_2\text{O}$, $^{12}\text{C}^2\text{H}_2\text{O}$ and $^{13}\text{C}^2\text{H}_2\text{O}$, followed by reduction with sodium cyanoborohydride, which is available as $\text{NaB}^1\text{H}_3\text{CN}$ and $\text{NaB}^2\text{H}_3\text{CN}$. As shown in **Figure 4**, with combinations of different isotopic forms, dimethylation can achieve duplex or triplex labeling. By applying $^{12}\text{C}^1\text{H}_2\text{O}$ with $\text{NaB}^1\text{H}_3\text{CN}$, $^{12}\text{C}^2\text{H}_2\text{O}$ with $\text{NaB}^1\text{H}_3\text{CN}$ and $^{13}\text{C}^2\text{H}_2\text{O}$ with $\text{NaB}^2\text{H}_3\text{CN}$, triplex labeling of “light, intermediate and heavy” can be achieved with a mass shift of 4 Da per derivatized site^{68,82}. Instead of tagging at a rare amino acid, like Cys in ICAT, dimethylation achieves isotopic labeling at the amine group which exists in almost all proteolytic peptides enabling identification and quantification of proteins based on multiple peptides at the cost of a more complex peptide mixture. Dimethylation increases peptide ionization efficiency, as the resulting tertiary amine is easier ionized in electrospray, and improves completeness of the b-ion series⁸³. Another interesting observation is that peptides that are differentially dimethylated with different numbers of deuterium atoms show a smaller retention time shift compared to that observed in ICAT⁷⁷. Presumably, deuterium atoms on a hydrophilic trimethylamine group have less interactions with the reversed-phase stationary phase than those in the ethyleneglycol linkers of ICAT tags^{44,65,68}. Selective N-terminal dimethylation is an adaptation of dimethylation⁸⁴, that exploits the pK_a difference between the N-terminal amine group and the amine group at the side chain of Lys residues. Diethylation, using the same labeling principle as dimethylation, but replacing formaldehyde with acetaldehyde thus offering a higher multiplexing capacity, was recently reported^{69,85}. By comparison with the deuterium-based triplex-dimethylation, the ^{13}C -based triplex-diethylation method exhibited better quantitative accuracy and precision⁸⁵, mainly because of the absence of the deuterium effect on retention time.

Figure 4. Triplex dimethylation of a peptide with a C-terminal Lys residue. ^{13}C in formaldehyde is marked with an asterisk.

2.3 ICPL, mTRAQ and mTMT tags

The Isotope-Coded Protein Label (ICPL) was first introduced in 2005⁶⁶. The tag contains a pyridine ring that ionizes well and an N-hydroxysuccinimide (NHS) ester to permit the efficient labeling of N-terminal alpha-amine groups and the epsilon-amine group of Lys^{86,87}. In the original version of the ICPL method (**Figure 5**), the ^1H atoms on the pyridine ring of the “light tag (+0)” were substituted with ^2H to generate the “heavy tag (+4)”, allowing duplex experiments. To avoid retention time shifts due to the deuterium effect, another version of ICPL with ^{13}C was developed⁸⁷. In the original ICPL work-flow, samples were labeled at the protein level, avoiding ratio deviations caused by different digestion efficiency and sample loss during the digestion process. However, this strategy is limited by the fact that only 60–70% of identified proteins can subsequently be quantified^{40,88}, presumably because not all identified peptides contain the ICPL tag, as not all primary amines are equally accessible in intact proteins. It is by far most common to conduct labeling after digestion, since every proteolytic peptide contains at least one amine group at the N-terminus. As a result, every peptide conveys quantitative information, which significantly increased the proportion of quantified proteins to 98% for ICPL⁸⁸.

The mTRAQ label was first used in discovery proteomics for global relative quantification in 2010⁴⁹. As shown in **Figure 5**, mTRAQ has a tertiary amine in the structure for efficient ionization and is available in three isotope forms (+0, +4 and +8)

resulting in mass shifts of 140, 144, and 148 Da per derivatized site, respectively. The mTMT tag, proposed in 2019⁷⁰, has the same chemical structure as the isobaric TMT tag that will be discussed later. In the original report, the TMTzero (TMT0) and ‘super heavy’ TMT (shTMT) tags were used to differentially label two samples with a mass difference of 11 Da (**Figure 5**). Both mTRAQ and mTMT have an amine reactive NHS ester, which enables the efficient labeling of primary amine groups in peptides. Both tags use only ^{15}N and ^{13}C atoms in the heavy labels.

Figure 5. The structure and isotope distribution of 2-plex ICPL, 3-plex mTRAQ and 2-plex mTMT tags.

2.4 Neutron-encoded isotopic labeling

Neutron encoding is used for quantifying peptides based on the mDa mass difference obtained by incorporating either ^2H , ^{13}C or ^{15}N isotopes (see in **Figure 6**), which extends the multiplexing possibilities of both chemical and metabolic labeling strategies in MS1- and MS2-based quantitative proteomics^{89,90}. MS1-based

quantification approaches generally use a mass spacing of at least 4 Da between

Figure 6. The concept of quantification based on neutron encoding. ^{13}C and ^{15}N are marked with an asterisk. The 6.4 mDa mass difference can be differentiated at resolutions higher than 240k in the m/z 100-1000 range. The peak intensities represent relative quantification.

labeling channels to avoid overlap between the isotope envelopes of differentially labeled peptides. However, this spacing restricts the number of isotopes that can be added and thus constrains the multiplexing capacity. Inspired by increasing the multiplexing capacity with the mDa difference obtained by neutron encoding in the isobaric label 8-plex TMT (see section III.A.1)^{90,91}, Hebert et al. first introduced neutron encoding as an MS1-based quantification approach in 2013 for metabolic labeling with amino acids in cell culture⁸⁹. In that report, they proposed NeuCode SILAC that achieved multiplexed labeling by using two kinds of Lys, one containing six ^{13}C and two ^{15}N atoms ('heavy 1', +8.0142 Da) and another containing eight ^2H atoms ('heavy 2', +8.0502 Da) for the metabolic labeling of yeast proteins. Compared to unlabeled Lys, both isotopically labeled Lys involve a ~8 Da mass increase but the small mass difference of 36 mDa between them can be distinguished at the MS1 level at a resolution above 200k, which is achievable in Fourier Transform Ion Cyclotron Resonance (FT-ICR) and recent Orbitrap mass analyzers. A scanning scheme including three sections was applied: (1) a high-resolution MS1 scan, which is used for quantification; (2) a modest-resolution MS1 scan, in which the mDa difference is not detectable and from which the precursors are selected for fragmentation and (3) multiple MS2 scans with modest-resolution to generate data for peptide identification.

The most attractive feature of this approach is that precursor pairs of the same peptide are isolated and fragmented together to produce indistinguishable MS2 spectra at a low resolution. In this way, the complexity of the MS1 spectra does not affect precursor ion selection and light and heavy forms of the same peptide are selected in the same window. This benefits the overall cycle time and increases the intensity of the overlapping fragment ions produced from both channels facilitating identification. Two factors determine the multiplexing capacity of the neutron encoding-based approach: (1) the number of available isotopologues; and (2) the attainable resolution. The more isotopologues are considered, the higher the required resolution. By adding four new Lys isotopologues ($\text{K}_{\#13\text{C}\#2\text{H}\#15\text{N}}$: K_{422} , K_{521} , K_{341} , and K_{440}) as shown in **Figure 7**, a 6-plex neutron-encoded SILAC experiment was performed^{92,93}.

	1	2	3	4	5	6	7	8
Lys ₆₀₂ (+0 mDa)	$^{13}\text{C}_1$	$^{13}\text{C}_1^1\text{H}_1$	$^{15}\text{N}_1^1\text{H}_2$	$^{13}\text{C}_1^1\text{H}_2$	$^{13}\text{C}_1^1\text{H}_2$	$^{13}\text{C}_1^1\text{H}_2$	$^{13}\text{C}_1^1\text{H}_2$	$^{15}\text{N}_1^1\text{H}_2$
Lys ₄₂₂ (+5.8 mDa)	$^{12}\text{C}_1$	$^{12}\text{C}_1^1\text{H}_1$	$^{15}\text{N}_1^1\text{H}_2$	$^{13}\text{C}_1^1\text{H}_2$	$^{13}\text{C}_1^1\text{H}_2$	$^{13}\text{C}_1^1\text{H}_2$	$^{13}\text{C}_1^2\text{H}_2$	$^{15}\text{N}_1^1\text{H}_2$
Lys ₅₂₁ (+12 mDa)	$^{13}\text{C}_1$	$^{13}\text{C}_1^1\text{H}_1$	$^{15}\text{N}_1^1\text{H}_2$	$^{13}\text{C}_1^1\text{H}_2$	$^{13}\text{C}_1^1\text{H}_2$	$^{13}\text{C}_1^1\text{H}_2$	$^{12}\text{C}_1^2\text{H}_2$	$^{14}\text{N}_1^1\text{H}_2$
Lys ₃₄₁ (+17.8 mDa)	$^{12}\text{C}_1$	$^{12}\text{C}_1^1\text{H}_1$	$^{14}\text{N}_1^1\text{H}_2$	$^{13}\text{C}_1^1\text{H}_2$	$^{13}\text{C}_1^1\text{H}_2$	$^{13}\text{C}_1^2\text{H}_2$	$^{12}\text{C}_1^2\text{H}_2$	$^{15}\text{N}_1^1\text{H}_2$
Lys ₄₄₀ (+24.1 mDa)	$^{12}\text{C}_1$	$^{12}\text{C}_1^1\text{H}_1$	$^{14}\text{N}_1^1\text{H}_2$	$^{13}\text{C}_1^1\text{H}_2$	$^{13}\text{C}_1^1\text{H}_2$	$^{13}\text{C}_1^2\text{H}_2$	$^{13}\text{C}_1^2\text{H}_2$	$^{14}\text{N}_1^1\text{H}_2$
Lys ₀₈₀ (+35.8 mDa)	$^{12}\text{C}_1$	$^{12}\text{C}_1^1\text{H}_1$	$^{14}\text{N}_1^1\text{H}_2$	$^{12}\text{C}_1^2\text{H}_2$	$^{12}\text{C}_1^2\text{H}_2$	$^{12}\text{C}_1^2\text{H}_2$	$^{12}\text{C}_1^2\text{H}_2$	$^{14}\text{N}_1^1\text{H}_2$

	1	2	3	4	5	6	7
DiPyrO ₀₀₄ (+0 mDa)	$^{15}\text{N}_1$	$^{15}\text{N}_1$	$^{15}\text{N}_1$	^{18}O	$^{15}\text{N}_1$	$^{12}\text{C}_1^1\text{H}_3$	$^{12}\text{C}_1^1\text{H}_3$
DiPyrO ₂₂₂₀ (+20.9 mDa)	$^{15}\text{N}_1$	$^{15}\text{N}_1$	$^{14}\text{N}_1$	^{18}O	$^{14}\text{N}_1$	$^{13}\text{C}_1^2\text{H}_1^1\text{H}_2$	$^{13}\text{C}_1^2\text{H}_1^1\text{H}_2$
DiPyrO ₀₈₀₀ (+45.3 mDa)	$^{14}\text{N}_1$	$^{14}\text{N}_1$	$^{14}\text{N}_1$	^{18}O	$^{14}\text{N}_1$	$^{12}\text{C}_1^2\text{H}_3$	$^{12}\text{C}_1^2\text{H}_3$

Figure 7. Structure and isotope distribution of 6-plex NeuCode SILAC lysine and the 3-plex DiPyrO tag.

The NeuCode approach has also been applied to chemical labeling MS1-based quantitation, using ^{13}C and ^{15}N in the 4-plex amine reactive acetyl-Arg-(acetyl)Lys-Gly-NHS tag having a 12.6 mDa mass difference between channels⁹⁴. However, the performance of the tag was suboptimal with a low number of identified peptides presumably caused by (1) the bulkiness of the tag leading to fragmentation of the tag itself generating sequence-uninformative fragment ions; and (2) sequestering protons on the Arg residue in the tag restricting fragmentation of the peptide backbone⁹⁴. To avoid these problems, various neutron-encoded chemical tags, that are more compact and facilitate peptide fragmentation, have been reported, such as neutron-encoded carbamylation⁹⁵ (**Figure 6**), mdDiLeu^{96,97} and DiPyrO⁹⁸. As asserted in the work on DiPyrO (**Figure 7**), deuterium is often needed to construct tags for a higher multiplexing capacity while ensuring the compactness of tag, which may affect the chromatographic retention of the same peptide labeled with a different number of deuterium atoms. Neutron encoding-based quantification can be improved in two ways: (1) by working with amine reactive tags, digestion with LysC instead of trypsin permits all peptides to accept two neutron-encoded tags, which doubles the mass difference between channels thus effectively reducing the requirements in terms of resolving power; (2) by combining the strategy of mass differences between isotopologues of the same element and of different elements to improve multiplexing capacity. For example, the acetyl-Arg-(acetyl)Lys-Gly-NHS tag allowed 12-plex labeling by triplicating the NeuCode-based 4-plex labeling⁹⁴: +0 Da (+0, +12.6, +25.2, +37.8 mDa), +4 Da (+0, +12.6, +25.2, +37.8 mDa), +8 Da (+0, +12.6, +25.2, +37.8 mDa). Overall, compared to the conventional MS1-based quantification approaches, the neutron-encoded MS1 approaches have advantages in terms of sampling depth, since redundant MS2 spectra of differentially labeled peptides are avoided, and in terms of multiplexing capability. However, they require very high resolution at the MS1 level, which increases the cycle time in Fourier Transform-based mass analyzers.

2.5 Summary of MS1-based quantification

To conclude, all MS1-based quantification approaches have the common strength of simultaneously analyzing multiple samples in a single LC-MS run. Therefore, relative quantitation between samples is performed in the same run rather than in different runs, which improves quantification precision. However, multiplexing capacity is usually limited for two reasons. First, isotopic labeling in MS1-based quantification approaches multiplies the complexity of MS1 spectra with the number of labeled samples, since the same peptides derived from different samples have distinct masses. The increase in MS1 complexity is a disadvantage when using DDA for data acquisition, since only a limited number of precursor ions can be selected for fragmentation within the timeframe of a chromatographic peak, resulting in more peaks being neglected for complex samples. Moreover, abundant peptides appearing as multiplets may be triggered multiple times, which leads to redundant MS2 spectra of the same peptide and consumes precious time that could be used to fragment peptides from lower abundant proteins⁶⁸. The second limitation with respect to increasing the multiplexing capacity lies in the requirement for more atoms to convey isotopes and the mass shift of at least 4 Da between labeling channels to prevent overlap of isotope envelopes. For both reasons, a large isotopic tag is required, which may affect the fragmentation properties negatively and result in a decline in identification rate⁹⁹⁻¹⁰¹. The neutron encoding approaches improve the multiplexing capacity without interfering with precursor ion selection. However, its feasibility is highly resolution-dependent (> 200k) and requires state-of-the-art FT-ICR or Orbitrap mass analyzers. Besides, being restricted by the compactness of the tag, and hence the limited number of locations that can be labeled with $^{13}\text{C}/^{15}\text{N}$, as well as the limited commercial availability of the tags, neutron encoding-based MS1 tags are currently not in widespread use.

3 MS2-based quantification

As discussed above, MS1-based quantification has a multiplexing capacity that is generally limited to triplex labeling. However, exploring biological questions often

requires quantitative comparison of proteomes across multiple conditions at different time points resulting in more than three samples. MS2-based quantification approaches readily allow for higher multiplexing capacity without most of the drawbacks of MS1-based quantification methods.

MS2-based quantification generally relies on labeling of the same peptides derived from different samples with different isobaric tags. As a result, differentially labeled peptides have identical masses, which not only avoids an increase in MS1 spectrum complexity but also increases signal intensity of peptide ions by combining the signals from each channel. Isobaric labeling further reduces the risk of acquiring redundant MS2 spectra from the same peptide. This makes isobaric labeling highly compatible with the DDA acquisition mode, where only a certain number of precursor ions can be isolated for fragmentation within a given time window. Except for the fragment ion-based quantification variant, isobarically labeled peptides of different samples produce the same sequence-derived fragment ions of the peptide backbone, which affords a better signal-to-noise (S/N) ratio and thus improves peptide identification.

Upon fragmentation, isobarically labeled peptides release unique quantification ions, often called reporter ions, for each labeling channel. The relative peak intensities of these ions are proportional to the corresponding peptide levels¹⁵. By comparing the intensities of the quantification ions, the relative quantitative information can be retrieved from the MS2 spectra which generally have a better S/N ratio than MS1 spectra^{57,102}. Deuterium-based labeling is avoided in most isobaric tags, since coelution of differentially labeled peptides containing the same sequence is critical for quantification accuracy. Based on the type of quantification ion, the current isobaric labeling methods fall into three general categories: (1) reporter ion-based quantification (e.g. TMT^{15,31}, iTRAQ¹⁴), (2) peptide fragment ion-based quantification (e.g. IPTL³², triplex-IPTL³⁵) and (3) peptide-coupled reporter ion-based quantification (e.g. TMTc^{42,103} and EASI⁵¹).

3.1 Reporter ion-based quantification

As reviewed by Arul et al.¹⁰⁴, reporter ion-based quantification is the most widely used isobaric labeling-based strategy, and includes TMT, iTRAQ and DiLeu. The structural design of reporter-ion-based isobaric tags consists of three components: (1) a reporter-ion moiety, containing distinct masses for different labeling channels thus representing the quantitative information of the constituent peptides in MS2, (2) a balancer group, containing complementary isotopes to the reporter-ion to ensure the same overall mass of the isobaric tags and (3) a reactive group, targeting a specific functional group on the peptide.

Figure 8. The concept of isobaric tags, with reporter ion, balancer and amine reactive groups. Fragmentation occurs between the balancer and the reporter ion, which carries the quantitative information. The isotope distribution of the original 2-plex TMT, the 4-plex iTRAQ and the 10-plex TMT (with neutron encoding) are shown as representative examples. For 10-plex TMT ^{13}C and ^{15}N are marked with an asterisk.

3.1.1 TMT and iTRAQ tags. The concept of isobaric labeling was first proposed in 2003 with the original version of the 2-plex TMT¹⁵. It represented the next breakthrough in quantitative proteomics after ICAT. In the original paper, two structures, 6-guanidinohexanoic acid-Met-Met-Gly-NHS and 6-guanidinohexanoic acid-Met-Pro-Met-Gly-NHS (**Figure 8**), were proposed. Fragmentation at the N-terminal side of the proline residue leads to formation of the reporter ion, permitting the simultaneous acquisition of fragment ions of the peptide backbone and reporter ions. In 2008, an upgraded 6-plex TMT design was reported³¹. Compared to the first-generation TMT¹⁵, the structure of 6-plex TMT is more compact and optimized to facilitate the formation of the reporter ion. It was further extended to 10-plex labeling by neutron encoding^{90,91}. As in all recent labeling methods, only ¹³C and ¹⁵N isotopes are used while deuterium is avoided (**Figure 8**). The low-mass reporter ions with a minimal mass shift of 6.3 mDa can be differentiated by mass analyzers at a resolution of 50k⁹⁰ at m/z 100 which is lower than the resolution of 200k as required for neutron-encoded MS1-based quantification. By substituting every ¹²C with ¹³C and every ¹⁴N with ¹⁵N, 18-plex can be achieved based on the chemical skeleton of the 6-plex TMT^{90,105}. However, presumably due to restrictions of chemical synthesis and reagent cost, 11-plex is currently the highest multiplexing capacity for this TMT design. Driven by the demand for even higher multiplexing capacity and presumably lower synthesis cost, a 16-plex proline-based isobaric Tandem Mass Tag (TMTpro) was recently introduced, which uses labeled proline as reporter ion rather than dimethylpiperidine and incorporates two beta-alanine residues as the balancer group^{105,106} (**Figure 9**). In the TMTpro tag, incorporation of 9 heavy isotopes (¹³C and ¹⁵N) results in 9 tags with a ~1 Da mass shift and applying neutron encoding permits 7 more tags with a mass difference of 6.3 mDa.

Shortly after publication of the first-generation duplex TMT, 4-plex iTRAQ was reported, designed based on the principles of the original TMT tag^{14,50} (**Figure 8**). To further improve multiplexing capacity, a larger balancer group was incorporated in the structure of the 8-plex iTRAQ^{107,108}. However, compared to the 4-plex iTRAQ, lower protein and peptide identification rates were reported, which might be caused by

formation of sequence-uninformative fragment ions derived from additional internal fragmentation of the larger tag itself^{90,91,99,100,109}.

Among the commercially available reporter-ion-based, isobaric tags, the TMT series has gained in popularity in recent years¹¹⁰, most likely because of its continuously developing multiplexing capabilities^{31,90,106}.

Figure 9. Structure and isotope distribution of the 16-plex TMTpro tag. ¹³C and ¹⁵N are marked with an asterisk.

3.1.2 DiLeu, DiART and IBT tags. Since iTRAQ and TMT tags require complex, multistep syntheses with moderate to low yields^{109,111}, several isobaric tags have been prepared as alternatives, such as the N,N-dimethyl leucine (DiLeu)^{109,112,113}, the deuterium isobaric amine reactive tag (DiART)^{111,114} and the isobaric tags (IBT)¹¹⁵. Inspired by the formation of the intense a_1 ion in dimethylated peptides^{83,116,117}, Xiang et al. proposed the 4-plex DiLeu tag, which has a dimethylated Leu as the core

	1	2	3	4	5	6	7
115a (115.12476)	$^{13}\text{C}_1$	^{18}O	$^{12}\text{C}_1^1\text{H}_1$	$^{15}\text{N}_1$	$^{12}\text{C}_1^1\text{H}_3$	$^{12}\text{C}_1^1\text{H}_3$	$^{12}\text{C}_1^1\text{H}_2$
115b (115.13108)	$^{13}\text{C}_1$	^{18}O	$^{13}\text{C}_1^1\text{H}_1$	$^{14}\text{N}_1$	$^{12}\text{C}_1^1\text{H}_3$	$^{12}\text{C}_1^1\text{H}_3$	$^{12}\text{C}_1^1\text{H}_2$
116a (116.12812)	$^{12}\text{C}_1$	^{18}O	$^{12}\text{C}_1^1\text{H}_1$	$^{15}\text{N}_1$	$^{12}\text{C}_1^1\text{H}_3$	$^{12}\text{C}_1^1\text{H}_3$	$^{13}\text{C}_1^1\text{H}_2$
116b (116.13444)	$^{12}\text{C}_1$	^{18}O	$^{12}\text{C}_1^1\text{H}_1$	$^{14}\text{N}_1$	$^{13}\text{C}_1^1\text{H}_3$	$^{13}\text{C}_1^1\text{H}_3$	$^{12}\text{C}_1^1\text{H}_2$
116c (116.14028)	$^{12}\text{C}_1$	^{18}O	$^{12}\text{C}_1^1\text{H}_1$	$^{14}\text{N}_1$	$^{12}\text{C}_1^2\text{H}_1^1\text{H}_2$	$^{12}\text{C}_1^2\text{H}_1^1\text{H}_2$	$^{12}\text{C}_1^1\text{H}_2$
117a (117.13147)	$^{13}\text{C}_1$	^{16}O	$^{12}\text{C}_1^1\text{H}_1$	$^{15}\text{N}_1$	$^{13}\text{C}_1^1\text{H}_3$	$^{13}\text{C}_1^1\text{H}_3$	$^{12}\text{C}_1^1\text{H}_2$
117b (117.13731)	$^{13}\text{C}_1$	^{16}O	$^{12}\text{C}_1^1\text{H}_1$	$^{15}\text{N}_1$	$^{12}\text{C}_1^2\text{H}_1^1\text{H}_2$	$^{12}\text{C}_1^2\text{H}_1^1\text{H}_2$	$^{12}\text{C}_1^1\text{H}_2$
117c (117.14363)	$^{13}\text{C}_1$	^{16}O	$^{13}\text{C}_1^1\text{H}_1$	$^{14}\text{N}_1$	$^{12}\text{C}_1^2\text{H}_1^1\text{H}_2$	$^{12}\text{C}_1^2\text{H}_1^1\text{H}_2$	$^{12}\text{C}_1^1\text{H}_2$
118a (118.13483)	$^{12}\text{C}_1$	^{16}O	$^{12}\text{C}_1^1\text{H}_1$	$^{15}\text{N}_1$	$^{13}\text{C}_1^1\text{H}_3$	$^{13}\text{C}_1^1\text{H}_3$	$^{13}\text{C}_1^1\text{H}_2$
118b (118.14067)	$^{12}\text{C}_1$	^{16}O	$^{12}\text{C}_1^1\text{H}_1$	$^{15}\text{N}_1$	$^{12}\text{C}_1^2\text{H}_1^1\text{H}_2$	$^{12}\text{C}_1^2\text{H}_1^1\text{H}_2$	$^{13}\text{C}_1^1\text{H}_2$
118c (118.14699)	$^{12}\text{C}_1$	^{16}O	$^{12}\text{C}_1^1\text{H}_1$	$^{14}\text{N}_1$	$^{13}\text{C}_1^2\text{H}_1^1\text{H}_2$	$^{13}\text{C}_1^2\text{H}_1^1\text{H}_2$	$^{12}\text{C}_1^1\text{H}_2$
118d (118.15283)	$^{12}\text{C}_1$	^{16}O	$^{12}\text{C}_1^1\text{H}_1$	$^{14}\text{N}_1$	$^{12}\text{C}_1^2\text{H}_2^1\text{H}_1$	$^{12}\text{C}_1^2\text{H}_2^1\text{H}_1$	$^{12}\text{C}_1^1\text{H}_2$

Figure 10. The structure and isotope distribution of 4 and 12-plex DiLeu. The 4-plex DiLeu consists of 115a, 116c, 117b and 118d (marked in bold).

structure¹¹², as shown in **Figure 10**. Essentially, DiLeu is based on the widely used dimethylation strategy in MS1-based quantification⁶⁵ adapted for MS2-based isobaric labeling. DiLeu-labeling was reported to give better fragmentation of the peptide backbone and more intense reporter ions when compared to iTRAQ-labeled peptides¹¹². The DiLeu tag has only a few carbon or nitrogen atoms and deuterium labels were therefore also incorporated to extend the multiplexing capacity, despite the fact that deuterium labeling affects the retention time of the labeled peptides. This effect was, however, minimized by placing the deuterium atoms in the hydrophilic dimethylamine group, which presumably has less interaction with the reversed-phase stationary

phase^{44,65,68}. While the deuterium-induced retention time shift constitutes a disadvantage for this isobaric tag, the commercial availability of different isotopic forms of Leu, the rather straightforward synthesis and the intense reporter ion make DiLeu a valuable alternative to TMT or iTRAQ.

Similar to the development of TMT, neutron encoding was employed to increase the multiplexing capacity of DiLeu from 4-plex to 12-plex (**Figure 10**), however, with a small mass difference of 5.8 mDa between the reporter ions¹⁰⁹. By further reducing the mass difference between the reporter ions to 3 mDa by stepwise dimethylation and neutron encoding, the multiplexing capacity of DiLeu was recently improved to 21-plex without expanding the compact structure of the tag¹¹³. However, the required resolution to discriminate this tiny mass difference increases from 30k to 60k (at 400 *m/z* for Orbitrap mass analyzers) thus requiring longer cycle times.

	1	2	3	4	5	6	7	8	9	10	11	12
114	$^{12}\text{C}_1\text{H}_3$	$^{14}\text{N}_1$	$^{12}\text{C}_1\text{H}_3$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_3$	$^{12}\text{C}_1\text{H}_3$	$^{13}\text{C}_1$	$^{15}\text{N}_1$	$^{13}\text{C}_1\text{H}_1$	$^{13}\text{C}_1\text{H}_3$	$^{13}\text{C}_1$
115N	$^{12}\text{C}_1\text{H}_3$	$^{15}\text{N}_1$	$^{12}\text{C}_1\text{H}_3$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_3$	$^{12}\text{C}_1\text{H}_3$	$^{13}\text{C}_1$	$^{14}\text{N}_1$	$^{13}\text{C}_1\text{H}_1$	$^{13}\text{C}_1\text{H}_3$	$^{13}\text{C}_1$
115C	$^{12}\text{C}_1\text{H}_3$	$^{14}\text{N}_1$	$^{12}\text{C}_1\text{H}_3$	$^{13}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_3$	$^{12}\text{C}_1\text{H}_3$	$^{12}\text{C}_1$	$^{15}\text{N}_1$	$^{13}\text{C}_1\text{H}_1$	$^{13}\text{C}_1\text{H}_3$	$^{13}\text{C}_1$
116N	$^{12}\text{C}_1\text{H}_3$	$^{15}\text{N}_1$	$^{12}\text{C}_1\text{H}_3$	$^{13}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_3$	$^{12}\text{C}_1\text{H}_3$	$^{12}\text{C}_1$	$^{14}\text{N}_1$	$^{13}\text{C}_1\text{H}_1$	$^{13}\text{C}_1\text{H}_3$	$^{13}\text{C}_1$
116C	$^{13}\text{C}_1\text{H}_3$	$^{14}\text{N}_1$	$^{13}\text{C}_1\text{H}_3$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_3$	$^{12}\text{C}_1\text{H}_3$	$^{13}\text{C}_1$	$^{15}\text{N}_1$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_3$	$^{13}\text{C}_1$
117N	$^{13}\text{C}_1\text{H}_3$	$^{15}\text{N}_1$	$^{13}\text{C}_1\text{H}_3$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_3$	$^{12}\text{C}_1\text{H}_3$	$^{12}\text{C}_1$	$^{14}\text{N}_1$	$^{13}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_3$	$^{13}\text{C}_1$
117C	$^{13}\text{C}_1\text{H}_3$	$^{14}\text{N}_1$	$^{13}\text{C}_1\text{H}_3$	$^{13}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_3$	$^{12}\text{C}_1\text{H}_3$	$^{12}\text{C}_1$	$^{15}\text{N}_1$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_3$	$^{13}\text{C}_1$
118N	$^{13}\text{C}_1\text{H}_3$	$^{15}\text{N}_1$	$^{13}\text{C}_1\text{H}_3$	$^{13}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_3$	$^{12}\text{C}_1\text{H}_3$	$^{12}\text{C}_1$	$^{14}\text{N}_1$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_3$	$^{13}\text{C}_1$
118C	$^{13}\text{C}_1\text{H}_3$	$^{14}\text{N}_1$	$^{13}\text{C}_1\text{H}_3$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_1$	$^{13}\text{C}_1\text{H}_3$	$^{13}\text{C}_1\text{H}_3$	$^{12}\text{C}_1$	$^{15}\text{N}_1$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_3$	$^{12}\text{C}_1$
119	$^{13}\text{C}_1\text{H}_3$	$^{15}\text{N}_1$	$^{13}\text{C}_1\text{H}_3$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_1$	$^{13}\text{C}_1\text{H}_3$	$^{13}\text{C}_1\text{H}_3$	$^{12}\text{C}_1$	$^{14}\text{N}_1$	$^{12}\text{C}_1\text{H}_1$	$^{12}\text{C}_1\text{H}_3$	$^{12}\text{C}_1$

Figure 11. The structure and isotope distribution of 10-plex IBT. ^{13}C and ^{15}N are marked with an asterisk.

Instead of relying on neutron encoding, some dimethylated Leu-based tags have been described that increase multiplexing capacity by incorporating a larger balancer group to replace the carbonyl group in 4-plex DiLeu, which can only accommodate two

isotopes (^{13}C and ^{18}O). For example, 6-plex DiART^{111,114} incorporates a beta-alanine and both 8-plex DiLeu¹¹⁸ and 10-plex IBT¹¹⁵ (**Figure 11**) use an alanine as the balancer between the dimethylated leucine and the amine reactive group. In contrast to 4-plex or 12-plex DiLeu, peptides labeled with 8-plex DiLeu and DiART exhibited considerable retention time shifts between labeling channels due to the more hydrophilic deuterium being inserted in the balancer group (alanine or beta-alanine), which constitutes a disadvantage for an isobaric, reporter-ion-based tag¹¹⁸. As variations on the same theme, DiAla and DiVal tags have also been synthesized and compared to the DiLeu tag¹¹⁹. Currently, the IBT tag is the only isobaric tag, apart from TMT, that achieves 10-plex labeling using only ^{13}C and ^{15}N isotopes.

3.1.3 Novel designs and applications of reporter-ion-based tags. Besides the aforementioned approaches for relative quantification of peptides based on reporter-ions, novel concepts and strategies that provide a blueprint for further developments and a wider range of applications of isobaric tags have been proposed. The Combinatorial Isobaric Mass Tags¹²⁰ (CMTs) increase the multiplexing capacity by the concept that every isobaric tag produces two sets of reporter-ions that are independent from each other and where quantitative information can be inferred from the combination of distinct reporter ions (**Figure 12**). While only showing the results of 6-plex labeling, the authors anticipate a 16-plex labeling capacity with a mass shift of ~1 Da between reporter-ions and an up to 28-plex labeling capacity with a mass shift of ~6 mDa using only five heavy isotopes. However, the authors pointed out that the formation of reporter ions was affected by the presence or absence of a mobile proton on the precursor peptide¹²¹. Although there is room for improvement with respect to the structure of this tag, the concept of combinatorial quantification is promising for increasing the multiplexing capacity of quantitative proteomics.

Figure 12. Structure of CMTs and reporter ions. Isotope distribution of 6-plex CMTs and the principle of reporter ion deconvolution. ^{13}C and ^{15}N are marked with an asterisk.

In another development, a series of quantitative approaches targeting specific subclasses of peptides or proteins was reported (**Figure 13**), namely the Cleavable Isobaric Labeled Affinity Tag (CILAT)¹²² containing a thiol group targeting the ortho-quinone produced by oxidation of Tyr residues with tyrosinase, the cystTMT tag targeting Cys residues^{123,124}, and the GlycoTMT tag¹²⁵ and the iTRAQ hydrazide (iTRAQH)¹²⁶ targeting reactive aldehyde groups of carbohydrates after oxidation of cis-diols with periodate. The above labels target specific residues or reactive sites other than primary amine groups, such as cystTMT targeting thiol groups in Cys residues with

iodoacetamide analogs or GlycoTMT targeting aldehydes in glycans with aminooxyl analogs. This is especially useful for studying post translation modifications (PTMs), since PTM-containing peptides are often low abundant and a large amount of starting material is required for MS detection. Accordingly, when using standard isobaric tags to analyze PTMs, a large amount of tag is needed for complete labeling, while the selective reaction strategies reduce the required amount.

Figure 13. Examples of TMT or iTRAQ based tags for studying glycosylated proteins and proteins containing reactive amino acid derivatives, ortho-quinone in the oxidized Tyr residue and thiol group of Cys residue.

3.1.4 Strategies to correct for ratio distortion. While reporter-ion-based quantitative approaches exhibit excellent throughput due to their high multiplexing capacity and flexible labeling capability targeting various functional groups, it has become increasingly apparent that reporter-ion-based quantification suffers from ratio distortion^{42,99,103,127,128} caused by peptide cofragmentation, as shown in **Figure 14**. Data acquisition in DDA mode generally utilizes a window of several Thomson (Th)¹²⁹ to isolate precursor ions for fragmentation. While isolation and fragmentation of a single precursor ion results in well-defined MS2 spectra, in which the ratio of reporter ions reflects the ratio of a given peptide in the differentially labeled samples, this is no longer the case when precursor ions of distinct peptides, that have similar retention times and a mass difference that is smaller than the isolation window, are co-isolated. As a result, the produced reporter ions from different peptides are indistinguishable and their

intensities no longer reflect the ratios between the peptides in the different samples thus leading to incorrect protein ratios.

Figure 14. Ratio distortion caused by cofragmentation of multiple peptide precursor ions.

This issue is more serious for more complex samples and depends further on the efficiency of the chromatographic separation. Although only a minority of peptides and proteins are generally affected, ratio distortion decreases the reliability of the entire analysis as it affects protein ratios in an unpredictable way^{103,127}. In the past decade, a number of approaches have been reported to tackle the ratio distortion problem. They can be classified into the following categories: (1) extending the LC gradient in combination with prefractionation^{127,130}, (2) narrowing the precursor isolation window^{42,51,127,131}, (3) delayed fragmentation at the apex of the LC peak¹³¹, (4) additional gas phase reactions¹²⁸, (5) subjecting primary fragment ions to an additional isolation and MS3 fragmentation step^{59,127,132,133}, (6) ion mobility separation to separate coeluting peptides¹³⁴⁻¹³⁶, and (7) correcting ratio distortion by estimating the extent of interference or by detecting and discarding MS2 spectra that are produced from multiple precursor ions^{137,138}.

According to Ting et al. and Wenger et al.^{127,128}, prefractionation and narrowing the isolation window reduce the problem of ratio distortion only to a limited extent, even

though this is somewhat surprising considering that the chance of having peptides of similar m/z coelute should decrease. This indicates that we are still only scratching the surface of the depth of the peptide mixtures in shotgun proteomics. A delayed fragmentation scheme combined with a narrowed isolation window on Orbitrap instruments was applied to alleviate ratio distortion by 32%, because fragmentation was triggered closer to the apex of a chromatographic peak. This avoided cofragmentation to some extent and afforded better S/N ratios resulting in higher quality MS2 spectra¹³¹. Reducing the average peptide charge state by proton transfer ion-ion reactions in the so-called QuantMode method increases the m/z difference between peptides, thereby reducing the possibility of cofragmentation¹²⁸, for example by avoiding interference between a doubly charged precursor ion and another triply charged precursor ion. However, this method is not widely used, since it relies on scan routines that are difficult to optimize and control and that require advanced instrumentation.

In the initially reported MS3-based approach to correct for ratio distortion, the most intense fragment ion in the MS2 spectrum is selected for further fragmentation. Since b- and y-ions of overlapping precursors are less likely to overlap as well, the resulting reporter ions are less prone to be affected by ratio distortion¹²⁷. Limited to the isolation of only one fragment ion for secondary fragmentation, the first generation of MS3-based methods suffered from modest intensities of reporter-ions in the MS3 spectra. In light of this, the improved MultiNotch MS3 method synchronously isolates multiple fragment ions for secondary fragmentation (synchronous precursor selection, SPS), which increased the intensities of the reporter ions in the MS3 spectra by more than 10-fold¹³², significantly improving the dynamic range and reducing reporter ion signal variance. Nevertheless, only intensity-based selection is employed which means that selecting fragment ions derived from multiple precursors cannot be avoided. The MultiNotch MS3 method was implemented on an Orbitrap Fusion mass spectrometer, which has an ion trap as well as a dedicated collision cell for multistage fragmentation. Since an additional scan event is required, MS3-based approaches have longer DDA cycle times and as a result fewer protein identifications in comparison to MS2-based

quantitative methods. Recently, a real-time database search platform, Orbiter, has been reported to shorten the cycle time by canceling the SPS MS3 scan when there are no peptide matches for a given MS2 spectrum⁵⁹. Compared to the previous MultiNotch MS3 method, the real-time database search-based SPS MS3 achieved a 2-fold faster acquisition rate resulting in more quantified proteins.

High field asymmetric waveform ion mobility spectrometry (FAIMS) is an atmospheric pressure ion mobility technique that separates gas-phase ions based on their behavior in strong and weak electric fields. FAIMS is easily interfaced with electrospray ionization and has been widely used as a means of on-line fractionation. More details can be found in the following reviews and publications¹³⁹⁻¹⁴². Pfammatter et al. utilized FAIMS to reduce ratio distortion by adding a separation dimension based on ion mobility to decrease the chance of precursor ion cofragmentation^{134,135,143}. Indeed, FAIMS was shown to robustly improve TMT-based quantification accuracy and precision without sacrificing peptide/protein identifications¹³⁶.

All of the aforementioned approaches aim at mitigating ratio distortion of reporter-ions by avoiding peptide cofragmentation during data acquisition. An alternative strategy is post-acquisition data processing, such as determining whether precursors were cofragmented based on the elution profiles in LC of the precursor ions. The extent of convolution of precursor ions can then be used to reduce the weighted contribution of chimeric MS2 spectra to the quantification result or to discard a given MS spectrum^{137,138}. Although it reduces the number of quantifiable peptides, this type of strategy is still the most straightforward, since it does not require any specialized equipment and additional ion separation or fragmentation routines.

3.2 Peptide backbone fragment ion-based quantification

The root cause of ratio distortion in reporter ion-based methods, such as TMT or iTRAQ, is that the reporter-ion is not specific for a given peptide. This makes the reporter-ion unable to accurately reflect the quantitative information once different peptides are cofragmented. To circumvent this pitfall, employing specific fragment ions of the

peptide backbone for quantification has been explored. In peptide backbone fragment ion-based methods, even though distinct peptides labeled with the same labels may be cofragmented, the peptide-specific fragment ions can still be attributed to the correct peptide thus the quantification information can also be properly assigned.

Figure 15. Schematic overview of the original IPTL approach.

In 2009, Koehler et al. first reported Isobaric Peptide Termini Labeling (IPTL)³² and proposed the concept of quantifying peptides based on sequence-specific fragment ions. The original IPTL method used 2-methoxy-4,5-dihydro-1H-imidazole (MDHI) and the tetradeuterated form 2-methoxy-4,5-dihydro-1H-imidazol-4,4,5,5-d₄ (MDHI-d₄) to specifically modify the epsilon-amino group of Lys followed by derivatization of the N-terminal amine group with succinic anhydride (SA) or tetradeuterated succinic anhydride-d₄ (SA-d₄) (**Figure 15**). As a result, the peptides, derived from two samples, are differentially but isobarically labeled, but generate sets of distinct fragment ions upon fragmentation. Since N- and C-termini contain different labels, the peptide and

protein ratios can be inferred by comparing the intensities of individual y- and b-series fragment ions. Even in the case that more than one peptide is selected in the same precursor isolation window, the resulting b- and y-ions can usually be correctly attributed to the corresponding peptide. IPTL potentially permits more accurate and precise quantification than reporter-ion-based methods due to multiple quantification data points per spectrum, for each y- and b-ion. IPTL does not refer to a label with a specific structure, but rather to a labeling strategy that can achieve isobaric labeling by modifying peptide termini. Based on this principle, a considerable number of methods and applications have been reported in the past ten years by combining various isotope labeling methods, such as rapid-IPTL³³, triplex-IPTL³⁵ and triplex-QITL³⁷ based on selective succinylation and dimethylation, IVTAL³⁴, G-IVTAL¹⁴⁴, QITL¹⁴⁵, diDO-IPTL¹⁴⁶ based on SILAC^{144,147} and proteolytic ¹⁸O labeling^{145,146} or pIDL¹⁴⁸, PITL¹⁴⁹ and SWATH-pseudo-IPTL³⁶ based on so-called pseudo-isobaric dimethyl labeling. Even though many methods have been reported, the IPTL-based approaches are not as extensively used as the reporter-ion-based approaches, presumably for two reasons: (1) their limited multiplexing capacity, especially when avoiding deuterium labeling and (2) the increasing complexity of MS2 spectra with increasing levels of multiplexing due to multiplets of fragment ions. Most the above-mentioned IPTL-based methods can achieve multiplexing up to triplex, which is far less than for reporter-ion-based methods (e.g. TMT has recently been extended to 16-plex labeling in a single LC-MS run¹⁰⁵). Recently, Liu et al.⁵² reported the so-called pseudo-isobaric dimethyl labeling (m-pIDL) method, which increases the multiplexing capacity to 6-plex. m-pIDL relies on a wide isolation window of 10 Th, which increases the complexity of MS2 spectra even further. Furthermore, the use of deuterium in the tags carries the risk of varying the retention time of labeled peptides and may lead to inaccurate quantification.

3.3 Peptide-coupled reporter-ion based quantification

In reporter-ion-based approaches, such as TMT, different peptides derived from the same sample are labeled with the identical tag, which releases indistinguishable reporter

ions upon tandem MS, with the associated risk of ratio distortion upon cofragmentation. It is worth to note that, when the fragmentation occurs on the tag and after losing the reporter ion, the remaining part termed the ‘peptide-coupled reporter-ion’, containing the complete peptide sequence of the precursor with the balancer group attached, contains the complementary quantitative information to the reporter ion (see **Figure 1**). In contrast to the reporter-ions, peptide-coupled reporter-ions produced from different peptides are peptide-specific and their mass differences between labeling channels can be discriminated in high-resolution MS2 spectra (e.g. at 17.5k in an Orbitrap). Only for the extreme cases where the difference between peptide-coupled reporter-ions is too small to be differentiated at the set MS2 resolution, the quantitative information cannot be accurately deduced from the peptide-coupled reporter-ions. Compared to the IPTL-based approaches, the biggest merit of peptide-coupled reporter-ion based quantification is that differentially labeled peptides have both identical precursor masses and identical fragments ions. While reporter-ion-based approaches belong to the mainstream of shotgun proteomics to date, there are only a few recent reports on peptide-coupled reporter-ion-based approaches, namely the TMTc¹⁰³, TMTc+⁴² and EASI⁵¹ tag. A limitation of peptide-coupled reporter-ions is that they are usually located in the range between m/z 500-1500, which makes it impossible to use neutron encoding for extending the multiplexing capacity, since a resolution of more than 1 million would be needed to resolve the differentially labeled ions.

3.3.1 TMTc, TMTc+. Peptide-coupled reporter-ion based quantification was first proposed by Wühr et al.¹⁰³ and called TMTc. However, two limitations leading to unsatisfactory quantification were encountered: 1) the modest efficiency of forming peptide-coupled reporter-ions and 2) the complicated isotope envelope of the peptide-coupled reporter-ions, which required deconvolution to extract the quantitative information. In this first report, the TMT tag, originally designed for the efficient formation of reporter-ions, was utilized for labeling and an isolation window of 2 Th was used to isolate precursor ions for fragmentation. A more recent implementation

namely TMTc+ used a 0.4 Th precursor isolation window to facilitate data processing, since the monoisotopic precursor ion peak was isolated specifically obviating deconvolution of overlapping isotope envelopes from different labeling channels⁴². Due to the original design of the TMT tag, which favors charge sequestration at the dimethylpiperidine ring, the yield of peptide-coupled reporter-ion formation was modest. This showed that the peptide-coupled reporter ion-based approach requires specifically designed tags.

3.3.2 EASI tag. The recently introduced 6-plex EASI tag⁵¹ utilizes a sulfoxide-based fragmentation site to increase the efficiency of peptide-coupled reporter-ion formation and applies an asymmetric isolation window (0.4 Th and -0.15 Th offset) to simplify the isotope envelope of peptide-coupled reporter-ions. As shown in **Figure 16**, the EASI tag consists of a neutral loss group, a balancer group and an amine-reactive NHS ester. Design of the EASI tag is based on previous work showing that the C-S bond adjacent to the sulfoxide group fragments easily at low normalized collision energy (NCE) without fragmentation of the peptide backbone^{161,162}. Fragment ions of the peptide backbone are subsequently produced at higher collision energy. Facilitated by the enhanced peptide-coupled reporter-ions, EASI tag achieved significant improvements in terms of sensitivity leading to better quantification accuracy and precision.

Figure 16. Functional design and isotope distribution of the 6-plex EASI tag. ¹³C isotope locations are marked with an asterisk.

3.4 Summary of MS2-based quantification

To conclude, the most widely used quantification strategy based on isotope labeling is isobaric labeling-based MS2 quantification using reporter ions. From the first TMT approach proposed in 2003 to the recently reported TMTpro^{15,106}, the isobaric labeling capacity has increased eight times from duplex to 16-plex with neutron encoding⁹⁰. During this time remedies such as MS3-based quantification¹²⁷, ion mobility-based quantification¹³⁴ and peptide-coupled reporter ion-based quantification¹⁰³ have been proposed to circumvent the problem of ratio distortion of reporter ion-based approaches. Quantification based on isobaric labeling has become an indispensable strategy in quantitative proteomics, due to its excellent throughput, precision and accuracy.

4 Combined MS1 and MS2-based quantification

Driven by the need for a higher multiplexing capacity, hybrid labeling strategies combining isotopic and isobaric labeling in the same sample preparation work-flow have been developed to increase the throughput of isobaric labeling approaches by two to three times¹⁶³. The basic idea is to use the mass difference at the MS1 level to further parallelize the analysis of two or three sets of isobarically labeled samples in a single LC-MS run. Published approaches, comprise the 18-plex Hyperplexing¹⁶⁴ (triplex SILAC combined with 6-plex TMT), 12-plex cPILOT¹⁶³ (duplex N-terminal selective dimethylation combined with 6-plex TMT) and 24-plex DiLeu cPILOT¹⁶⁵ (duplex N-terminal selective dimethylation combined with 12-plex DiLeu), metabolically or chemically incorporate distinct isotopic labels followed by chemical derivatization with isobaric tags. As these strategies are all based on the orthogonal combination of MS1 and MS2-based quantification strategies, they inherently suffer from the issues of MS1- and MS2-based quantification¹⁶⁶, such as the multiplied complexity of spectra at the MS1 level and ratio distortion caused by peptide cofragmentation at the MS2 level. It should be pointed out that, for a given LC-MS platform, more complex samples are more seriously affected by these limitations.

5 Isotope labeling-based quantification in DIA

As discussed, increasing the tag size¹⁶⁷ and neutron-encoding⁹¹ were applied to increase the throughput of isobaric labeling in DDA up to a maximum capacity of analyzing 16 isobarically labeled samples in a single LC-MS run¹⁰⁵. However, the stochastic nature of selecting precursor ions for tandem MS in DDA leads to an imperfect overlap in peptide identifications, which means that there is a considerable likelihood that some peptides might be missed in some runs depending on their relative abundance in a given sample¹⁶⁸. Since only the peptides that are measured in all runs can be reliably compared across many runs in studies comprising hundreds or more samples, it would be beneficial to use a strategy that does not rely on data-dependent and thus stochastic precursor ion selection. This has led to the development of data-independent acquisition (DIA) modes of operation, where all precursor ions in a predefined m/z window are fragmented notwithstanding their relative intensity in the MS1 spectrum.

DIA approaches have gained in popularity notably in large-scale biomedical studies, for example related to clinical biomarker discovery, as they largely avoid the missing value problem of DDA. Another advantage of DIA is that a so-called digital fingerprint of a sample is acquired that can later be interrogated for novel features (e.g. biomarker candidates) as new insights emerge from biology without the need to analyze the sample again.

A severe drawback of DIA is, however, that most of the current DIA-based approaches can only analyze one sample per LC-MS run, which limits throughput. It would thus be ideal to combine DIA with the multiplexing capability of DDA. However, multiplexed labeling is rarely used in DIA, presumably because current approaches lead to more complex MS2 spectra, severe ratio distortion, and/or a reduction in quantification accuracy and precision. Reporter ion-based TMT or iTRAQ approaches are not suitable for DIA due to massive cofragmentation. For isotopic labeling and peptide fragment ion-based isobaric labeling strategies, such as SILAC^{27,28} and IPTL³², the peptides from different labeling channels have sets of distinct fragments ions, which multiplies the complexity of MS2 spectra with the number of labeled samples making identification

of peptides extremely challenging, since MS2 spectra of DIA are already highly convoluted. Still, there are initial attempts at developing multiplexed stable isotope labeling strategies for DIA, which are briefly discussed below.

Motivated by the need for higher throughput in DIA, researchers have devised novel labeling strategies. The recently reported DIA multiplexed approaches, NeuCoDIA¹⁶⁹ and MdFDIA¹⁷⁰ incorporate isotopes during cell culture (metabolic labeling), while the mdDiLeu approach uses chemical labeling¹⁷¹. They all have the same principle of incorporating a mDa mass difference by neutron encoding in peptides. Upon tandem MS, the multiplets of fragment ions containing the neutron-encoded tag represent the quantification information, which is comparable to the quantification principle of neutron-encoded MS1 based quantification discussed in section 2.4 on Neutron-encoded isotopic labeling. These methods rely on ultrahigh resolution ($>120k$ in the m/z 100-1000 range) to discriminate the mDa difference at the fragment ion level with, as a consequence, a reduced data acquisition rate in Orbitrap mass analyzers. Based on the DIA cycle time calculator reported by Scheltema et al.¹⁷², using 20 isolation windows on an Orbitrap Fusion mass spectrometer gives a cycle time of 2.2-4.3 s at an MS2 resolution of 30k, while the cycle time increases to around 8 s at a resolution of 120k, the lower end for discriminating neutron-encoded fragment ions. The low scanning rate negatively affects identification and quantification and is by and large not suitable for large-scale comparative proteomics studies.

6 Scope and outline of the thesis

This thesis aims to contribute to the advancement of stable isotope labeling-based quantitative proteomics mainly along two lines of research: (1) In data-dependent acquisition (DDA) mode, eliminating the ratio distortion caused by peptide cofragmentation and increasing the multiplexing capacity. (2) In data-independent acquisition (DIA) mode, achieving multiplexed quantification with general DIA LC-MS settings.

Chapter 1 gives an overview of chemical, stable-isotope-based quantitative proteomics

approaches discussing the principles, benefits and limitations of various tags including tags that were developed as part of this thesis.

In **Chapter 2**, to improve the multiplexing capacity of IPTL while avoiding deuterium-containing tags, I proposed the selective maleylation-directed isobaric peptide termini labeling (SMD-IPTL) method¹⁵⁰ that not only retains all the advantages of the original IPTL approaches but also improves the multiplexing capacity to at least 4-plex with the potential to be extended to 7-plex labeling by using ^{13}C or ^{15}N -labeled cysteine, alanine and acetic anhydride. The SMD-IPTL approach is based on the selective maleylation at the N-termini of LysC peptides. The newly introduced maleyl derivatives can be further modified with different isotopic forms of acetylcysteine, while at the C-termini the complementary isotopically labeled acetylalanine is inserted to balance the overall mass of the modified peptide.

In **Chapter 3**, to reduce the complexity of MS2 spectra of IPTL methods and to further improve multiplexing capacity, I developed a collision-induced dissociation (CID)-cleavable, isobaric acetyl-isoleucine-proline-glycine (Ac-IPG) tag¹⁵⁴. The Ac-IPG tag is based on selective N-terminal dimethylation^{35,84} followed by derivatization of the epsilon-amine group at the C-terminal Lys residue of LysC peptides with isobaric Ac-IPG tags having complementary isotope distributions on the Pro-Gly and Ac-Ile parts. Fragmentation occurs between Ile and Pro¹⁵⁵⁻¹⁵⁹ in addition to fragmentation of the peptide backbone upon CID. While the resulting y-ions can be distinguished between labeling channels based on the distinct isotopes on the Pro-Gly part and thus contain the quantitative information for the respective peptides, b-ions of the different labeling channels have identical m/z values, which allows database searching with commonly used algorithms. The Ac-IPG tag conserves the merits of quantifying peptides based on specific peptide fragment ions while reducing the complexity of MS2 spectra compared to conventional IPTL methods thus facilitating peptide identification. The Ac-IPG tag was used for triplex labeling, but the multiplexing capacity can be extended to 10-plex with ^{13}C - and ^{15}N -labeled acetic anhydride, isoleucine, proline and glycine via the reported synthesis route¹⁵⁴.

In **Chapter 4**, to establish a new peptide-coupled reporter ion-based tag that avoids ratio distortion in DDA and improves the throughput of DIA-based methods without sacrificing data acquisition rate, I proposed a novel isobaric Ac-AG tag, which conceals quantitative information in isobarically labeled peptides and reveals it upon tandem MS in the form of peptide-coupled reporter ions. The key feature of the tag is that the bond between Ac-Ala and Gly fragments prior to fragmentation of the peptide backbone at a low NCE generating an intense peptide-coupled reporter-ion while fragment ions of the peptide backbone are generated at a higher NCE. By combining two NCEs in the same MS/MS scan, the ions required for both identification and quantification are acquired in the same MS2 spectrum. Since there is no isotopic label at the N-terminus and y-ions contain the entire Ac-AG tag, all fragment ions of the peptide backbone originating from different labeling channels have the same respective masses, which means that the complexity of MS2 spectra does not increase with the number of differentially labeled samples. Since Ac-Ala does not have a good ionization site, it is lost as a neutral molecule and the peptide-coupled reporter-ion with the attached Gly part from the Ac-AG tag has the same charge state as the precursor ion^{51,160}. Therefore, the ratios between labeling channels are simply reflected by the intensities of the peptide-coupled reporter ions in DDA mode, whereas the ratios can also be deduced from the peptide-coupled reporter ion isotope envelope in DIA mode after deconvolution. The Ac-AG tag is thus the first isobaric tag that can be applied for multiplexed quantitation in both DDA and DIA mode without having to resort to ultra-high MS2 resolution.

In **Chapter 5**, to further increase the throughput in the DIA mode without reducing data acquisition rate and simplify the synthesis of isotopic tags, I proposed the isotopically labeled Ac-IP (acetyl-isoleucine-proline) tag that enables differentially labeled peptides have distinct precursor ions carrying the quantitative information but identical MS2 spectra. The isotopically labeled Ac-Ile part leaves as a neutral loss upon CID in addition to fragmentation of the peptide backbone generates information for peptide identification. Notably, the multiplexing capacity can be extended to a 10-plex to bring the multiplexing capacity of DIA closer to that of DDA while maintaining the merits of

reproducibility and unbiased precursor ion selection. Labeling steps include selective N-terminal dimethylation followed by derivatization of the epsilon-amino group of the C-terminal Lys residue of LysC peptides with the Ac-IP tags. In the survey scan, the shape of the isotope envelope of the precursor ions changes with different mixing ratios, which means that the quantitative information can be deduced from the distinct shapes.

References

- (1) Aebersold, R.; Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **2016**, *537*, 347-355.
- (2) Larance, M.; Lamond, A. I. Multidimensional proteomics for cell biology. *Nature Reviews Molecular Cell Biology* **2015**, *16*, 269-280.
- (3) Geyer, P. E.; Holdt, L. M.; Teupser, D.; Mann, M. Revisiting biomarker discovery by plasma proteomics. *Mol Syst Biol* **2017**, *13*.
- (4) Anjo, S. I.; Santa, C.; Manadas, B. SWATH-MS as a tool for biomarker discovery: From basic research to clinical applications. *Proteomics* **2017**, *17*.
- (5) Borrebaeck, C. A. K. Precision diagnostics: moving towards protein biomarker signatures of clinical utility in cancer. *Nature Reviews Cancer* **2017**, *17*, 199-204.
- (6) Jiang, Y.; Sun, A. H.; Zhao, Y.; Ying, W. T.; Sun, H. C.; Yang, X. R.; Xing, B. C.; Sun, W.; Ren, L. L.; Hu, B.; Li, C. Y.; Zhang, L.; Qin, G. R.; Zhang, M. H.; Chen, N.; Zhang, M. L.; Huang, Y.; Zhou, J. N.; Zhao, Y.; Liu, M. W., et al. Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature* **2019**, *567*, 257-+.
- (7) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R., 3rd. Protein analysis by shotgun/bottom-up proteomics. *Chem Rev* **2013**, *113*, 2343-2394.
- (8) Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. Top Down Proteomics: Facts and Perspectives. *Biochemical & Biophysical Research Communications* **2014**, *445*, 683-693.
- (9) Toby, T. K.; Fornelli, L.; Kelleher, N. L. In *Annual Review of Analytical Chemistry, Vol 9*, Bohn, P. W.; Pemberton, J. E., Eds.; Annual Reviews: Palo Alto, 2016, pp 499-519.
- (10) Bondarenko, P. V.; Chelius, D.; Shaler, T. A. Identification and Relative Quantitation of Protein Mixtures by Enzymatic Digestion Followed by Capillary Reversed-Phase Liquid Chromatography–Tandem Mass Spectrometry. *Analytical Chemistry* **2002**, *74*, 4741-4749.
- (11) Asara, J. M.; Christofk, H. R.; Freemark, L. M.; Cantley, L. C. A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *Proteomics* **2008**, *8*, 994-999.
- (12) Christin, C.; Bischoff, R.; Horvatovich, P. Data processing pipelines for comprehensive profiling of proteomics samples by label-free LC-MS for biomarker discovery. *Talanta* **2011**, *83*, 1209-1224.
- (13) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology* **1999**, *17*, 994-999.
- (14) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlett-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents. *Molecular and Cellular Proteomics* **2004**, *3*, 1154-1169.
- (15) 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. *Analytical Chemistry* **2003**, *75*, 1895-1904.
- (16) Neilson, K. A.; Ali, N. A.; Muralidharan, S.; Mirzaei, M.; Mariani, M.; Assadourian, G.; Lee, A.; van Sluyter, S. C.; Haynes, P. A. Less label, more free: Approaches in label-free quantitative mass spectrometry. *Proteomics* **2011**, *11*, 535-553.
- (17) Zhu, W.; Smith, J. W.; Huang, C.-M. Mass Spectrometry-Based Label-Free Quantitative Proteomics. *Journal of Biomedicine and Biotechnology* **2010**, *2010*, 840518.
- (18) Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* **2007**, *389*, 1017-1031.
- (19) Suits, F.; Hoekman, B.; Rosenling, T.; Bischoff, R.; Horvatovich, P. Threshold-Avoiding Proteomics Pipeline. *Analytical Chemistry* **2011**, *83*, 7786-7794.
- (20) Röst, H. L.; Sachsenberg, T.; Aiche, S.; Bielow, C.; Weissner, H.; Aicheler, F.; Andreotti, S.; Ehrlich, H.-C.; Gutenbrunner, P.; Kenar, E.; Liang, X.; Nahnsen, S.; Nilse, L.; Pfeuffer, J.; Rosenberger, G.; Rurik, M.; Schmitt, U.; Veit, J.; Walzer, M.; Wojnar, D., et al. OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nature Methods* **2016**, *13*, 741-748.
- (21) Cox, J.; Hein, M. Y.; Lubner, C. A.; Paron, I.; Nagaraj, N.; Mann, M. Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol Cell Proteomics* **2014**, *13*, 2513-2526.
- (22) 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.
- (23) Liu, H. B.; Sadygov, R. G.; Yates, J. R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Analytical Chemistry* **2004**, *76*, 4193-4201.
- (24) Colangelo, C. M.; Williams, K. R. Isotope-coded affinity tags for protein quantification. *Methods in molecular biology (Clifton, N.J.)* **2006**, *328*, 151-158.
- (25) Yi, E. C.; Li, X. J.; Cooke, K.; Lee, H.; Raught, B.; Page, A.; Aneliunas, V.; Hieter, P.; Goodlett, D. R.; Aebersold, R. Increased quantitative proteome coverage with (13)C/(12)C-based, acid-cleavable isotope-coded affinity tag reagent and modified data acquisition scheme. *Proteomics* **2005**, *5*, 380-387.
- (26) Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proceedings of the National Academy of Sciences* **1999**, *96*, 6591-6596.
- (27) Jiang, H.; English, A. M. Quantitative analysis of the yeast proteome by incorporation of isotopically labeled leucine. *Journal of proteome research* **2002**, *1*, 345-350.
- (28) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; 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 and Cellular Proteomics* **2002**, *1*, 376-386.
- (29) Zhu, H.; Pan, S.; Gu, S.; Bradbury, E. M.; Chen, X. Amino acid residue specific stable isotope labeling for quantitative proteomics. *Rapid communications in mass spectrometry* **2002**, *16*, 2115-2123.
- (30) Ow, S. Y.; Cardona, T.; Taton, A.; Magnuson, A.; Lindblad, P.; Stensjö, K.; Wright, P. C. Quantitative Shotgun Proteomics of Enriched Heterocysts from *Nostoc* sp. PCC 7120 Using 8-Plex Isobaric Peptide Tags. *Journal of Proteome Research* **2008**, *7*, 1615-1628.
- (31) Dayon, L.; Hainard, A.; Licker, V.; Turck, N.; Kuhn, K.; Hochstrasser, D. F.; Burkhard, P. R.; Sanchez, J. C. Relative Quantification of Proteins in Human Cerebrospinal Fluids by MS/MS Using 6-Plex Isobaric Tags. *Analytical Chemistry* **2008**, *80*, 2921-2931.
- (32) Koehler, C. J.; Strozynski, M.; Kozielski, F.; Treumann, A.; Thiede, B. Isobaric Peptide Termini Labeling for MS/MS-Based Quantitative Proteomics. *Journal of Proteome Research* **2009**, *8*, 4333-4341.

- (33) Koehler, C. J.; Arntzen, M. O.; Strozynski, M.; Treumann, A.; Thiede, B. Isobaric Peptide Termini Labeling Utilizing Site-Specific N-Terminal Succinylation. *Analytical Chemistry* **2011**, *83*, 4775-4781.
- (34) Nie, A. Y.; Zhang, L.; Yan, G. Q.; Yao, J.; Zhang, Y.; Lu, H. J.; Yang, P. Y.; He, F. C. In Vivo Termini Amino Acid Labeling for Quantitative Proteomics. *Analytical Chemistry* **2011**, *83*, 6026-6033.
- (35) Koehler, C. J.; Arntzen, M. O.; de Souza, G. A.; Thiede, B. An Approach for Triplex-Isobaric Peptide Termini Labeling (Triplex-IPTL). *Analytical Chemistry* **2013**, *85*, 2478-2485.
- (36) Zhang, S.; Chen, L. F.; Shan, Y. C.; Sui, Z. G.; Wu, Q.; Zhang, L. H.; Liang, Z.; Zhang, Y. K. Pseudo isobaric peptide termini labelling for relative proteome quantification by SWATH MS acquisition. *Analyst* **2016**, *141*, 4912-4918.
- (37) Jiang, H. C.; Yin, H. R.; Xie, L. Q.; Zhang, Y.; Zhang, L.; Yang, P. Y.; Lu, H. J. A novel triplex isobaric termini labeling quantitative approach for simultaneously supplying three quantitative sources. *Analytica Chimica Acta* **2018**, *1001*, 70-77.
- (38) Zhou, Y.; Shan, Y.; Zhang, L.; Zhang, Y. Recent advances in stable isotope labeling based techniques for proteome relative quantification. *Journal of Chromatography A* **2014**, *1365*, 1-11.
- (39) Rauniyar, N.; Yates, J. R. Isobaric Labeling-Based Relative Quantification in Shotgun Proteomics. *Journal of Proteome Research* **2014**, *13*, 5293-5309.
- (40) Chahrour, O.; Cobice, D.; Malone, J. Stable isotope labelling methods in mass spectrometry-based quantitative proteomics. *J Pharmaceut Biomed* **2015**, *113*, 2-20.
- (41) Krijgsveld, J.; Ketting, R. F.; Mahmoudi, T.; Johansen, J.; Artal-Sanz, M.; Verrijzer, C. P.; Plasterk, R. H.; Heck, A. J. Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics. *Nat Biotechnol* **2003**, *21*, 927-931.
- (42) Sonnett, M.; Yeung, E.; Wuhr, M. Accurate, Sensitive, and Precise Multiplexed Proteomics Using the Complement Reporter Ion Cluster. *Analytical Chemistry* **2018**, *90*, 5032-5039.
- (43) Koehler, C. J.; Arntzen, M. O.; Thiede, B. The impact of carbon-13 and deuterium on relative quantification of proteins using stable isotope diethyl labeling. *Rapid Communications in Mass Spectrometry* **2015**, *29*, 830-836.
- (44) Zhang, R.; Sioma, C. S.; Thompson, R. A.; Xiong, L.; Regnier, F. E. Controlling Deuterium Isotope Effects in Comparative Proteomics. *Analytical Chemistry* **2002**, *74*, 3662-3669.
- (45) Di Palma, S.; Raijmakers, R.; Heck, A. J. R.; Mohammed, S. Evaluation of the Deuterium Isotope Effect in Zwitterionic Hydrophilic Interaction Liquid Chromatography Separations for Implementation in a Quantitative Proteomic Approach. *Analytical Chemistry* **2011**, *83*, 8352-8356.
- (46) Boutilier, J. M.; Warden, H.; Doucette, A. A.; Wentzell, P. D. Chromatographic behaviour of peptides following dimethylation with H₂/D₂-formaldehyde: Implications for comparative proteomics. *Journal of Chromatography B* **2012**, *908*, 59-66.
- (47) Turowski, M.; Yamakawa, N.; Meller, J.; Kimata, K.; Ikegami, T.; Hosoya, K.; Tanaka, N.; Thornton, E. R. Deuterium Isotope Effects on Hydrophobic Interactions: The Importance of Dispersion Interactions in the Hydrophobic Phase. *Journal of the American Chemical Society* **2003**, *125*, 13836-13849.
- (48) Zhang, R.; Sioma, C. S.; Wang, S.; Regnier, F. E. Fractionation of Isotopically Labeled Peptides in Quantitative Proteomics. *Analytical Chemistry* **2001**, *73*, 5142-5149.
- (49) Kang, U. B.; Yeom, J.; Kim, H.; Lee, C. Quantitative Analysis of mTRAQ-Labeled Proteome Using Full MS Scans. *Journal of Proteome Research* **2010**, *9*, 3750-3758.
- (50) Mertins, P.; Udeshi, N. D.; Clauser, K. R.; Mani, D.; Patel, J.; Ong, S. e.; Jaffe, J. D.; Carr, S. A. iTRAQ Labeling is Superior to mTRAQ for Quantitative Global Proteomics and Phosphoproteomics. *Molecular and Cellular Proteomics* **2012**, *11*, 1-12.
- (51) Winter, S. V.; Meier, F.; Wichmann, C.; Cox, J.; Mann, M.; Meissner, F. EASI-tag enables accurate multiplexed and interference-free MS₂-based proteome quantification. *Nature Methods* **2018**, *15*, 527-530.
- (52) Liu, J. H.; Zhou, Y.; Shan, Y. C.; Zhao, B. F.; Hu, Y. C.; Sui, Z. G.; Liang, Z.; Zhang, L. H.; Zhang, Y. K. A Multiplex Fragment-Ion-Based Method for Accurate Proteome Quantification. *Analytical Chemistry* **2019**, *91*, 3921-3928.
- (53) Yao, X. D.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. Proteolytic O-18 labeling for comparative proteomics: Model studies with two serotypes of adenovirus. *Analytical Chemistry* **2001**, *73*, 2836-2842.
- (54) Sakai, J.; Kojima, S.; Yanagi, K.; Kanaoka, M. 18O-labeling quantitative proteomics using an ion trap mass spectrometer. *Proteomics* **2005**, *5*, 16-23.
- (55) Stahl, D. C.; Swiderik, K. M.; Davis, M. T.; Lee, T. D. Data-controlled automation of liquid chromatography/tandem mass spectrometry analysis of peptide mixtures. *Journal of the American Society for Mass Spectrometry* **1996**, *7*, 532-540.
- (56) Hu, A.; Noble, W.; Wolf-Yadlin, A. Technical advances in proteomics: new developments in data-independent acquisition [version 1; peer review: 3 approved]. *F1000Research* **2016**, *5*.
- (57) Venable, J. D.; Dong, M. Q.; Wohlschlegel, J.; Dillin, A.; Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nature Methods* **2004**, *1*, 39-45.
- (58) Pelletier, A. R.; Chung, Y.-E.; Ning, Z.; Wong, N.; Figeys, D.; Lavallée-Adam, M. MealTime-MS: A Machine Learning-Guided Real-Time Mass Spectrometry Analysis for Protein Identification and Efficient Dynamic Exclusion. *Journal of the American Society for Mass Spectrometry* **2020**, *31*, 1459-1472.
- (59) Schweppe, D. K.; Eng, J. K.; Yu, Q.; Bailey, D.; Rad, R.; Navarrete-Perea, J.; Huttlin, E. L.; Erickson, B. K.; Paulo, J. A.; Gygi, S. P. Full-Featured, Real-Time Database Searching Platform Enables Fast and Accurate Multiplexed Quantitative Proteomics. *Journal of Proteome Research* **2020**, *19*, 2026-2034.
- (60) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Molecular and Cellular Proteomics* **2012**, *11*, O111.016717.
- (61) Bern, M.; Finney, G.; Hoopmann, M. R.; Merrihew, G.; Toth, M. J.; MacCoss, M. J. Deconvolution of Mixture Spectra from Ion-Trap Data-Independent-Acquisition Tandem Mass Spectrometry. *Analytical Chemistry* **2010**, *82*, 833-841.
- (62) Collins, B. C.; Gillet, L. C.; Rosenberger, G.; Röst, H. L.; Vichalkovski, A.; Gstaiger, M.; Aebersold, R. Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nature Methods* **2013**, *10*, 1246-1253.
- (63) Rardin, M. J.; Schilling, B.; Cheng, L. Y.; MacLean, B. X.; Sorensen, D. J.; Sahu, A. K.; MacCoss, M. J.; Vitek, O.; Gibson, B. W. MS1 Peptide Ion Intensity Chromatograms in MS2 (SWATH) Data Independent Acquisitions. Improving Post Acquisition Analysis of Proteomic Experiments. *Mol Cell Proteomics* **2015**, *14*, 2405-2419.

- (64) Huang, T.; Bruderer, R.; Muntel, J.; Xuan, Y.; Vitek, O.; Reiter, L. Combining Precursor and Fragment Information for Improved Detection of Differential Abundance in Data Independent Acquisition. *2020*, *19*, 421-430.
- (65) Hsu, J. L.; Huang, S. Y.; Chow, N. H.; Chen, S. H. Stable-isotope dimethyl labeling for quantitative proteomics. *Analytical Chemistry* **2003**, *75*, 6843-6852.
- (66) Schmidt, A. A novel strategy for quantitative proteomics using isotope-coded protein labels (vol 5, pg 4, 2005). *Proteomics* **2005**, *5*, 826-826.
- (67) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **2003**, *75*, 4646-4658.
- (68) Boersema, P. J.; Aye, T. T.; van Veen, T. A. B.; Heck, A. J. R.; Mohammed, S. Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. *Proteomics* **2008**, *8*, 4624-4632.
- (69) Choi, Y.; Jeong, K.; Shin, S.; Lee, J. W.; Lee, Y.-s.; Kim, S.; Kim, S. A.; Jung, J.; Kim, K. P.; Kim, V. N.; Kim, J.-S. MS1-Level Proteome Quantification Platform Allowing Maximally Increased Multiplexity for SILAC and In Vitro Chemical Labeling. *Analytical Chemistry* **2020**, *92*, 4980-4989.
- (70) Paulo, J. A.; Gygi, S. P. mTMT: An Alternative, Nonisobaric, Tandem Mass Tag Allowing for Precursor-Based Quantification. *Analytical Chemistry* **2019**, *91*, 12167-12172.
- (71) Miyagi, M.; Rao, K. C. Proteolytic ¹⁸O-labeling strategies for quantitative proteomics. *Mass Spectrom Rev* **2007**, *26*, 121-136.
- (72) Zhao, Y.; Jia, W.; Sun, W.; Jin, W.; Guo, L.; Wei, J.; Ying, W.; Zhang, Y.; Xie, Y.; Jiang, Y.; He, F.; Qian, X. Combination of Improved ¹⁸O Incorporation and Multiple Reaction Monitoring: A Universal Strategy for Absolute Quantitative Verification of Serum Candidate Biomarkers of Liver Cancer. *Journal of Proteome Research* **2010**, *9*, 3319-3327.
- (73) Liu, J.; Shan, Y.; Zhou, Y.; Liang, Z.; Zhang, L.; Zhang, Y. Advances and applications of stable isotope labeling-based methods for proteome relative quantitation. *TrAC Trends in Analytical Chemistry* **2020**, *124*, 115815.
- (74) Han, D. K.; Eng, J.; Zhou, H. L.; Aebersold, R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nature Biotechnology* **2001**, *19*, 946-951.
- (75) Smolka, M. B.; Zhou, H. L.; Purkayastha, S.; Aebersold, R. Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis. *Analytical Biochemistry* **2001**, *297*, 25-31.
- (76) Yi, E. C.; Li, X.-j.; Cooke, K.; Lee, H.; Raught, B.; Page, A.; Aneliunas, V.; Hieter, P.; Goodlett, D. R.; Aebersold, R. Increased quantitative proteome coverage with ¹³C/¹²C-based, acid-cleavable isotope-coded affinity tag reagent and modified data acquisition scheme. *Proteomics* **2005**, *5*, 380-387.
- (77) Hansen, K. C.; Schmitt-Ulms, G.; Chalkley, R. J.; Hirsch, J.; Baldwin, M. A.; Burlingame, A. L. Mass Spectrometric Analysis of Protein Mixtures at Low Levels Using Cleavable ¹³C-Isotope-coded Affinity Tag and Multidimensional Chromatography. **2003**, *2*, 299-314.
- (78) Li, J. X.; Steen, H.; Gygi, S. P. Protein profiling with cleavable isotope-coded affinity tag (cICAT) reagents - The yeast salinity stress response. *Mol Cell Proteomics* **2003**, *2*, 1198-1204.
- (79) Oda, Y.; Owa, T.; Sato, T.; Boucher, B.; Daniels, S.; Yamanaka, H.; Shinohara, Y.; Yokoi, A.;

- Kuromitsu, J.; Nagasu, T. Quantitative chemical proteomics for identifying candidate drug targets. *Analytical Chemistry* **2003**, *75*, 2159-2165.
- (80) Parker, K. C.; Williamson, B.; Marchese, J.; Graber, A.; He, F.; Jacobson, A.; Juhasz, P.; Martin, S. Depth of proteome issues - A yeast isotope-coded affinity tag reagent study. *Mol Cell Proteomics* **2004**, *3*, 625-659.
- (81) Yu, L. R.; Conrads, T. P.; Uo, T.; Issaq, H. J.; Morrison, R. S.; Veenstra, T. D. Evaluation of the acid-cleavable isotope-coded affinity tag reagents: Application to camptothecin-treated cortical neurons. *Journal of Proteome Research* **2004**, *3*, 469-477.
- (82) Boersema, P. J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J. R. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nature Protocols* **2009**, *4*, 484-494.
- (83) Fu, Q.; Li, L. De Novo Sequencing of Neuropeptides Using Reductive Isotopic Methylation and Investigation of ESI QTOF MS/MS Fragmentation Pattern of Neuropeptides with N-Terminal Dimethylation. *Analytical Chemistry* **2005**, *77*, 7783-7795.
- (84) Qin, H.; Wang, F.; Zhang, Y.; Hu, Z.; Song, C.; Wu, R.; Ye, M.; Zou, H. Isobaric cross-sequence labeling of peptides by using site-selective N-terminus dimethylation. *Chemical Communications* **2012**, *48*, 6265-6267.
- (85) Jung, J.; Jeong, K.; Choi, Y.; Kim, S. A.; Kim, H.; Lee, J. W.; Kim, V. N.; Kim, K. P.; Kim, J.-S. Deuterium-Free, Three-Plexed Peptide Diethylation for Highly Accurate Quantitative Proteomics. *Journal of Proteome Research* **2019**, *18*, 1078-1087.
- (86) Com, E.; Clavreul, A.; Lagarrigue, M.; Michalak, S.; Menei, P.; Pineau, C. Quantitative proteomic Isotope-Coded Protein Label (ICPL) analysis reveals alteration of several functional processes in the glioblastoma. *Journal of Proteomics* **2012**, *75*, 3898-3913.
- (87) Yu, F.-j.; Lam, T. c.; Liu, L.-q.; Chun, R. K.-m.; Cheung, J. K.-w.; Li, K.-k.; To, C.-h. Isotope-coded protein label based quantitative proteomic analysis reveals significant up-regulation of apolipoprotein A1 and ovotransferrin in the myopic chick vitreous. *Sci Rep-Uk* **2017**, *7*, 12649.
- (88) Leroy, B.; Rosier, C.; Erculisse, V.; Leys, N.; Mergeay, M.; Wattiez, R. Differential proteomic analysis using isotope-coded protein-labeling strategies: Comparison, improvements and application to simulated microgravity effect on *Cupriavidus metallidurans* CH34. *Proteomics* **2010**, *10*, 2281-2291.
- (89) Hebert, A. S.; Merrill, A. E.; Bailey, D. J.; Still, A. J.; Westphall, M. S.; Strieter, E. R.; Pagliarini, D. J.; Coon, J. J. Neutron-encoded mass signatures for multiplexed proteome quantification. *Nature Methods* **2013**, *10*, 332-334.
- (90) McAlister, G. C.; Huttlin, E. L.; Haas, W.; Ting, L.; Jedrychowski, M. P.; Rogers, J. C.; Kuhn, K.; Pike, I.; Grothe, R. A.; Blethrow, J. D.; Gygi, S. P. Increasing the Multiplexing Capacity of TMTs Using Reporter Ion Isotopologues with Isobaric Masses. *Analytical Chemistry* **2012**, *84*, 7469-7478.
- (91) Werner, T.; Becher, I.; Sweetman, G.; Doce, C.; Savitski, M. M.; Bantscheff, M. High-Resolution Enabled TMT 8-plexing. *Analytical Chemistry* **2012**, *84*, 7188-7194.
- (92) Merrill, A. E.; Hebert, A. S.; MacGilvray, M. E.; Rose, C. M.; Bailey, D. J.; Bradley, J. C.; Wood, W. W.; El Masri, M.; Westphall, M. S.; Gasch, A. P.; Coon, J. J. NeuCode Labels for Relative Protein Quantification. **2014**, *13*, 2503-2512.
- (93) Overmyer, K. A.; Tyanova, S.; Hebert, A. S.; Westphall, M. S.; Cox, J.; Coon, J. J. Multiplexed proteome analysis with neutron-encoded stable isotope labeling in cells and mice. *Nature Protocols* **2018**, *13*, 293-306.
- (94) Hebert, A. S.; Merrill, A. E.; Stefely, J. A.; Bailey, D. J.; Wenger, C. D.; Westphall, M. S.;

Pagliarini, D. J.; Coon, J. J. Amine-reactive Neutron-encoded Labels for Highly Plexed Proteomic Quantitation. *Mol Cell Proteomics* **2013**, *12*, 3360-3369.

(95) Ulbrich, A.; Merrill, A. E.; Hebert, A. S.; Westphall, M. S.; Keller, M. P.; Attie, A. D.; Coon, J. J. Neutron-Encoded Protein Quantification by Peptide Carbamylation. *Journal of the American Society for Mass Spectrometry* **2014**, *25*, 6-9.

(96) Hao, L.; Johnson, J.; Lietz, C. B.; Buchberger, A.; Frost, D.; Kao, W. J.; Li, L. Mass Defect-Based N,N-Dimethyl Leucine Labels for Quantitative Proteomics and Amine Metabolomics of Pancreatic Cancer Cells. *Analytical Chemistry* **2017**, *89*, 1138-1146.

(97) Zhong, X. F.; Frost, D. C.; Li, L. J. High-Resolution Enabled 5-plex Mass Defect-Based N,N-Dimethyl Leucine Tags for Quantitative Proteomics. *Analytical Chemistry* **2019**, *91*, 7991-7995.

(98) Frost, D. C.; Buchberger, A. R.; Li, L. J. Mass Defect-Based Dimethyl Pyrimidinyl Ornithine (DiPyrO) Tags for Multiplex Quantitative Proteomics. *Analytical Chemistry* **2017**, *89*, 10798-10805.

(99) Ow, S. Y.; Salim, M.; Noirel, J.; Evans, C.; Rehman, I.; Wright, P. C. iTRAQ Underestimation in Simple and Complex Mixtures: "The Good, the Bad and the Ugly". *Journal of Proteome Research* **2009**, *8*, 5347-5355.

(100) 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. *Analytical Chemistry* **2010**, *82*, 6549-6558.

(101) Pappireddi, N.; Martin, L.; Wühr, M. A Review on Quantitative Multiplexed Proteomics. **2019**, *20*, 1210-1224.

(102) Zhang, G. A.; Neubert, T. A. Automated comparative proteomics based on multiplex tandem mass spectrometry and stable isotope labeling. *Mol Cell Proteomics* **2006**, *5*, 401-411.

(103) Wühr, M.; Haas, W.; McAlister, G. C.; Peshkin, L.; Rad, R.; Kirschner, M. W.; Gygi, S. P. Accurate Multiplexed Proteomics at the MS2 Level Using the Complement Reporter Ion Cluster. *Analytical Chemistry* **2012**, *84*, 9214-9221.

(104) Arul, A. B.; Robinson, R. A. S. Sample Multiplexing Strategies in Quantitative Proteomics. *Analytical Chemistry* **2019**, *91*, 178-189.

(105) Thompson, A.; Wölmer, N.; Koncarevic, S.; Selzer, S.; Böhm, G.; Legner, H.; Schmid, P.; Kienle, S.; Penning, P.; Höhle, C.; Berfelde, A.; Martinez-Pinna, R.; Farztdinov, V.; Jung, S.; Kuhn, K.; Pike, I. TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-Based Isobaric 16-Plex Tandem Mass Tag Reagent Set. *Analytical Chemistry* **2019**, *91*, 15941-15950.

(106) Li, J. M.; Van Vranken, J. G.; Vaite, L. P.; Schweppe, D. K.; Huttlin, E. L.; Etienne, C.; Nandhikonda, P.; Viner, R.; Robitaille, A. M.; Thompson, A. H.; Kuhn, K.; Pike, I.; Bomgarden, R. D.; Rogers, J. C.; Gygi, S. P.; Paulo, J. A. TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples. *Nature Methods* **2020**, *17*, 399-404.

(107) Choe, L.; D'Ascenzo, M.; Relkin, N. R.; Pappin, D.; Ross, P.; Williamson, B.; Guertin, S.; Pribil, P.; Lee, K. H. 8-Plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. **2007**, *7*, 3651-3660.

(108) Aggarwal, S.; Yadav, A. K. Dissecting the iTRAQ Data Analysis. *Methods Mol Biol* **2016**, *1362*, 277-291.

(109) Frost, D. C.; Greer, T.; Li, L. High-Resolution Enabled 12-Plex DiLeu Isobaric Tags for

Quantitative Proteomics. *Analytical Chemistry* **2015**, *87*, 1646-1654.

(110) Dayon, L.; Affolter, M. Progress and pitfalls of using isobaric mass tags for proteome profiling. *Expert Review of Proteomics* **2020**, *17*, 149-161.

(111) Zeng, D. X.; Li, S. W. Revival of deuterium-labeled reagents for protein quantitation. *Chemical Communications* **2009**, 3369-3371.

(112) Xiang, F.; Ye, H.; Chen, R. B.; Fu, Q.; Li, L. J. N,N-Dimethyl Leucines as Novel Isobaric Tandem Mass Tags for Quantitative Proteomics and Peptidomics. *Analytical Chemistry* **2010**, *82*, 2817-2825.

(113) Frost, D. C.; Feng, Y.; Li, L. 21-plex DiLeu Isobaric Tags for High-Throughput Quantitative Proteomics. *Analytical Chemistry* **2020**, *92*, 8228-8234.

(114) Zhang, J.; Wang, Y.; Li, S. Deuterium Isobaric Amine-Reactive Tags for Quantitative Proteomics. *Analytical Chemistry* **2010**, *82*, 7588-7595.

(115) Ren, Y.; He, Y.; Lin, Z.; Zi, J.; Yang, H.; Zhang, S.; Lou, X.; Wang, Q.; Li, S.; Liu, S. Reagents for Isobaric Labeling Peptides in Quantitative Proteomics. *Analytical Chemistry* **2018**, *90*, 12366-12371.

(116) Hsu, J.-L.; Huang, S.-Y.; Shiea, J.-T.; Huang, W.-Y.; Chen, S.-H. Beyond Quantitative Proteomics: Signal Enhancement of the a1 Ion as a Mass Tag for Peptide Sequencing Using Dimethyl Labeling. *Journal of Proteome Research* **2005**, *4*, 101-108.

(117) Colzani, M.; Schutz, F.; Potts, A.; Waridel, P.; Quadroni, M. Relative protein quantification by isobaric SILAC with immonium ion splitting (ISIS). *Mol Cell Proteomics* **2008**, *7*, 927-937.

(118) Frost, D. C.; Greer, T.; Xiang, F.; Liang, Z.; Li, L. Development and characterization of novel 8-plex DiLeu isobaric labels for quantitative proteomics and peptidomics. *Rapid Communications in Mass Spectrometry* **2015**, *29*, 1115-1124.

(119) Yu, Q.; Shi, X.; Greer, T.; Lietz, C. B.; Kent, K. C.; Li, L. Evaluation and Application of Dimethylated Amino Acids as Isobaric Tags for Quantitative Proteomics of the TGF- β /Smad3 Signaling Pathway. *Journal of Proteome Research* **2016**, *15*, 3420-3431.

(120) Braun, C. R.; Bird, G. H.; Wühr, M.; Erickson, B. K.; Rad, R.; Walensky, L. D.; Gygi, S. P.; Haas, W. Generation of Multiple Reporter Ions from a Single Isobaric Reagent Increases Multiplexing Capacity for Quantitative Proteomics. *Analytical Chemistry* **2015**, *87*, 9855-9863.

(121) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and localized protons: a framework for understanding peptide dissociation. *J Mass Spectrom* **2000**, *35*, 1399-1406.

(122) Li, S.; Zeng, D. CILAT – a new reagent for quantitative proteomics. *Chemical Communications* **2007**, 2181-2183.

(123) Murray, C. I.; Uhrigshardt, H.; O'Meally, R. N.; Cole, R. N.; Van Eyk, J. E. Identification and Quantification of S-Nitrosylation by Cysteine Reactive Tandem Mass Tag Switch Assay. *Mol Cell Proteomics* **2012**, *11*.

(124) Bächor, R.; Waliczek, M.; Stefanowicz, P.; Szewczuk, Z. Trends in the Design of New Isobaric Labeling Reagents for Quantitative Proteomics. **2019**, *24*, 701.

(125) Hahne, H.; Neubert, P.; Kuhn, K.; Etienne, C.; Bomgarden, R.; Rogers, J. C.; Kuster, B. Carbonyl-Reactive Tandem Mass Tags for the Proteome-Wide Quantification of N-Linked Glycans. *Analytical Chemistry* **2012**, *84*, 3716-3724.

(126) Palmese, A.; De Rosa, C.; Chiappetta, G.; Marino, G.; Amoresano, A. Novel method to investigate protein carbonylation by iTRAQ strategy. *Analytical and Bioanalytical Chemistry* **2012**, *404*, 1631-1635.

- (127) Ting, L.; Rad, R.; Gygi, S. P.; Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nature Methods* **2011**, *8*, 937-940.
- (128) Wenger, C. D.; Lee, M. V.; Hebert, A. S.; McAlister, G. C.; Phanstiel, D. H.; Westphall, M. S.; Coon, J. J. Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nature Methods* **2011**, *8*, 933-935.
- (129) Couderc, F.; Berjeaud, J. M.; Promé, J. C.; Cooks, R. G.; Rockwood, A. L. Letters to the editor. *Rapid Communications in Mass Spectrometry* **1991**, *5*, 92-93.
- (130) Ow, S. Y.; Salim, M.; Noirel, J.; Evans, C.; Wright, P. C. Minimising iTRAQ ratio compression through understanding LC-MS elution dependence and high-resolution HILIC fractionation. *Proteomics* **2011**, *11*, 2341-2346.
- (131) Savitski, M. M.; Sweetman, G.; Askenazi, M.; Marto, J. A.; 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. *Analytical Chemistry* **2011**, *83*, 8959-8967.
- (132) McAlister, G. C.; Nusinow, D. P.; Jedrychowski, M. P.; Wuhr, M.; Huttlin, E. L.; Erickson, B. K.; Rad, R.; Haas, W.; Gygi, S. P. MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes. *Analytical Chemistry* **2014**, *86*, 7150-7158.
- (133) Dayon, L.; Sonderegger, B.; Kussmann, M. Combination of Gas-Phase Fractionation and MS3 Acquisition Modes for Relative Protein Quantification with Isobaric Tagging. *Journal of Proteome Research* **2012**, *11*, 5081-5089.
- (134) 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*, 4653-4665.
- (135) Pfammatter, S.; Bonneil, E.; McManus, F. P.; Prasad, S.; Bailey, D. J.; 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. *Mol Cell Proteomics* **2018**, *17*, 2051-2067.
- (136) Schweppe, D. K.; Prasad, S.; Belford, M. W.; Navarrete-Perea, J.; Bailey, D. J.; Huguet, R.; Jedrychowski, M. P.; Rad, R.; McAlister, G.; Abbatiello, S. E.; Wouters, E. R.; Zabrouskov, V.; Dunyach, J. J.; Paulo, J. A.; Gygi, S. P. Characterization and Optimization of Multiplexed Quantitative Analyses Using High-Field Asymmetric-Waveform Ion Mobility Mass Spectrometry. *Analytical Chemistry* **2019**, *91*, 4010-4016.
- (137) Iwasaki, M.; Tabata, T.; Kawahara, Y.; Ishihama, Y.; Nakagawa, M. Removal of Interference MS/MS Spectra for Accurate Quantification in Isobaric Tag-Based Proteomics. *Journal of Proteome Research* **2019**, *18*, 2535-2544.
- (138) Li, H.; Hwang, K. B.; Mun, D. G.; Kim, H.; Lee, H.; Lee, S. W.; Paek, E. Estimating Influence of Cofragmentation on Peptide Quantification and Identification in iTRAQ Experiments by Simulating Multiplexed Spectra. *Journal of Proteome Research* **2014**, *13*, 3488-3497.
- (139) Barnett, D. A.; Ells, B.; Guevremont, R.; Purves, R. W. Application of ESI-FAIMS-MS to the analysis of tryptic peptides. *Journal of the American Society for Mass Spectrometry* **2002**, *13*, 1282-1291.
- (140) Hebert, A. S.; Prasad, S.; Belford, M. W.; Bailey, D. J.; McAlister, G. C.; Abbatiello, S. E.; Huguet, R.; Wouters, E. R.; Dunyach, J.-J.; Brademan, D. R.; Westphall, M. S.; Coon, J. J.

- Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer. *Analytical Chemistry* **2018**, *90*, 9529-9537.
- (141) Bekker-Jensen, D. B.; Martínez-Val, A.; Steigerwald, S.; Rüther, P.; Fort, K. L.; Arrey, T. N.; Harder, A.; Makarov, A.; Olsen, J. V. A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients*. *Mol Cell Proteomics* **2020**, *19*, 716-729.
- (142) Swearingen, K. E.; Moritz, R. L. High-field asymmetric waveform ion mobility spectrometry for mass spectrometry-based proteomics. *Expert Review of Proteomics* **2012**, *9*, 505-517.
- (143) Bonneil, E.; Pfammatter, S.; Thibault, P. Enhancement of mass spectrometry performance for proteomic analyses using highfield asymmetric waveform ion mobility spectrometry (FAIMS). *J Mass Spectrom* **2015**, *50*, ii-ii.
- (144) Xie, L. Q.; Nie, A. Y.; Yang, S. J.; Zhao, C.; Zhang, L.; Yang, P. Y.; Lu, H. J. Global in vivo terminal amino acid labeling for exploring differential expressed proteins induced by dialyzed serum cultivation. *Analyst* **2014**, *139*, 4497-4504.
- (145) Yang, S. J.; Nie, A. Y.; Zhang, L.; Yan, G. Q.; Yao, J.; Xie, L. Q.; Lu, H. J.; Yang, P. Y. A novel quantitative proteomics workflow by isobaric terminal labeling. *Journal of Proteomics* **2012**, *75*, 5797-5806.
- (146) Waldbauer, J.; Zhang, L.; Rizzo, A.; Muratore, D. diDO-IPTL: A Peptide-Labeling Strategy for Precision Quantitative Proteomics. *Analytical Chemistry* **2017**, *89*, 11498-11504.
- (147) Cao, T.; Zhang, L.; Zhang, Y.; Yan, G. Q.; Fang, C. Y.; Bao, H. M.; Lu, H. J. Site-Specific Quantification of Protein Ubiquitination on MS2 Fragment Ion Level via Isobaric Peptide Labeling. *Analytical Chemistry* **2017**, *89*, 11468-11475.
- (148) Zhou, Y.; Shan, Y. C.; Wu, Q.; Zhang, S.; Zhang, L. H.; Zhang, Y. K. Mass Defect-Based Pseudo-Isobaric Dimethyl Labeling for Proteome Quantification. *Analytical Chemistry* **2013**, *85*, 10658-10663.
- (149) Zhang, S.; Wu, Q.; Shan, Y. C.; Zhou, Y.; Zhang, L. H.; Zhang, Y. K. Partially isobaric peptide termini labeling assisted proteome quantitation based on MS and MS/MS signals. *Journal of Proteomics* **2015**, *114*, 152-160.
- (150) Tian, X.; de Vries, M. P.; Visscher, S. W. J.; Permentier, H. P.; Bischoff, R. Selective Maleylation-Directed Isobaric Peptide Termini Labeling for Accurate Proteome Quantification. *Analytical Chemistry* **2020**, *92*, 7836-7844.
- (151) Arntzen, M. Ø.; Koehler, C. J.; Barsnes, H.; Berven, F. S.; Treumann, A.; Thiede, B. IsobariQ: Software for Isobaric Quantitative Proteomics using IPTL, iTRAQ, and TMT. *Journal of Proteome Research* **2011**, *10*, 913-920.
- (152) Xie, L.-Q.; Zhang, L.; Nie, A.-Y.; Yan, G.-Q.; Yao, J.; Zhang, Y.; Yang, P.-Y.; Lu, H.-J. ITMSQ: A software tool for N- and C-terminal fragment ion pairs based isobaric tandem mass spectrometry quantification. *Proteomics* **2015**, *15*, 3755-3764.
- (153) Zhang, S.; Wu, Q.; Shan, Y. C.; Sui, Z. G.; Zhang, L. H.; Zhang, Y. K. A paired ions scoring algorithm based on Morpheus for simultaneous identification and quantification of proteome samples prepared by isobaric peptide termini labeling strategies. *Proteomics* **2015**, *15*, 1781-1788.
- (154) Tian, X.; de Vries, M. P.; Permentier, H. P.; Bischoff, R. A Collision-Induced Dissociation Cleavable Isobaric Tag for Peptide Fragment Ion-Based Quantification in Proteomics. *Journal of Proteome Research* **2020**, *19*, 3817-3824.
- (155) Dongre, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H. Influence of peptide composition,

gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model. *Journal of the American Chemical Society* **1996**, *118*, 8365-8374.

(156) Mák, M.; Mezö, G.; Skribanek, Z.; Hudecz, F. Stability of Asp-Pro bond under high and low energy collision induced dissociation conditions in the immunodominant epitope region of Herpes simplex virion glycoprotein D. *Rapid Communications in Mass Spectrometry* **1998**, *12*, 837-842.

(157) Hogan, J. M.; McLuckey, S. A. Charge state dependent collision-induced dissociation of native and reduced porcine elastase. *J Mass Spectrom* **2003**, *38*, 245-256.

(158) Huang, Y.; Triscari, J. M.; Pasa-Tolic, L.; Anderson, G. A.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. Dissociation Behavior of Doubly-Charged Tryptic Peptides: Correlation of Gas-Phase Cleavage Abundance with Ramachandran Plots. *Journal of the American Chemical Society* **2004**, *126*, 3034-3035.

(159) Tiwary, S.; Levy, R.; Gutenbrunner, P.; Salinas Soto, F.; Palaniappan, K. K.; Deming, L.; Berndt, M.; Brant, A.; Cimermanic, P.; Cox, J. High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis. *Nature Methods* **2019**, *16*, 519-525.

(160) Tian, X.; de Vries, M. P.; Permentier, H. P.; Bischoff, R. A Versatile Isobaric Tag Enables Proteome Quantification in Data-Dependent and Data-Independent Acquisition Modes. *Analytical Chemistry* **2020**, *92*, 16149-16157.

(161) Kao, A. H.; Chiu, C. L.; Vellucci, D.; Yang, Y. Y.; Patel, V. R.; Guan, S. H.; Randall, A.; Baldi, P.; Rychnovsky, S. D.; Huang, L. Development of a Novel Cross-linking Strategy for Fast and Accurate Identification of Cross-linked Peptides of Protein Complexes. *Mol Cell Proteomics* **2011**, *10*.

(162) Stadlmeier, M.; Bogen, J.; Wallner, M.; Wuhr, M.; Carell, T. A Sulfoxide-Based Isobaric Labelling Reagent for Accurate Quantitative Mass Spectrometry. *Angew Chem Int Edit* **2018**, *57*, 2958-2962.

(163) Evans, A. R.; Robinson, R. A. S. Global combined precursor isotopic labeling and isobaric tagging (cPILOT) approach with selective MS3 acquisition. *Proteomics* **2013**, *13*, 3267-3272.

(164) Dephoure, N.; Gygi, S. P. Hyperplexing: A Method for Higher-Order Multiplexed Quantitative Proteomics Provides a Map of the Dynamic Response to Rapamycin in Yeast. *Sci Signal* **2012**, *5*.

(165) Frost, D. C.; Rust, C. J.; Robinson, R. A. S.; Li, L. J. Increased N,N-Dimethyl Leucine Isobaric Tag Multiplexing by a Combined Precursor Isotopic Labeling and Isobaric Tagging Approach. *Analytical Chemistry* **2018**, *90*, 10664-10669.

(166) Aggarwal, S.; Talukdar, N. C.; Yadav, A. K. Advances in Higher Order Multiplexing Techniques in Proteomics. *Journal of Proteome Research* **2019**, *18*, 2360-2369.

(167) Pierce, A.; Unwin, R. D.; Evans, C. A.; Griffiths, S.; Carney, L.; Zhang, L.; Jaworska, E.; Lee, C.-F.; Blinco, D.; Okoniewski, M. J.; Miller, C. J.; Bitton, D. A.; Spooner, E.; Whetton, A. D. Eight-channel iTRAQ Enables Comparison of the Activity of Six Leukemogenic Tyrosine Kinases. **2008**, *7*, 853-863.

(168) Brenes, A.; Hukelmann, E.; Bensaddek, D.; Lamond, A. I. Multibatch TMT Reveals False Positives, Batch Effects and Missing Values. *Molecular and Cellular Proteomics* **2019**, *18*, 1967-1980.

(169) Minogue, C. E.; Hebert, A. S.; Rensvold, J. W.; Westphall, M. S.; Pagliarini, D. J.; Coon, J. J. Multiplexed Quantification for Data-Independent Acquisition. *Analytical Chemistry* **2015**, *87*, 2570-2575.

(170) Di, Y.; Zhang, Y.; Zhang, L.; Tao, T.; Lu, H. MdFDIA: A Mass Defect Based Four-Plex Data-

Independent Acquisition Strategy for Proteome Quantification. *Analytical Chemistry* **2017**, *89*, 10248-10255.

(171) Zhong, X.; Frost, D. C.; Yu, Q.; Li, M.; Gu, T.-J.; Li, L. Mass Defect-Based DiLeu Tagging for Multiplexed Data-Independent Acquisition. *Analytical Chemistry* **2020**, *92*, 11119-11126.

(172) Scheltema, R. A.; Hauschild, J.-P.; Lange, O.; Hornburg, D.; Denisov, E.; Damoc, E.; Kuehn, A.; Makarov, A.; Mann, M. The Q Exactive HF, a Benchtop Mass Spectrometer with a Pre-filter, High Performance Quadrupole and an Ultra-High Field Orbitrap Analyzer. *Mol Cell Proteomics* **2014**, mcp.M114.043489.

Chapter 2

Selective maleylation-directed isobaric peptide termini labeling for accurate proteome quantification

Xiaobo Tian,[†] Marcel P. de Vries,[‡] Susan W.J. Visscher,[†] Hjalmar P. Permentier,[†]
and Rainer Bischoff[†]

[†]*Department of Analytical Biochemistry and Interfaculty Mass Spectrometry Center, Groningen
Research Institute of Pharmacy, University of Groningen, A. Deusinglaan 1, 9713 AV,
Groningen, The Netherlands*

[‡]*Department of Pediatrics, University Medical Center Groningen, University of Groningen,
Hanzeplein 1, 9713 GZ, Groningen, The Netherlands*

Manuscript published in *Anal. Chem.* 2020, 92, 11, 7836–7844

ABSTRACT

Isobaric peptide termini labeling (IPTL) is an attractive protein quantification method, because it provides more accurate and reliable quantification information than traditional isobaric labeling methods (TMT, iTRAQ) by making use of the entire fragment ion series instead of only a single reporter ion. The multiplexing capacity of published IPTL implementations is, however, limited to three. Here, we present a selective maleylation-directed isobaric peptide termini labeling (SMD-IPTL) approach for quantitative proteomics of LysC protein digests. SMD-IPTL extends the multiplexing capacity to 4-plex with the potential for higher levels of multiplexing using commercially available $^{13}\text{C}/^{15}\text{N}$ -labeled amino acids. SMD-IPTL is achieved in a one-pot reaction in three consecutive steps: 1) selective maleylation at the N-terminus; 2) labeling at the $\epsilon\text{-NH}_2$ group of the C-terminal Lys with isotopically labeled acetyl alanine; 3) thiol Michael addition of an isotopically labeled acetyl cysteine at the maleylated N-terminus. The isobarically labeled peptides are fragmented into sets of b- and y-ion clusters upon LC-MS/MS, which not only convey sequence information but also quantitative information for every labeling channel and avoid the issue of ratio distortion observed with reporter-ion-based approaches. We demonstrate the SMD-IPTL approach with a 4-plex labeled sample of BSA and yeast lysates mixed at different ratios. Utilizing SMD-IPTL for labeling and a narrow precursor isolation window of 0.8 Th with an offset of -0.2 Th, accurate ratios were measured across a 10-fold mixing range of BSA in a background of yeast proteome. With the yeast proteins mixed at ratios of 1:5:1:5, BSA was detected at ratios 0.94:2.46:4.70:9.92 when spiked at 1:2:5:10 ratios with an average standard deviation of peptide ratios of 0.34.

1 Introduction

Proteome quantification gives information on the relative amounts of large numbers of proteins between samples¹⁻³. The existing mass spectrometry-based proteome-wide quantitative methods can be classified into label-free proteomics^{4,5} and label-based proteomics⁶⁻⁸. Even though advanced data-acquisition schemes and algorithms have

been developed for label-free proteomics, the limited throughput and signal variation due to, among others, variable sample loss during work-up and changing ionization efficiency between injections argue in favor of multiplexed, label-based proteomics. Multiplexed quantification approaches (e.g., ICAT^{6,9,10}, SILAC¹¹⁻¹³, iTRAQ^{7,14}, TMT^{8,15} and IPTL¹⁶⁻²¹) exploit different combinations of heavy and light isotopes to differentially label peptides, which enables simultaneous sample work-up and LC-MS/MS analysis of multiple samples in a single experiment. The commonly used isotopes are ^{13}C , ^{15}N , ^{18}O and ^2H , with ^2H being less popular due to the potential risk of altering peptide retention time²²⁻²⁵.

The existing multiplexing strategies can also be classified into two categories, MS1 quantification and MS2 quantification, based on the stage at which peptides are quantified²⁶. For MS1 quantification, also called isotopic quantification, such as SILAC¹² and ICAT⁶, the same peptide from different samples will be labeled with different isotopic tags, which results in the same peptide showing multiple precursor ions at the MS1 level. Relative quantification is achieved by comparing the intensities or peak areas of the precursor ions at the MS1 level. Therefore, any isotopic quantification method will at least double the complexity of the MS1 spectrum, which further aggravates the already challenging issue of limited sampling capacity of precursor ions for MS/MS fragmentation across a chromatographic peak. On the contrary, the MS2 quantification methods use isobarically labeled peptides, so the same peptide originating from different samples will have the same mass. After fragmentation, the isobarically labeled peptide will release a unique reporter ion (TMT, iTRAQ) or peptide fragment ions (IPTL) which can be used to reveal the quantification information. The MS2 quantification methods not only allow for straightforward quantification of multiple samples in a single MS2 spectrum but also further reduce the required instrument time. The most widely used isobaric quantification methods are those using reporter ion tags (TMT and iTRAQ) because of the multiplex capacity and well-developed data processing software. However, reporter ion-based quantification methods suffer from ratio distortion²⁶⁻³¹, which is particularly serious in complex

samples, arising from the co-fragmentation of multiple peptides passing the precursor ion selection window. These peptides release identical reporter ions which are indistinguishable in MS2. To correct for ratio distortion, several methods have been proposed, such as additional gas phase purification³¹ and MultiNotch MS^{32,33}. An alternative isobaric method which gives rise to multiple quantification ions per peptide and is therefore less affected by the co-fragmentation problem is isobaric peptide termini labeling (IPTL). IPTL was first reported in 2009 by Koehler et al.¹⁶, but the approach is not as extensively used as TMT and iTRAQ presumably due to the limited multiplexing capacity, especially when avoiding deuterium labeling.

The initially reported IPTL method showed relative quantification of two samples of LysC digested proteins, where the peptides were crosswise modified at the C- and N-terminus with a pair of complementary isotopic tags resulting in isobarically labeled peptides that were fragmented into product ion clusters. The peptide and protein ratios can be inferred by comparing the intensities of the individual y- and b-series fragment ions. Even in case of co-fragmentation of two or more peptides, fragment ions can usually be correctly attributed. IPTL potentially permits more accurate and reliable quantification with multiple quantification data points per spectrum, for each y- and b-ion, and suffers less from co-fragmentation. Consequently, a number of optimized methods and applications have been reported in the past ten years, such as selective succinylation¹⁷ and dimethylation^{19,32-35} based IPTL (triplex-IPTL¹⁹, triplex-QITL²¹), SILAC^{34,35} or proteolytic ¹⁸O labeling^{33,36} combined with IPTL (IVTAL¹⁸, G-IVTL³⁴, QITL³³, diDO-IPTL³⁶), and pseudo-isobaric dimethyl labeling^{20,37-40} (pIDL³⁷, PITL³⁸, SWATH-pseudo-IPTL²⁰, MdFDIA⁴¹). Most of the methods are for duplex labeling or triplex at most, which means that the multiplexing capacity of IPTL is still far less than that of TMT, which has been extended to 16 labeling channels in a single LC-MS/MS run⁴². Recently, Liu et al.⁴⁰ reported the pseudo-isobaric dimethyl labeling (m-pIDL) method, which increased the multiplex capacity to 6-plex. m-pIDL does not suffer from co-fragmentation while relying on a wide isolation window of 10 Th. A potential limitation is the utilization of deuterium in the isotopic tags, which carries the risk of

changing the retention time of labeled peptides.

To improve the IPTL multiplex capacity with non-deuterium tags, we propose the selective maleylation-directed isobaric peptide termini labeling (SMD-IPTL) method, which is based on the selective maleylation at the N-termini of LysC digested peptides. The performance of SMD-IPTL was assessed at the 4-plex level by spiking different amounts of BSA into a yeast proteome background. SMD-IPTL can be extended to the 7-plex level using commercially available ¹³C or ¹⁵N-labeled cysteine and alanine.

2 Experiments

The details of chemicals and materials, synthesis of isotopically labeled acetyl-cysteine and acetyl-alanine p-nitrophenol ester, LC purification, LysC digestion, LC/MS/MS analysis and database searching and quantification can be found in the Supporting Information.

2.1 Optimization of selective maleylation at the peptide N-terminus. Solutions of different pH values (7.0, 6.5, 6.0, 5.5, 5.0) were prepared with 100 mM sodium acetate and acetic acid¹⁷. Subsequently, the peptide WLYRAK was dissolved in solutions of different pH values (7.0, 6.5, 6.0, 5.5, 5.0) to a concentration of 10 μM. Then, 4 μg/μL maleic anhydride was freshly prepared in acetonitrile and 2 μL was added to 100 μL of each WLYRAK solution. The reaction tube was kept shaking at room temperature for 30 min. The reaction was tracked with LC-MS. Maleylation on LysC peptides was further optimized by infusing 50 μg/μL maleic anhydride with a syringe pump at a flow rate of 0.4 μL/min into 25 μg LysC peptides dissolved in 1 mL sodium acetate-acetic acid solution of pH 5.5 for 1 h.

2.2 Duplex labeling of maleylated WLYRAK. 100 μL maleylated peptide solution was dried in a vacuum concentrator, followed by adding 100 μL of 50 mM sodium tetraborate and the pH was adjusted to 9 with 500 mM NaOH. Then, 100 mM ¹³C₁-acetyl-alanine p-nitrophenol ester (¹³C₁-Ac-Ala-PNP) containing one ¹³C label in the acetyl group or acetyl-alanine p-nitrophenol ester (Ac-Ala-PNP) which contains no ¹³C label was prepared in DMF. Subsequently, 2 μL p-nitrophenol ester was added to the

maleylated peptide solution and incubated for 1 h at room temperature. To ensure complete labeling, 2 μ L of p-nitrophenol ester was added again and incubated for 30 min more. Afterwards, the pH of acetyl-alanine labeled solutions was adjusted to 9. Subsequently, 5 μ L 400 mM $^{13}\text{C}_1$ -acetyl-cysteine ($^{13}\text{C}_1$ -Ac-Cys-OH) which contains one ^{13}C label in the acetyl group was added to the Ac-Ala-PNP labeled solution. Conversely, 5 μ L 400 mM acetyl-cysteine (Ac-Cys-OH) was added to the $^{13}\text{C}_1$ -Ac-Ala-PNP labeled solution. The reaction solutions were bubbled with argon for 5 min and incubated overnight at 55 °C. Finally, potentially formed esters at the hydroxyl groups of Ser, Thr or Tyr and excess PNP ester were hydrolyzed by treatment with 5 % hydroxylamine⁴³ for 5 min at 55 °C prior to desalting the samples by SPE using the STAGE (STop And Go Extraction) TIPS Desalting Procedure⁴⁴ followed by LC-MS analysis.

2.3 Quadruplex isobaric labeling of LysC peptides. 400 μ L maleylated LysC peptides of BSA or yeast protein was mixed with 50 μ L 100 mM sodium tetraborate and the pH was adjusted to 9 with 5 M NaOH. The solution was split into 4 tubes for labeling reactions of 4-channels. 2 μ L 100 mM Ac-Ala-PNP, $^{13}\text{C}_1$ -Ac-Ala-PNP, $^{13}\text{C}_2$ -Ac-Ala-PNP and $^{13}\text{C}_3$ -Ac-Ala-PNP was respectively added to the 4 tubes of maleylated LysC peptide solution and incubated for 1 h at room temperature. To ensure complete labeling, 2 μ L of p-nitrophenol ester was added again and incubated for 30 min more. Afterwards, the pH of acetyl-alanine p-nitrophenol ester labeled solutions was adjusted to 9. 10 μ L 400 mM $^{13}\text{C}_3$ -Ac-Cys-OH was added to the Ac-Ala-PNP labeled solution, 10 μ L 400 mM $^{13}\text{C}_2$ -Ac-Cys-OH was added to the $^{13}\text{C}_1$ -Ac-Ala-PNP labeled solution, 10 μ L 400 mM $^{13}\text{C}_1$ -Ac-Cys-OH was added to the $^{13}\text{C}_2$ -Ac-Ala-PNP labeled solution and 10 μ L 400 mM Ac-Cys-OH was added to the $^{13}\text{C}_3$ -Ac-Ala-PNP labeled solution. The reaction solutions were bubbled with argon for 5 min and incubated overnight at 55 °C. Finally, potentially formed esters at the hydroxyl groups of Ser, Thr or Tyr and excess PNP ester were hydrolyzed by treatment with 5 % hydroxylamine⁴³ for 5 min at 55 °C prior to desalting the samples by SPE using the STAGE (STop And Go Extraction) TIPS Desalting Procedure⁴⁴ followed by LC-MS analysis.

3 Results and discussion

3.1 Improving the multiplexing capacity of IPTL. SMD-IPTL not only retains all of the merits of IPTL^{16,17,19} but also improves it by: I) increasing the multiplex capacity to 4-plex or more with readily available isotopically-labeled amino acids; and II) utilizing non-deuterium labeled tags to ensure that isobarically-labeled peptides have the same retention time.

Inspired by selective succinylation¹⁷ and thiol Michael addition on maleic derivatives⁴⁵, we assumed that maleic anhydride has comparable selective reactivity for the peptide N-terminus as succinic anhydride. The introduced maleic derivatives at the N-terminus can be used for further modification. We decided to use readily available ^{13}C - and ^{15}N -labeled amino acids as tag building blocks. As shown in **Figure 1A**, the isobaric labeling can be achieved in a one-pot reaction with three consecutive steps: 1) selective maleylation at the N-terminus; 2) labeling with acetyl-alanine p-nitrophenol ester at the ϵ -NH₂ group of C-terminal Lys; 3) thiol Michael addition at the double bond of the newly maleylated N-terminus. As a result, the LysC peptides will be isobarically and differentially labeled at the N- and C-terminus with a complementary pair of isotopically labeled acetyl-cysteine and acetyl-alanine, respectively. With commercially available isotopically labeled cysteine and alanine, seven isobaric combinations can be potentially made, without the need for deuterium labels, as shown in **Figure 1B**. The isobarically labeled peptides derived from different samples are mixed prior to LC-MS/MS analysis (**Figure 1C**) and have identical retention times and masses. The MS2 peak intensities of y- and b-ion series fragments for each labeling channel are extracted and their intensity ratios represent the difference in the amount of the peptides in each sample in the mixture, as shown in the insert in **Figure 1C**.

Figure 1. Scheme of SMD-IPTL. (A) The labeling steps of SMD-IPTL. (B) The 7 possible combinations of isotopically labeled acetyl-cysteine and acetyl-alanine. The atom marked with asterisk denotes ^{13}C or ^{15}N . (C) The LC-MS/MS process for a mixture of 7-plex labeled samples.

3.2 Optimization of the selective maleylation reaction. Crosswise labeling of the N-terminal $\alpha\text{-NH}_2$ group and Lys $\epsilon\text{-NH}_2$ group with a pair of complementary tags is the prerequisite for IPTL. The challenge of IPTL therefore lies in the selective labeling of these two forms of $-\text{NH}_2$ group with readily available tags. Both selective succinylation¹⁷ and selective dimethylation³² have been used in IPTL, exploiting the pK_a difference between α - and ϵ -amino groups to selectively label the $\alpha\text{-NH}_2$ group at a specific pH. Maleic anhydride has a similar structure and reactivity as succinic anhydride, but the carbon double bond in the maleic anhydride permits further derivatization after maleylation, which can be used to insert an isotopic tag at the peptide N-terminus⁴⁵. We used the peptide WLYRAK to investigate the reactivity and selectivity of maleylation at various pH values (pH 7.0, 6.5, 6.0, 5.5, 5.0) and in water. As shown in **Figure 2A**, selectivity of maleylation for the $\alpha\text{-NH}_2$ group increases as the pH decreases. However, the reaction is not complete at pH 5.0, so that pH 5.5 presents the best compromise between overall reaction yield and selectivity with an overall yield of more than 90% and a double maleylation of 5-8% (see **Figure S5** for the MS2 spectrum of the N-maleylated peptide Ma-WLYRAK). The selective maleylation

reaction was further optimized by infusing 50 $\mu\text{g}/\mu\text{L}$ maleic anhydride with a syringe pump at a flow rate of 0.4 $\mu\text{L}/\text{min}$ into the peptide solution at pH 5.5. As shown in the **Figure S6**, double maleylation was further reduced to less than 2%. The amount of double-maleylated peptide did not increase when prolonging the infusion time. Based on these results, sodium acetate at pH 5.5 and slow infusion of maleic anhydride with a syringe pump at 0.4 $\mu\text{L}/\text{min}$ were used for optimal selective maleylation.

Figure 2. Optimization of maleylation and isobaric labeling of peptide WLYRAK. (A) Maleylation in various pH buffers. XICs of m/z 418.74, 467.74 and 516.74 were combined for all pH values. Peak **a** in the extracted ion chromatograms is the unmodified peptide; peak **b** is the peptide selectively maleylated on the $\alpha\text{-NH}_2$ group; peak **c** is the double-maleylated peptide on both $\alpha\text{-NH}_2$ and $\epsilon\text{-NH}_2$ groups. (B) One-pot isobaric labeling reaction steps. XICs of m/z 467.74, 524.27 and 605.78 were combined for all steps. Peak **b** is the single-maleylated peptide (Ma-WLYRAK); peak **d** is the peptide after labeling with acetyl-alanine at the $\epsilon\text{-NH}_2$ group (Ma-WLYRAK-Ac-Ala); peak **e** is the peptide after thiol Michael addition of acetyl-cysteine (Ac-Cys-Ma-WLYRAK-Ac-Ala). Mass spectra are shown to the right of the chromatograms. (C) Enlarged view of the y_3 ion in the MS2 spectrum for different mixing ratios of **f**: $^{13}\text{C}_1$ -Ac-Cys-WLYRAK-Ac-Ala and **g**: Ac-Cys-WLYRAK-Ac-Ala- $^{13}\text{C}_1$. (D) MS2 spectrum of 1:1 mixed $^{13}\text{C}_1$ -Ac-Cys-WLYRAK-Ac-Ala and Ac-Cys-WLYRAK-Ac-Ala- $^{13}\text{C}_1$. The peak marked with asterisk is the precursor ion having lost the Ac-Ala group.

3.3 One-pot isobaric labeling of peptide WLYRAK. Following selective maleylation at the α -NH₂ group, ¹³C₁-Ac-Ala-PNP or Ac-Ala-PNP were used to react with the ϵ -NH₂ group of the C-terminal Lys at pH 9. As shown in **Figure 2B**, Ma-WLYRAK can be completely converted to Ma-WLYRAK-Ac-Ala-¹³C₁ or Ma-WLYRAK-Ac-Ala. Subsequently, after adjusting the pH to 9, the complementary isotopic form of acetyl cysteine was incorporated to generate the ¹³C₁-Ac-Cys-WLYRAK-Ac-Ala and Ac-Cys-WLYRAK-Ac-Ala-¹³C₁. Although the newly inserted N-terminal maleyl group is less reactive to thiol than maleimide, Tian et al. demonstrated that the maleyl group can efficiently react with cysteine and mercaptoethanol⁴⁵. However, we found that the thiol Michael addition was slow, taking more than 30 h to complete. After screening various additives, including triethylamine⁴⁶, hexylamine, proline⁴⁷, dimethylphenylphosphine⁴⁶, and sodium tetraborate⁴⁸, we found that sodium tetraborate enables full labeling of the N-terminal maleyl group in 15 h (**Figure S7**).

To demonstrate the feasibility of SMD-IPTL, isobarically labeled ¹³C₁-Ac-Cys-WLYRAK-Ac-Ala and Ac-Cys-WLYRAK-Ac-Ala-¹³C₁ were mixed at various ratios (1:1, 1:5, 1:10, 1:20, 5:1, 10:1, 20:1) followed by LC-MS/MS analysis. As shown in **Figure 2D**, the fragment ions cover the entire b-ion series and several y-ions, which means that fragmentation of the peptide backbone works well after inserting tags on both termini. Notably, every y- and b-ion appeared as a peak pair in the spectrum, as shown in **Figure 2C** (enlarged view of the y₃ ion for different mixing ratios). The relative intensity of the light and heavy peaks is consistent with the corresponding mixing ratio. **Figure S8** shows the correlation between experimental and theoretical ratios over the tested mixing ratios, which indicates that accurate quantification can be achieved over a 20-fold dynamic range for peptide WLYRAK.

3.4 Reducing isotope interference of fragment ions by using a narrow precursor isolation window. In the MS2 spectra of IPTL, every labeling channel has a set of unique fragment ions, which represents the main advantage of IPTL by providing more accurate and reliable quantification information, since the entire fragment ion series

contains ratio information instead of only the reporter ion in the TMT and iTRAQ reporter ion-based approaches. **Figure 3A, B, C, D** show the y₅ ion of isobarically labeled SEIAHRFK (a BSA-derived peptide after LysC digestion) from 4 labeling channels with an interval of 1 Da between the channels. However, with a precursor isolation window (IW) of 2 Th, which is the default setting in data-dependent acquisition (DDA) proteomics on Orbitrap mass spectrometers, deducing quantification ratios from intensities of fragment ions is complicated due to the interference of the natural ¹³C contribution to the isotopologue pattern of the fragment ions²⁸. As shown in **Figure 3E**, setting the isolation window to 2 Th in the LC-MS/MS analysis of a 1:1:1:1 mixed BSA sample resulted in an MS2 spectrum (**Figure 3F**) in which the ratio of the y₅ ions did not correspond to the expected 1:1:1:1 ratio, since there is a small additional ¹³C peak next to the 4 major peaks. The use of a narrow precursor isolation window^{29,49} has been reported to reduce the ¹³C contribution to the fragment ions. As shown in **Figure 3G**, when the isolation window was set at 0.8 Th with -0.2 Th offset to specifically select the monoisotopic peak, the ratio between the y₅ ions perfectly matched the expected 1:1:1:1 ratio (**Figure 3H**). However, on the Q Exactive plus mass spectrometer, the precursor ion selection automatically shifts the isolation window to center on the highest intensity peak, rather than the monoisotopic peak, which means that for peptides above ~2 kDa multiple isotopologues are co-fragmented. This means that for the labeled peptide VPQVSTPTLVEVSRSLGK (2273.14 Da), even with a narrow isolation window of 0.8 Th and an offset of -0.2 Th (**Figure 3I**), the ¹³C contribution still affected the MS2 spectrum. As shown in **Figure 3J, K, L**, the ¹³C contribution increases with fragment ion mass. Therefore, for the MS2 spectra derived from peptides where the isolation window is centered on the first ¹³C isotopologue, only the small fragment ions y₂, b₁, b₂ were used for data processing. The y₁ ion was never used for quantification, because all peptides in a LysC digest have the same y₁ ion, which may distort ratios by peptide co-fragmentation similar to reporter ion-based approaches⁴⁰. Fragment ions other than the main b- and y-ion series, such as those having H₂O and NH₃ loss, were also ignored, due to the increased likelihood of

convolution between them.

Figure 3. Reducing interference of the endogenous ^{13}C contribution to the fragment ions by narrowing the width of the precursor isolation window (IW). Peptides are from the 4-plex labeled LysC digested BSA. (A) the y_5 ion of $^{13}\text{C}_3$ -Ac-Cys-Ma-SEIAHRFK-Ac-Ala; (B) the y_5 ion of $^{13}\text{C}_2$ -Ac-Cys-Ma-SEIAHRFK-Ac-Ala- $^{13}\text{C}_1$; (C) the y_5 ion of $^{13}\text{C}_1$ -Ac-Cys-Ma-SEIAHRFK-Ac-Ala- $^{13}\text{C}_2$; (D) the y_5 ion of Ac-Cys-Ma-SEIAHRFK-Ac-Ala- $^{13}\text{C}_3$; (E) an IW of 2 Th to select the precursor ion of isobarically labeled SEIAHRFK mixed at a ratio of 1:1:1:1; (F) the y_5 ion of isobarically labeled SEIAHRFK mixed at a ratio of 1:1:1:1 and fragmented with an IW of 2 Th; (G) an IW of 0.8 Th with a -0.2 Th offset to select the precursor ion of isobarically labeled SEIAHRFK mixed at ratio of 1:1:1:1; (H) the y_5 ion of isobarically labeled SEIAHRFK mixed at a ratio of 1:1:1:1 and fragmented with an IW of 0.8 Th with a -0.2 Th offset; (I) an IW of 0.8 Th with a -0.2 Th offset to select the precursor ion of isobarically labeled VPQVSTPTLVEVSRSLGK mixed at a ratio of 1:1:1:1; (J), (K) and (L), respectively show the b_1 , y_6 , and y_{15} ions of isobarically labeled VPQVSTPTLVEVSRSLGK mixed at a ratio of 1:1:1:1 and fragmented with an IW of 0.8 Th with a -0.2 Th offset.

After preselection of the fragment ions suitable for quantification, the next step is to calculate the ratios at the spectrum, peptide and protein levels sequentially. As shown in **Figure 4A**, the log2-normalized ratio has a better convergence to zero as the intensity of fragment ions increases, which means that more intense fragments are more reliable and should contribute more to the calculated ratio. Consequently, the ratio at the

spectrum level was calculated as the normalized ratio of the sum of all fragment ion intensities from the same labeling channel (**Figure S3** and **Figure S4**). The same trend of intensities is apparent at the peptide level (**Figure 4B**). Thus, the ratio at the peptide level was calculated by using the spectra with the three highest total peak intensities and the ratio at the protein level was calculated by using the peptides with the three highest total peak intensities. According to this calculation, based on medians of the log2-normalized measured ratios, the 1:5:1:5 and 1:2:5:10 mixed 4-plex labeled BSA ratios were determined to be 1.10:5.28:1.05:4.50 and 1.06:2.60:4.63:9.51 at the peptide level, as shown in **Figure 4C** and **Figure 4D**, respectively.

Figure 4. Normalized ratios distribution of 4-plex labeled BSA mixed at various ratios. (A) the normalized ratios of all assigned fragment ions from 1:1:1:1 mixed 4-plex labeled BSA; (B) the normalized ratios of identified peptides from 1:1:1:1 mixed 4-plex labeled BSA; (C) the normalized ratios of identified peptides from 1:5:1:5 mixed 4-plex labeled BSA; (D) the normalized ratios of identified peptides from 1:2:5:10 mixed 4-plex labeled BSA.

3.5 4-plex labeling of LysC peptides of the yeast proteome. To evaluate the efficiency and reproducibility of SMD-IPTL labeling reactions and the quantification accuracy on a complex sample, 4-plex labeling was performed on the LysC peptides from a yeast proteome. Three replicate maleylation reactions were performed and the maleylation

yield of each peptide was determined as the percentage of intensity of the N-terminal maleylated form to the total intensity of all possible forms including unmodified, N-terminal maleylated, C-terminal maleylated and both N- and C-terminal maleylated peptide. As shown in **Table 1** and **Figure S9**, around 95% of the peptides had a labeling yield higher than 90% and the average coefficient of variation (CV) of 3 replicates was 14.0%. The efficiency and reproducibility of subsequent labeling steps of the 4 channels with 4 pairs of complementary acetyl-cysteine and acetyl-alanine p-nitrophenol ester can be found in **Figure S10**. More than 98% of identified peptides had a total yield above 98% and an average CV of 1.3%.

Table 1 Efficiency evaluation of 3 replicates of maleylation on LysC yeast peptides

	Number of peptides					CV
	yield >95%	90%-95%	85%-90%	80%-85%	yield <80%	
Replicate 1	2175	95	39	21	72	14.0%
Replicate 2	2056	82	53	21	69	13.1%
Replicate 3	2083	90	33	16	84	14.8%

To investigate the quantification accuracy of SMD-IPTL in a complex sample, the 4-plex labeled LysC yeast peptides were mixed at a ratio of 1:1:1:1 and analyzed with an isolation window of 0.8 Th and an offset of -0.2 Th. As shown in the **Figure 5A**, the log2-normalized ratios are mainly distributed over the range from -0.4 to 0.4. The medians of log2-normalized ratios of the 4 labeling channels are 0.03, 0.03, -0.06 and -0.03, respectively (**Figure 5A**). On the protein level (337 proteins were identified), 97.0% of proteins were quantified within a 2-fold range and 91.7% of proteins were quantified within a 1.5-fold range.

3.6 Quantifying BSA in a background of yeast proteins. To investigate the dynamic range of SMD-IPTL in a complex background, a BSA-yeast proteome sample was prepared. 4-plex labeled LysC peptides of BSA at a ratio of 1:2:5:10 and 4-plex labeled LysC peptides of yeast proteins at the ratio of 1:5:1:5 were mixed separately. The yeast

peptides were then mixed with peptides of BSA, according to the scheme shown in **Figure 5B**. The mixed BSA-yeast sample was analyzed with an isolation window of 0.8 Th and an offset of -0.2 Th. To assess the quantification of the yeast proteins, as shown in the **Figure 5B**, the medians of log2-normalized protein ratios of the 4 labeling channels were determined as -0.01, 2.37, 0.15, 2.29, respectively, close to the theoretical medians of 0, 2.32, 0, 2.32. In the channels of Ac-Cys-Ma-peptides-Ac-Ala-¹³C₃ and ¹³C₂-Ac-Cys-Ma-peptides-Ac-Ala-¹³C₁, four times lower amounts of yeast peptides were mixed than in the other two channels, and the quantified protein ratios therefore had a larger variation. The 1:5:1:5 mixed channels of yeast had an average standard deviation of all log2 ratios of 0.40. For 4-plex labeled BSA, the log2-normalized protein ratios were -0.09:1.30:2.23:3.31 which is close to the log2-theoretical ratios of 0.00:1.00:2.32:3.32.

Figure 5. Various ratios of 4-plex labeled yeast and BSA-yeast samples. (A) normalized ratios of yeast sample mixed at a ratio of 1:1:1:1 (150 ng:150 ng:150 ng:150 ng); (B) normalized ratio of a BSA-yeast proteome sample which consists of the 1:2:5:10 (1 ng:2 ng:5 ng:10 ng) mixed 4-plex labeled BSA and the 1:5:1:5 (50 ng:250 ng:50 ng:250 ng) mixed 4-plex labeled yeast.

While SMD-IPTL exhibits good quantification capability across a 10-fold dynamic range, there is still room for improvement with respect to peptide identification, which is a prerequisite for calculating theoretical fragment ions as required for relative peptide quantification. In the MS2 spectra of SMD-IPTL, the number of fragment ions is multiplied by the number of labeling channels, which provides more accurate quantification information. However, this also makes identification more challenging due to the more complex MS2 spectra, which are not properly handled by existing protein identification software⁵⁰⁻⁵², since unmatched fragment ion peaks reduce the peptide score. The development of identification algorithms capable of handling IPTL data would further improve the identification of peptides in SMD-IPTL samples and hence the number of quantified proteins. In addition, data acquisition software that is capable of specifically selecting the monoisotopic peak for all peptides in combination with a narrow precursor isolation window would further improve the quantification accuracy of SMD-IPTL. Besides, we observed that the peptide ionization efficiency decreases slightly after isobaric labeling, which might be caused by the modifications on the amino groups of the peptides. Finally, the multiplexing capacity of SMD-IPTL can be further improved by extending the range of isotopic forms of Cys and Ala and the use of mass spectrometers with sufficient resolution to distinguish ¹³C and ¹⁵N isotopes in fragment ions⁵³.

4 Conclusions

SMD-IPTL not only retains all the advantages of IPTL approaches but also improves the multiplexing capacity to at least 4-plex labeling with available non-deuterium isotopically labeled amino acids. SMD-IPTL is readily expandable to 7-plexing following the same strategy. Isobaric labeling of peptides is achieved in a one-pot reaction comprising three steps. We applied a precursor ion isolation window of 0.8 Th with an offset of -0.2 Th during data acquisition, which significantly helped deducing protein ratios from intensities of fragment ions. Finally, the proteome quantification capability of SMD-IPTL was demonstrated with the 4-plex labeling of a yeast proteome

sample spiked with BSA over a 10-fold dynamic range. Further improvements in data acquisition and data analysis software are needed to make the SMD-IPTL approach more widely used in the future.

Acknowledgements

We gratefully acknowledge the China Scholarship Council (CSC) for a PhD fellowship to X. Tian. We thank Prof. Bernd Thiede for sharing the quantification software of isobariQ. X. Tian thanks Jos Hermans for helping with the LC-MS analyses. This work is partially funded by the Open Technology Programme of Toegepaste en Technische Wetenschappen (TTW) with project number 15230 which is financed by the Netherlands Organisation for Scientific Research (NWO).

Notes

The authors declare no competing financial interest. The raw mass spectrometry proteomics data of yeast and BAS-yeast sample have been deposited to the ProteomeXchange Consortium via the PRIDE⁵⁴ partner repository with the dataset identifier PXD018589.

References

- (1) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chemical Reviews* **2013**, *113*, 2343-2394.
- (2) Rauniyar, N.; Yates, J. R. Isobaric Labeling-Based Relative Quantification in Shotgun Proteomics. *Journal of Proteome Research* **2014**, *13*, 5293-5309.
- (3) Chahrour, O.; Cobice, D.; Malone, J. Stable isotope labelling methods in mass spectrometry-based quantitative proteomics. *J Pharmaceut Biomed* **2015**, *113*, 2-20.
- (4) Bondarenko, P. V.; Chelius, D.; Shaler, T. A. Identification and Relative Quantitation of Protein Mixtures by Enzymatic Digestion Followed by Capillary Reversed-Phase Liquid Chromatography–Tandem Mass Spectrometry. *Analytical Chemistry* **2002**, *74*, 4741-4749.
- (5) Asara, J. M.; Christofk, H. R.; Freimark, L. M.; Cantley, L. C. A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *Proteomics* **2008**, *8*, 994-999.
- (6) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology* **1999**, *17*, 994-999.
- (7) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlett-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents. *Molecular and Cellular Proteomics* **2004**, *3*, 1154-1169.
- (8) 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. *Analytical Chemistry* **2003**, *75*, 1895-1904.
- (9) Colangelo, C. M.; Williams, K. R. Isotope-coded affinity tags for protein quantification. *Methods in molecular biology (Clifton, N.J.)* **2006**, *328*, 151-158.
- (10) Yi, E. C.; Li, X. J.; Cooke, K.; Lee, H.; Raught, B.; Page, A.; Aneliunas, V.; Hieter, P.; Goodlett, D. R.; Aebersold, R. Increased quantitative proteome coverage with (13)C/(12)C-based, acid-cleavable isotope-coded affinity tag reagent and modified data acquisition scheme. *Proteomics* **2005**, *5*, 380-387.
- (11) Jiang, H.; English, A. M. Quantitative analysis of the yeast proteome by incorporation of isotopically labeled leucine. *Journal of proteome research* **2002**, *1*, 345-350.
- (12) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; 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 and Cellular Proteomics* **2002**, *1*, 376-386.
- (13) Zhu, H.; Pan, S.; Gu, S.; Bradbury, E. M.; Chen, X. Amino acid residue specific stable isotope labeling for quantitative proteomics. *Rapid communications in mass spectrometry* **2002**, *16*, 2115-2123.
- (14) Ow, S. Y.; Cardona, T.; Taton, A.; Magnuson, A.; Lindblad, P.; Stensjö, K.; Wright, P. C. Quantitative Shotgun Proteomics of Enriched Heterocysts from *Nostoc* sp. PCC 7120 Using 8-Plex Isobaric Peptide Tags. *Journal of Proteome Research* **2008**, *7*, 1615-1628.
- (15) Dayon, L.; Hainard, A.; Licker, V.; Turck, N.; Kuhn, K.; Hochstrasser, D. F.; Burkhard, P. R.; Sanchez, J. C. Relative Quantification of Proteins in Human Cerebrospinal Fluids by MS/MS Using 6-Plex Isobaric Tags. *Analytical Chemistry* **2008**, *80*, 2921-2931.
- (16) Koehler, C. J.; Strozynski, M.; Kozielski, F.; Treumann, A.; Thiede, B. Isobaric Peptide Termini

- Labeling for MS/MS-Based Quantitative Proteomics. *Journal of Proteome Research* **2009**, *8*, 4333-4341.
- (17) Koehler, C. J.; Arntzen, M. O.; Strozynski, M.; Treumann, A.; Thiede, B. Isobaric Peptide Termini Labeling Utilizing Site-Specific N-Terminal Succinylation. *Analytical Chemistry* **2011**, *83*, 4775-4781.
 - (18) Nie, A. Y.; Zhang, L.; Yan, G. Q.; Yao, J.; Zhang, Y.; Lu, H. J.; Yang, P. Y.; He, F. C. In Vivo Termini Amino Acid Labeling for Quantitative Proteomics. *Analytical Chemistry* **2011**, *83*, 6026-6033.
 - (19) Koehler, C. J.; Arntzen, M. O.; de Souza, G. A.; Thiede, B. An Approach for Triplex-Isobaric Peptide Termini Labeling (Triplex-IPTL). *Analytical Chemistry* **2013**, *85*, 2478-2485.
 - (20) Zhang, S.; Chen, L. F.; Shan, Y. C.; Sui, Z. G.; Wu, Q.; Zhang, L. H.; Liang, Z.; Zhang, Y. K. Pseudo isobaric peptide termini labelling for relative proteome quantification by SWATH MS acquisition. *Analyst* **2016**, *141*, 4912-4918.
 - (21) Jiang, H. C.; Yin, H. R.; Xie, L. Q.; Zhang, Y.; Zhang, L.; Yang, P. Y.; Lu, H. J. A novel triplex isobaric termini labeling quantitative approach for simultaneously supplying three quantitative sources. *Analytica Chimica Acta* **2018**, *1001*, 70-77.
 - (22) Koehler, C. J.; Arntzen, M. O.; Thiede, B. The impact of carbon-13 and deuterium on relative quantification of proteins using stable isotope diethyl labeling. *Rapid Communications in Mass Spectrometry* **2015**, *29*, 830-836.
 - (23) Zhang, R.; Sioma, C. S.; Thompson, R. A.; Xiong, L.; Regnier, F. E. Controlling Deuterium Isotope Effects in Comparative Proteomics. *Analytical Chemistry* **2002**, *74*, 3662-3669.
 - (24) Di Palma, S.; Rajmakers, R.; Heck, A. J. R.; Mohammed, S. Evaluation of the Deuterium Isotope Effect in Zwitterionic Hydrophilic Interaction Liquid Chromatography Separations for Implementation in a Quantitative Proteomic Approach. *Analytical Chemistry* **2011**, *83*, 8352-8356.
 - (25) Boutilier, J. M.; Warden, H.; Doucette, A. A.; Wentzell, P. D. Chromatographic behaviour of peptides following dimethylation with H₂/D₂-formaldehyde: Implications for comparative proteomics. *Journal of Chromatography B* **2012**, *908*, 59-66.
 - (26) Mertins, P.; Udesi, N. D.; Clauser, K. R.; Mani, D.; Patel, J.; Ong, S. e.; Jaffe, J. D.; Carr, S. A. iTRAQ Labeling is Superior to mTRAQ for Quantitative Global Proteomics and Phosphoproteomics. *Molecular and Cellular Proteomics* **2012**, *11*, 1-12.
 - (27) Ting, L.; Rad, R.; Gygi, S. P.; Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nature Methods* **2011**, *8*, 937-940.
 - (28) Wühr, M.; Haas, W.; McAlister, G. C.; Peshkin, L.; Rad, R.; Kirschner, M. W.; Gygi, S. P. Accurate Multiplexed Proteomics at the MS2 Level Using the Complement Reporter Ion Cluster. *Analytical Chemistry* **2012**, *84*, 9214-9221.
 - (29) Sonnett, M.; Yeung, E.; Wühr, M. Accurate, Sensitive, and Precise Multiplexed Proteomics Using the Complement Reporter Ion Cluster. *Analytical Chemistry* **2018**, *90*, 5032-5039.
 - (30) Bantscheff, M.; Boesche, M.; Eberhard, D.; Matthieson, T.; Sweetman, G.; Kuster, B. Robust and Sensitive iTRAQ Quantification on an LTQ Orbitrap Mass Spectrometer. *Molecular and Cellular Proteomics* **2008**, *7*, 1702-1713.
 - (31) Wenger, C. D.; Lee, M. V.; Hebert, A. S.; McAlister, G. C.; Phanstiel, D. H.; Westphall, M. S.; Coon, J. J. Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nature Methods* **2011**, *8*, 933-935.
 - (32) Qin, H.; Wang, F.; Zhang, Y.; Hu, Z.; Song, C.; Wu, R.; Ye, M.; Zou, H. Isobaric cross-sequence

labeling of peptides by using site-selective N-terminus dimethylation. *Chemical Communications* **2012**, 48, 6265-6267.

(33) Yang, S. J.; Nie, A. Y.; Zhang, L.; Yan, G. Q.; Yao, J.; Xie, L. Q.; Lu, H. J.; Yang, P. Y. A novel quantitative proteomics workflow by isobaric terminal labeling. *Journal of Proteomics* **2012**, 75, 5797-5806.

(34) Xie, L. Q.; Nie, A. Y.; Yang, S. J.; Zhao, C.; Zhang, L.; Yang, P. Y.; Lu, H. J. Global in vivo terminal amino acid labeling for exploring differential expressed proteins induced by dialyzed serum cultivation. *Analyst* **2014**, 139, 4497-4504.

(35) Cao, T.; Zhang, L.; Zhang, Y.; Yan, G. Q.; Fang, C. Y.; Bao, H. M.; Lu, H. J. Site-Specific Quantification of Protein Ubiquitination on MS2 Fragment Ion Level via Isobaric Peptide Labeling. *Analytical Chemistry* **2017**, 89, 11468-11475.

(36) Waldbauer, J.; Zhang, L.; Rizzo, A.; Muratore, D. diDO-IPTL: A Peptide-Labeling Strategy for Precision Quantitative Proteomics. *Analytical Chemistry* **2017**, 89, 11498-11504.

(37) Zhou, Y.; Shan, Y. C.; Wu, Q.; Zhang, S.; Zhang, L. H.; Zhang, Y. K. Mass Defect-Based Pseudo-Isobaric Dimethyl Labeling for Proteome Quantification. *Analytical Chemistry* **2013**, 85, 10658-10663.

(38) Zhang, S.; Wu, Q.; Shan, Y. C.; Zhou, Y.; Zhang, L. H.; Zhang, Y. K. Partially isobaric peptide termini labeling assisted proteome quantitation based on MS and MS/MS signals. *Journal of Proteomics* **2015**, 114, 152-160.

(39) Di, Y.; Zhang, Y.; Zhang, L.; Tao, T.; Lu, H. J. MdFDIA: A Mass Defect Based Four-Plex Data-Independent Acquisition Strategy for Proteome Quantification. *Analytical Chemistry* **2017**, 89, 10248-10255.

(40) Liu, J. H.; Zhou, Y.; Shan, Y. C.; Zhao, B. F.; Hu, Y. C.; Sui, Z. G.; Liang, Z.; Zhang, L. H.; Zhang, Y. K. A Multiplex Fragment-Ion-Based Method for Accurate Proteome Quantification. *Analytical Chemistry* **2019**, 91, 3921-3928.

(41) Di, Y.; Zhang, Y.; Zhang, L.; Tao, T.; Lu, H. MdFDIA: A Mass Defect Based Four-Plex Data-Independent Acquisition Strategy for Proteome Quantification. *Analytical Chemistry* **2017**, 89, 10248-10255.

(42) Thompson, A.; Wölmer, N.; Koncarevic, S.; Selzer, S.; Böhm, G.; Legner, H.; Schmid, P.; Kienle, S.; Penning, P.; Höhle, C.; Berfelde, A.; Martinez-Pinna, R.; Farztdinov, V.; Jung, S.; Kuhn, K.; Pike, I. TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-Based Isobaric 16-Plex Tandem Mass Tag Reagent Set. *Analytical Chemistry* **2019**, 91, 15941-15950.

(43) Abello, N.; Kerstjens, H. A. M.; Postma, D. S.; Bischoff, R. Selective Acylation of Primary Amines in Peptides and Proteins. *Journal of Proteome Research* **2007**, 6, 4770-4776.

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

(45) Tian, S. S.; Zheng, S. Z.; Han, Y. P.; Guo, Z. C.; Zhai, G. J.; Bai, X.; He, X.; Fan, E. G.; Zhang, Y. K.; Zhang, K. Maleic Anhydride Labeling-Based Approach for Quantitative Proteomics and Successive Derivatization of Peptides. *Analytical Chemistry* **2017**, 89, 8259-8265.

(46) Chan, J. W.; Hoyle, C. E.; Lowe, A. B.; Bowman, M. Nucleophile-Initiated Thiol-Michael Reactions: Effect of Organocatalyst, Thiol, and Ene. *Macromolecules* **2010**, 43, 6381-6388.

(47) Kumar, A.; Akanksha. Amino acid catalyzed thio-Michael addition reactions. *Tetrahedron* **2007**, 63, 11086-11092.

(48) Hussain, S.; Bharadwaj, S. K.; Chaudhuri, M. K.; Kalita, H. Borax as an efficient metal-free catalyst for hetero-Michael reactions in an aqueous medium. *Eur J Org Chem* **2007**, 374-378.

(49) Winter, S. V.; Meier, F.; Wichmann, C.; Cox, J.; Mann, M.; Meissner, F. EASI-tag enables accurate multiplexed and interference-free MS2-based proteome quantification. *Nature Methods* **2018**, 15, 527-530.

(50) Xie, L.-Q.; Zhang, L.; Nie, A.-Y.; Yan, G.-Q.; Yao, J.; Zhang, Y.; Yang, P.-Y.; Lu, H.-J. ITMSQ: A software tool for N- and C-terminal fragment ion pairs based isobaric tandem mass spectrometry quantification. *Proteomics* **2015**, 15, 3755-3764.

(51) Arntzen, M. Ø.; Koehler, C. J.; Barsnes, H.; Berven, F. S.; Treumann, A.; Thiede, B. IsobariQ: Software for Isobaric Quantitative Proteomics using IPTL, iTRAQ, and TMT. *Journal of Proteome Research* **2011**, 10, 913-920.

(52) Zheng, N. R.; Shan, Y. C.; Deng, Y. L.; Zhang, Y. K. Novel Algorithm for Identification and Quantification of Proteins Based on Strategy of Isobaric Peptide Termini Labelling. *Chinese J Anal Chem* **2017**, 45, 1441-1447.

(53) Werner, T.; Sweetman, G.; Savitski, M. F.; Mathieson, T.; Bantscheff, M.; Savitski, M. M. Ion Coalescence of Neutron Encoded TMT 10-Plex Reporter Ions. *Analytical Chemistry* **2014**, 86, 3594-3601.

(54) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, Ş.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F., et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **2018**, 47, D442-D450.

Supporting Information

Materials and Methods

Chemicals and Materials. Maleic anhydride, Bovine serum albumin (BSA), N-Acetyl-L-cysteine ($^{12}\text{C}_5\text{H}_9\text{NO}_3\text{S}$), N-Acetyl-L-alanine ($^{12}\text{C}_5\text{H}_9\text{NO}_3$), L-Cysteine ($^{12}\text{C}_3\text{H}_7\text{NO}_2\text{S}$), L-Alanine ($^{12}\text{C}_3\text{H}_7\text{NO}_2$), L-Alanine-1- ^{13}C ($^{13}\text{C}_1^{12}\text{C}_2\text{H}_7\text{NO}_2$), Acetic anhydride ($^{12}\text{C}_4\text{H}_6\text{O}_3$), Acetic anhydride-1,1'- $^{13}\text{C}_2$ ($^{13}\text{C}_2^{12}\text{C}_2\text{H}_6\text{O}_3$), Acetic anhydride- $^{13}\text{C}_4$ ($^{13}\text{C}_4\text{H}_6\text{O}_3$), 1-(3-Dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (EDC·HCl), N,N-Diisopropylethylamine (DIPEA), 4-Nitrophenol, Tris(2-carboxyethyl) phosphine hydrochloride (TCEP), 2-Iodoacetamide, hydroxylamine hydrate, N,N-Dimethylformamide (DMF), Acetonitrile, Acetone, Dichloromethane (DCM), Acetic acid, Sodium Acetate, Sodium tetraborate decahydrate (Borax), Sodium hydroxide (NaOH), Triethylammonium bicarbonate (TEAB) were purchased from Sigma-Aldrich. L-Cysteine-1- ^{13}C ($^{13}\text{C}_1^{12}\text{C}_2\text{H}_7\text{NO}_2\text{S}$) was purchased from Cambridge Isotope Laboratories Inc. Trypsin, LysC and yeast protein extracts were purchased from Promega. Synthetic peptide WLYRAK was purchased from GL Biochem (Shanghai, China). Solid phase extraction disk Empore Octadecyl C18 was purchased from 3M.

Synthesis of isotopically labeled acetyl-cysteine and acetyl-alanine p-nitrophenol ester. The synthesis scheme of isotopically labeled acetyl-cysteine can be found in **Scheme S1**. 100 μL of 2 M cysteine or $^{13}\text{C}_1$ -cysteine was preheated to 70 °C on a thermomixer for 5 min. Then, 20 μL acetic anhydride or acetic anhydride-1,1'- $^{13}\text{C}_2$ or acetic anhydride- $^{13}\text{C}_4$ was added slowly and the reaction mixture was kept shaking for 20 min. Afterwards, the solution was adjusted to pH 9-10 with 5 M NaOH and incubated at 55 °C overnight. Then, 20 μL 300 mM tris(2-carboxyethyl) phosphine (TCEP) was added and the reaction mixture was incubated for 10 min. Finally, acetyl-cysteine was purified on reverse phase (RP)-HPLC with detection at 195 nm. The collected HPLC fraction was freeze-dried to afford a white fluffy solid which was stored at -20 °C until

use. The LC-MS analysis of acetyl-cysteine can be found in **Figure S1**.

The synthesis scheme of acetyl-alanine p-nitrophenol ester (Ac-Ala-PNP) can be found in **Scheme S2**. 100 μL of 2 M alanine or $^{13}\text{C}_1$ -alanine was preheated to 70 °C on a thermomixer for 5 min. Then, 20 μL acetic anhydride or acetic anhydride-1,1'- $^{13}\text{C}_2$ or acetic anhydride- $^{13}\text{C}_4$ was added slowly and the reaction mixture was kept shaking for 20 min. Subsequently, the solution was dried with a vacuum rotary evaporator. 100 μL of 100 mM hydrochloric acid was added to dissolve the oily product and the solution was dried again. This step was repeated 5 times to remove acetic acid as much as possible. Afterwards, 500 μL DCM, 88 mg 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC·HCl), and 65 mg p-nitrophenol were added. The reaction was kept stirring for 3 h, and after drying, the acetyl-alanine p-nitrophenol ester was purified on RP-HPLC with detection at 254 nm. The collected HPLC fraction was freeze-dried to afford a white fluffy solid which was stored at -20 °C until use. The LC-MS analysis of acetyl-alanine p-nitrophenol ester can be found in **Figure S2**.

LC purification. For LC purification, the HPLC system (Shimadzu, Kyoto, Japan) was equipped with a SIL-20AC autosampler, two LC-20AT pumps and an SPD-20A absorbance detector. The synthesized acetyl-cysteine and acetyl-alanine p-nitrophenol ester were separated on an XBridge BEH C18 OBD Prep Column (250 mm $\times$ 10 mm, 5 μm particles, 130 Å pore size, Waters, Ireland) with a 30 min gradient of 2-15% or 2-90% acetonitrile in water/0.1% formic acid at a flow rate of 2 mL/min. The collected fraction was freeze-dried and stored at -20 °C.

LysC Digestion. 100 μg BSA or yeast protein was dissolved in 100 μL 100 mM TEAB. Then, 5 μL 200 mM TCEP was added and the mixture was incubated for an hour at 37 °C. Afterwards, 5 μL 375 mM freshly prepared 2-iodoacetamide was added and the mixture was incubated for 30 min in the dark at room temperature. Finally, 800 μL acetone was added and the mixture was kept at -20 °C overnight for precipitation. The precipitated protein was centrifuged at 20800 g for 30 min. Subsequently, the

supernatant was decanted and alkylated proteins were stored at -20 °C until digestion. 100 µg alkylated BSA or yeast protein was dissolved in 200 µL 100 mM TEAB and 2 µg LysC was added. The pH was adjusted to 7.5 to 8 with 1 M TEAB and the digestion was kept overnight at 37 °C. The resulting LysC peptide mixtures of BSA and yeast protein were freeze-dried and stored at -80 °C before conducting the labelling reactions.

LC/MS/MS analysis. LC-MS/MS analyses were performed on an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific) coupled to an LC system (Shimadzu, Kyoto, Japan) equipped with a SIL-20AC HT autosampler, and two LC-20AD pumps. The sample was separated on an ACQUITY UPLC BEH C18 column (2.1 mm × 50 mm, 1.7 µm particles, Waters) at a column temperature of 60 °C with a 10 min gradient of 2-90% or 2-60% acetonitrile in water/0.1% formic acid at a flow rate of 300 µL/min. NanoLC-MS/MS analyses were performed on a Q-Exactive Plus mass spectrometer (Thermo Scientific, San José, CA) equipped with an UltiMate 3000 RSLCnano system (Thermo Scientific). The peptides were first loaded onto a trap column (Acclaim PepMap 100 pre-column, 75 µm, 2 cm, C18, 3 mm, 100 Å, Thermo Scientific) and then separated on an analytical column (PepMap RSLC C18, 50 cm 75 µm ID, 2 µm, 100 Å, Thermo Scientific). A flow rate of 300 nL/min and a column temperature of 40 °C were utilized. Solvent A was 0.1% formic acid in water and solvent B was 0.1% formic acid in acetonitrile. A gradient was applied from 2% to 50% or 65% B in the first 60 min, then 85% B for 5 min and equilibration for 15 min with 2% B. Mass spectrometry data were measured using a data-dependent top 10 method. Full MS scans were acquired between m/z 300 and 1650 with resolution of 70000 (at m/z 200); the automatic gain control (AGC) target was 3E6 and maximum injection time was 50 ms. Precursor ions were selected with a window of 2 Th or 0.8 Th with -0.2 Th as offset. Selected ions were fragmented in the HCD collision cell with a stepped normalized collision energy (nce) of 28, and tandem mass spectra were acquired in the Orbitrap mass analyzer with a resolution of 17500 (at m/z 200). The AGC target was 1E5 and the maximum injection

time was 100 ms. The charge exclusion discards ions with a single charge and charges larger than 4 and dynamic exclusion time was 20 s.

Database Searching and Quantification. LC-MS/MS raw files were analyzed with PEAKS Studio X+ (Bioinformatic Solutions) and searched against the Uniprot reference database of yeast (UP000002311, 6049 entries, downloaded on Jan. 20, 2020) into which the BSA entry (P02769) was inserted manually. LysC was selected as enzyme, digestion mode as specific and max missed cleavage sites as 0. A tolerance of 20 ppm for the precursor ion and 0.02 Da for the MS/MS fragment ions was applied. Carbamidomethylation (+57.02) on cysteine was set as fixed modification and oxidation (+15.99) on methionine as variable modification. For 4-plex labelling experiments, fixed modifications (N-terminus/C-terminal Lys) were set to the 4 corresponding channels $^{13}\text{C}_3\text{-Ac-Cys-Maleylation}$ (+264.04) and Ac-Ala (+113.05), $^{13}\text{C}_2\text{-Ac-Cys-Maleylation}$ (+263.04) and $^{13}\text{C}_1\text{-Ac-Ala}$ (114.05), $^{13}\text{C}_1\text{-Ac-Cys-Maleylation}$ (+262.03) and $^{13}\text{C}_2\text{-Ac-Ala}$ (115.05), $\text{Ac-Cys-Maleylation}$ (+261.03) and $^{13}\text{C}_3\text{-Ac-Ala}$ (116.06), respectively. The 4 labelling channels were searched separately because PEAKS cannot deconvolute the multiplexed fragment ions. The results were filtered with an FDR of 0.05 for peptides.

The PEAKS output on the supporting peptides (*protein-peptides.csv*) and the DB search peptide-spectrum matches (*DB search psm.csv*) were exported for every labelling channel. All the following steps were performed using in-house built Python scripts (available at <https://github.com/tianxiaobo002/SMD-IPTL-4-plex-data-processing>). Only the unique peptides extracted from all *protein-peptides.csv* files were used for quantification. For every peptide spectrum match (PSM) that was derived from unique peptides in the *DB search psm.csv* files, the corresponding theoretical fragment ions were calculated and grouped by scan number. The raw data was converted into a Python readable mgf file with RawConverter⁴⁵ to circumvent altering peak intensities and modifying the mass of precursor ions by PEAKS. All MS/MS spectra were extracted from the resulting mgf file and also grouped by the scan number. Afterwards, based on

the scan number, the original peak intensities in the mgf file were related to the corresponding theoretical fragment ions and the matched intensities were used to quantify peptides and ultimately proteins. The y1 ions from the C-terminal lysine and the fragment ions losing H₂O and NH₃ were excluded. The ratio of each PSM, peptide and protein were calculated based on the total peak intensities of the same labelling channel. For fragment ion spectra that were produced from monoisotopic peaks and in which no ¹³C interference existed, all matched b- and y-ions except y₁ were used for quantification and the minimum number of matched ions was set to 3. For fragment ion spectra that were not produced from monoisotopic peaks only and in which ¹³C interference was observed in all fragment ions, only the small ions (b₁, b₂ and y₂) were used for peptide quantification and the minimum number of matched ions was set to 1. Details and examples can be found in **Figure S3** and **Figure S4**. Each peptide ratio was calculated from three spectra with the highest total peak intensity and each protein ratio was calculated from the three peptides with the highest total peak intensity. SMD-IPTL quantification outputs of the 4-plex labelled BSA, yeast, BSA-yeast samples at the fragment ion, PSM, peptide and protein level are available at <https://github.com/tianxiaobo002/SMD-IPTL-quantification-outputs-at-multiple-levels>.

Supplementary Figures

Scheme S1. The synthesis approach for isotopic acetyl-cysteine (+0, +1, +2, +3). “*” means ¹³C.

Scheme S2. The synthesis approach for isotopic acetyl-alanine p-nitrophenol ester (+0, +1, +2, +3).

“*” means ¹³C.

Figure S1. (A) XIC of m/z 164.03, Ac-Cys-OH (+0). (B) mass spectrum of Ac-Cys-OH (+0). (C) XIC of m/z 165.04, Ac-Cys-OH (+1). (D) mass spectrum of Ac-Cys-OH (+1). (E) XIC of m/z 166.04, Ac-Cys-OH (+2). (F) mass spectrum of Ac-Cys-OH (+2). (G) XIC of m/z 167.04, Ac-Cys-OH (+3). (H) mass spectrum of Ac-Cys-OH (+3).

Figure S2. (A) XIC of m/z 253.08, Ac-Ala-PNP (+0). (B) mass spectrum of Ac-Ala-PNP (+0). (C) XIC of m/z 254.08, Ac-Ala-PNP (+1). (D) mass spectrum of Ac-Ala-PNP (+1). (E) XIC of m/z 255.08, Ac-Ala-PNP (+2). (F) mass spectrum of Ac-Ala-PNP (+2). (G) XIC of m/z 256.09, Ac-Ala-PNP (+3). (H) mass spectrum of Ac-Ala-PNP (+3).

Rules of calculating ratios at multiple levels:

1. Ratios at the spectrum level

- When the MS2 spectrum was derived from the monoisotopic precursor ion peak, all fragment ion peaks were processed except for y1 ions and fragment ions that are due to the loss of H₂O or NH₃.
- When the MS2 spectrum was derived from the non-monoisotopic precursor ion peak, only y2, b1 and b2 fragment ion peaks were processed.
- The ratio of each PSM was calculated as the ratio of the total peak intensities, which came from the same labelling channel and normalized with the weighted the average peak intensity. For example, for a sample that was mixed a ratios 1:5:1:5, the weighted average peak intensity = total peak intensities of 4 channels/(1+5+1+5).

2. Ratios at the peptide level

Each peptide ratio was calculated from the three spectra with the highest total peak intensity.

3. Ratios at the protein level

Each protein ratio was calculated from the three peptides with the highest total peak intensity

Calculating the ratio for spectrum 6302

		Peak intensity of individual labelling channel				fragment_ion_mass	peptide	protein
scan_number	ion_name	Ac-Cys-plus0	Ac-Cys-plus1	Ac-Cys-plus2	Ac-Cys-plus3			
6302	b1	1926488.9	1655704.9	768647.9	1377537.6	349.0705	SEIAHRFK	P02769
6302	b2	932790.6	1330674.0	1045959.2	845849.9	478.1131	SEIAHRFK	P02769
6302	y2	894223.6	685712.6	350580.7	667127.2	407.2294	SEIAHRFK	P02769
6302	y3	1373063.6	2036021.4	1511427.8	1534253.9	563.3305	SEIAHRFK	P02769
6302	y4	3417818.2	3935459.2	2863230.8	3579711.0	700.3894	SEIAHRFK	P02769
6302	y5	6751701.5	6397088.0	5554731.0	5303868.5	771.4265	SEIAHRFK	P02769
6302	y6	2109349.0	2044102.5	1485072.5	1293936.5	884.5079	SEIAHRFK	P02769
6302	y7	1062791.1	851464.3	981342.7	959399.4	1013.5505	SEIAHRFK	P02769
Total intensity of individual labelling channel		18468226.5	18936226.9	14560992.6	15561684.0			

Normalizer = (18468226.5 + 18936226.9 + 14560992.6 + 15561684.0)/4 = 16881782.5

Ratio of scan 6302 = 18468226.5/16881782.5 : 18936226.9/16881782.5 : 14560992.6/16881782.5 : 15561684.0/16881782.5 = 1.09 : 1.12 : 0.86 : 0.92

Calculating the ratio for the peptide SEIAHRFK

		Peak intensity of individual labelling channel				spectrum_total_intensity	peptide	protein
scan_number		Ac-Cys-plus0	Ac-Cys-plus1	Ac-Cys-plus2	Ac-Cys-plus3			
6302		18468226.5	18936226.9	14560992.6	15561684.0	67527130.0	SEIAHRFK	P02769
6442		6786562.9	6725740.3	6344137.0	6480702.6	26337142.8	SEIAHRFK	P02769
6153		2524740.8	3350597.1	5089866.6	4165686.7	15130891.2	SEIAHRFK	P02769
6837		1627526.6	1734630.0	1372834.2	1382953.0	6117943.8	SEIAHRFK	P02769
6563		618966.3	623450.6	638752.2	614468.3	2495637.4	SEIAHRFK	P02769
6696		234631.5	237421.6	253047.6	251116.0	976216.7	SEIAHRFK	P02769
6013		115872.3	124693.1	185870.0	170535.3	596970.7	SEIAHRFK	P02769
6976		89637.2	92037.8	78490.5	85802.4	345967.9	SEIAHRFK	P02769
7110		53692.6	45405.2	37569.2	35310.1	171977.1	SEIAHRFK	P02769
7368		20397.0	26997.8	13237.3	18148.0	78780.1	SEIAHRFK	P02769
5876		9557.3	17132.6	22541.2	10906.0	60137.1	SEIAHRFK	P02769
7507		14671.9	8948.4	10560.9	9836.9	44018.1	SEIAHRFK	P02769
7654		9225.8	11623.6	4299.4	10781.9	35930.7	SEIAHRFK	P02769
7833		3828.7	5826.1	8355.8	5137.7	23148.3	SEIAHRFK	P02769
Total intensity of individual labelling channel (top 3 spectra)		27779530.2	29012564.3	25994996.2	26208073.3			

Normalizer = (27779530.2 + 29012564.3 + 25994996.2 + 26208073.3)/4 = 27248791.0

Ratio of peptide (SEIAHRFK) = 27779530.2/27248791.0 : 29012564.3/27248791.0 : 25994996.2/27248791.0 : 26208073.3/27248791.0 = 1.02 : 1.06 : 0.95 : 0.96

Calculating ratio for the protein P02769

		Peak intensity of individual labelling channel				peptide_total_intensity	protein
peptide		Ac-Cys-plus0	Ac-Cys-plus1	Ac-Cys-plus2	Ac-Cys-plus3		
SEIAHRFK		27779530.2	29012564.3	25994996.2	26208073.3	108995164.0	P02769
IETMREK		22902162.8	22351350.5	22079420.2	28039112.6	95372046.1	P02769
AEFVEVTK		17930378.4	19098917.3	16838666.9	16967158.9	70835121.5	P02769
DLGEEHFK		13459303.0	14218860.2	13667705.6	16433524.9	57779393.7	P02769
DDSPDLPK		12360930.5	15269821.3	12136748.7	13196130.3	52963630.8	P02769
HLVDEPNLIK		12509464.4	15523802.3	11698462.4	11303102.5	51034831.6	P02769
GACLLPK		8236883.2	7515355.2	5062859.6	5724957.3	26540055.3	P02769
FGERRALK		5898108.8	6205452.6	5763294.5	5428501.0	23295356.9	P02769
AWSVARLSQK		5064246.7	5855756.9	4939133.2	4810813.0	20669949.8	P02769
QTALVELLK		3810215.9	4206969.3	3421911.6	4584276.0	16023372.8	P02769
Total intensity of individual labelling channel (top 3 peptides)		129951223.9	139258849.9	121603198.9	132695649.8		

Normalizer = (129951223.9 + 139258849.9 + 121603198.9 + 132695649.8)/4 = 130877230.6

Ratio of protein (P02769) = 129951223.9/130877230.6 : 139258849.9/130877230.6 : 121603198.9/130877230.6 : 132695649.8/130877230.6 = 0.99 : 1.06 : 0.93 : 1.01

Figure S4. The examples of calculating ratios at the spectrum (PSM), peptide and protein level.

Figure S3. The workflow of data processing of a 4-plex labelled sample. The Python scripts can be found at <https://github.com/tianxiaobo002/SMD-IPTL-4-plex-data-processing>.

Figure S5. MS/MS spectrum of Ma-WLYRAK

Figure S6. Slow infusion of maleic anhydride to a peptide solution in 100 mM sodium acetate, pH 5.5. XICs of m/z 418.74, 467.74 and 516.74 were combined for all time points. Peak **a** is the unmodified peptide WLYRAK; peak **b** is the selectively maleylated peptide at the N-terminal α -NH₂ group, Ma-WLYRAK; peak **c** is the double-maleylated peptide at both the α -NH₂ and ϵ -NH₂ groups, Ma-WLYRAK-Ma.

Figure S7. Various catalysts and additives were screened for the thiol Michael addition of acetyl-cysteine to Ma-WLYRAK. XICs of m/z 467.74 and 549.26 were combined for all additives. Peak **a** is the selectively labelled peptide Ma-WLYRAK; peak **b** is the labelled peptide with acetyl-cysteine, Ac-Cys-Ma-WLYRAK. Reaction mixtures were analyzed after 15 h.

Figure S8. Correlation of measured ratios of peak intensities and theoretical ratios of peak intensities at various mixing ratios plotted logarithmically.

Figure S9. Maleylation of Lys-C yeast peptides (n=3). The maleylation yield of each peptide was determined as the percentage of intensity of the N-terminal maleylated form to the total intensity of all possible forms including unmodified, N-terminal maleylated, C-terminal maleylated and both N- and C-terminal maleylated peptide. A labelling yield of 0% was observed for a few peptides, which might be caused by incorrect peptide identification.

Figure S10. 4-plex labelling of the Ma-Lys-C peptides of yeast. The yield is the overall yield of two steps including labelling at the C-terminal Lys and thiol Michael addition at newly maleylated N-terminus. The yield of each peptide was determined as the percentage of intensity of N-terminal Ac-Cys-Ma labelled and C-terminal Ac-Ala labelled forms to the total intensity of all possible forms.

Chapter 3

A collision-induced dissociation cleavable isobaric tag for peptide fragment-ion-based quantification in proteomics

Xiaobo Tian,[†] Marcel P. de Vries,[‡] Hjalmar P. Permentier,[†] and Rainer Bischoff[†]

[†]Department of Analytical Biochemistry and Interfaculty Mass Spectrometry Center, Groningen Research Institute of Pharmacy, University of Groningen, A. Deusinglaan 1, 9713 AV, Groningen, The Netherlands

[‡]Department of Pediatrics, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713 GZ, Groningen, The Netherlands

Manuscript published in *J. Proteome Res.* 2020, 19, 9, 3817–3824

ABSTRACT

Quantifying peptides based on unique peptide fragment ions avoids the issue of ratio distortion that is commonly observed for reporter ion-based quantification approaches. Herein, we present a collision-induced dissociation (CID)-cleavable, isobaric acetyl-isoleucine-proline-glycine (Ac-IPG) tag, which conserves the merits of quantifying peptides based on unique fragments while reducing the complexity of the b-ion series compared to conventional fragment ion-based quantification methods thus facilitating data processing. Multiplex labeling is based on selective N-terminal dimethylation followed by derivatization of the ϵ -amino group of the C-terminal Lys residue of LysC peptides with isobaric Ac-IPG tags having complementary isotope distributions on Pro-Gly and Ac-Ile. Upon fragmentation between Ile and Pro, the resulting y-ions, with the neutral loss of Ac-Ile, can be distinguished between the different labeling channels based on different numbers of isotope labels on the Pro-Gly part and thus contain the information for relative quantification, while b-ions of different labeling channels have the same m/z values. The proteome quantification capability of this method was demonstrated by triplex labeling of a yeast proteome spiked with bovine serum albumin (BSA) over a 10-fold dynamic range. With the yeast proteins as background, BSA was detected at ratios of 1.14 : 5.06 : 9.78 when spiked at 1:5:10 ratios. The raw mass data is available on the ProteomeXchange with the identifier PXD 018790.

1 Introduction

Quantitative information at the proteome level aids in revealing the roles of proteins in biological, physiological and pathological processes¹. The most widely used strategies in mass spectrometry-based proteome-wide quantitative approaches, other than label-free quantification, are based on stable isotope labeling, which allows us to analyze multiple samples simultaneously in a single LC-MS run²⁻⁴. The major advantage of multiplexed labeling is that it allows quantifying peptides/proteins under the same experimental conditions, which circumvents signal variation due to, among others, variable sample loss during work-up and changing ionization efficiency between

injections⁵. As reviewed by Arul et al.⁴, the currently most widely used multiplexing methods, like iTRAQ^{6,7} and TMT⁸⁻¹⁰, are based on isobaric labeling. Peptides from different samples are quantified relative to each other based on isotopically distinct reporter ions in the MS2 spectra. In that way, the complexity of MS1 and MS2 spectra is not increased by the number of labeled samples. However, reporter ion-based quantification methods suffer from the issue of ratio distortion^{5,11-14}, which is caused by co-fragmentation of multiple precursor ions having similar m/z values and retention times. When more than one peptide is fragmented concurrently, the produced reporter ions from different peptides are indistinguishable in the MS2 spectrum resulting in intensities that no longer reflect the ratios of the individual constituent peptides. To address the issue of ratio distortion, narrowing the width of the precursor ion isolation window^{5,13}, gas-phase purification¹² and MultiNotch MS3¹¹ have been reported. These methods alleviate the problem of ratio distortion to some extent but are not widely used, as they require sophisticated equipment combined with scan routines that are difficult to optimize and control.

An alternative isobaric labeling strategy, that does not suffer from ratio distortion, is isobaric peptide termini labeling (IPTL), which quantifies peptides based on isotopically labeled peptide fragments (b- and y-ions). In IPTL¹⁵⁻²⁰, LysC peptides are isobarically labeled with a pair of complementary tags at their N- and C-termini and fragmented into different sets of b- and y-ion clusters upon MS/MS. The fragment ions not only convey sequence information, but each fragment ion cluster contains information to determine the peptide ratio between samples in contrast to single reporter ions in iTRAQ or TMT. While co-fragmentation occurs also in IPTL, fragment ions can be attributed to the correct peptide. Recently, we reported the SMD-IPTL method²¹, in which the multiplexing capacity of IPTL was increased to four-plex with the option to extend it to seven-plex. While SMD-IPTL provides accurate and precise quantitative data, it increases the complexity of MS2 spectra with challenges for subsequent data analysis²². For example, established protein identification software considers additional

peaks in the isotope envelop of fragment ions as unmatched peaks reducing the peptide score^{11,23-25}.

To reduce the complexity of MS2 spectra of IPTL methods and to further improve multiplexing capacity, we propose a novel labeling approach that relies on a CID-cleavable isobaric Ac-IPG tag. In this method, quantitative information is concealed in the Ac-IPG tag which is located on the ϵ -amino group of the C-terminal Lys residue of LysC peptides and revealed upon tandem MS in the form of y-ion clusters containing different numbers of stable isotopes. The b-ions of different labeling channels have the same mass, because the N-terminus is not differentially labeled with stable isotopes. This allows multiplexing and relative peptide and protein quantification based on y-ion clusters while identification is simpler than for conventional IPTL approaches because the MS2 spectra have regular b-ion series.

2 Experiments

2.1 Chemicals and Materials. 2-Chlorotriyl chloride resin, Fmoc-Gly-OH, Fmoc-¹³C₁-Gly-OH, Fmoc-¹³C₂-Gly-OH, Fmoc-Pro-OH, Fmoc-Ile-OH, sodium cyanoborohydride, formaldehyde solution (36.5-38% in H₂O), bovine serum albumin (BSA), acetic anhydride (¹²C₄H₆O₃), acetic anhydride-1,1'-¹³C₂ (¹³C₂¹²C₂H₆O₃), acetic anhydride-¹³C₄ (¹³C₄H₆O₃), 1-[bis(dimethylamino) methylene]-1H-1,2,3-triazolo [4,5-b] pyridinium 3-oxid hexafluorophosphate (HATU), 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (EDC·HCl), piperidine, pyridine, N,N-diisopropylethylamine (DIPEA), trifluoroethanol, p-nitrophenol (PNP), tris(2-carboxyethyl) phosphine hydrochloride (TCEP), 2-iodoacetamide, N,N-dimethylformamide (DMF), acetonitrile (ACN), acetone, dichloromethane (DCM), acetic acid, sodium acetate, sodium hydroxide (NaOH), hydroxylamine hydrate, triethylammonium bicarbonate (TEAB), and trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich. LysC and yeast extracts were purchased from Promega. Synthetic peptide GTDWLANK and FDWA was purchased from GL Biochem (Shanghai, China). Solid phase extraction disk Empore Octadecyl C18 was purchased from 3M.

2.2 Synthesis of isobarically labeled acetyl-isoleucine-proline-glycine-p-nitrophenol ester. The synthesis scheme of triplex isobaric Ac-IPG-PNP tags can be found in **Scheme S1**. 30 mg of 2-chlorotriyl chloride (CTC) resin (1.0-1.8 mmol g⁻¹) was swollen with 1.5 mL of DCM/DMF (1/1) on a thermomixer at 800 rpm for 30 min at room temperature. The swollen resin was washed three times with 0.75 mL of DMF. Then, 1 mL of 0.8 M DIPEA in DMF, 0.5 mL of DMF and 0.2 mmol (60 mg) of Fmoc-Gly-OH, Fmoc-¹³C₁-Gly-OH or Fmoc-¹³C₂-Gly-OH were added. The reaction mixture was shaken for 90 min at room temperature. The solution was filtered out and the remaining isotopically labeled Fmoc-Gly-CTC resin was washed three times with 0.75 mL of DMF. Afterwards, the resin was treated with 1.5 mL of 20% piperidine in DMF for 5 min and 15 min sequentially. Then, 0.5 mL of 0.4 M HATU in DMF, 0.5 mL of 0.8 M DIPEA in DMF, 0.5 mL of DMF and 0.2 mmol (68 mg) Fmoc-Pro-OH were added. The reaction mixture was shaken for 90 min at room temperature. The resulting isotopically labeled Fmoc-Pro-Gly-CTC resin was washed three times with 0.75 mL of DMF. Afterwards, the resin was treated with 1.5 mL of 20% piperidine in DMF for 5 min and 15 min sequentially. Then, 0.5 mL of 0.4 M HATU in DMF, 0.5 mL of 0.8 M DIPEA in DMF, 0.5 mL of DMF and 0.2 mmol (71 mg) Fmoc-Ile-OH were added. The reaction mixture was shaken for 90 min at room temperature. Then, the resin was treated with 1.5 mL of 20% piperidine in DMF for 5 min and 15 min sequentially to afford **3a-c**. 0.5 mL of DCM, 5 μ L of pyridine and 5 μ L of acetic anhydride were added to **3c**; 0.5 mL of DCM, 5 μ L of pyridine and 5 μ L of acetic anhydride-1,1'-¹³C₂ were added to **3b**; 0.5 mL of DCM, 5 μ L of pyridine and 5 μ L of acetic anhydride-¹³C₄ were added to **3a**. The reaction mixtures were shaken for 30 min at room temperature and then the isobarically labeled **4a-c** were released with 0.75 mL of 20% trifluoroethanol in DCM. Subsequently, the dried isobarically labeled **4a-c** were reacted with 80 mg EDC·HCl and 40 mg p-nitrophenol (PNP) in 0.7 mL DCM to generate **5a-c** correspondingly. The reactions were shaken for 3 h, and after drying, the Ac-Ile-Pro-Gly-PNP esters were purified by RP-HPLC with detection at 254 nm. The collected HPLC fraction was freeze-dried to afford a white fluffy solid which was stored at -20 °C until use. Around

3 mg was obtained for each tag. The LC-MS/MS analysis of Ac-Ile-Pro-Gly-PNP esters can be found in **Figure S2**. 5 mg non-isotopically labeled Ac-IPG-PNP was synthesized for NMR analysis (**Figure S3** and **Figure S4**).

2.3 LC purification. For LC purification, the HPLC system (Shimadzu, Kyoto, Japan) was equipped with a SIL-20AC autosampler, two LC-20AT pumps and an SPD-20A absorbance detector. The synthesized Ac-IPG-PNP esters were separated on an XBridge BEH C18 OBD Prep Column (250 mm × 10 mm, 5 μm particles, 130 Å pore size, Waters, Ireland) with a 30 min gradient of 2-90% acetonitrile in water/0.1% formic acid at a flow rate of 2 mL min⁻¹.

2.4 NMR. ¹H NMR and ¹³C NMR spectra were recorded on a Bruker Ascend 600 MHz NMR spectrometer (operating at 600 and 150 MHz, respectively).

2.5 Alkylation and LysC digestion. 100 μg BSA or yeast protein was dissolved in 100 μL 100 mM TEAB. Then, 5 μL 200 mM TCEP was added and the mixture was incubated for 1 h at 37 °C. Afterwards, 5 μL 375 mM freshly prepared 2-iodoacetamide was added and the mixture was incubated for 30 min in the dark at room temperature. Finally, 800 μL acetone was added and the mixture was kept at -20 °C overnight for precipitation. The precipitated protein was centrifuged at 20800 g for 30 min. Subsequently, the supernatant was decanted and alkylated proteins were stored at -20 °C until digestion. 100 μg alkylated BSA or yeast protein was dissolved in 200 μL 100 mM TEAB and 2 μg LysC dissolved in 1 % formic acid was added. The pH was adjusted to 7.5-8 with 1 M TEAB and the digestion was kept overnight at 37 °C. The reaction was terminated by acidification using 10 % TFA and resulting LysC peptide mixtures of BSA and yeast proteins were freeze-dried and stored at -80 °C before conducting the labeling reactions.

2.6 Selective N-terminal dimethylation of peptides. 100 μg of peptide GTDWLANK, LysC peptides of BSA or LysC peptides of yeast proteins were dissolved in 100 μL 1% (aq.) acetic acid. 10 μL of 4% formaldehyde in H₂O was added to the peptide solution and the mixture was shaken for 2 min before adding 15 μL of 600 mM sodium cyanoborohydride dissolved in 1% (aq.) acetic acid^{18,26}. The reaction was shaken for 15 min at room temperature. Excess reagents were removed by SPE using STAGE (STop And Go Extraction) TIPS Desalting Procedure²⁷ and the collected N-terminally dimethylated peptides were freeze-dried and stored at -80 °C.

2.7 Triplex labeling of N-terminally dimethylated peptides. 5 μg of N-terminally dimethylated GTDWLANK or LysC peptides of BSA or LysC peptides of yeast proteins were dissolved in 50 μL 200 mM TEAB buffer of pH 8.5. Then, 4 μL of 50 mM Ac-IP-¹³C₂-G-PNP, ¹³C₁-Ac-IP-¹³C₁-G-PNP and ¹³C₂-Ac-IPG-PNP in DMF were added to 3 peptide solutions, respectively. The reaction mixtures were shaken for 2 h at room temperature. To ensure complete labeling, 2 μL of p-nitrophenol ester was added again and incubated for 1 h more. Any esterification on the hydroxyl groups of Ser, Thr or Tyr and excess PNP ester were hydrolyzed in the presence of 5% hydroxylamine hydrate at room temperature for 5 min and then desalted by SPE using the STAGE (STop And Go Extraction) TIPS Desalting Procedure prior to LC-MS analysis²⁷. 750 μL of 2% acetonitrile (0.1% TFA) was added to remove excess Ac-IPG-COOH before eluting peptides from the STAGE tips, as shown in **Figure S9**.

2.8 LC/MS/MS analysis. All mixed samples of isobarically labeled peptide were analyzed on a Q Exactive Plus and other samples were analyzed on an Orbitrap Velos Pro as detailed below.

LC-MS/MS analyses were performed on an Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific) coupled to an LC system (Shimadzu, Kyoto, Japan) equipped with a SIL-20AC HT autosampler, and two LC-20AD pumps. The sample was separated on an ACQUITY UPLC BEH C18 column (2.1 mm × 50 mm, 1.7 μm

particles, Waters) at a column temperature of 60 °C with a 10 min gradient of 2-90% acetonitrile in water/0.1% formic acid at a flow rate of 300 $\mu\text{L min}^{-1}$.

NanoLC-MS/MS analyses were performed on a Q Exactive Plus mass spectrometer (Thermo Scientific, San José, CA) equipped with an UltiMate 3000 RSLCnano system (Thermo Scientific). The peptides (around 300 ng) were first loaded onto a trap column (Acclaim PepMap 100 pre-column, 75 μm , 2 cm, C18, 3 mm, 100 Å, Thermo Scientific) and then separated on an analytical column (PepMap RSLC C18, 50 cm 75 μm ID, 2 μm , 100 Å, Thermo Scientific). A flow rate of 300 nL min^{-1} and a column temperature of 40 °C were utilized. Solvent A was 0.1% formic acid in water and solvent B was 0.1% formic acid in acetonitrile. A gradient was applied from 2% to 50% or 65% B in the first 60 min, then 85% B for 5 min and equilibration for 15 min with 2% B. Mass spectrometry data were measured using a data-dependent top 10 method. Full MS scans were acquired between m/z 300 and 1650 with a resolution of 70000 (at m/z 200); the automatic gain control (AGC) target was 3E6 and maximum injection time was 50 ms. Precursor ions were selected with a window of 2 Th or 0.8 Th with -0.2 Th as offset. Selected ions were fragmented in the HCD collision cell with stepped normalized collision energies (NCE) of 18, 20, 22, 24, 26, 28, 30 and 32. The tandem (MS2) mass spectra were acquired in the Orbitrap mass analyzer with a resolution of 17500 (at m/z 200). The AGC target was 1E5 and the maximum injection time was 100 ms. The charge exclusion discards ions with a single charge and charges larger than 4 and dynamic exclusion time was 20 s.

2.9 Database Searching and Quantification. LC-MS/MS raw files were analyzed with the Andromeda search engine in SearchGUI^{28,29} and searched against the UniProt reference database of yeast (UP000002311, 6049 entries, downloaded on Jan. 20, 2020) into which the BSA entry (P02769) was inserted manually. LysC was selected as enzyme, digestion mode as specific and max missed cleavage sites as 0. A tolerance of 20 ppm for the precursor ion and 0.05 Da for the MS/MS fragment ions was applied. Carbamidomethylation (+57.02) on cysteine and dimethylation (+28.03) on N-terminus

were set as fixed modification, oxidation (+15.99) on methionine as variable modification. For triplex labeling experiments, variable modifications on Lys were set as triplex isobaric Ac-IPG tags with two ^{13}C isotopes (+311.17) with neutral losses of $^{13}\text{C}_2\text{-Ac-Ile}$ (-157.10), $^{13}\text{C}_1\text{-Ac-Ile}$ (-156.10) and Ac-Ile (-155.10) in MS2 spectra. The 3 labeling channels were searched in one run and the search results were processed with PeptideShaker³⁰.

The peptide spectrum matches (PSM) were exported from the PeptideShaker identified features. All the following steps were performed using in-house built Python scripts. For every PSM of unique peptides, the corresponding theoretical fragment ions with neutral loss of Ac-Ile were calculated and grouped by retention time. The raw data was converted into a Python readable mgf file with RawConverter³¹ to circumvent that the precursor ion m/z value is adjusted to the monoisotopic mass by Andromeda in SearchGUI. All MS/MS spectra were extracted from the resulting mgf file and also grouped by retention time. Afterwards, based on the retention time, the original peak intensities in the mgf file were related to the corresponding theoretical fragment ions and the matched intensities were used to quantify peptides and ultimately proteins. The y_1 ions with neutral loss from the C-terminal lysine and the fragment ions losing H_2O and NH_3 were excluded²¹. The ratios for each PSM, peptide and protein were calculated based on the total peak intensities of the same labeling channel. For fragment ion spectra that were produced from monoisotopic peaks and in which no ^{13}C interference existed, all matched y -ions except y_1 were used for quantification and the minimum number of matched ions was set to 3. For fragment ion spectra that were not produced from monoisotopic peaks only and in which ^{13}C interference was observed in all y -ions, only the small y_2 -ions were used for peptide quantification. Examples of calculating ratios can be found in **Figure S10**. Each peptide ratio was calculated from the three spectra with the highest total peak intensity and each protein ratio was calculated from the three peptides with the highest total intensity. The Python scripts and quantification outputs are available at <https://github.com/tianxiaobo002/Ac-IPG-scripts-and-quantification-outputs-at-multiple-levels>. The raw mass spectrometry proteomics data

of yeast and BSA-yeast samples have been deposited to the ProteomeXchange Consortium via the PRIDE³² partner repository with the dataset identifier PXD 018790.

3 Results and discussion

3.1 Design of Ac-IPG tag and labeling strategy. The CID-cleavable Ac-IPG tag targets the ϵ -amino group of the C-terminal Lys residues of LysC peptides after blocking their N-termini by dimethylation. The Ac-IPG tag consists of three parts: (1) Ac-Ile, which will form a neutral loss fragment upon CID; (2) the Pro-Gly tag, which will remain at the C-terminal Lys upon CID, and (3) an amine reactive p-nitrophenol ester (PNP) (**Figure 1A**). With different combinations of commercially available ^{13}C - and ^{15}N -labeled acetic anhydride, isoleucine, proline and glycine, the Ac-IPG tag has a multiplexing capacity of 3 with an interval of 1 Da with the potential of being extended to the 10-plex level (**Figure S1**). Although deuterium labeled amino acids can extend the multiplexing capacity further, these are generally avoided due to the possibility of varying retention times of the same peptide modified with a different number of deuterium atoms^{33,34}. The peptide bond N-terminal to Pro is known to be fragile upon CID³⁵⁻³⁹. This was exploited in the Ac-IPG tag to induce neutral loss of Ac-Ile upon CID in addition to fragmentation along the peptide backbone. The different isotope distributions of the Pro-Gly tag render the fragment ions distinguishable between the labeling channels and the corresponding peak intensities represent the information for relative quantification (inset in **Figure 1C**).

In order to reduce complexity of the MS2 spectra, the N-termini of peptides were first modified by selective N-terminal dimethylation (**Figure 1B**). This serves two purposes: (1) it blocks the N-terminus of the LysC peptide to permit modifying the ϵ -amino group of the C-terminal Lys with a single isobaric tag per peptide; (2) it preserves ionization efficiency and improves completeness of the b-ion series by converting a primary into a tertiary amine⁴⁰⁻⁴³. As there is no isotope label at the N-terminus, b-ions originating from different samples will have the same mass, reducing complexity of the MS2 spectra. The differentially labeled samples are mixed prior to LC-MS/MS analysis, as

Figure 1. Schematic view of the Ac-IPG approach. A) Functional design of the Ac-IPG-PNP tag (^{13}C isotope locations of the triplex Ac-IPG-PNP tag are marked with asterisk); B) triplex isobaric labeling steps; C) LC-MS/MS for a mixture of triplex labeled samples.

shown in **Figure 1C**. The monoisotopic peak, selected with an isolation window of 0.8 Th and -0.2 Th offset^{5,13,14}, is fragmented into sets of b-ions and y-ion clusters upon CID from which peak intensities of the y-ions are extracted for each labeling channel. Their intensity ratios represent the relative quantification information of each sample in the mixture for that particular peptide.

3.2 Peptide derivatization and optimization of fragmentation energy. To assess the labeling conditions, the fragmentation properties and quantification at the peptide level, we first prepared N-terminally dimethylated GTDWLANK (N-dime-GTDWLANK) and applied FDWA as control peptide. N-terminal dimethylation by reductive amination was highly selective and complete (**Figure S5**) and the yield was higher than 98% (**Figure 2A**). The intensities of both peptides were increased after N-terminal dimethylation, which agrees with the fact that the newly formed tertiary amine ionizes more easily than the original primary amine. N-dime-GTDWLANK was reacted with three isobaric Ac-IPG tags separately in 200 mM TEAB buffer, pH 8.5 at room temperature for 2 h. LC-MS of the products showed that the peptide was efficiently labeled (**Figure 2A**) with a yield of around 98%. There was only a slight decline in intensity upon labeling with the Ac-IPG tag but the intensity of N-dime-GTDWLANK-GPI-Ac was still higher than that of the original unmodified peptide. Fragmentation of N-dime-GTDWLANK-GPI-Ac was studied at different normalized collision energies (NCE)⁴⁴. **Figure 2C** shows the MS2 spectrum at an NCE of 28 (MS2 spectra at NCE values ranging from 18 to 32 are shown in **Figure S6**). Neutral loss of Ac-Ile was complete at an NCE of 18 without any fragmentation of the peptide backbone. The intensity of fragment ions with neutral loss of Ac-Ile increased from NCE 22 to 28 but dropped again when the NCE was increased to 32. NCEs of 26, 28 and 30 were further evaluated with N-dime-BSA peptides-GPI-Ac and an NCE of 28 resulted in the highest number of PSMs (**Figure S15**), which is consistent with the results at the peptide level. Having established the optimal NCE at 28, we determined whether the CID-cleavable isobaric Ac-IPG tag allows for accurate quantification by comparing the intensities of

y-ions of three differentially labeled N-dime-GTDWLANK-GPI-Ac peptides. Separate analyses (**Figure S7**) of the three forms shows them having identical b-ions but different y-ions. When the three differentially labeled forms were mixed at ratios 1:1:1, 1:2:5 and 1:10:20 and analyzed by LC-MS/MS (**Figure 2B** and **Figure S8**), the corresponding measured ratios were 0.94 : 0.99 : 1.06, 1.01 : 2.04 : 4.94 and 1.30 : 9.60 : 20.11, respectively, which closely match the mixing ratios.

Figure 2. Derivatization and quantification of GTDWLANK. A) LC-MS of selective N-terminal dimethylation and derivatization of the C-terminal Lys with isobaric Ac-IPG tags. From top to bottom: extracted ion chromatogram (XIC) of m/z 452.73 and 538.23, the combined XICs of m/z 452.73, 466.74, 480.75 and 566.26 and the combined XICs of m/z 466.74, 566.26 and 622.33. Peak a is unmodified GTDWLANK. Peak b is N-dime-GTDWLANK. Peak c is N-dime-GTDWLANK-¹³C₁-GPI-Ac-¹³C₁, Peak d is unmodified FDWA. Peak e is N-dime-FDWA. Mass spectra are shown to the right of the chromatograms. B) Precursor ion selection of triplex labeled N-dime-GTDWLANK-GPI-Ac and the corresponding y₇ ion in the MS2 spectra at ratios 1:1:1, 1:2:5 and 1:10:20; C) MS2 spectrum of N-dime-GTDWLANK-¹³C₂-GPI-Ac with an NCE of 28.

Figure 3. Analysis of Ac-IPG triplex tagged BSA-derived LysC peptides. A) Log2-normalized ratio distribution at the fragment ion level at a mixing ratio of 1:1:1; B) Log2-normalized ratio distribution at the peptide level at a mixing ratio of 1:1:1; C) Log2-normalized ratios at the peptide level at a mixing ratio of 1:5:10. Expected values for log2-normalized mixing ratios are shown as dotted lines.

3.3 Peptide and protein identification and quantification. Peptide and protein identification and quantification were assessed with N-terminally dimethylated LysC peptides of bovine serum albumin (BSA) that were labeled with triplex isobaric Ac-IPG tags and mixed at different ratios prior to LC-MS/MS analysis. As shown in **Figure 3A**, the log2-normalized intensity ratios of y-ions of triplex-labeled BSA peptides mixed at a 1:1:1 ratio cluster around 0 with better convergence for more intense ions. The overall ratios between different labeling channels at the PSM level were calculated from the sums of all y-ion intensities from the same labeling channel. Since multiple quantification ions, in the form of y-ions, are present in each PSM, random fluctuations of peak intensities of single ions have less influence on the overall ratio (see **Figure S10** for details). Ratios at the peptide level were calculated from the three most intense

PSMs and ratios at the protein level were calculated based on the three most intense peptides. Examples of calculations at the PSM, peptide and protein levels are shown in **Figure S10**. As a result of averaging multiple data points, the log2-normalized ratio distribution at the PSM level (**Figure S11**) and at the peptide level (**Figure 3B**) show good convergence on 0. Subsequently, the triplex labeled LysC peptides of BSA were mixed at a ratio of 1:5:10 (**Figure 3C**), and the medians of log2-normalized peptide ratios were found to be 0.12 : 2.32 : 3.30 (measured ratio: 1.09 : 4.99 : 9.85).

Figure 4. Analysis of Ac-IPG tagged yeast-derived LysC peptides. A) Log2-normalized protein ratio distribution of triplex labeled LysC yeast peptides mixed at a ratio of 1:1:1; B) Log2-normalized

protein ratio distribution of triplex labeled LysC yeast peptides mixed with a ratio of 1:2:5; C) Log2-normalized protein ratio of triplex labeled LysC BSA peptides mixed with a ratio of 1:5:10. Expected values for log2-normalized mixing ratios are shown as dotted lines.

3.4 Labeling efficiency and reproducibility at the level of a complex proteomics sample. The efficiency and reproducibility of the isobaric labeling steps and the quantification accuracy in complex samples were assessed by performing triplicate experiments on LysC peptides from a yeast proteome sample. The yield of selective N-terminal dimethylation was determined as the percentage of intensity of the N-terminally dimethylated form of each identified peptide to the total intensity of all possible forms (unmodified, N-terminally dimethylated, C-terminally dimethylated and both N- and C-terminally dimethylated peptides). As shown in **Figure S12**, more than 98% of all peptides had a labeling yield exceeding 95% with an average coefficient of variation (CV) of 12.4% (n=3). The yield of labeling the C-terminal Lys residues with isobaric Ac-IPG tags was calculated as the ratio of the intensity of the Ac-IPG labeled form to the sum of the Ac-IPG labeled form and the unmodified form. More than 93% of Ac-IPG labeled peptides had a labeling yield higher than 95% in the triplex LysC yeast peptides mixed at a ratio of 1:1:1 (**Figure S13**). The average CV of the labeling yield of the three channels of $^{13}\text{C}_2$ -Ac-IPG, $^{13}\text{C}_1$ -Ac-IP- $^{13}\text{C}_1$ -G and Ac-IP- $^{13}\text{C}_2$ -G was 8.7% (n=3) (**Figure 4A**).

3.5 Quantification of a specific protein within a complex proteomics background. The quantification capability and dynamic range of the Ac-IPG approach in a complex background was assessed by preparing a BSA-yeast proteomics sample consisting of triplex labeled LysC-digested BSA mixed at 1:5:10 and triplex labeled LysC yeast peptides mixed at 1:2:5, as shown in **Figure S14**. 485 yeast proteins were quantified and the medians of the log2-normalized protein ratios of the three labeling channels were determined as 0.00 : 1.09 : 2.29, which is close to the expected values of 0.00 : 1.00 : 2.32 (**Figure 4B**). For triplex labeled BSA mixed at 1:5:10, the log2-normalized

protein ratios were 0.19 : 2.34 : 3.29, respectively, which is close to the expected values of 0.00 : 2.32 : 3.32 (**Figure 4C**).

4 Conclusions

In conclusion, we propose a novel quantification approach relying on a CID-cleavable Ac-IPG tag and selective N-terminal dimethylation. This approach has the advantage that MS2 spectra are less complex than for conventional IPTL, while avoiding the issue of ratio distortion observed for reporter ion-based quantification methods, such as TMT and iTRAQ. The proteome quantification capability of the CID-cleavable Ac-IPG tag method was demonstrated by triplex labeling of a yeast proteome spiked with BSA over a 10-fold dynamic range. The multiplexing capacity can be readily expanded to 10-plex with commercially available ^{13}C - and ^{15}N -labeled acetic anhydride, isoleucine, proline and glycine via the reported synthesis route.

Acknowledgements

We gratefully acknowledge the China Scholarship Council (CSC) for a PhD fellowship to X. Tian. X. Tian thanks Jos Hermans for help with the LC-MS analyses and C. Ye for help with the NMR analyses. This work is partially funded by the Open Technology Program of Toegepaste en Technische Wetenschappen (TTW) with project number 15230, which is financed by the Netherlands Organization for Scientific Research (NWO).

Notes

The authors declare no competing financial interest.

References

- (1) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chemical Reviews* **2013**, *113*, 2343-2394.
- (2) Rauniyar, N.; Yates, J. R. Isobaric Labeling-Based Relative Quantification in Shotgun Proteomics. *Journal of Proteome Research* **2014**, *13*, 5293-5309.
- (3) Chahrour, O.; Cobice, D.; Malone, J. Stable isotope labelling methods in mass spectrometry-based quantitative proteomics. *J Pharmaceut Biomed* **2015**, *113*, 2-20.
- (4) Arul, A. B.; Robinson, R. A. S. Sample Multiplexing Strategies in Quantitative Proteomics. *Analytical Chemistry* **2019**, *91*, 178-189.
- (5) Sonnett, M.; Yeung, E.; Wühr, M. Accurate, Sensitive, and Precise Multiplexed Proteomics Using the Complement Reporter Ion Cluster. *Analytical Chemistry* **2018**, *90*, 5032-5039.
- (6) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlett-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents. *Molecular and Cellular Proteomics* **2004**, *3*, 1154-1169.
- (7) Ow, S. Y.; Cardona, T.; Taton, A.; Magnuson, A.; Lindblad, P.; Stensjö, K.; Wright, P. C. Quantitative Shotgun Proteomics of Enriched Heterocysts from *Nostoc* sp. PCC 7120 Using 8-Plex Isobaric Peptide Tags. *Journal of Proteome Research* **2008**, *7*, 1615-1628.
- (8) 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. *Analytical Chemistry* **2003**, *75*, 1895-1904.
- (9) Dayon, L.; Hainard, A.; Licker, V.; Turck, N.; Kuhn, K.; Hochstrasser, D. F.; Burkhard, P. R.; Sanchez, J. C. Relative Quantification of Proteins in Human Cerebrospinal Fluids by MS/MS Using 6-Plex Isobaric Tags. *Analytical Chemistry* **2008**, *80*, 2921-2931.
- (10) Thompson, A.; Wölmer, N.; Koncarevic, S.; Selzer, S.; Böhm, G.; Legner, H.; Schmid, P.; Kienle, S.; Penning, P.; Höhle, C.; Berfelde, A.; Martinez-Pinna, R.; Farztdinov, V.; Jung, S.; Kuhn, K.; Pike, I. TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-Based Isobaric 16-Plex Tandem Mass Tag Reagent Set. *Analytical Chemistry* **2019**, *91*, 15941-15950.
- (11) Ting, L.; Rad, R.; Gygi, S. P.; Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nature Methods* **2011**, *8*, 937-940.
- (12) Wenger, C. D.; Lee, M. V.; Hebert, A. S.; McAlister, G. C.; Phanstiel, D. H.; Westphall, M. S.; Coon, J. J. Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nature Methods* **2011**, *8*, 933-935.
- (13) Wühr, M.; Haas, W.; McAlister, G. C.; Peshkin, L.; Rad, R.; Kirschner, M. W.; Gygi, S. P. Accurate Multiplexed Proteomics at the MS2 Level Using the Complement Reporter Ion Cluster. *Analytical Chemistry* **2012**, *84*, 9214-9221.
- (14) Winter, S. V.; Meier, F.; Wichmann, C.; Cox, J.; Mann, M.; Meissner, F. EASI-tag enables accurate multiplexed and interference-free MS2-based proteome quantification. *Nature Methods* **2018**, *15*, 527-530.
- (15) Koehler, C. J.; Strozynski, M.; Kozielski, F.; Treumann, A.; Thiede, B. Isobaric Peptide Termini Labeling for MS/MS-Based Quantitative Proteomics. *Journal of Proteome Research* **2009**, *8*, 4333-4341.
- (16) Koehler, C. J.; Arntzen, M. O.; Strozynski, M.; Treumann, A.; Thiede, B. Isobaric Peptide Termini Labeling Utilizing Site-Specific N-Terminal Succinylation. *Analytical Chemistry* **2011**, *83*,

4775-4781.

- (17) Nie, A. Y.; Zhang, L.; Yan, G. Q.; Yao, J.; Zhang, Y.; Lu, H. J.; Yang, P. Y.; He, F. C. In Vivo Termini Amino Acid Labeling for Quantitative Proteomics. *Analytical Chemistry* **2011**, *83*, 6026-6033.
- (18) Koehler, C. J.; Arntzen, M. O.; de Souza, G. A.; Thiede, B. An Approach for Triplex-Isobaric Peptide Termini Labeling (Triplex-IPTL). *Analytical Chemistry* **2013**, *85*, 2478-2485.
- (19) Liu, J. H.; Zhou, Y.; Shan, Y. C.; Zhao, B. F.; Hu, Y. C.; Sui, Z. G.; Liang, Z.; Zhang, L. H.; Zhang, Y. K. A Multiplex Fragment-Ion-Based Method for Accurate Proteome Quantification. *Analytical Chemistry* **2019**, *91*, 3921-3928.
- (20) Jiang, H. C.; Yin, H. R.; Xie, L. Q.; Zhang, Y.; Zhang, L.; Yang, P. Y.; Lu, H. J. A novel triplex isobaric termini labeling quantitative approach for simultaneously supplying three quantitative sources. *Analytica Chimica Acta* **2018**, *1001*, 70-77.
- (21) Tian, X.; de Vries, M. P.; Visscher, S. W. J.; Permentier, H. P.; Bischoff, R. Selective maleylation-directed isobaric peptide termini labelling for accurate proteome quantification. *Analytical Chemistry* **2020**, *92*, 7836-7844.
- (22) Jiang, H.; Yin, H.; Xie, L.; Zhang, Y.; Zhang, L.; Yang, P.-Y.; Lu, H. A novel triplex isobaric termini labeling quantitative approach for simultaneously supplying three quantitative sources. *Analytica Chimica Acta* **2018**, *1001*, 70-77.
- (23) Arntzen, M. Ø.; Koehler, C. J.; Barsnes, H.; Berven, F. S.; Treumann, A.; Thiede, B. IsobariQ: Software for Isobaric Quantitative Proteomics using IPTL, iTRAQ, and TMT. *Journal of Proteome Research* **2011**, *10*, 913-920.
- (24) Xie, L.-Q.; Zhang, L.; Nie, A.-Y.; Yan, G.-Q.; Yao, J.; Zhang, Y.; Yang, P.-Y.; Lu, H.-J. ITMSQ: A software tool for N- and C-terminal fragment ion pairs based isobaric tandem mass spectrometry quantification. *Proteomics* **2015**, *15*, 3755-3764.
- (25) Zheng, N. R.; Shan, Y. C.; Deng, Y. L.; Zhang, Y. K. Novel Algorithm for Identification and Quantification of Proteins Based on Strategy of Isobaric Peptide Termini Labelling. *Chinese J Anal Chem* **2017**, *45*, 1441-1447.
- (26) Qin, H.; Wang, F.; Zhang, Y.; Hu, Z.; Song, C.; Wu, R.; Ye, M.; Zou, H. Isobaric cross-sequence labeling of peptides by using site-selective N-terminus dimethylation. *Chemical Communications* **2012**, *48*, 6265-6267.
- (27) Rappsilber, J.; Ishihama, Y.; Mann, M. Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. *Analytical Chemistry* **2003**, *75*, 663-670.
- (28) Vaudel, M.; Barsnes, H.; Berven, F. S.; Sickmann, A.; Martens, L. SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. *Proteomics* **2011**, *11*, 996-999.
- (29) Barsnes, H.; Vaudel, M. SearchGUI: A Highly Adaptable Common Interface for Proteomics Search and de Novo Engines. *Journal of Proteome Research* **2018**, *17*, 2552-2555.
- (30) Vaudel, M.; Burkhardt, J. M.; Zahedi, R. P.; Oveland, E.; Berven, F. S.; Sickmann, A.; Martens, L.; Barsnes, H. PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nature Biotechnology* **2015**, *33*, 22-24.
- (31) He, L.; Diedrich, J.; Chu, Y. Y.; Yates, J. R. Extracting Accurate Precursor Information for Tandem Mass Spectra by RawConverter. *Analytical Chemistry* **2015**, *87*, 11361-11367.
- (32) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.;

Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yılmaz, Ş.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F., et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **2018**, *47*, D442-D450.

(33) Zhang, R.; Sioma, C. S.; Wang, S.; Regnier, F. E. Fractionation of Isotopically Labeled Peptides in Quantitative Proteomics. *Analytical Chemistry* **2001**, *73*, 5142-5149.

(34) Zhang, R.; Sioma, C. S.; Thompson, R. A.; Xiong, L.; Regnier, F. E. Controlling Deuterium Isotope Effects in Comparative Proteomics. *Analytical Chemistry* **2002**, *74*, 3662-3669.

(35) Dongre, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H. Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model. *Journal of the American Chemical Society* **1996**, *118*, 8365-8374.

(36) Mák, M.; Mezö, G.; Skribanek, Z.; Hudecz, F. Stability of Asp-Pro bond under high and low energy collision induced dissociation conditions in the immunodominant epitope region of Herpes simplex virion glycoprotein D. *Rapid Communications in Mass Spectrometry* **1998**, *12*, 837-842.

(37) Hogan, J. M.; McLuckey, S. A. Charge state dependent collision-induced dissociation of native and reduced porcine elastase. *J Mass Spectrom* **2003**, *38*, 245-256.

(38) Huang, Y.; Triscari, J. M.; Pasa-Tolic, L.; Anderson, G. A.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. Dissociation Behavior of Doubly-Charged Tryptic Peptides: Correlation of Gas-Phase Cleavage Abundance with Ramachandran Plots. *Journal of the American Chemical Society* **2004**, *126*, 3034-3035.

(39) Tiwary, S.; Levy, R.; Gutenbrunner, P.; Salinas Soto, F.; Palaniappan, K. K.; Deming, L.; Berndt, M.; Brant, A.; Cimermanic, P.; Cox, J. High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis. *Nature Methods* **2019**, *16*, 519-525.

(40) Hsu, J. L.; Huang, S. Y.; Chow, N. H.; Chen, S. H. Stable-isotope dimethyl labeling for quantitative proteomics. *Analytical Chemistry* **2003**, *75*, 6843-6852.

(41) Guo, K.; Ji, C.; Li, L. Stable-Isotope Dimethylation Labeling Combined with LC-ESI MS for Quantification of Amine-Containing Metabolites in Biological Samples. *Analytical Chemistry* **2007**, *79*, 8631-8638.

(42) Cho, K.-C.; Kang, J. W.; Choi, Y.; Kim, T. W.; Kim, K. P. Effects of peptide acetylation and dimethylation on electrospray ionization efficiency. *J Mass Spectrom* **2016**, *51*, 105-110.

(43) Fu, Q.; Li, L. De Novo Sequencing of Neuropeptides Using Reductive Isotopic Methylation and Investigation of ESI QTOF MS/MS Fragmentation Pattern of Neuropeptides with N-Terminal Dimethylation. *Analytical Chemistry* **2005**, *77*, 7783-7795.

(44) Olsen, J. V.; 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.

Supporting Information

Supplementary Figures

Scheme S1. Synthesis of the triplex Ac-IPG tag. i) DIPEA, isotopically labeled Fmoc-Gly-COOH (+0, +1, +2); ii, iv and vi) 20 % piperidine in DMF; iii) HATU, DIPEA, Fmoc-Pro-OH; v) HATU, DIPEA, Fmoc-Ile-OH; vii) isotopically labeled acetic anhydride (+0, +2, +4); viii) 20 % trifluoroethanol in DCM; ix) EDCI·HCl, p-nitrophenol.

Figure S1. A) Isotope distributions of 10-plex Ac-IPG tags. The atoms marked with “*” denote ^{13}C or ^{15}N . All isotopic forms can be synthesized from commercially available isotopically labeled amino acids. B) Schematic LC-MS/MS process for a mixture of 10-plex labeled samples.

Figure S2. LC-MS/MS of triplex Ac-IPG-PNP. (A) XIC of m/z 451.21, (+2)-Ac-IPG-(+0)-PNP. (B) Mass spectrum of (+2)-Ac-IPG-(+0)-PNP. (C) MS/MS spectrum of (+2)-Ac-IPG-(+0)-PNP. (D) XIC of m/z 451.21, (+1)-Ac-IPG-(+1)-PNP. (E) Mass spectrum of (+1)-Ac-IPG-(+1)-PNP. (F) MS/MS spectrum of (+1)-Ac-IPG-(+1)-PNP. (G) XIC of m/z 451.21, (+0)-Ac-IPG-(+2)-PNP. (H) Mass spectrum of (+0)-Ac-IPG-(+2)-PNP. (I) MS/MS spectrum of (+0)-Ac-IPG-(+2)-PNP.

Figure S3. ^1H NMR of Ac-IPG-PNP. ^1H -NMR (600 MHz, CDCl_3): δ = 0.89 (t, J = 7.44 Hz, 3H), 0.96 (d, J = 6.81 Hz, 3H), 1.10-1.20 (m, 1H), 1.50-1.65 (m, 1H), 1.75-1.83 (m, 1H), 1.90-1.97 (m, 1H), 2.03 (s, 3H), 2.00-2.10 (m, 1H), 2.10-2.18 (m, 1H), 2.40-2.50 (m, 1H), 3.62-3.90 (m, 2H), 4.25-4.30 (m, 2H), 4.65 (t, J = 8.90 Hz, 1H), 4.67 (dd, J = 8.08, 2.82 Hz, 1H), 6.10 (d, J = 9.09 Hz, 1H), 7.32 (d, J = 9.08 Hz, 2H), 7.55-7.63 (m, 1H), 8.30 (d, J = 9.17 Hz, 2H) ppm.

Figure S4. ^{13}C NMR of Ac-IPG-PNP. ^{13}C -NMR (150 MHz, CDCl_3): δ = 11.05, 15.26, 23.21, 24.56, 25.05, 26.86, 37.82, 41.51, 48.34, 54.87, 59.87, 122.25, 126.08, 145.56, 154.93, 167.42, 170.03, 171.48, 172.90 ppm.

Figure S5. Selective N-terminal dimethylation of peptide GTDWLANK. (A) XIC of m/z 452.73. Peak a is the unmodified peptide. (B) Mass spectrum of unmodified peptide. (C) Combined XICs of m/z 452.73, 466.74 (N-terminally dimethylated) and 480.75 (both N- and C-terminally dimethylated). Peak b is the N-terminally dimethylated peptide. (D) Mass spectrum of N-terminally dimethylated peptide. (E) MS/MS spectrum of unmodified peptide. (F) MS/MS spectrum of N-terminally dimethylated peptide.

Figure S6. MS2 spectra of N-dime-GTDWLANK- $^{13}\text{C}_2$ -GPI-Ac. A) MS2 spectrum at HCD NCE 18; B) MS2 spectrum at HCD NCE 20; C) MS2 spectrum at HCD NCE 22; D) MS2 spectrum at HCD NCE 24; E) MS2 spectrum at HCD NCE 26; F) MS2 spectrum at HCD NCE 28; G) MS2 spectrum at HCD NCE 30. H) MS2 spectrum at HCD NCE 32.

Figure S7. Separate analyses of three different forms of isobarically labeled N-dime-GTDWLANK-GPI-Ac. (A) Mass spectrum of N-dime-GTWLANK- $^{13}\text{C}_2$ -GPI-Ac. (B), (C), (D), (E), (F) respectively show the b_2 , b_6 , y_2 , y_6 and y_7 ion of N-dime-GTWLANK- $^{13}\text{C}_2$ -GPI-Ac. (G) Mass spectrum of N-dime-GTWLANK- $^{13}\text{C}_1$ -GPI-Ac- $^{13}\text{C}_1$. (H), (I), (J), (K), (L) respectively show the b_2 , b_6 , y_2 , y_6 and y_7 ion of N-dime-GTWLANK- $^{13}\text{C}_1$ -GPI-Ac- $^{13}\text{C}_1$. (M) Mass spectrum of N-dime-GTWLANK-GPI-Ac- $^{13}\text{C}_2$. (N), (O), (P), (Q), (R) respectively show the b_2 , b_6 , y_2 , y_6 and y_7 ion of N-dime-GTWLANK-GPI-Ac- $^{13}\text{C}_2$.

Figure S8. MS2 spectra of triplex labeled N-dime-GTDWLANK-GPI-Ac mixed at various ratios. (A), (B), (C) respectively show the y_2 , y_6 and y_7 ions of triplex labeled N-dime-GTDWLANK-GPI-Ac mixed with a ratio of 1:1:1; (D), (E), (F) respectively show the y_2 , y_6 and y_7 ions of triplex labeled N-dime-GTDWLANK-GPI-Ac mixed with a ratio of 1:2:5; (G), (H), (I) respectively show the y_2 , y_6 and y_7 ions of triplex labeled N-dime-GTDWLANK-GPI-Ac mixed with a ratio of 5:2:1; (J), (K), (L) respectively show the y_2 , y_6 and y_7 ions of triplex labeled N-dime-GTDWLANK-GPI-Ac mixed with a ratio of 20:10:1; (M), (N), (O) respectively show the y_2 , y_6 and y_7 ions of triplex labeled N-dime-GTDWLANK-GPI-Ac mixed with a ratio of 1:10:20.

Figure S9. Removing excess Ac-IPG-COOH by 5 times flushing with 2% AcCN, 0.1% TFA (150 μ L x 5) during SPE before eluting peptides. The XICs of m/z 172.96 and 330.18 were combined for every panel. A) first flushing; B) second flushing; C) third flushing; D) fourth flushing; E) fifth flushing. More than 99.5% Ac-IPG-COOH was removed after the fifth flushing step.

Normalizer (weighted average) = (124942.4 + 557515.4 + 1049260.1)/(1+5+10) =108232.4
Spectrum ratio = 124942.4/ 108232.4 : 557515.4/108232.4 : 1049260.1/108232.4 = 1.1:5.1:9.7

Normalizer = (4274330 + 19992482.9 + 34777550.7)/(1+5+10) =3690272.7
Spectrum ratio = 4274330/3690272.7 : 19992482.9/3690272.7 : 34777550.7/3690272.7 = 1.1:5.4:9.4

Figure S10. Examples of calculating ratios at the PSM, peptide and protein level for triplex labeled LysC peptides of BSA mixed at a ratio of 1:5:10. (Python scripts and more quantification outputs are available at <https://github.com/tianxiaobo002/Ac-IPG-scripts-and-quantification-outputs-at-multiple-levels>)

Figure S11. The log2-normalized ratio distribution at the PSM level of BSA at a mixing ratio of 1:1:1.

Figure S12. Yield distribution of triplicate samples of selective N-terminal dimethylation of the LysC yeast peptides.

Figure S13. Ac-IPG labeling yield in the triplex LysC yeast peptides mixed at a ratio of 1:1:1.

Figure S14. Scheme of preparing the BSA-yeast proteomics sample: triplex labeled BSA was mixed at a ratio of 1:5:10 (1 ng: 5 ng: 10 ng) and LysC yeast sample was mixed at a ratio of 1:2:5 (50 ng: 100 ng: 250 ng).

Figure S15. The influence of NCE on fragmentation and identification of LysC peptides of BSA.

Chapter 4

4

A versatile isobaric tag enables proteome quantification in data dependent and data independent acquisition mode

Xiaobo Tian,[†] Marcel P. de Vries,[‡] Hjalmar P. Permentier,[†] and Rainer Bischoff[†]

[†]Department of Analytical Biochemistry and Interfaculty Mass Spectrometry Center, Groningen Research Institute of Pharmacy, University of Groningen, A. Deusinglaan 1, 9713 AV, Groningen, The Netherlands

[‡]Department of Pediatrics, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713 GZ, Groningen, The Netherlands

Manuscript published in *Anal. Chem.* 2020, 92, 24, 16149–16157

ABSTRACT

Quantifying proteins based on peptide-coupled reporter ions is a multiplexed quantitative strategy in proteomics that alleviates the problem of ratio distortion caused by peptide co-fragmentation, as commonly observed in other reporter ion-based approaches, such as TMT and iTRAQ. Fueled by improvements in mass spectrometry and data processing, data-independent acquisition (DIA) is an attractive alternative to data dependent acquisition (DDA) due to its better reproducibility. While multiplexed labeling is widely used in DDA, it is rarely used in DIA, presumably because current approaches lead to more complex MS2 spectra, severe ratio distortion, or to a reduction in quantification accuracy and precision. Herein, we present a versatile acetyl-alanine-glycine (Ac-AG) tag which conceals quantitative information in isobarically labeled peptides and reveals it upon tandem MS in the form of peptide-coupled reporter ions. Since the peptide-coupled reporter ion is precursor-specific, while fragment ions of the peptide backbone originating from different labeling channels are identical, the Ac-AG tag is compatible with both DDA and DIA. By isolating the monoisotopic peak of the precursor ion in DDA, intensities of the peptide-coupled reporter ions represent the relative ratios between constituent samples, whereas in DIA, the ratio can be inferred after deconvoluting the peptide-coupled reporter ion isotopes. The proteome quantification capability of the Ac-AG tag was demonstrated by triplex labeling of a yeast proteome spiked with bovine serum albumin (BSA) over a 10-fold dynamic range. Within this complex proteomics background, BSA spiked at 1:5:10 ratios was detected at ratios of 1.00 : 4.87 : 10.13 in DDA and 1.16 : 5.20 : 9.64 in DIA.

1 Introduction

Advancements in liquid chromatography and mass spectrometry over the last decades have led to the maturation and widespread application of a variety of mass spectrometry-based proteome quantification approaches. Label-free quantification is hampered by the effect of changing ionization efficiencies between LC-MS runs and the stochastic nature of precursor ion selection in the data-dependent acquisition (DDA)

mode. Multiplex stable isotope labeling circumvents these issues and has been widely applied to label samples derived from different experimental conditions. Analysis of multiple samples labeled with distinct isotopes in a single LC-MS run leads to more reliable quantification. As reviewed by Arul et al.¹, isotope label-based quantification methods can be classified into MS1 quantification, which is usually based on isotopic labeling resulting in different precursor masses of the same peptide derived from different samples, and MS2 quantification, which generally relies on isobaric labeling, resulting in identical precursor masses while quantifying based on fragment ions. Isobaric labeling strategies quantify constituent proteins of mixed samples by distinct fragment ions upon MS/MS without increasing the complexity of the MS1 spectra. These methods significantly improve the accuracy and precision as well as the throughput of quantitative proteomics in the DDA mode^{2,3}.

Isobaric strategies can be further classified into three categories: 1) reporter ion-based quantification (e.g., TMT^{2,4,5} and iTRAQ^{6,7}); 2) peptide fragment ion-based quantification, including isobaric peptide termini labeling (IPTL)-based approaches⁸⁻¹⁰; 3) peptide-coupled reporter ion-based quantification, like TMTc^{3,11} and EASI tag¹². Among these, the reporter ion-based methods (TMT and iTRAQ) are most widely used, because of their high multiplexing capacity and well-developed data-processing software. However, it has become increasingly apparent that reporter ion-based quantification methods suffer from ratio distortion^{3,11,13,14} caused by peptide cofragmentation in which the produced reporter ions from different peptides are indistinguishable resulting in intensities that no longer reflect the ratios between the peptides and thus lead to inaccurate protein ratios. Various solutions to alleviate the ratio distortion problem have been described, such as narrowing the width of the precursor ion isolation window^{3,11}, gas-phase purification¹⁴ and MultiNotch MS3^{13,15}, but these methods are not widely used as they rely on specific, state-of-the-art instrumentation, use scan routines that are difficult to optimize and control or require MS3 scanning leading to a longer cycle time. Peptide fragment ion-based quantification, an approach that is not affected by cofragmentation⁸, has thus attracted attention: a

growing number of IPTL-based approaches have been reported, such as Triplex-QITL¹⁶, SMD-IPTL¹⁰ and m-pIDL¹⁷. In these methods, every labeling channel has a set of distinct fragment ions resulting in more precise quantification. However, the increased complexity¹⁸ of MS2 spectra requires custom-made data analysis tools, which has restricted widespread use. In contrast to the two methods mentioned above, quantifying proteins based on a precursor-specific, peptide-coupled reporter ion is not affected by peptide cofragmentation nor does it increase the complexity of the MS2 spectra. This approach was first proposed by Wühr et al.¹¹ in 2012 but two limitations leading to unsatisfactory quantification accuracy and precision were pointed out: 1) the modest efficiency of forming peptide-coupled reporter ions and 2) the complicated isotope envelope of the peptide-coupled reporter ions. The improved TMTc+ approach³ applied a narrow precursor ion isolation window of 0.4 Th rather than the generally used 2 Th to select the monoisotopic peak, which significantly improved the quantification accuracy. The recently introduced EASI tag¹² utilizes a sulfoxide-based tag to increase the formation efficiency of peptide-coupled reporter ions and applied an asymmetric isolation window to specifically isolate the monoisotopic peak to further simplify the isotope envelope of peptide-coupled reporter ions.

Besides the commonly applied DDA strategy, data-independent acquisition (DIA) approaches are becoming increasingly popular, because they avoid the stochastic nature of DDA with its bias towards fragmenting higher intensity peptide ions and the resulting missing value problem between runs¹⁹. In contrast to DDA, isolation of precursors for fragmentation in DIA does not rely on the peak intensities in MS1; instead, all precursors present within the isolation window will be fragmented and analyzed together. As a result, the problem of missing values of specific peptides between injections in a sample batch is largely avoided and reproducibility is improved. On the other hand, MS2 spectra are more complex, as the fragment ions derived from multiple, co-isolated precursor ions are present in the same spectrum, which makes analyzing DIA data more challenging than DDA data. It is worth noting that, fueled by improvements in scan rate and resolution in mass spectrometry and the development of

sophisticated data processing algorithms¹⁹⁻²¹, the analysis of DIA data has been improved over the last decade, especially with respect to the reproducibility from run to run²⁰.

However, DIA-based methods are not suitable for most current multiplexing approaches, which makes them inferior in terms of throughput to DDA-based methods, such as those based on TMT or iTRAQ, for which the reporter ions lose their quantification capability due to massive cofragmentation. For isotopic labeling and peptide fragment ion-based isobaric labeling strategies, such as stable isotope labeling by amino acids in cell culture^{22,23} (SILAC) and IPTL, the peptides of different labeling channels have sets of distinct fragments ions, which multiplies the complexity of MS2 spectra with the number of labeled samples making identification of peptides more challenging. The recently published NeuCoDIA²⁴ and MdFDIA²⁵ methods apply neutron-encoding isotope labeling to achieve multiplexed DIA quantification, but they rely on ultrahigh resolution (resolution of MS2 >120 k), which is not yet widespread and comes at the cost of a decreasing data acquisition rate. Thus, there is a need for a practical approach which can quantify multiplexed proteomics samples in DIA mode without compromising the data acquisition rate.

To establish a new peptide-coupled reporter ion-based tag that avoids ratio distortion in DDA and improves the throughput of DIA-based methods without sacrificing data acquisition rate, we propose a novel isobaric Ac-AG tag, which conceals quantitative information in isobarically labeled peptides and reveals it upon tandem MS in the form of peptide-coupled reporter ions. The ratios between labeling channels are simply reflected by the intensities of the peptide-coupled reporter ions in DDA mode, whereas the ratios can also be deduced from the peptide-coupled reporter ion isotope envelope in DIA mode after deconvolution. The Ac-AG tag is thus the first isobaric tag that can be applied for multiplexed quantitation in both DDA and DIA mode without having to resort to ultra-high MS2 resolution.

2 Experiments

Details of chemicals and materials, the synthesis of triplex Ac-AG-tags (acetyl-alanine-glycine-p-nitrophenol, Ac-AG-PNP, see **Figure 1C**), reduction/alkylation and LysC digestion, selective N-terminal dimethylation of peptides and LC-MS/MS analysis can be found in the Supporting Information.

2.1 Triplex labeling of N-terminally dimethylated peptides. 5 µg of N-terminally dimethylated GTDWLANK, LysC peptides of BSA or LysC peptides of yeast proteins were dissolved in 50 µL of 200 mM triethylammonium bicarbonate (TEAB) buffer of pH 8.5. Then, 4 µL of 50 mM Ac-A-¹³C₂-G-PNP, ¹³C₁-Ac-A-¹³C₁-G-PNP and ¹³C₂-Ac-AG-PNP in DMF were added to 3 peptide solutions, respectively. The reaction mixtures were shaken for 2 h at room temperature. To ensure complete labeling, 2 µL of the respective Ac-AG-PNP reagent was added again and incubated for 1 h more. Any esterification on the hydroxyl groups of Ser, Thr or Tyr and excess PNP ester were hydrolyzed in the presence of 5% hydroxylamine hydrate at 55 °C for 5 min and samples were then desalted by SPE using the STAGE (STop And Go Extraction) TIPS Desalting Procedure prior to LC-MS analysis²⁶. 500 µL of 2% acetonitrile in water with 0.1% trifluoroacetic acid (TFA) was added to remove excess Ac-AG-COOH before eluting peptides from the STAGE tips with 60% acetonitrile in water with 0.1% TFA.

2.2 Database Searching and Quantification. LC-MS/MS raw files measured in DDA mode were analyzed with PEAKS_X+ (Bioinformatics Solutions Inc., Waterloo, Ontario, Canada) and searched against the Uniprot reference database of yeast (UP000002311, 6049 entries, downloaded on Jan. 20, 2020) into which the BSA entry (P02769) was inserted manually. LysC was selected as enzyme, digestion mode as specific and max missed cleavage sites as 0. A tolerance of 20 ppm for the precursor ion and 0.02 Da for the MS/MS fragment ions was applied. Carbamidomethylation (+57.02) on cysteine and dimethylation (+28.03) on the N-terminus were set as fixed modifications and oxidation (+15.99) on methionine as variable modification. For

triplex labeling experiments, variable modifications on Lys were set as triplex isobaric Ac-AG tags with two ¹³C isotopes (+172.08). The results were filtered with a false discovery rate (FDR) of 0.5% for peptides. The filtered results were exported as spectral library for DIA analysis using of the default parameters of PEAKS X+.

For identification of peptides in DIA, the raw data was searched against the prepared spectral library and the Uniprot reference database of yeast (UP000002311) complemented with BSA was selected as the PEAKS reference database. A tolerance of 20 ppm for the precursor ion and 0.02 Da for the MS/MS fragment ions was applied. The DIA results were also filtered with an FDR of 0.5% for peptides.

The PEAKS output of the peptides matched to proteins (*protein-peptides.csv*) and the peptide-spectrum matches (*DB search psm.csv*) were exported for both DDA and DIA searches. All the following steps were performed using in-house built Python scripts (available at https://github.com/tianxiaobo002/Ac-AG-tag_scripts_and_quantification_outputs). Only the unique peptides extracted from the *protein-peptides.csv* files were used for quantification. For every peptide spectrum match (PSM) that was derived from unique peptides in the *DB search psm.csv* files, the theoretical peptide-coupled reporter ion was calculated and grouped by the scan number. The raw data was converted into a Python readable mgf file with RawConverter²⁷. All MS/MS spectra were extracted from the resulting mgf file and also grouped by the scan number. Afterwards, based on the scan number, the measured peak intensities in the mgf file were related to the corresponding theoretical peptide-coupled reporter ions and the matched intensities were used to quantify peptides and ultimately proteins. The ratio of each PSM, peptide and protein were calculated based on the peak intensities of the respective labeling channel. For quantification in DDA, the intensities of peptide-coupled reporter ions simply represent the relative quantification information. Examples of calculating ratios at the PSM, peptide and protein levels can be found in **Figure S1**. Each peptide ratio was calculated from the three PSM with the highest total peak intensity and each protein ratio was calculated from the three peptides with the

highest total peak intensity. For quantification in DIA, the only difference from the quantification in DDA is the correction of the ^{13}C contribution in the isotope envelope of peptide-coupled reporter ions before calculating ratios at the PSM level. Based on the peptide sequence derived from the PSM, the molecular formula is established and from this the native isotope envelope distribution of peptide-coupled reporter ions can be calculated, and used to correct the ^{13}C contribution. The corrected intensities were used for all following calculations. Examples of correcting ^{13}C contributions can be found in **Figure S2**.

The Python scripts and quantification outputs of triplex Ac-AG labeled BSA, yeast and BSA-yeast samples at the PSM, peptide and protein level are available at https://github.com/tianxiaobo002/Ac-AG-tag_scripts_and_quantification_outputs. The raw mass spectrometry proteomics data of yeast and BSA-yeast samples have been deposited to the ProteomeXchange Consortium via the PRIDE²⁸ partner repository with the dataset identifier PXD021187.

3 Results and discussion

3.1 Design of the Ac-AG tag and labeling strategy. During development of the SMD-IPTL¹⁰ method we observed that the amide bond between the incorporated Ac-Ala and the side chain of the C-terminal Lys residue (produced by LysC protein digestion) fragments prior to the peptide backbone at lower stepped normalized collision energies (NCEs) generating a fragment ion covering the full sequence. We therefore assumed that an Ac-Ala-Gly tag on the side chain of Lys would have similar dissociation features and that the peptide bond at the C-terminal side of Gly would remain stable upon collision-induced dissociation (CID)²⁹⁻³³. Fragmentation between the Gly and Ac-Ala parts would thus generate a peptide-coupled reporter-ion at low NCEs and fragment ions of the peptide backbone would be generated at higher NCEs.¹² In this way, the ions required for identification and quantification can be produced in the same MS2 spectrum by applying two NCEs.¹² Multiple NCE values can be set in the Orbitrap acquisition software; when more than one value is set, the mass spectrometer will

perform a stepwise fragmentation of the precursor ion and all fragments created in the steps are collected and sent to the Orbitrap analyzer for detection in one scan. Since ‘Ac-Ala does not have a good ionization site, it is expected to give rise to a neutral loss, and the peptide-coupled reporter ion with the attached Gly part from the Ac-AG tag will have the same charge state as the precursor ion¹².

With different combinations of commercially available ^{13}C -labeled acetic anhydride, alanine and glycine, triplex isobaric Ac-AG tags were prepared as shown in **Figure 1C** and **Scheme S1** (Supporting Information). The p-nitrophenol (PNP) ester was selected instead of the more widely used N-hydroxysuccinimide (NHS) ester, because it has comparable reactivity with the epsilon amino group of Lys but a much better stability in the presence of water, which is important during purification of the synthesized Ac-AG-PNP tag. As derivatization of the C-terminal Lys with the Ac-AG tag might decrease the peptide ionization efficiency, the N-terminal amino groups of the peptides were dimethylated converting them into tertiary amines, which improves ionization³⁴⁻³⁷. Dimethylation has also been reported to improve the completeness of the b-ion series³⁷. As shown in **Figure 1A**, the isobaric labeling method includes two steps: 1) selective N-terminal dimethylation^{38,39}; and 2) labeling with the Ac-AG-PNP ester at the epsilon amino group of the C-terminal Lys residue. As there is no isotope label at the N-terminus and y-ions are expected to contain the full Ac-AG tag, all fragment ions of the peptide backbone originating from different labeling channels will have the same respective masses, which means that the complexity of MS2 spectra does not increase with the number of differentially labeled samples. Only the peptide-coupled reporter ions with different isotopic forms of Gly, produced after loss of the complementarily labeled Ac-Ala part, will have different masses between labeling channels.

The differentially labeled samples are mixed prior to LC-MS/MS analysis and the data acquisition can be performed in DDA or DIA mode, as shown in **Figure 1B**. In the analysis with DDA, a narrow precursor isolation window (e.g. 0.6 Th) allows isolation of the monoisotopic peak for fragmentation and the resulting peptide-coupled reporter ions provide the quantitative readout with their respective intensities representing the

relative ratios of the constituent peptides/proteins³. In the case of DIA, the peptide-coupled reporter ions of different labeling channels form an isotope envelope due to the respective ¹³C contributions, since the entire isotope envelope of the precursor is selected for fragmentation. The quantification information can be inferred from DIA data after deconvolution (see **Figure S2** for details).

Figure 1. Schematic view of the Ac-AG approach. A) Isobaric labeling steps; B) LC-MS/MS of a mixture of triplex-labeled samples with DDA or DIA approaches; C) Functional design of the triplex-labeled Ac-AG-PNP tag (¹³C isotope locations are marked with asterisk).

3.2 Derivatization with the Ac-AG tag at the peptide level. To evaluate the efficiency of the labeling steps and investigate the influence of modification with the Ac-AG tag in combination with N-terminal dimethylation on ionization efficiency and charge state

distribution, GTDWLANK was selected as model peptide. FDWA was included as a reference peptide, since it cannot react with the Ac-AG tag and will thus only show the effects on ionization efficiency and charge state distribution due to N-terminal dimethylation. By normalizing ionization efficiency with FDWA before and after labeling with Ac-AG tag, the influence of Ac-AG tag can be evaluated. The extracted ion chromatograms (XICs) of GTDWLANK, N-terminally dimethylated (N-dime)-GTDWLANK and N-dime-GTDWLANK-GA-Ac were compared. As shown in **Figure S3**, GTDWLANK was efficiently converted to N-dime-GTDWLANK-GA-Ac with yields higher than 98% for both selective N-terminal dimethylation and labeling with the Ac-AG tag. We observed that the intensities of both model and reference peptides increased after N-terminal dimethylation and that there was a slight decline in intensity of the model peptide upon labeling with the Ac-AG tag. However, the intensity was still higher than that of the original unmodified peptide. The charge state distribution of the model peptide before and after labeling with the Ac-AG tag showed that the doubly-charged ion remains the most abundant (96.5%) species after modification (**Table S1**).

3.3 Optimizing the fragmentation energy for Ac-AG labeled peptides. The peptide-coupled reporter ion is used for quantification and fragment ions of the peptide backbone are required for identification. Formation of both types of fragment ions at sufficient yields is thus a prerequisite for Ac-AG tag-based quantification. Fragmentation of N-dime-GTDWLANK-GA-Ac was studied at different NCEs ranging from 18 to 32 (**Figure S4**). At an NCE of 18, the peptide-coupled reporter ion produced from the neutral loss of Ac-Ala was the dominant ion compared to the y- and b-ions from peptide backbone fragmentation (**Figure 2A**). As the NCE was increased stepwise from 18 to 28, the intensity of the peptide-coupled reporter ion gradually decreased and the ion was no longer detectable at NCE of 28 (**Figure S4**). On the contrary, the intensities of ions from peptide backbone fragmentation increased from NCE 18 to 28 but dropped again when the NCE was increased to 32. Based on these results, an NCE of 18 was selected as the energy for producing the peptide-coupled

A versatile isobaric tag enables proteome quantification in data dependent and data independent acquisition mode

reporter ion and an NCE of 28 for forming b- and y-ion fragments. **Figure 2B** shows the MS2 spectrum at a combined NCE of 18 and 28.

Figure 2. LC-MS/MS analysis of triplex-labeled GTDWLANK. A) MS2 spectrum of triplex-labeled N-dime-GTDWLANK-GA-Ac with an NCE of 18 acquired in the parallel reaction monitoring mode; B) MS2 spectrum of triplex-labeled N-dime-GTDWLANK-GA-Ac with a combined NCE of 18 and 28 acquired in the DDA mode; C) precursor isolation with a window of 0.6 Th and peptide-coupled reporter ions in the DDA MS2 spectra at ratios of 1:10:20, 1:2:5, 1:1:1, 5:2:1 and 20:10:1.

The three labeled forms of N-dime-GTDWLANK-GA-Ac were first analyzed separately and found to have identical fragment ions of the peptide backbone, but a distinct peptide-coupled reporter ion (**Figure S5**). To demonstrate the feasibility of the Ac-AG tag approach, triplex-labeled N-dime-GTDWLANK-GA-Ac peptides were mixed at ratios of 1:1:1, 1:2:5, 1:10:20, 5:2:1 and 20:10:1 followed by LC-MS/MS

analysis. The measured ratios were 0.93:1.03:1.04, 0.98:1.98:5.04, 1.29:9.95:19.75, 5.05:1.94:1.02 and 20.15:9.69:1.16, respectively (**Figure 2C**).

Figure 3. Analysis of Ac-AG triplex-tagged BSA-derived LysC peptides. A) Log2-normalized ratio distribution at the PSM level at a mixing ratio of 1:1:1; B) Log2-normalized ratio distribution at the peptide level at a mixing ratio of 1:1:1; C) Log2-normalized ratio at the peptide level at a mixing ratio of 10:5:1; D) Log2-normalized ratio at the peptide level at a mixing ratio of 1:5:10. Expected values for log2-normalized mixing ratios are shown as dotted lines.

3.4 DDA analysis of triplex-labeled LysC peptides of BSA. Triplex Ac-AG labeled N-terminally dimethylated LysC peptides of BSA were prepared, mixed at various ratios, and analyzed with LC-MS/MS in DDA mode with a narrow, 0.6 Th, precursor isolation window. To investigate the effect of different fragmentation energies on identification, NCEs of 24, 26, 28 and 30 were evaluated and the PSMs were counted to establish the optimal NCE for forming fragments of the peptide backbone. As shown in **Figure S6**, an NCE of 28 produced more PSM than other NCEs, which is consistent

with the result of N-dime-GTDWLANK-GA-Ac. As shown in **Figure 3A**, the log₂-normalized ratios at the PSM level of the triplex-labeled BSA sample mixed at a ratio of 1:1:1 converge to zero for more intense PSM. The corresponding peptide ratios were calculated based on the three PSMs with the highest total peak intensity. As shown in **Figure 3B**, the log₂-normalized peptide ratios of the triplex-labeled BSA sample mixed at a ratio of 1:1:1 also converge to zero. Likewise, the protein ratio can be calculated from the three peptides with the highest total peak intensity. The ratios of the 1:1:1 mixed BSA sample were determined as 0.93 : 1.07 : 1 at the protein level. Calculation details at the PSM, peptide and protein level can be found in **Figure S1**. Based on the medians of the log₂-normalized peptide ratios, the 10:5:1 and 1:5:10 mixed triplex-labeled BSA sample were determined to be 9.51 : 5.24 : 1.24 and 1.16 : 5.03 : 9.78, as shown in **Figure 3C** and **Figure 3D**, respectively.

3.5 DDA analysis of triplex-labeled LysC peptides of the yeast proteome. To confirm the efficiency and reproducibility of the isobaric labeling steps and the quantification accuracy in complex samples, triplex-labeled LysC peptides of yeast proteins were prepared. An evaluation of the efficiency and reproducibility of selective N-terminal dimethylation at the proteome scale was reported in our previous work⁴⁰. for N-terminal dimethylation, more than 98% of LysC yeast peptides had a labeling yield exceeding 95%. To assess the labeling yield of the Ac-AG tag at the C-terminal Lys, Ac-AG was set as variable modification in the database search to simultaneously identify Ac-AG labeled peptides and peptides without the Ac-AG modification. Assuming that the peptides with and without modification have comparable ionization efficiencies (as shown in **Figure S3**), the Ac-AG labeling efficiency can be calculated, and more than 99% of all peptides had a labeling yield higher than 98%. Moreover, only 50 proteins were identified when no Ac-AG modification selected and most of them were identified by the C-terminal peptide of protein. The charge state distribution of unlabeled LysC yeast peptides and N-dime-LysC peptides-GA-Ac were compared showing that the doubly-charged peptide precursor ions remain most abundant after

labeling with N-terminal dimethylation and the Ac-AG tag (**Figure S7**); the number of identified peptides remain comparable. Around 90% (1784 out of 1996) of PSMs had peptide-coupled reporter ions with a complete isotope envelope which can be used for quantification. Around 96% (9835 out of 10255) matched y-ions ions did not show the neutral loss due to fragmentation of the Ac-AG tag. The log₂-normalized protein ratios of the triplex-labeled LysC yeast sample mixed at a ratio of 1:1:1 are shown in **Figure 4A**. Based on the medians of log₂-normalized protein ratios (0.03, 0.01 and 0.01), the medians of measured ratios of 1 : 1 : 1 mixed yeast sample were measured as 1.02 : 1.00 : 1.00. The three channels have a standard deviation of measured ratios of 0.12, 0.11 and 0.12, respectively.

Figure 4. Analysis of Ac-AG tagged peptides from the LysC-digested yeast proteome. A) Log₂-normalized protein ratio distribution of triplex-labeled LysC yeast peptides mixed at a ratio of 1:1:1; B) Log₂-normalized protein ratio distribution of triplex-labeled LysC yeast peptides mixed at a ratio of 1:2:5 combined with LysC BSA peptides mixed at a ratio of 1:5:10. Expected values for log₂-normalized mixing ratios are shown as dotted lines.

3.6 Quantifying BSA in a background of yeast proteins in DDA mode. To examine the quantification capabilities and dynamic range of the Ac-AG tag in a complex background, a BSA-yeast proteomics sample consisting of triplex-labeled LysC-digested BSA mixed at 1:5:10 and triplex-labeled LysC yeast peptides mixed at 1:2:5 was prepared (**Figure S8A**) followed by LC-MS/MS analysis. In the BSA-yeast sample, 660 proteins were identified and 600 proteins were quantified by the peptide-coupled reporter ions. Around 60 identified proteins could not be quantified, due to the low quality of PSMs lacking a complete isotope envelope of peptide-coupled reporter ions or incorrect peptide assignments. As shown in **Figure 4B**, the medians of log2-normalized protein ratios of the LysC yeast sample mixed at a ratio of 1:2:5 were determined to be 0.00, 1.05 and 2.30 which are close to the expected values of 0.00, 1.00 and 2.32, respectively. For the triplex-labeled BSA sample mixed at a ratio of 1:5:10, the measured ratio was determined to be 1.00 : 4.87 : 10.13.

3.7 Exploiting the Ac-AG tag for multiplex quantification in the DIA mode. The recently reported DIA multiplexed approaches, NeuCoDIA²⁴ and MdFDIA²⁵ incorporate isotopes by cell culture, while the mdDiLeu approach uses chemical labeling⁴¹. These methods employ neutron encoding⁴², which makes them rely on ultrahigh resolution (> 120 k) at the fragment ion level with, as a consequence, a reduced data acquisition rate. In contrast, the Ac-AG tag not only can be used to chemically label peptides derived from any type of sample and but the peptide-coupled reporter ions produced with a combined NCE of 18 and 28 as the fragmentation condition in DDA can be properly differentiated with a resolution of 17.5 k for the MS2 spectra.

In contrast to the DDA mode with a narrow precursor ion isolation window, wide isolation windows are used in DIA meaning that the complete isotope envelope of precursor ions is selected for fragmentation. As a result, the isotope envelopes of peptide-coupled reporter ions of adjacent labeling channels overlap due to the contribution of heavier (¹³C) isotopes. An isolation window of 29 Th was applied with

a 2 Th overlap between adjacent isolation windows to capture the entire isotope envelope of a given precursor ion in at least one isolation window. As shown in **Figure 5A**, the complete isotope envelope was observed for the peptide-coupled reporter ion of N-dime-GTDWLANK-GA-Ac-(+2), as well as for different mixing ratios of triplex-labeled N-dime-GTDWLANK-GA-Ac. The intensities of the individual isotopic peaks of the isotope envelope of the peptide-coupled reporter ion contain the information that is needed for quantification. Different with DDA, it is necessary to deconvolve the overlapping isotopic envelopes of peptide-coupled reporter ions from the different labeling channels prior to using the corrected intensities of the monoisotopic peaks for ratio calculations at the PSM level. The peptide sequence, determined based on the PSM, is converted to the corresponding molecular formula of the peptide-coupled reporter ion allowing to calculate its native isotope envelope distribution. This distribution is subsequently used to correct the intensity of the respective monoisotopic peaks. (see **Figure S2** for further details). The corrected intensities of all monoisotopic peaks were used for the following ratio calculations at the PSM, peptide and protein levels.

To determine whether the DIA approach gives accurate quantitative results, we analyzed samples of triplex-labeled LysC peptides of BSA mixed at ratios of 1:1:1, 1:2:5, 1:10:20, 5:2:1 and 20:10:1 in DIA mode with a 29 Th isolation window. All the measured ratios were close to the original mixing ratios, as shown in **Figure S9**. We subsequently analyzed a BSA-yeast proteome sample consisting of triplex-labeled LysC-digested BSA mixed at ratios of 1:5:10 and triplex-labeled LysC yeast peptides mixed at ratios of 5:2:1 (**Figure S8B**). Using the spectral library established based on identification results from the previous DDA analysis, 450 proteins were identified and 416 proteins were quantified. As shown in **Figure 5B**, based on the medians (2.31, 0.97 and 0.12) of log2-normalized protein ratios, the medians of measured ratios of 5 : 2 : 1 mixed yeast sample were determined as 4.96 : 1.96 : 1.09. For the triplex-labeled BSA sample mixed at a ratio of 1:5:10, the measured ratio was determined to be 1.16 : 5.2 : 9.64.

Figure 5. Analysis of triplex-labeled N-dime-GTDWLANK-GA-Ac (A) and BSA-yeast in the DIA mode (B). A) From left to right, peptide-coupled reporter ion of a single channel of the precursor ion of N-dime-GTDWLANK-GA-Ac (+2) and the peptide-coupled reporter ions in the DIA MS2 spectra at ratios of 1:2:5, 5:2:1, 1:10:20, and 20:10:1; B) Analysis of the BSA-yeast sample in the DIA mode. Log2-normalized protein ratio distribution of triplex-labeled LysC yeast peptides mixed at a ratio of 5:2:1 combined with LysC BSA peptides mixed at a ratio of 1:5:10. Expected values for log2-normalized mixing ratios are shown as dotted lines.

Based on the aforementioned results, we believe that the Ac-AG tag is an attractive alternative of peptide-coupled reporter ion quantitation to TMTc³ and EASI tag¹². Although the Ac-AG tag exhibits good quantification capability across a 10-fold dynamic range in both DDA and DIA mode, there is still room for improvement: 1) The labeling capacity; even though the capacity can likely be increased with commercially available isotopically labeled amino acids, it compares unfavorably to the 16-plex that

can be achieved with TMTpro², although TMTpro cannot be used in DIA mode. 2) The width of the precursor isolation window in DDA mode; the currently used isolation window of 0.6 Th is not small enough to clearly isolate the monoisotopic peak from a triply-charged or quadruply-charged precursor ion where the intervals between isotopic peaks are 0.33 and 0.25, respectively. So, for quantifying triply- or quadruply-charged peptides, isotope cluster deconvolution is also required. 3) Although the Ac-AG tag exhibits good quantification capability across a 10-fold dynamic range in DIA mode, we have to clarify that the deconvolution rules applied in this manuscript is based on the correct intensities of individual peaks in the isotope envelope of peptide-coupled reporter ions. Inaccurate ratios can be produced when more than one peptide-coupled reporter ion isotope envelope overlaps.

4 Conclusions

We describe an isobaric labeling approach based on the Ac-AG tag that maintains the advantages of existing peptide-coupled reporter ion-based quantification methods in DDA but also allows multiplexing in the DIA mode without sacrificing the rate of data acquisition and complicating MS2 spectra. While the labeling capacity is currently limited to triplex, it is potentially expandable with commercially available isotopically labeled amino acids. A combined NCE of 18 and 28 is used during data acquisition to generate both the peptide-coupled reporter ion for quantification and the peptide backbone fragment ions for identification in the same MS2 spectrum. The proteome quantification capabilities in DDA and DIA were demonstrated by triplex labeling of a yeast proteome spiked with BSA over a 10-fold dynamic range. The Ac-AG tag is easily prepared and makes multiplexed quantification in both DDA and DIA modes accessible to a wider community of proteome researchers.

Acknowledgements

We gratefully acknowledge the China Scholarship Council (CSC) for a PhD fellowship to X. Tian. X. Tian thanks Jos Hermans for helping with the LC-MS analyses. This

work is partially funded by the Open Technology Programme of Toegepaste en Technische Wetenschappen (TTW) with project number 15230 which is financed by the Netherlands Organisation for Scientific Research (NWO).

Notes

The authors declare no competing financial interest. The raw mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE²⁸ partner repository with the dataset identifier PXD021187.

References

- (1) Arul, A. B.; Robinson, R. A. S. Sample Multiplexing Strategies in Quantitative Proteomics. *Analytical Chemistry* **2019**, *91*, 178-189.
- (2) Thompson, A.; Wölmer, N.; Koncarevic, S.; Selzer, S.; Böhm, G.; Legner, H.; Schmid, P.; Kienle, S.; Penning, P.; Höhle, C.; Berfelde, A.; Martinez-Pinna, R.; Farztdinov, V.; Jung, S.; Kuhn, K.; Pike, I. TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-Based Isobaric 16-Plex Tandem Mass Tag Reagent Set. *Analytical Chemistry* **2019**, *91*, 15941-15950.
- (3) Sonnett, M.; Yeung, E.; Wuhr, M. Accurate, Sensitive, and Precise Multiplexed Proteomics Using the Complement Reporter Ion Cluster. *Analytical Chemistry* **2018**, *90*, 5032-5039.
- (4) 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. *Analytical Chemistry* **2003**, *75*, 1895-1904.
- (5) Dayon, L.; Hainard, A.; Licker, V.; Turck, N.; Kuhn, K.; Hochstrasser, D. F.; Burkhard, P. R.; Sanchez, J. C. Relative Quantification of Proteins in Human Cerebrospinal Fluids by MS/MS Using 6-Plex Isobaric Tags. *Analytical Chemistry* **2008**, *80*, 2921-2931.
- (6) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlett-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents. *Molecular and Cellular Proteomics* **2004**, *3*, 1154-1169.
- (7) Ow, S. Y.; Cardona, T.; Taton, A.; Magnuson, A.; Lindblad, P.; Stensjö, K.; Wright, P. C. Quantitative Shotgun Proteomics of Enriched Heterocysts from *Nostoc* sp. PCC 7120 Using 8-Plex Isobaric Peptide Tags. *Journal of Proteome Research* **2008**, *7*, 1615-1628.
- (8) Koehler, C. J.; Strozynski, M.; Kozielski, F.; Treumann, A.; Thiede, B. Isobaric Peptide Termini Labeling for MS/MS-Based Quantitative Proteomics. *Journal of Proteome Research* **2009**, *8*, 4333-4341.
- (9) Koehler, C. J.; Arntzen, M. O.; Strozynski, M.; Treumann, A.; Thiede, B. Isobaric Peptide Termini Labeling Utilizing Site-Specific N-Terminal Succinylation. *Analytical Chemistry* **2011**, *83*, 4775-4781.
- (10) Tian, X.; de Vries, M. P.; Visscher, S. W. J.; Permentier, H. P.; Bischoff, R. Selective Maleylation-Directed Isobaric Peptide Termini Labeling for Accurate Proteome Quantification. *Analytical Chemistry* **2020**, *92*, 7836-7844.
- (11) Wühr, M.; Haas, W.; McAlister, G. C.; Peshkin, L.; Rad, R.; Kirschner, M. W.; Gygi, S. P. Accurate Multiplexed Proteomics at the MS2 Level Using the Complement Reporter Ion Cluster. *Analytical Chemistry* **2012**, *84*, 9214-9221.
- (12) Winter, S. V.; Meier, F.; Wichmann, C.; Cox, J.; Mann, M.; Meissner, F. EASI-tag enables accurate multiplexed and interference-free MS2-based proteome quantification. *Nature Methods* **2018**, *15*, 527-530.
- (13) Ting, L.; Rad, R.; Gygi, S. P.; Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nature Methods* **2011**, *8*, 937-940.
- (14) Wenger, C. D.; Lee, M. V.; Hebert, A. S.; McAlister, G. C.; Phanstiel, D. H.; Westphall, M. S.; Coon, J. J. Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nature Methods* **2011**, *8*, 933-935.
- (15) McAlister, G. C.; Nusinow, D. P.; Jedrychowski, M. P.; Wuhr, M.; Huttlin, E. L.; Erickson, B. K.; Rad, R.; Haas, W.; Gygi, S. P. MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed

Detection of Differential Expression across Cancer Cell Line Proteomes. *Analytical Chemistry* **2014**, *86*, 7150-7158.

(16) Jiang, H. C.; Yin, H. R.; Xie, L. Q.; Zhang, Y.; Zhang, L.; Yang, P. Y.; Lu, H. J. A novel triplex isobaric termini labeling quantitative approach for simultaneously supplying three quantitative sources. *Analytica Chimica Acta* **2018**, *1001*, 70-77.

(17) Liu, J. H.; Zhou, Y.; Shan, Y. C.; Zhao, B. F.; Hu, Y. C.; Sui, Z. G.; Liang, Z.; Zhang, L. H.; Zhang, Y. K. A Multiplex Fragment-Ion-Based Method for Accurate Proteome Quantification. *Analytical Chemistry* **2019**, *91*, 3921-3928.

(18) Xie, L.-Q.; Zhang, L.; Nie, A.-Y.; Yan, G.-Q.; Yao, J.; Zhang, Y.; Yang, P.-Y.; Lu, H.-J. ITMSQ: A software tool for N- and C-terminal fragment ion pairs based isobaric tandem mass spectrometry quantification. *Proteomics* **2015**, *15*, 3755-3764.

(19) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Molecular and Cellular Proteomics* **2012**, *11*, O111.016717.

(20) Bern, M.; Finney, G.; Hoopmann, M. R.; Merrihew, G.; Toth, M. J.; MacCoss, M. J. Deconvolution of Mixture Spectra from Ion-Trap Data-Independent-Acquisition Tandem Mass Spectrometry. *Analytical Chemistry* **2010**, *82*, 833-841.

(21) Collins, B. C.; Gillet, L. C.; Rosenberger, G.; Röst, H. L.; Vichalkovski, A.; Gstaiger, M.; Aebersold, R. Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nature Methods* **2013**, *10*, 1246-1253.

(22) Jiang, H.; English, A. M. Quantitative analysis of the yeast proteome by incorporation of isotopically labeled leucine. *Journal of proteome research* **2002**, *1*, 345-350.

(23) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; 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 and Cellular Proteomics* **2002**, *1*, 376-386.

(24) Minogue, C. E.; Hebert, A. S.; Rensvold, J. W.; Westphall, M. S.; Pagliarini, D. J.; Coon, J. J. Multiplexed Quantification for Data-Independent Acquisition. *Analytical Chemistry* **2015**, *87*, 2570-2575.

(25) Di, Y.; Zhang, Y.; Zhang, L.; Tao, T.; Lu, H. MdFDIA: A Mass Defect Based Four-Plex Data-Independent Acquisition Strategy for Proteome Quantification. *Analytical Chemistry* **2017**, *89*, 10248-10255.

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

(27) He, L.; Diedrich, J.; Chu, Y. Y.; Yates, J. R. Extracting Accurate Precursor Information for Tandem Mass Spectra by RawConverter. *Analytical Chemistry* **2015**, *87*, 11361-11367.

(28) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, Ş.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F., et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **2018**, *47*, D442-D450.

(29) Dongre, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H. Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile

proton model. *Journal of the American Chemical Society* **1996**, *118*, 8365-8374.

(30) Mák, M.; Mezö, G.; Skribanek, Z.; Hudecz, F. Stability of Asp-Pro bond under high and low energy collision induced dissociation conditions in the immunodominant epitope region of Herpes simplex virion glycoprotein D. *Rapid Communications in Mass Spectrometry* **1998**, *12*, 837-842.

(31) Hogan, J. M.; McLuckey, S. A. Charge state dependent collision-induced dissociation of native and reduced porcine elastase. *J Mass Spectrom* **2003**, *38*, 245-256.

(32) Huang, Y.; Triscari, J. M.; Pasa-Tolic, L.; Anderson, G. A.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. Dissociation Behavior of Doubly-Charged Tryptic Peptides: Correlation of Gas-Phase Cleavage Abundance with Ramachandran Plots. *Journal of the American Chemical Society* **2004**, *126*, 3034-3035.

(33) Tiwary, S.; Levy, R.; Gutenbrunner, P.; Salinas Soto, F.; Palaniappan, K. K.; Deming, L.; Berndt, M.; Brant, A.; Cimermanic, P.; Cox, J. High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis. *Nature Methods* **2019**, *16*, 519-525.

(34) Hsu, J. L.; Huang, S. Y.; Chow, N. H.; Chen, S. H. Stable-isotope dimethyl labeling for quantitative proteomics. *Analytical Chemistry* **2003**, *75*, 6843-6852.

(35) Guo, K.; Ji, C.; Li, L. Stable-Isotope Dimethylation Labeling Combined with LC-ESI MS for Quantification of Amine-Containing Metabolites in Biological Samples. *Analytical Chemistry* **2007**, *79*, 8631-8638.

(36) Cho, K.-C.; Kang, J. W.; Choi, Y.; Kim, T. W.; Kim, K. P. Effects of peptide acetylation and dimethylation on electrospray ionization efficiency. *J Mass Spectrom* **2016**, *51*, 105-110.

(37) Fu, Q.; Li, L. De Novo Sequencing of Neuropeptides Using Reductive Isotopic Methylation and Investigation of ESI QTOF MS/MS Fragmentation Pattern of Neuropeptides with N-Terminal Dimethylation. *Analytical Chemistry* **2005**, *77*, 7783-7795.

(38) Qin, H.; Wang, F.; Zhang, Y.; Hu, Z.; Song, C.; Wu, R.; Ye, M.; Zou, H. Isobaric cross-sequence labeling of peptides by using site-selective N-terminus dimethylation. *Chemical Communications* **2012**, *48*, 6265-6267.

(39) Koehler, C. J.; Arntzen, M. O.; de Souza, G. A.; Thiede, B. An Approach for Triplex-Isobaric Peptide Termini Labeling (Triplex-IPTL). *Analytical Chemistry* **2013**, *85*, 2478-2485.

(40) Tian, X.; de Vries, M. P.; Permentier, H. P.; Bischoff, R. A Collision-Induced Dissociation Cleavable Isobaric Tag for Peptide Fragment Ion-Based Quantification in Proteomics. *Journal of Proteome Research* **2020**, *19*, 3817-3824.

(41) Zhong, X.; Frost, D. C.; Yu, Q.; Li, M.; Gu, T.-J.; Li, L. Mass Defect-Based DiLeu Tagging for Multiplexed Data-Independent Acquisition. *Analytical Chemistry* **2020**, *92*, 11119-11126.

(42) Hebert, A. S.; Merrill, A. E.; Bailey, D. J.; Still, A. J.; Westphall, M. S.; Strieter, E. R.; Pagliarini, D. J.; Coon, J. J. Neutron-encoded mass signatures for multiplexed proteome quantification. *Nature Methods* **2013**, *10*, 332-334.

Supporting Information

Materials and Methods

Chemicals and Materials. 2-Chlorotrityl chloride resin, Fmoc-Gly-OH, Fmoc-¹³C₁-Gly-OH, Fmoc-¹³C₂-Gly-OH, Fmoc-Ala-OH, sodium cyanoborohydride, formaldehyde solution (36.5-38% in H₂O), bovine serum albumin (BSA), acetic anhydride (¹²C₄H₆O₃), acetic anhydride-1,1'-¹³C₂ (¹³C₂¹²C₂H₆O₃), acetic anhydride-¹³C₄ (¹³C₄H₆O₃), 1-[bis(dimethylamino) methylene]-1H-1,2,3-triazolo [4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU), 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (EDC·HCl), piperidine, pyridine, N,N-diisopropylethylamine (DIPEA), trifluoroethanol, p-nitrophenol (PNP), tris(2-carboxyethyl) phosphine hydrochloride (TCEP), 2-iodoacetamide, N,N-dimethylformamide (DMF), acetonitrile (ACN), acetone, dichloromethane (DCM), acetic acid, sodium acetate, sodium hydroxide (NaOH), hydroxylamine hydrate, triethylammonium bicarbonate (TEAB), and trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich (Netherlands). LysC and yeast extracts were purchased from Promega. Synthetic peptide GTDWLANK and FDWA were purchased from GL Biochem (Shanghai, China). Solid phase extraction disk Empore Octadecyl C18 was purchased from 3M.

Synthesis of isobarically labeled acetyl-alanine-glycine-p-nitrophenol ester. The synthesis scheme of triplex isobaric acetyl-alanine-glycine-p-nitrophenol (Ac-AG-PNP) ester tags can be found in **Scheme S1**. 30 mg of 2-chlorotrityl chloride (CTC) resin (1.0-1.8 mmol g⁻¹) was swollen with 1.5 mL of DCM/DMF (1/1) on a thermomixer at 800 rpm for 30 min at room temperature. The swollen resin was washed three times with 0.75 mL of DMF. Then, 1 mL of 0.8 M DIPEA in DMF, 0.5 mL of DMF and 0.2 mmol (60 mg) of Fmoc-Gly-OH, Fmoc-¹³C₁-Gly-OH or Fmoc-¹³C₂-Gly-OH were added. The reaction mixture was shaken for 90 min at room temperature. The solution was filtered out and the remaining isotopically labeled Fmoc-Gly-CTC resin was

washed three times with 0.75 mL of DMF. Afterwards, the resin was treated with 1.5 mL of 20% piperidine in DMF for 5 min and 15 min sequentially. Then, 0.5 mL of 0.4 M HATU in DMF, 0.5 mL of 0.8 M DIPEA in DMF, 0.5 mL of DMF and 0.2 mmol (62 mg) Fmoc-Ala-OH were added. The reaction mixture was shaken for 90 min at room temperature. The resulting isotopically labeled Fmoc-Ala-Gly-CTC resin was washed three times with 0.75 mL of DMF. Then, the resin was treated with 1.5 mL of 20% piperidine in DMF for 5 min and 15 min sequentially to afford **2a-c**. Then 0.5 mL of DCM, 5 µL of pyridine and 5 µL of acetic anhydride were added to **2c**; 0.5 mL of DCM, 5 µL of pyridine and 5 µL of acetic anhydride-1,1'-¹³C₂ were added to **2b**; 0.5 mL of DCM, 5 µL of pyridine and 5 µL of acetic anhydride-¹³C₄ were added to **2a**. The reaction mixtures were shaken for 30 min at room temperature and then the isobarically labeled **3a-c** were released with 0.75 mL of 20% trifluoroethanol in DCM. Subsequently, the dried isobarically labeled **4a-c** were reacted with 80 mg EDC·HCl and 40 mg p-nitrophenol (PNP) in 0.7 mL DCM to generate **5a-c** correspondingly. The reactions were shaken for 3 h, and after drying, the Ac-Ala-Gly-PNP esters were purified by RP-HPLC with detection at 254 nm. The collected HPLC fraction was freeze-dried to afford a white fluffy solid which was stored at -20 °C until use. Around 5 mg was obtained for each tag and the overall yield was around 50%. The LC-MS/MS analysis of Ac-Ala-Gly-PNP esters can be found in **Figure S10**. For LC purification, the HPLC system (Shimadzu, Kyoto, Japan) was equipped with a SIL-20AC autosampler, two LC-20AT pumps and an SPD-20A absorbance detector. The synthesized Ac-AG-PNP esters were separated on an XBridge BEH C18 OBD Prep Column (250 mm × 10 mm, 5 µm particles, 130 Å pore size, Waters, Ireland) with a 30 min gradient of 2-90% acetonitrile in water/0.1% formic acid at a flow rate of 2 mL min⁻¹.

Reduction/Alkylation and LysC Digestion. 100 µg of BSA or yeast protein was dissolved in 100 µL of 100 mM TEAB. Then, 5 µL of 200 mM TCEP was added and the mixture was incubated for 1 h at 37 °C. Afterwards, 5 µL of 375 mM freshly prepared 2-iodoacetamide was added and the mixture was incubated for 30 min in the

dark at room temperature. Finally, 800 μL of acetone was added and the mixture was kept at $-20\text{ }^{\circ}\text{C}$ overnight for precipitation. The precipitated protein was centrifuged at 20800 g for 30 min. Subsequently, the supernatant was decanted and alkylated proteins were stored at $-20\text{ }^{\circ}\text{C}$ until digestion. 100 μg alkylated BSA or yeast protein was dissolved in 200 μL of 100 mM TEAB and 2 μg LysC dissolved in 1 % formic acid was added. The pH was adjusted to 7.5-8 with 1 M TEAB and the digestion was kept overnight at $37\text{ }^{\circ}\text{C}$. The reaction was terminated by acidification using 10% TFA and the resulting LysC peptide mixtures of BSA and yeast proteins were freeze-dried and stored at $-80\text{ }^{\circ}\text{C}$ before conducting the labeling reactions.

Selective N-terminal dimethylation of peptides. 100 μg of peptide GTDWLANK, LysC peptides of BSA or LysC peptides of yeast proteins were dissolved in 100 μL of 1% (aq.) acetic acid. 10 μL of 4% formaldehyde in water was added to the peptide solution and the mixture was shaken for 2 min before adding 15 μL of 600 mM sodium cyanoborohydride dissolved in 1% (aq.) acetic acid^{18, 26}. The reaction was shaken for 15 min at room temperature. Excess reagents were removed by SPE using the STAGE (STop And Go Extraction) TIPS Desalting Procedure²⁷ and the collected N-terminally dimethylated peptides were freeze-dried and stored at $-80\text{ }^{\circ}\text{C}$.

LC-MS/MS analysis. All samples of isobarically labeled peptides were analyzed by nanoLC-MS/MS on a Q Exactive Plus and other samples were analyzed by LC-MS/MS on an Orbitrap Velos Pro as detailed below.

LC-MS/MS analyses were performed on an Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific) coupled to an LC system (Shimadzu, Kyoto, Japan) equipped with a SIL-20AC HT autosampler, and two LC-20AD pumps. The sample was separated on an ACQUITY UPLC BEH C18 column (2.1 mm $\times$ 50 mm, 1.7 μm particles, Waters) at a column temperature of $60\text{ }^{\circ}\text{C}$ with a 10 min gradient of 2-90% acetonitrile in water/0.1% formic acid at a flow rate of 300 $\mu\text{L min}^{-1}$.

NanoLC-MS/MS analyses were performed on a Q Exactive Plus mass spectrometer

(Thermo Scientific, San José, CA) equipped with an UltiMate 3000 RSLCnano system (Thermo Scientific). The peptides (around 300 ng) were first loaded onto a trap column (Acclaim PepMap100 pre-column, 0.3mm, 5 mm, C18, 5 μm , 100 $\text{\AA}$, Thermo Scientific) with 0.1% formic acid at a flowrate of 20 $\mu\text{L min}^{-1}$ and then separated on an analytical column (PepMap RSLC C18, 50 cm 75 $\mu\text{m ID}$, 2 μm , 100 $\text{\AA}$, Thermo Scientific). A flow rate of 300 nL min^{-1} and a column temperature of $40\text{ }^{\circ}\text{C}$ were utilized. Solvent A was 0.1% formic acid in water and solvent B was 0.1% formic acid in acetonitrile. A gradient was applied from 2% to 50% or 65% B in the first 60 min, then 85% B for 5 min and equilibration for 15 min with 2% B.

For analyses in the DDA mode, mass spectrometry data were measured using a data-dependent top 10 method. Full MS scans were acquired between m/z 300 and 1650 with a resolution of 70000 (at m/z 200). The automatic gain control (AGC) target was 3E6 and maximum injection time was 50 ms. Precursor ions were selected with a window of 2 Th or 0.6 Th. Selected ions were fragmented in the HCD collision cell with stepped normalized collision energies (NCEs) of 18, 20, 22, 24, 26, 28, 30 and 32. For quantification experiments, a combined NCE of 18 and 28 were set in the Thermo Xcalibur Instrument Setup (version 3.1.66.10). MS2 spectra were acquired in the Orbitrap mass analyzer with a resolution of 17500 (at m/z 200). The AGC target was 1E5 and the maximum injection time was 100 ms. The charge exclusion discards ions with a single charge and charges higher than 4. The dynamic exclusion time was 20 s. For analyses in the DIA mode, each DIA analysis consisted of two alternating scan experiments, a survey scan (m/z 350-1200), which was analyzed at a resolution of 70000 (at m/z 200), an AGC target of 3E6 and a maximum injection time of 50 ms, followed by 20 sequential precursor ion selection windows of 29 Th (centered on 462.5, 487.5, 512.5, 537.5, 562.5, 587.5, 612.5, 637.5, 662.5, 687.5, 712.5, 737.5, 762.5, 787.5, 812.5, 837.5, 862.5, 887.5, 912.5 and 937.5) at a resolution of 17,500 covering an m/z range of 450–950. A combined NCE of 18 and 28 was set in the Thermo Xcalibur Instrument Setup (version 3.1.66.10) in analogy to DDA. The AGC target was 1E6 with the maximum injection time set to 'auto'. The 2 Th overlap between windows was

applied to ensure isolation and fragmentation of entire isotope envelopes of precursors. The cycle time of an MS1 survey scan and 20 MS2 scans takes around 2 s, which ensures that sufficient data points are acquired for quantification of LC peaks. The width of LC peak span 20-30 seconds and 10-15 data points are supposed to be acquired for every LC peak.

Supplementary Figures

Figure S1. Examples of calculating ratios for PSMs, peptides and protein of triplex labeled BSA mixed at a ratio of 5:2:1. (Python scripts and more quantification outputs are available at https://github.com/tianxiaobo002/Ac-AG-tag_scripts_and_quantification_outputs).

(1) At PSM level, the quantification ratios (72202240 : 28389638 : 13828063) of measured

A versatile isobaric tag enables proteome quantification in data dependent and data independent acquisition mode

intensities were normalized with the weighted average (14302492) produced by dividing the total intensity (114419941) of three channels by mixing ratios (5+2+1).

(2) At peptide level, top three PSMs (scan 6134, 6246, 6353) were used to calculate the quantification ratio of AEFVEVTK. The quantification ratio (73857000 : 29146054 : 14232920) of three selected PSMs were normalized with the weighted average (14654496) produced by dividing the total intensity (117235975) of three channels by mixing ratios (5+2+1).

(3) At protein level, top three peptides (AEFVEVTK, LVNELTEFAK, TVMENFVAFVDK) were selected to calculate the quantification ratio of BSA. The quantification ratio (85335695 : 35057840 : 16998987) of three selected peptides were normalized with the weighted average (17174065) produced by dividing the total intensity (137392523) of three channels by mixing ratios (5+2+1).

Figure S2. Correcting the ¹³C contribution in the peptide-coupled reporter-ion of triplex labeled N-dime-GTDWLANK-GA-Ac mixed at a ratio of 20:10:1.

Figure S3. Derivatization of GTDWLANK. A) Starting material of the peptide GTDWLANK and the reference peptide FDWA (extracted ion chromatograms (XICs) of m/z 452.73 and 538.23); B) N-terminally dimethylated GTDWLANK and N-terminally dimethylated FDWA, XICs of m/z 452.73 (GTDWLANK), 466.74 (N-dime-GTDWLANK), 480.75 (dime-GTDWLANK-dime, doubly dimethylated form), 538.23 (FDWA) and 566.26 (N-dime-FDWA); C) N-dime-GTDWLANK-GA-Ac and N-dime-FDWA, XICs of m/z 466.74 (N-dime-GTDWLANK), 552.78 (N-dime-GTDWLANK-GA-Ac) and 566.26 (N-dime-FDWA).

Figure S4. MS2 spectra of triplex labeled N-dime-GTDWLANK-GA-Ac (mixed at 1:1:1) at various NCEs. A) MS2 spectrum at NCE 18; B) MS2 spectrum at NCE 20; C) MS2 spectrum at NCE 22; D) MS2 spectrum at NCE 24; E) MS2 spectrum at NCE 26; F) MS2 spectrum at NCE 28; G) MS2 spectrum at NCE 30. H) MS2 spectrum at NCE 32.

Figure S5. Separate analysis of the three channels of triplex isobarically labeled N-dime-GTDWLANK-GA-Ac. (A), (B), (C), (D), respectively show the b_2 , y_2 , y_5 and quantification ion of N-dime-GTWLANK-(+2)-GA-Ac; (E) (F), (G), (H), respectively show the b_2 , y_2 , y_5 and quantification ion of N-dime-GTWLANK-(+1)-GA-Ac(+1); (I) (J), (K), (L), respectively show the b_2 , y_2 , y_5 and quantification ion of N-dime-GTWLANK-GA-Ac-(+2).

Figure S6. Number of Peptide Spectral Matches (PSMs) of N-dime-BSA peptides-GA-Ac at NCEs of 24, 26, 28 and 30.

Figure S7. Charge state distribution of unlabeled and Ac-AG labeled LysC yeast peptides.

Figure S8. Scheme of preparing the BSA-yeast proteomics sample: A) triplex labeled BSA was mixed at a ratio of 1:5:10 (1 ng: 5 ng: 10 ng) and the LysC yeast sample was mixed at a ratio of 1:2:5 (50 ng: 100 ng: 250 ng); B) triplex labeled BSA was mixed at a ratio of 1:5:10 (1 ng: 5 ng: 10 ng) and LysC yeast sample was mixed at a ratio of 5:2:1 (250 ng: 100 ng: 50 ng).

Figure S9. Triplex-labeled LysC BSA mixed at various ratios and analyzed in the DIA mode. A), B), C), D) and E) are measured ratios of LysC BSA sample mixed at ratios of 1:1:1, 1:2:5, 5:2:1, 1:5:10 and 10:5:1, respectively.

Figure S10. LC-MS/MS of the triplex Ac-AG-PNP tag. (A) XIC of m/z 312.11, Ac-AG-(+2)-PNP. (B) Mass spectrum of Ac-AG-(+2)-PNP. (C) MS/MS spectrum of Ac-AG-(+2)-PNP. (D) XIC of m/z 312.11, (+1)-Ac-AG-(+1)-PNP. (E) Mass spectrum of (+1)-Ac-AG-(+1)-PNP. (F) MS/MS spectrum of (+1)-Ac-AG-(+1)-PNP. (G) XIC of m/z 312.11, (+2)-Ac-AG-PNP. (H) Mass spectrum of (+2)-Ac-AG-PNP. (I) MS/MS spectrum of (+2)-Ac-AG-PNP.

Scheme S1. Synthesis approach for the triplex Ac-AG-PNP tags. i) 0.8 M DIPEA in DMF, Fmoc-Gly-COOH or Fmoc-¹³C₁-Gly-COOH or Fmoc-¹³C₂-Gly-COOH; ii) 20 % piperidine in DMF; iii) 0.4 M HATU, 0.8 M DIPEA in DMF and Fmoc-Ala-COOH; iv) 20 % piperidine in DMF; v) Acetic

A versatile isobaric tag enables proteome quantification in data dependent and data independent acquisition mode

anhydride or Acetic anhydride-1,1'-¹³C₂ or Acetic anhydride-¹³C₄; vi) 20 % trifluoroethanol in DCM;

vii) EDC·HCl and p-nitrophenol.

Table S1. Charge state distribution of GTDWLANK before and after labeling with the Ac-AG tag

Peptide	Intensity of singly-charged ion in absolute numbers and in percentage of the total intensity		Intensity of the doubly-charged ion in absolute numbers and in percentage of the total intensity	
GTDWLANK	6.91E+04	3.6%	1.87E+06	96.4%
N-dime-GTDWLANK	6.81E+04	2.4%	2.82E+06	97.6%
N-dime-GTDWLANK-GA-Ac	1.24E+05	3.5%	3.42E+06	96.5%

Chapter 5

The isotopic Ac-IP tag enables multiplexed proteome quantification in data-independent acquisition mode

Xiaobo Tian,[†] Marcel P. de Vries,[‡] Hjalmar P. Permentier,[†] and Rainer Bischoff[†]

[†]Department of Analytical Biochemistry and Interfaculty Mass Spectrometry Center, Groningen Research Institute of Pharmacy, University of Groningen, A. Deusinglaan 1, 9713 AV, Groningen, The Netherlands

[‡]Department of Pediatrics, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713 GZ, Groningen, The Netherlands

Manuscript published in *Anal. Chem.* 2021, 93, 23, 8196–8202

ABSTRACT

Data-independent acquisition (DIA) is an increasingly used approach for quantitative proteomics. However, most current isotope labeling strategies are not suitable for DIA, as they lead to more complex MS2 spectra or severe ratio distortion. As a result, DIA suffers from a lower throughput than data-dependent acquisition (DDA) due to a lower level of multiplexing. Herein, we synthesized an isotopically labeled acetyl-isoleucine-proline (Ac-IP) tag for multiplexed quantification in DIA. Differentially labeled peptides have distinct precursor ions carrying the quantitative information but identical MS2 spectra, since the isotopically labeled Ac-Ile part leaves as a neutral loss upon collision-induced dissociation, while fragmentation of the peptide backbone generates regular fragment ions for identification. The Ac-IP labeled samples can be analyzed using general DIA LC-MS settings and the data obtained can be processed with established approaches. Relative quantification requires deconvolution of the isotope envelope of the respective precursor ions. Suitability of the Ac-IP tag is demonstrated with a triplex-labeled yeast proteome spiked with bovine serum albumin (BSA) that was mixed at 10 : 5 : 1 ratios resulting in measured ratios of 9.7 : 5.3 : 1.1.

1 Introduction

Multiplexing with isotopic labels is an attractive alternative to label-free quantification in mass spectrometry-based proteomics, since it increases throughput and reduces run-to-run variability¹. Remarkable advancements have been made in terms of multiplexing capacity over the last decades^{2,3}. Current multiplexing approaches can be classified into MS1-based methods^{4,5} that rely on isotopic labeling and quantify peptides based on distinct precursor ion masses and MS2-based methods that rely on isobaric labeling strategies and quantify peptides based on reporter ions^{6,7}, fragments of the peptide backbone^{8,9}, or peptide-coupled reporter-ions¹⁰⁻¹². In the widely used data-dependent acquisition (DDA) mode, where a certain number of precursor ions is selected for fragmentation during an LC-MS run in order of decreasing intensity¹³, current MS1-based quantification methods analyze up to three samples in a single experiment, since

higher multiplexing is constrained by the increasing complexity of MS1 spectra and the need for a mass difference of at least 4 Da between labeling channels to avoid overlap of isotope envelopes¹⁴. In addition, the stochastic nature of selecting precursor ions for tandem MS in DDA leads to an imperfect overlap in peptide identifications, which means that there is a considerable likelihood that some peptides might be missed in some LC-MS runs depending on their relative abundance in given samples¹⁵. Since only the peptides that are measured in all runs can be reliably quantified and thus compared across many runs in studies comprising hundreds of samples, it would be beneficial to use a strategy that is not data-dependent and thus avoids the stochastic precursor ion selection. This has led to the development of data-independent acquisition (DIA).

Since the Sequential Windowed Acquisition of all Theoretical fragment ion Mass Spectra (SWATH-MS)¹⁶ was proposed, DIA has become more widely used due to the advancement of mass spectrometry technology and sophisticated data processing algorithms¹⁶⁻¹⁹. Instead of sampling a certain number of precursor ions, DIA selects all precursors within a predefined isolation window for fragmentation. This results in better reproducibility and avoids the stochastic nature of precursor ion selection in DDA. However, most DIA-based approaches can only analyze one sample per LC-MS run and most current isotope labeling-based quantification strategies, that are commonly used in DDA, complicate the MS2 spectra or severely distort the quantitative ratio in DIA. For example, reporter-ion-based approaches, such as TMT⁷ or iTRAQ⁶, are not suitable for DIA since most peptide are cofragmented with others. For the approaches that generating a set of distinct fragment ions for different labeling channel, such as the isotopic labeling, SILAC²⁰ and fragment ion-based method, IPTL⁸, the complexity of MS2 spectra in DIA is multiplied by the number of labeled samples, which makes it extremely challenging to identify peptides, since MS2 spectra of DIA are already highly convoluted. As a result, DIA has a lower throughput than DDA due to a lower level of multiplexing. To increase the multiplexing capacity of DIA, a number of approaches have been reported, such as NeuCoDIA²¹, MdFDIA²² and mdDiLeu²³. These

approaches rely on neutron encoding²⁴, which is based on the mDa mass difference due to incorporating either ¹³C or ¹⁵N, and require ultrahigh resolution (> 120k) at the MS2 level resulting in a reduced data acquisition rate on Orbitrap instruments. Recently, we reported a novel isobaric Ac-AG tag that reveals quantification information based on peptide-coupled reporter-ions at a modest resolution of 17.5k at the MS2 level in the DIA mode²⁵. Since in DDA mode only a limited number of precursor ions can be isolated for tandem MS within the time window of a chromatographic peak, isobaric labeling is advantageous, since it enables multiplex labeling without increasing the complexity of MS1 spectra. This constraint does not apply to the DIA mode where precursor selection for MS2 is independent of the complexity of the corresponding MS1 spectra, which means that isobaric labeling provides no particular advantage in this case, while the need for complementary isotope distributions between balancer and reporter ions complicates synthesis and limits the ultimately achievable multiplexing capacity.

Since isobaric labeling is not required in DIA, we developed a collision-induced dissociation (CID) cleavable, isotopically labeled acetyl-isoleucine-proline (Ac-IP) tag and demonstrated it with a triplex-labeled yeast-BSA (bovine serum albumin) mixed sample. While quantification is achieved at the precursor ion (MS1) level, differentially Ac-IP-labeled peptides generate identical MS2 spectra independent of the number of labeled samples and thus do not increase the complexity of MS2 spectra in DIA.

2 Experiments

Details of chemicals and materials, the synthesis of triplex acetyl-isoleucine-proline-N-hydroxysuccinimide (Ac-IP-NHS tags, see **Figure 1A**), reduction/alkylation and LysC digestion, selective N-terminal dimethylation of peptides and LC-MS/MS analysis can be found in the Supporting Information.

2.1 Triplex labeling of N-terminally dimethylated peptides. 5 µg of N-terminally dimethylated GTDWLANK, LysC peptides of BSA or LysC peptides of yeast proteins

were dissolved in 50 µL of 200 mM triethylammonium bicarbonate (TEAB) buffer of pH 8.5. Then, 4 µL of 50 mM Ac-IP-NHS, ¹³C₁-Ac-IP-NHS and ¹³C₂-Ac-IP-NHS in DMF were added to the three peptide solutions, respectively. The reaction mixtures were shaken for 2 h at room temperature. To ensure complete labeling, 2 µL of the respective Ac-IP-NHS reagent was added again and incubated for 1 h more. Esterification of the hydroxyl groups of Ser, Thr or Tyr and excess N-hydroxysuccinimide (NHS) ester were hydrolyzed in the presence of 5% hydroxylamine hydrate at 55 °C for 5 min and samples were then desalted by SPE using the STAGE (STop And Go Extraction) TIPS Desalting Procedure²⁶ prior to LC-MS analysis. 600 µL of 2% acetonitrile in water with 0.1% trifluoroacetic acid (TFA) was added to remove excess Ac-IP-COOH before eluting peptides from the STAGE tips with 60% acetonitrile in water with 0.1% TFA (**Figure S1** the Supporting Information).

2.2 Database Searching and Quantification. To create the spectral library, LC-MS/MS raw files measured in DDA mode were analyzed with PEAKS X+ (Bioinformatics Solutions Inc., Waterloo, Ontario, Canada) and searched against the UniProt reference database of yeast (UP000002311, 6049 entries, downloaded on Jan. 20, 2020) to which the BSA entry (P02769) was added manually. LysC was selected as enzyme, the digestion mode was set to specific and the maximal number of missed cleavage sites to 1. A tolerance of 20 ppm for the precursor ion and 0.02 Da for the MS/MS fragment ions was applied. Carbamidomethylation (+57.02) on cysteine and dimethylation (+28.03) on the N-terminus were set as fixed modifications and oxidation (+15.99) on methionine as variable modification. The neutral losses of triplex Ac-IP tags were manually added in the 'PTMNeutralLossScorer.properties' file as follows: 'Ac-IP on Lys@K_155.094629_TRUE, Ac-IP (¹³C+1) on Lys@K_156.097983_TRUE, Ac-IP (¹³C+2) on Lys@K_157.101338_TRUE'. For triplex labeling experiments, variable modifications on Lys were set as Ac-IP (+252.1474), ¹³C₁-Ac-IP (+253.1508) and ¹³C₂-Ac-IP (+254.1542). The identification results were filtered with a false discovery rate (FDR) of 0.5% for peptides. The filtered results were exported as spectral

library for DIA analysis using the default parameters of PEAKS X+.

For identification of peptides in DIA, the raw files measured in DIA mode were searched against the prepared spectral library and the UniProt reference database of yeast (UP000002311) complemented with BSA was selected as the PEAKS reference database. A tolerance of 20 ppm for the precursor ion and 0.02 Da for the MS/MS fragment ions was applied. The DIA results were also filtered with an FDR of 0.5% for peptides.

The PEAKS output of the peptides matched to proteins (*protein-peptides.csv*) and the peptide-spectrum matches (*DB search psm.csv*) were exported from the DIA identifications. All the following steps were performed using in-house built Python scripts (available at https://github.com/tianxiaobo002/Ac-IP_tag_scripts_and_quantification_outputs). Only the unique peptides extracted from the *protein-peptides.csv* files were used for quantification. For every peptide spectrum match (PSM) that was derived from unique peptides in the *DB search psm.csv* files, the theoretical precursor masses were calculated and grouped by the scan number. The MS1 spectra of DIA raw data were converted to the mgf format with RawConverter²⁷ for processing in Python. All MS1 spectra were extracted from the resulting mgf file and also grouped by the scan number. For a given PSM (based on the scan number of that PSM), the corresponding nearest preceding MS1 spectrum was selected from which the measured peak intensities of precursor masses were used to determine ratios for that PSM. Examples of deconvolution and ratio calculation at the PSM level can be found in **Figure S2**. When more than one PSM matched to the same peptide, the ratio derived from the PSM with the highest '-10lgP' value was used to represent the ratio for that peptide. The top three peptides with the highest '-10lgP' were used to calculate the ratio for the corresponding protein. Examples of calculating ratios for peptides and proteins can be found in **Figure S3**.

The quantification scripts and quantification outputs of triplex Ac-IP labeled BSA, yeast and BSA-yeast samples at the PSM, peptide and protein level are available at https://github.com/tianxiaobo002/Ac-IP_tag_scripts_and_quantification_outputs. The

raw mass spectrometry data of the yeast and the BSA-yeast samples have been deposited with the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD022606.

3 Results and discussion

3.1 Design of the Ac-IP tag and labeling strategy. The Ac-IP tag consists of three parts: (1) Ac-Ile, which contains distinct isotopes and will leave as a neutral loss fragment upon CID, (2) proline, which facilitates dissociation of the Ile-Pro bond and (3) an amine reactive NHS ester reacting with the epsilon-amino group of the C-terminal Lys residues in peptides generated by LysC digestion (see **Figure 1A** for the structure design, **Scheme 1** for the synthesis route and **Figure S4** for LC-MS/MS results and **Figure S5-7** for NMR analyses of the tags in the Supporting Information). Since the Ile-Pro bond fragments easily^{9,28}, the isotopically labeled Ac-Ile part of the tag leaves as a neutral loss fragment upon CID in addition to the fragmentation of the peptide backbone. Precursors of peptides derived from different samples thus have distinct masses carrying the quantitative information while generating identical MS2 spectra that can be analyzed by established DIA identification approaches²⁹. It is worth pointing out that the signals from the same peptide backbone fragment ions of every labeling channel add up resulting in a better signal-to-noise (S/N) ratio in the MS2 spectra and thus facilitating peptide identification.

Since labeling with the Ac-IP tag removes a high proton affinity site (epsilon-amino group of Lys), which might reduce the ionization efficiency of modified peptides (see **Figure S8**), we first dimethylated the N-terminal amine group³⁰ converting it into a tertiary amine to facilitate ionization³¹ and to prevent it from reacting with the Ac-IP tag (**Figure 1B**). Another advantage of N-terminal dimethylation is that it leads to a more complete b-ion series upon CID³¹, which aids in interpretation of the MS2 spectra. Since the mass difference between the Ac-IP tags is 1 Da, differentially labeled peptides will form overlapping isotope envelopes with a changing intensity profile according to the different quantitative ratios of a given peptide in the different samples (**Figure 1C**). Once a peptide has been identified, its composition is used to calculate the theoretical intensity distribution of the isotope envelope, which serves to deconvolute the overlapping precursor isotope envelopes from the MS1 spectra to retrieve the underlying quantitative information (see **Figure S2** for details).

Figure 1. Design and concept of quantification based on the Ac-IP tag in the DIA mode. (A) Molecular structure of the triplex Ac-IP tag. (B) Sample preparation workflow: protein digestion with LysC followed by selective N-terminal dimethylation and labeling with the Ac-IP tag. (C) Schematic representation of sample mixing and results at the MS1 and the MS2 level prior to and after deconvolution for a mixture of triplex-labeled samples. The differentially labeled samples are pooled prior to LC-MS/MS in the DIA mode.

3.2 Derivatization and quantification at the peptide level. GTDWLANK was selected as the model peptide to assess the labeling efficiency of the Ac-IP approach and investigate the influence of modifications on ionization efficiency and charge state. As expected, ionization efficiency increased after N-terminal dimethylation³¹ (N-dime) while the intensities of precursor ions with or without the Ac-IP modification were comparable (**Figure S9**). Both reactions had yields exceeding 98%. The charge state distribution of N-dime-GTDWLANK-PI-Ac remained almost unchanged with the doubly-charged ion being most abundant (97.4%) (**Table S1**). Fragmentation of N-

dime-GTDWLANK-PI-Ac was evaluated at different normalized collision energies (NCEs) ranging from 18 to 32. As shown in **Figure S10**, in the range from 18 to 22, fragmentation occurred only at the fragile Ile-Pro bond of the Ac-IP tag giving only the “precursor minus Ac-Ile” fragment ion in MS2 spectra. With increasing NCE from 24 to 32, fragment ions of the peptide backbone became stronger accompanied by a gradual decrease in intensity of the ion “precursor minus Ac-Ile”. Neutral loss of Ac-Ile is complete and peptide backbone fragmentation is efficient at NCE 30, which was therefore selected for further experiments.

Figure 2. Investigations with the triplex-labeled N-dime-GTDWLANK-PI-Ac. (A) Precursor isotope envelope of single channel, N-dime-GTDWLANK-PI-Ac(+0). (B), (C), (D), (E) and (F) Precursor isotope envelopes of triplex-labeled N-dime-GTDWLANK-PI-Ac (+0, +1 and +2) mixed at ratios of 1:1:1, 1:2:5, 1:10:20, 5:2:1 and 20:10:1, respectively (measured ratios are shown on top of each panel). (G) MS2 spectrum of the triplex-labeled N-dime-GTDWLANK-PI-Ac mixed at 1:1:1 with a NCE of 30.

N-dime-GTDWLANK-PI-Ac (+0, +1 and +2) were analyzed individually or mixed at a ratio of 1:1:1. Differential labeling resulted in distinct peptide masses but identical MS2 spectra (**Figure S11**). The triplex-labeled N-dime-GTDWLANK-PI-Ac peptides were then mixed at ratios of 1:1:1, 1:2:5, 1:10:20, 5:2:1 and 20:10:1 followed by LC-MS/MS analysis. As shown in **Figure 2A-F**, the shape of the isotope envelope of the precursor ions changed with different mixing ratios while all mixtures produced

identical MS2 spectra (**Figure 2G**). Quantification information was retrieved from the isotope envelopes of the precursor ions after deconvolution (see example in **Figure S2**) and the measured ratios were in agreement with the actual mixing ratios.

3.3 Preparation of a spectral library and DIA quantification of triplex-labeled BSA sample.

Tripix-labeled LysC BSA peptides (+0, +1 and +2) were analyzed with DDA and the corresponding identifications, reported by PEAKS software³², were exported as a spectral library. Subsequently, triplex-labeled BSA peptides were mixed at ratios of 1:1:1, 1:5:10 and 10:5:1 and analyzed with DIA followed by searching data against the spectral library. The DIA scanning scheme includes a survey scan acquired at a resolution of 70k and twenty sequential precursor ion selection windows covering an m/z range of 450–950 acquired within a cycle time of about 2 s at a resolution of 17.5k at the MS2 level, making the method accessible to most high-resolution mass spectrometers. The log₂-normalized ratios of the triplex-labeled BSA sample mixed at a ratio of 1:1:1 converge to zero for peptide-spectrum matches (PSMs) with a higher significance ($-10\log P$ values) (**Figure 3A** and **Figure S12**) and for more intense peptides (**Figure 3B**). Consequently, the three peptides with the highest significance values, provided by the PEAKS identification algorithm, were selected and the total intensities of their respective labeling channels were used to calculate the normalized protein ratios (see **Figure S3** for details). Applying these rules, the BSA samples mixed at 1:1:1, 10:5:1 and 1:5:10 were measured as 1.02:1.00:0.98, 9.75:5.25:1.00 and 1.02:5.04:9.94, respectively (**Figure 3C** and **D**).

Figure 3. Evaluations of quantification performance based on triplex-labeled LysC peptides of BSA. (A) Log2-normalized ratio distribution of peptide spectral matches (PSMs) of BSA peptides prepared from LysC digestion and mixed at a ratio of 1:1:1 in relation to PSM significance ($-10\lg P$ values). Only the PSM with the highest significance value for each peptide, as provided by the PEAKS identification algorithm, was used. (B) Log2-normalized ratio distribution of BSA peptides prepared from LysC digestion and mixed at a ratio of 1:1:1 in relation to precursor intensity. (C) and (D) are ratio distributions at the peptide level of BSA peptides prepared from LysC digestion and mixed at ratios of 10:5:1 and 1:5:10 respectively. The dotted lines represent the expected values of log2 normalized ratios.

3.4 Quantification of a specific protein within a complex proteomics background.

To assess the labeling efficiency and the quantification accuracy in complex samples, triplex-labeled LysC peptides of yeast proteins were prepared. For N-terminal dimethylation⁹, more than 98% of LysC yeast peptides had a labeling yield exceeding 95% as calculated from DDA data with variable modification searches. The Ac-IP labeling efficiency was high with 2292 out of 2313 peptides (99%) having a labeling yield of more than 98%. Only 0.03% of the identified peptides were also identified when the modification at the C-terminal Lys residue was set to proline instead of Ac-IP, which indicates that in-source fragmentation of the Ile-Pro bond was negligible. The labeled peptide precursor ions were mostly doubly-charged (**Figure S13**) and the numbers of identified peptides were comparable before and after labeling. Notably, ~90% (24531 out of 27938) of matched y-ions had a neutral loss of Ac-Ile exceeding 98%. The spectra from the DDA runs of 908 identified proteins in the triplex Ac-IP labeled LysC yeast sample were exported as a spectral library and 688 proteins were identified by PEAKS when searching the DIA data of the 1:1:1 triplex-labeled sample. Out of these, 608 proteins were quantified based on the rules outlined above. 80 identified proteins could not be quantified due to peptide precursor ions with incomplete isotope envelopes. The medians of protein ratios of the 1:1:1 mixed yeast sample were measured as 1.01:0.96:1.02 (**Figure 4A**). In the mixed BSA (10:5:1)-yeast (5:2:1) sample, the medians of protein ratios of the LysC yeast sample were measured to be 4.96:1.95:1.09 (**Figure 4B**) and the ratios of triplex-labeled BSA to be 9.67:5.26:1.07 (**Figure 4C**).

Figure 4. Analysis of Ac-IP-tagged LysC yeast samples and BSA-yeast mixed sample. (A) Box plots showing the distribution of measured protein ratios in Log2 transformation of triplex-labeled LysC yeast samples mixed at 1:1:1; (B) and (C), quantification results of a mixed BSA (10:5:1)-yeast (5:2:1) proteomics sample, in which triplex-labeled BSA peptides prepared from LysC digestion were mixed at 10:5:1 and yeast peptides prepared from LysC digestion were mixed at 5:2:1. (B) and (C) are box plots showing the distribution of measured protein ratios in Log2

transformation of triplex-labeled LysC yeast samples and triplex-labeled BSA, respectively. The dotted lines represent the expected values of log2 normalized ratios.

Based on the aforementioned results, the Ac-IP tag has a number of features that make it particularly suitable for multiplex DIA. It allows to quantify peptides based on distinct precursor ions, which results in identical fragment ion spectra, and thus avoids further complicating MS2 spectra in DIA. Compared to previously reported approaches, the Ac-IP tag has the potential to be extended to a higher multiplexing capacity, again without rendering MS2 spectra more complex. While the proof-of-principle study was performed with triplex labeled samples, multiplexing can be extended to 10-plex based on commercially available ^{13}C - or ^{15}N -labeled acetic anhydride and isoleucine following the same synthetic procedure (see **Figure S14**). Compared to other tags that have been proposed for multiplex DIA, such as the MdFDIA, mdDiLeu and Ac-AG tag, the Ac-IP tag opens the possibility to obtain more accurate quantification by increasing the resolution at the MS1 level without extending the cycle time to a point that is no longer commensurate with DIA³³, as would be the case for tags that quantify peptides at the MS2 level (**Table S2**). The mass shift between the Ac-IP tags is 1 Da, which allows to work at the customary resolution of 70k for MS1 and 17.5k for MS2 spectra in an Orbitrap. Resorting to ultra-high resolution for MS1 spectra and making use of neutron encoding²⁴ would further boost the multiplexing capacity. In addition, while the Ac-IP tag exhibits good quantification capability across a 10-fold dynamic range in DIA mode, it should be pointed out that the deconvolution method used in this manuscript is based on the assumption of expected intensities for individual peaks in the isotope envelope of the precursor ions, which means that inaccurate ratios are produced when more than one unresolved precursor isotope envelope overlap or when there is interference from background ions. In these cases, it is necessary to resolve the isotope envelopes and assign them to different peptide precursor ions and higher resolution at the MS1 level may be needed for accurate quantification. Furthermore, in this study, we chose the widely used DIA identification method based on a spectral library search, which

inherently limits the number of DIA identifications to the number of proteins that are in the library. Using a DIA identification algorithm that does not require spectral libraries and that recognizes fragment ions with specific neutral losses would further enhance the potential of the Ac-IP tag approach.

4 Conclusions

In this paper, we describe a novel CID cleavable isotope tag for multiplexed quantification in DIA mode based on MS1 spectra. The peptides labeled with Ac-IP tags have distinct precursor masses containing the relative quantification information, while have identical MS2 spectra. The Ac-IP approach neither complicates the interpretation of MS2 spectra nor relies on an ultrahigh resolution power, and indeed enhances the fragment ion intensity. The feasibilities of quantifying peptides/proteins in the DIA mode of the Ac-IP tag were demonstrated by the quantification of a triplex-labeled BSA sample with 10-fold dynamic changes that was spiked in a complex background of the triplex-labeled LysC yeast sample. The distinctive feature of producing the same MS2 spectra from the different precursor ions, the potential for a higher multiplexing capacity in DIA and the straightforward synthesis would attract more attention to the multiplexed DIA direction.

Acknowledgements

We gratefully acknowledge the China Scholarship Council (CSC) for a PhD fellowship to X. Tian. X. Tian thanks Jos Hermans for helping with the LC-MS analyses. X.T. thanks B. Zhang and C. Ye for help with the NMR analyses. This work is partially funded by the Open Technology Programme of Toegepaste en Technische Wetenschappen (TTW) with project number 15230 which is financed by the Netherlands Organisation for Scientific Research (NWO).

Notes

The authors declare no competing financial interest.

References

- (1) Dayon, L.; Affolter, M. Progress and pitfalls of using isobaric mass tags for proteome profiling. *Expert Review of Proteomics* **2020**, *17*, 149-161.
- (2) Li, J. M.; Van Vranken, J. G.; Vaite, L. P.; Schweppe, D. K.; Huttlin, E. L.; Etienne, C.; Nandhikonda, P.; Viner, R.; Robitaille, A. M.; Thompson, A. H.; Kuhn, K.; Pike, I.; Bomgarden, R. D.; Rogers, J. C.; Gygi, S. P.; Paulo, J. A. TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples. *Nature Methods* **2020**, *17*, 399-404.
- (3) Liu, J.; Shan, Y.; Zhou, Y.; Liang, Z.; Zhang, L.; Zhang, Y. Advances and applications of stable isotope labeling-based methods for proteome relative quantitation. *TrAC Trends in Analytical Chemistry* **2020**, *124*, 115815.
- (4) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology* **1999**, *17*, 994-999.
- (5) Jiang, H.; English, A. M. Quantitative analysis of the yeast proteome by incorporation of isotopically labeled leucine. *Journal of proteome research* **2002**, *1*, 345-350.
- (6) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlett-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents. *Molecular and Cellular Proteomics* **2004**, *3*, 1154-1169.
- (7) Dayon, L.; Hainard, A.; Licker, V.; Turck, N.; Kuhn, K.; Hochstrasser, D. F.; Burkhard, P. R.; Sanchez, J. C. Relative Quantification of Proteins in Human Cerebrospinal Fluids by MS/MS Using 6-Plex Isobaric Tags. *Analytical Chemistry* **2008**, *80*, 2921-2931.
- (8) Koehler, C. J.; Strozynski, M.; Kozielski, F.; Treumann, A.; Thiede, B. Isobaric Peptide Termini Labeling for MS/MS-Based Quantitative Proteomics. *Journal of Proteome Research* **2009**, *8*, 4333-4341.
- (9) Tian, X.; de Vries, M. P.; Permentier, H. P.; Bischoff, R. A Collision-Induced Dissociation Cleavable Isobaric Tag for Peptide Fragment Ion-Based Quantification in Proteomics. *Journal of Proteome Research* **2020**, *19*, 3817-3824.
- (10) Wühr, M.; Haas, W.; McAlister, G. C.; Peshkin, L.; Rad, R.; Kirschner, M. W.; Gygi, S. P. Accurate Multiplexed Proteomics at the MS2 Level Using the Complement Reporter Ion Cluster. *Analytical Chemistry* **2012**, *84*, 9214-9221.
- (11) Sonnett, M.; Yeung, E.; Wühr, M. Accurate, Sensitive, and Precise Multiplexed Proteomics Using the Complement Reporter Ion Cluster. *Analytical Chemistry* **2018**, *90*, 5032-5039.
- (12) Winter, S. V.; Meier, F.; Wichmann, C.; Cox, J.; Mann, M.; Meissner, F. EASI-tag enables accurate multiplexed and interference-free MS2-based proteome quantification. *Nature Methods* **2018**, *15*, 527-530.
- (13) Pappireddi, N.; Martin, L.; Wühr, M. A Review on Quantitative Multiplexed Proteomics. *ChemBioChem* **2019**, *20*, 1210-1224.
- (14) Paulo, J. A.; Gygi, S. P. mTMT: An Alternative, Nonisobaric, Tandem Mass Tag Allowing for Precursor-Based Quantification. *Analytical Chemistry* **2019**, *91*, 12167-12172.
- (15) Brenes, A.; Hukelmann, E.; Bensaddek, D.; Lamond, A. I. Multibatch TMT Reveals False Positives, Batch Effects and Missing Values. *Molecular and Cellular Proteomics* **2019**, *18*, 1967-1980.

- (16) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Molecular and Cellular Proteomics* **2012**, *11*, O111.016717.
- (17) Bern, M.; Finney, G.; Hoopmann, M. R.; Merrihew, G.; Toth, M. J.; MacCoss, M. J. Deconvolution of Mixture Spectra from Ion-Trap Data-Independent-Acquisition Tandem Mass Spectrometry. *Analytical Chemistry* **2010**, *82*, 833-841.
- (18) Collins, B. C.; Gillet, L. C.; Rosenberger, G.; Röst, H. L.; Vichalkovski, A.; Gstaiger, M.; Aebersold, R. Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nature Methods* **2013**, *10*, 1246-1253.
- (19) Rardin, M. J.; Schilling, B.; Cheng, L. Y.; MacLean, B. X.; Sorensen, D. J.; Sahu, A. K.; MacCoss, M. J.; Vitek, O.; Gibson, B. W. MS1 Peptide Ion Intensity Chromatograms in MS2 (SWATH) Data Independent Acquisitions. Improving Post Acquisition Analysis of Proteomic Experiments. *Mol Cell Proteomics* **2015**, *14*, 2405-2419.
- (20) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; 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 and Cellular Proteomics* **2002**, *1*, 376-386.
- (21) Minogue, C. E.; Hebert, A. S.; Rensvold, J. W.; Westphall, M. S.; Pagliarini, D. J.; Coon, J. J. Multiplexed Quantification for Data-Independent Acquisition. *Analytical Chemistry* **2015**, *87*, 2570-2575.
- (22) Di, Y.; Zhang, Y.; Zhang, L.; Tao, T.; Lu, H. MdFDIA: A Mass Defect Based Four-Plex Data-Independent Acquisition Strategy for Proteome Quantification. *Analytical Chemistry* **2017**, *89*, 10248-10255.
- (23) Zhong, X.; Frost, D. C.; Yu, Q.; Li, M.; Gu, T.-J.; Li, L. Mass Defect-Based DiLeu Tagging for Multiplexed Data-Independent Acquisition. *Analytical Chemistry* **2020**, *92*, 11119-11126.
- (24) Hebert, A. S.; Merrill, A. E.; Bailey, D. J.; Still, A. J.; Westphall, M. S.; Strieter, E. R.; Pagliarini, D. J.; Coon, J. J. Neutron-encoded mass signatures for multiplexed proteome quantification. *Nature Methods* **2013**, *10*, 332-334.
- (25) Tian, X.; de Vries, M. P.; Permentier, H. P.; Bischoff, R. A Versatile Isobaric Tag Enables Proteome Quantification in Data-Dependent and Data-Independent Acquisition Modes. *Analytical Chemistry* **2020**, *92*, 16149-16157.
- (26) Rappsilber, J.; Ishihama, Y.; Mann, M. Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. *Analytical Chemistry* **2003**, *75*, 663-670.
- (27) He, L.; Diedrich, J.; Chu, Y. Y.; Yates, J. R. Extracting Accurate Precursor Information for Tandem Mass Spectra by RawConverter. *Analytical Chemistry* **2015**, *87*, 11361-11367.
- (28) Huang, Y.; Triscari, J. M.; Pasa-Tolic, L.; Anderson, G. A.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. Dissociation Behavior of Doubly-Charged Tryptic Peptides: Correlation of Gas-Phase Cleavage Abundance with Ramachandran Plots. *Journal of the American Chemical Society* **2004**, *126*, 3034-3035.
- (29) Muntel, J.; Gandhi, T.; Verbeke, L.; Bernhardt, O. M.; Treiber, T.; Bruderer, R.; Reiter, L. Surpassing 10 000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. *Molecular Omics* **2019**, *15*, 348-360.
- (30) Qin, H.; Wang, F.; Zhang, Y.; Hu, Z.; Song, C.; Wu, R.; Ye, M.; Zou, H. Isobaric cross-sequence

labeling of peptides by using site-selective N-terminus dimethylation. *Chemical Communications* **2012**, *48*, 6265-6267.

(31) Fu, Q.; Li, L. De Novo Sequencing of Neuropeptides Using Reductive Isotopic Methylation and Investigation of ESI QTOF MS/MS Fragmentation Pattern of Neuropeptides with N-Terminal Dimethylation. *Analytical Chemistry* **2005**, *77*, 7783-7795.

(32) Tran, N. H.; Qiao, R.; Xin, L.; Chen, X.; Liu, C.; Zhang, X.; Shan, B.; Ghodsi, A.; Li, M. Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. *Nature Methods* **2019**, *16*, 63-66.

(33) Scheltema, R. A.; Hauschild, J.-P.; Lange, O.; Hornburg, D.; Denisov, E.; Damoc, E.; Kuehn, A.; Makarov, A.; Mann, M. The Q Exactive HF, a Benchtop Mass Spectrometer with a Pre-filter, High Performance Quadrupole and an Ultra-High Field Orbitrap Analyzer. *Mol Cell Proteomics* **2014**, mcp.M114.043489.

Supporting Information

Materials and Methods

Chemicals and Materials. 2-Chlorotrityl chloride resin, Fmoc-Pro-OH, Fmoc-Ile-OH, sodium cyanoborohydride, formaldehyde solution (36.5-38% in H₂O), bovine serum albumin (BSA), acetic anhydride (¹²C₄H₆O₃), acetic anhydride-1,1'-¹³C₂ (¹³C₂¹²C₂H₆O₃), acetic anhydride-¹³C₄ (¹³C₄H₆O₃), 1-[bis(dimethylamino) methylene]-1H-1,2,3-triazolo [4,5-b] pyridinium 3-oxid hexafluorophosphate (HATU), 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (EDC·HCl), piperidine, pyridine, N,N-diisopropylethylamine (DIPEA), trifluoroethanol, N-hydroxysuccinimide (NHS), tris(2-carboxyethyl) phosphine hydrochloride (TCEP), 2-iodoacetamide, N,N-dimethylformamide (DMF), acetonitrile (ACN), acetone, dichloromethane (DCM), acetic acid, sodium hydroxide (NaOH), hydroxylamine hydrate, triethylammonium bicarbonate (TEAB), and trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich (The Netherlands). LysC and yeast extracts were purchased from Promega (The Netherlands). Synthetic peptides GTDWLANK and FDWA were purchased from GL Biochem (Shanghai, China). Solid phase extraction disk Empore Octadecyl C18 was purchased from 3M (United States).

Synthesis of triplex Ac-IP tag. The synthesis scheme of triplex Ac-IP-NHS is shown in **Scheme S1**. 30 mg of 2-chlorotrityl chloride (CTC) resin (1.0-1.8 mmol g⁻¹) was swollen with 1.5 mL of DCM/DMF (1/1) on a thermomixer at 800 rpm for 30 min at room temperature. The swollen resin was washed three times with 0.75 mL of DMF. Then, 1 mL of 0.8 M DIPEA in DMF, 0.5 mL of DMF and 0.2 mmol (67 mg) of Fmoc-Pro-OH was added. The reaction mixture was shaken for 90 min at room temperature. The solution was filtered out and the remaining Fmoc-Pro-CTC resin was washed three times with 0.75 mL of DMF. Afterwards, the resin was treated with 1.5 mL of 20% piperidine in DMF for 5 min and 15 min sequentially. Then, 0.5 mL of 0.4 M HATU in

DMF, 0.5 mL of 0.8 M DIPEA in DMF, 0.5 mL of DMF and 0.2 mmol (71 mg) Fmoc-Ile-OH were added. The reaction mixture was shaken for 90 min at room temperature. The resulting Fmoc-Ile-Pro-CTC resin was washed three times with 0.75 mL of DMF. Then, the resin was treated with 1.5 mL of 20% piperidine in DMF for 5 min and 15 min sequentially to afford intermediate **2**. Then three parallel reactions were performed: first, 0.5 mL of DCM, 5 µL of pyridine and 5 µL of acetic anhydride reacted with **2**; second, 0.5 mL of DCM, 5 µL of pyridine and 5 µL of acetic anhydride-1,1'-¹³C₂ reacted with **2**; third, 0.5 mL of DCM, 5 µL of pyridine and 5 µL of acetic anhydride-¹³C₄ reacted with **2**. The three reaction mixtures were shaken for 30 min at room temperature and then the isotopically labeled **3a-c** were released with 0.75 mL of 20% trifluoroethanol in DCM. Subsequently, the dried isotopically labeled **4a-c** were reacted with 80 mg EDC·HCl and 35 mg N-hydroxysuccinimide (NHS) in 0.7 mL DCM to generate **5a-c**, correspondingly. The reactions were shaken for 3 h, and after drying, the Ac-Ile-Pro-NHS esters were purified by RP-HPLC with detection at 214 nm. The collected HPLC fractions were freeze-dried to afford a white fluffy solid which was stored at -20 °C until use. Around 6 mg was obtained for each tag and the overall yield was around 50%. The LC-MS/MS analysis of triplex Ac-Ile-Pro-NHS esters is shown in **Supplementary Figure S1**. For LC purification, the HPLC system (Shimadzu, Kyoto, Japan) was equipped with a SIL-20AC autosampler, two LC-20AT pumps and an SPD-20A absorbance detector. The synthesized Ac-IP-NHS esters were separated on an XBridge BEH C18 OBD Prep Column (250 mm × 10 mm, 5 µm particles, 130 Å pore size, Waters, Ireland) with a 30 min gradient of 2-90% acetonitrile in water/0.1% formic acid at a flow rate of 2 mL min⁻¹.

Reduction/Alkylation and LysC Digestion. 100 µg of BSA or yeast protein was dissolved in 100 µL of 100 mM TEAB. Then, 5 µL of 200 mM TCEP was added and the mixture was incubated for 1 h at 37 °C. Afterwards, 5 µL of 375 mM freshly prepared 2-iodoacetamide was added and the mixture was incubated for 30 min in the dark at room temperature. Finally, 800 µL of acetone was added and the mixture was

kept at -20 °C overnight for precipitation. The precipitated protein was centrifuged at 20800 g for 30 min. Subsequently, the supernatant was decanted and alkylated proteins were stored at -20 °C until digestion. 100 µg alkylated BSA or yeast protein was dissolved in 200 µL of 100 mM TEAB and 2 µg LysC dissolved in 1 % formic acid was added. The pH was adjusted to 7.5-8 with 1 M TEAB and the digestion was kept overnight at 37 °C. The reaction was terminated by acidification using 10% TFA and the resulting LysC peptide mixtures of BSA and yeast proteins were freeze-dried and stored at -80 °C before conducting the labeling reactions.

Selective N-terminal dimethylation of peptides. 100 µg of peptide GTDWLANK, LysC peptides of BSA or LysC peptides of yeast proteins were dissolved in 100 µL of 1% (aq.) acetic acid. 10 µL of 4% formaldehyde in water was added to the peptide solution and the mixture was shaken for 2 min before adding 15 µL of 600 mM sodium cyanoborohydride dissolved in 1% (aq.) acetic acid. The reaction was shaken for 15 min at room temperature. Excess reagents were removed by SPE using the STAGE (STop And Go Extraction) TIPS Desalting Procedure and the collected N-terminally dimethylated peptides were freeze-dried and stored at -80 °C.

LC-MS/MS analysis. All multiplex-labeled proteomics samples were analyzed by nanoLC-MS/MS on a Q Exactive Plus mass spectrometer (Thermo Scientific, Bremen, Germany) while other samples were analyzed by LC-MS/MS on an Orbitrap Velos Pro (Thermo Scientific, Bremen, Germany) as detailed below.

LC-MS/MS analyses were performed on an Orbitrap Velos Pro mass spectrometer (Thermo Scientific) coupled to an LC system (Shimadzu, Kyoto, Japan) equipped with a SIL-20AC HT autosampler, and two LC-20AD pumps. The sample was separated on an ACQUITY UPLC BEH C18 column (2.1 mm × 50 mm, 1.7 µm particles, Waters, Ireland) at a column temperature of 60 °C with a 10 min gradient of 2-90% acetonitrile in water/0.1% formic acid at a flow rate of 300 µL min⁻¹.

NanoLC-MS/MS analyses were performed on a Q Exactive Plus mass spectrometer

(Thermo Scientific) equipped with an UltiMate 3000 RSLC nano system (Thermo Scientific, Germering, Germany). The peptides (around 300 ng) were first loaded onto a trap column (Acclaim PepMap 100 pre-column, 75 µm, 2 cm, C18, 3 µm, 100 Å, Thermo Scientific) and then separated on an analytical column (PepMap RSLC C18, 50 cm, 75 µm ID, 2 µm, 100 Å, Thermo Scientific) at a flow rate of 300 nL min⁻¹ and a column temperature of 40 °C. Solvent A was 0.1% formic acid in water and solvent B was 0.1% formic acid in acetonitrile. A gradient was applied from 2% to 50% or 65% B in the first 90 min, then 85% B for 5 min and equilibration for 15 min with 2% B.

For constructing the spectral library, mass spectrometry data were measured using a data-dependent acquisition (DDA) top 10 method. Full MS scans were acquired between *m/z* 300 and 1650 with a resolution of 70000 (at *m/z* 200). The automatic gain control (AGC) target was 3E6 and maximum injection time was 50 ms. Precursor ions were selected with a window of 2 Th. For studying fragmentation properties, selected ions were fragmented in the HCD collision cell with stepped normalized collision energies (NCEs) of 18, 20, 22, 24, 26, 28, 30 and 32. An NCE of 30 was selected for the final identification experiments. MS2 spectra were acquired in the Orbitrap mass analyzer with a resolution of 17500 (at *m/z* 200). The AGC target was 1E5 and the maximum injection time was 100 ms. Charge exclusion was set to discard singly-charged ions and ions with a more than 4 charges. The dynamic exclusion time was 20 s.

In data-independent acquisition (DIA) mode, each analysis consisted of two alternating scan experiments: a survey scan (*m/z* 350-1200), which was analyzed at a resolution of 70000 (at *m/z* 200), an AGC target of 3E6 and a maximum injection time of 50 ms, followed by MS2 scans with 20 sequential precursor ion selection windows of 29 Th (centered on 462.5, 487.5, 512.5, 537.5, 562.5, 587.5, 612.5, 637.5, 662.5, 687.5, 712.5, 737.5, 762.5, 787.5, 812.5, 837.5, 862.5, 887.5, 912.5 and 937.5) at a resolution of 17500 covering an *m/z* range of 450–950. An NCE of 30 was utilized in the DIA mode in analogy to DDA. The AGC target was 1E6 with the maximum injection time set to 'auto' and a 4 Th overlap between windows. The cycle including an MS1 survey scan

The isotopic Ac-IP tag enables multiplexed proteome quantification in data-independent acquisition mode

and 20 MS2 scans takes around 2 s, which means that 10-15 data points are acquired across each LC peak (peak width was 20-30 s).

Supplementary Figures

Table S1. Charge state distribution of different forms of modified GTDWLANK

Peptide	Singly-charged ions	Doubly-charged ions
	Percentage of total intensity	Percentage of total intensity
GTDWLANK	3.6%	96.4%
N-dime-GTDWLANK	2.4%	97.7%
N-dime-GTDWLANK-PI-Ac	2.6%	97.4%

Table S2. The influence of increasing resolution at MS1 and MS2 level on cycle time for the Q Exactive plus

MS 1 resolution	Maximum MS1 fill time	Number of DIA windows	MS 2 resolution	Maximum MS2 fill time	Cycle time
70k	50 ms	20	17.5k	100 ms	1.68-2.40 s
140k	50 ms	20	17.5k	100 ms	1.94-2.66 s
70k	50 ms	20	140k	100 ms	10.64-10.64 s

Figure S1. Removal of excess Ac-IP-COOH by washing six times with 100 μ L 2% ACN, 0.1% TFA during SPE before eluting peptides. The XICs of m/z 271.17 (Ac-IP-COOH) were compared. A-F) flow-through of first through sixth wash. More than 99.5% of Ac-IP-COOH was removed after the sixth washing step.

Figure S2. Example of deconvoluting the precursor ion isotope envelope and calculating the normalized ratio for a given peptide spectral match (PSM). The peptide **N-dime-TGTLTSETAHNMK-PI-Ac** is part of a yeast protein (P32324) derived from the triplex-labeled yeast sample mixed at 1:1:1.

Figure S3. Example of calculating peptide and protein ratios for triplex-labeled LysC yeast peptides mixed at 1:1:1. More than 10 PSMs were related to the peptide “TGTLTSETAHNMK” (only 5 are shown in the figure); MS2 spectrum 21061 had the highest significance value and the corresponding MS1 spectrum was 21043, marked in bold and selected to represent the ratio of this PSM. Fourteen peptides were related to protein P32324 (only 5 are shown in the figure). The peptides TGTLTSETAHNMK, DSVVAAFQWATK and AVQYLHEIK had the highest significance values, marked in bold, and thus selected to calculate the ratio of protein P32324.

Figure S4. LC-MS/MS of the triplex Ac-IP-NHS tag. (A) XIC of m/z 368.18, Ac-IP-NHS (+0). (B) Mass spectrum of Ac-IP-NHS (+0). (C) MS2 spectrum of Ac-IP-NHS (+0). (D) XIC of m/z 369.18, Ac-IP-NHS (+1). (E) Mass spectrum of Ac-IP-NHS (+1). (F) MS2 spectrum of Ac-IP-NHS (+1). (G) XIC of m/z 370.19, Ac-IP-NHS (+2). (H) Mass spectrum of Ac-IP-NHS (+2). (I) MS2 spectrum of Ac-IP-NHS (+2). Fragmentation was performed in the ion trap of an Orbitrap Velos Pro at a collision energy of 35.

The isotopic Ac-IP tag enables multiplexed proteome quantification in data-independent acquisition mode

Figure S5. ^1H NMR of Ac-IP-NHS. ^1H -NMR (500 MHz, CDCl_3): δ = 0.88-0.91 (t, J = 7.35 Hz, 3H), 0.99-1.00 (d, J = 6.78 Hz, 3H), 1.09-1.18 (m, 1H), 1.56-1.64 (m, 1H), 1.79-1.87 (m, 1H), 2.02 (s, 3H), 2.06-2.12 (m, 1H), 2.13-2.20 (m, 1H), 2.27-2.33 (m, 1H), 2.37-2.44 (m, 1H), 2.85 (s, 4H), 3.71-3.76 (m, 1H), 3.91-3.96 (m, 1H), 4.64-4.68 (t, J = 7.79 Hz, 1H), 4.80-4.83 (dd, J = 8.57, 5.25 Hz, 1H), 6.36-6.38 (d, J = 9.15 Hz, 1H) ppm.

Figure S6. ^{13}C NMR of Ac-IP-NHS. ^{13}C -NMR (126 MHz, CDCl_3): δ = 171.38, 170.45, 168.59, 167.41, 56.89, 54.88, 47.35, 37.79, 29.38, 25.60, 25.09, 24.46, 23.09, 15.08, 11.09 ppm.

The isotopic Ac-IP tag enables multiplexed proteome quantification in data-independent acquisition mode

Figure S7. COSY (A), HSQC (B) and HMBC (C) NMR of Ac-IP-NHS.

Figure S8. Influence of double Ac-IP labeling on ionization efficiency, before labeling (left) and after labeling (right). Ac-LWL was applied as an internal standard.

Figure S9. Derivatization of GTDWLANK. A) Starting material of the model peptide GTDWLANK and the reference peptide FDWA* (extracted ion chromatograms (XICs) of m/z 452.73 and 538.23); B) MS spectrum of GTDWLANK; C) N-terminally dimethylated (N-dime) GTDWLANK and N-dime-FDWA (XICs of m/z 452.73 (GTDWLANK), 466.74 (N-dime-GTDWLANK), 480.75 (N-dime-GTDWLANK-dime, doubly dimethylated form), 538.23 (FDWA) and 566.26 (N-dime-FDWA); D) MS spectrum of N-dime-GTDWLANK; E) N-dime-GTDWLANK-PI-Ac and N-dime-FDWA (XICs of m/z 466.74 (N-dime-GTDWLANK), 566.26 (N-dime-FDWA) and 592.82 (N-dime-GTDWLANK-PI-Ac); F) MS spectrum of N-dime-

The isotopic Ac-IP tag enables multiplexed proteome quantification in data-independent acquisition mode

GTDWLANK-PI-Ac-(+0).

*FDWA was included as a reference peptide, since the N-dime-FDWA cannot react with the Ac-IP tag. By normalizing the ionization efficiency with respect to N-dime-FDWA before and after labeling with the Ac-IP tag, the influence of modification can be assessed.

Figure S10. MS2 spectra of N-dime-GTDWLANK-PI-Ac at various NCEs. A) MS2 spectrum at an NCE of 18; B) MS2 spectrum at an NCE of 20; C) MS2 spectrum at an NCE of 22; D) MS2 spectrum at an NCE of 24; E) MS2 spectrum at an NCE of 26; F) MS2 spectrum at an NCE of 28; G) MS2 spectrum at an NCE of 30; H) MS2 spectrum at an NCE of 32.

Figure S11. Comparison of (+0)-Ac-IP, (+1)-Ac-IP, (+2)-Ac-IP labeled and 1:1:1 mixed triplex-labeled N-dime-GTDWLANK-PI-Ac. A) MS1 of the (+0)-Ac-IP labeled form; B) MS2 of the (+0)-Ac-IP labeled form; C) MS1 of the (+1)-Ac-IP labeled form; D) MS2 of the (+1)-Ac-IP labeled form; E) MS1 of the (+2)-Ac-IP labeled form; F) MS2 of the (+2)-Ac-IP labeled form; G) MS1 of triplex-labeled N-dime-GTDWLANK-PI-Ac mixed at a ratio of 1:1:1; H) MS2 of triplex-labeled N-dime-GTDWLANK-PI-Ac mixed at a ratio of 1:1:1.

Figure S12. Comparison of all PSMs (A) and of those PSMs with the highest -10lgP value (B) for a given peptide from a triplex-labeled LysC-digested BSA sample mixed at a ratio of 1:1:1.

Figure S13. Charge state distribution of identified unlabeled and Ac-IP labeled LysC yeast peptides.

Figure S14. Isotope distribution of a hypothetical 10-plex Ac-IP-NHS tag. The atoms marked with asterisks denote ¹³C or ¹⁵N. All isotopic forms can be synthesized from commercially available isotopically labeled amino acids and acetic anhydride.

Perspective and summary

In this thesis, we first discussed the development of stable isotope-based proteome quantification approaches using chemical labeling and compared their strengths and weaknesses. The limitations of each type of method are summarized, including: (1) for isotopic labeling, the limited multiplexing capacity (up to triplex) and the complexity of MS1 spectra that scales linearly with the number of labeled samples, which adversely affects data acquisition in the DDA mode, (2) for reporter ion-based isobaric tags, that the quantification may be distorted due to peptide cofragmentation, in which more than one peptide of the same sample is selected as precursor ion in the same isolation window and releasing indistinguishable reporter ions at the MS2 level, (3) for fragment ion-based isobaric labeling, the limited multiplexing capacity especially when avoiding deuterium labeling and the complexity of MS2 spectra which is multiplied with the number of labeled samples making identification of peptides challenging, (4) for the peptide-coupled reporter ion-based approaches, the modest efficiency of forming peptide-coupled reporter-ions and the complicated isotope envelope of the peptide-coupled reporter-ions, which requires deconvolution to extract the quantitative information.

To address or alleviate these limitations, it took 20 years from the first isotopic tag, ICAT¹, to the currently released most advanced isobaric tag, TMTpro². In the process, various remedies, techniques and concepts have been proposed, like the careful design of isotope tags, the exploitation of MS3-based quantitation and the assistance of ion mobility spectrometry. Still, these problems have not been overcome completely. There is still room for optimization and the development of new isotopic and isobaric tags and labeling strategies, notably for DIA. The design of isotope tags requires, among others, (1) to balance the formation of ions for quantification and fragment ions of the peptide backbone for identification, (2) to avoid formation of sequence uninformative fragment ions, (3) to increase multiplexing capacity, and (4) to control synthesis cost and increase practicality. Increasing multiplexing capacity at a reasonable cost is highly desirable to allow the analysis of large sample numbers, for example, in clinical proteomics studies. Besides, it should be noted that the widespread usage of a specific tag is not only

determined by its technical merits but also by its commercial availability. Last but surely not least, it is important to keep in mind that, besides the labeling strategies discussed here, the advancement of mass spectrometer technology and sophisticated data processing workflows are equally important to advance this field.

To address the above limitations in quantitative proteomics, I synthesized series chemical isotope labels by exploiting commercially available isotopic amino acids and optimized data processing.

In **Chapter 2**, the SMD-IPTL approach³ increased multiplexing capacity of IPTL, a fragment ion-based isobaric labeling approach, from 3-plex to 4-plex and has the potential to be extended to 7-plex with commercially available ¹³C- or ¹⁵N-labeled alanine and cysteine. Besides, a narrow precursor ion isolation window of 0.8 Th with an offset of -0.2 Th was applied during data acquisition, which significantly helped deducing protein ratios from intensities of fragment ions.

In **Chapter 3**, the Ac-IPG tag⁴ further increased the multiplexing capacity of fragment ion-based quantification and reduced the complexity of MS2 spectra by half, since only the y-series ions are distinct between labeling channels while the b-series ions are identical. Meanwhile, this approach maintains the IPTL feature of quantifying peptides based on peptide specific fragment ions, which avoids the issue of ratio distortion commonly observed for reporter ion-based quantification methods, such as TMT and iTRAQ.

In **Chapter 4**, we introduced the isobaric Ac-AG tag⁵ which quantifies peptides based on peptide-coupled reporter-ions and is compatible with both DDA and DIA modes. It is a valuable alternative to peptide-coupled reporter ion quantitation by the TMTc⁶ and EASI tags⁷. It is worthwhile to note that the Ac-AG tag generates extensive and intense peptide-coupled reporter ions. Around 90% of PSMs in a complex protein sample had peptide-coupled reporter ions with a complete isotope envelope which can be used for quantification. The ratios between labeling channels are simply reflected by the intensities of the peptide-coupled reporter ions in DDA mode, whereas the ratios can also be deduced from the peptide-coupled reporter ion isotope envelope in DIA mode

after deconvolution. The Ac-AG tag is thus a versatile isobaric tag that can be applied for multiplexed quantitation in both DDA and DIA mode without having to resort to ultra-high MS2 resolution.

In **Chapter 5**, the Ac-IP tag results in differentially labeled peptides that have distinct precursor ions carrying the quantitative information but identical MS2 spectra, since the differentially and isotopically labeled Ac-Ile part leaves as a neutral loss upon CID in addition to the fragmentation of the peptide backbone. In the survey scan, the shape of the isotope envelope of the precursor ions changes with different mixing ratios, and the quantitative information can be deduced by deconvolution. The multiplexing capacity may be extended to 10-plex to bring the multiplexing capacity of DIA closer to that of DDA while maintaining the merits of reproducibility and unbiased precursor ion selection.

To conclude, the work reported in this thesis proceeded along two lines of research: (1) In data-dependent acquisition (DDA) mode, eliminating the problem of ratio distortion caused by peptide cofragmentation and increasing the multiplexing capacity. (2) In data-independent acquisition (DIA) mode, achieving multiplexed quantification with general DIA LC-MS settings. It should be noted that identifications of peptides with the four approaches (SMD-IPTL, Ac-IPG tag, Ac-AG tag, and Ac-IP tag) were performed with readily available proteomics software (e.g. PEAKS and SearchGUI with Andromeda) with general settings. Quantifications were achieved with home-made Python scripts that have been shared on GitHub. All these approaches promote the multiplexed quantification in both DDA and DIA modes and make them accessible to a wider community of proteome researchers.

References:

- (1) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology* **1999**, *17*, 994-999.
- (2) Thompson, A.; Wölmer, N.; Koncarevic, S.; Selzer, S.; Böhm, G.; Legner, H.; Schmid, P.; Kienle, S.; Penning, P.; Höhle, C.; Berfelde, A.; Martinez-Pinna, R.; Farztdinov, V.; Jung, S.; Kuhn, K.; Pike, I. TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-Based Isobaric 16-Plex Tandem Mass Tag Reagent Set. *Analytical Chemistry* **2019**, *91*, 15941-15950.
- (3) Tian, X.; de Vries, M. P.; Visscher, S. W. J.; Permentier, H. P.; Bischoff, R. Selective Maleylation-Directed Isobaric Peptide Termini Labeling for Accurate Proteome Quantification. *Analytical Chemistry* **2020**, *92*, 7836-7844.
- (4) Tian, X.; de Vries, M. P.; Permentier, H. P.; Bischoff, R. A Collision-Induced Dissociation Cleavable Isobaric Tag for Peptide Fragment Ion-Based Quantification in Proteomics. *Journal of Proteome Research* **2020**, *19*, 3817-3824.
- (5) Tian, X.; de Vries, M. P.; Permentier, H. P.; Bischoff, R. A Versatile Isobaric Tag Enables Proteome Quantification in Data-Dependent and Data-Independent Acquisition Modes. *Analytical Chemistry* **2020**, *92*, 16149-16157.
- (6) Sonnett, M.; Yeung, E.; Wuhr, M. Accurate, Sensitive, and Precise Multiplexed Proteomics Using the Complement Reporter Ion Cluster. *Analytical Chemistry* **2018**, *90*, 5032-5039.
- (7) Winter, S. V.; Meier, F.; Wichmann, C.; Cox, J.; Mann, M.; Meissner, F. EASI-tag enables accurate multiplexed and interference-free MS2-based proteome quantification. *Nature Methods* **2018**, *15*, 527-530.

Perspectief en samenvatting

In dit proefschrift is eerst de ontwikkeling besproken van verschillende methodes voor de kwantificering van proteomen op basis van stabiele isotopen met behulp van chemische labeling, en zijn hun sterke en zwakke punten vergeleken. De beperkingen van elke methode zijn samengevat, te weten: (1) isotopisch labelen heeft een beperkte multiplex-capaciteit (tot maximaal triplex) en de complexiteit van MS1-spectra schaaft hierin lineair met het aantal gelabelde monsters, wat een nadelige invloed heeft op data-acquisitie in de DDA-modus; (2) voor isobarische tags gebaseerd op reporter-ionen kan de kwantificering worden verstoord door cofragmentatie van peptides, waarbij meer dan één peptide van hetzelfde monster wordt geselecteerd als precursor-ion in hetzelfde isolatievenster en waarbij niet te onderscheiden reporter-ionen worden gevormd op MS2-niveau; (3) isobare labeling op basis van fragmenten heeft ook een beperkte multiplexcapaciteit, vooral wanneer deuterium-labeling wordt vermeden, en de complexiteit van MS2-spectra vermenigvuldigt hier met het aantal gelabelde monsters, wat de identificatie van peptides tot een uitdaging maakt; (4) in methodes gebaseerd op peptide-gekoppelde reporter-ionen worden de reporter-ionen slechts met een lage efficiëntie gevormd en de complexe isotoop-distributie van deze ionen moet worden gedeconvolveerd om de kwantitatieve informatie af te leiden.

De ontwikkelingen om deze beperkingen aan te pakken of te verlichten vergden 20 jaar vanaf de eerste isotopische tag, ICAT¹, tot de momenteel meest geavanceerde isobarische tag, TMTpro². Gedurende deze periode zijn er verschillende oplossingen, technieken en concepten voorgesteld, zoals verbeterde ontwerpen van isotoop-tags, het gebruik van op MS3 gebaseerde kwantificering en de toepassing van ion-mobiliteit spectrometrie. Toch zijn de problemen nog niet volledig overwonnen; er is nog ruimte voor optimalisatie en voor de ontwikkeling van nieuwe isotopische en isobarische tags en labelingstrategieën, met name voor DIA-modus. Het ontwerp van isotoop-tags vereist onder meer (1) het in evenwicht te brengen van de vorming van ionen voor kwantificering en van fragment-ionen van de peptideketen voor identificatie, (2) het vermijden van de vorming van niet-informatieve fragment-ionen, (3) het vergroten van de multiplex-capaciteit en (4) het beperken van de synthesekosten en daarmee het

vergroten van de bruikbaarheid. Het vergroten van de multiplex-capaciteit tegen redelijke kosten is zeer wenselijk om de analyse van grote aantallen monsters mogelijk te maken, bijvoorbeeld in klinische proteomics-studies. Bovendien moet worden opgemerkt dat het wijdverbreide gebruik van een specifieke tag niet alleen wordt bepaald door zijn technische verdiensten, maar ook door zijn commerciële beschikbaarheid. Ten slotte is het belangrijk om in gedachten te houden dat, naast de hier besproken labelingsstrategieën, de vooruitgang van massaspectrometer-technologie en geavanceerde workflows voor dataverwerking net zo belangrijk zijn om het onderzoeksveld vooruit te helpen.

Om de bovenstaande beperkingen in kwantitatieve proteomics aan te pakken heb ik verschillende series van chemische isotoop-tags gesynthetiseerd met gebruik van commercieel verkrijgbare isotopische aminozuren, en methodes voor geoptimaliseerde dataverwerking ontwikkeld.

De in **Hoofdstuk 2** beschreven SMD-IPTL-benadering³ verhoogde de multiplex-capaciteit van IPTL, een op fragment-ionen gebaseerde isobare labelingsmethode, van 3-plex naar 4-plex en kan potentieel worden uitgebreid tot 7-plex met commercieel verkrijgbare ¹³C- of ¹⁵N-gelabeld alanine en cysteïne. Bovendien werd tijdens de data-acquisitie een smal precursor-ion isolatievenster van 0,8 Th met een offset van -0,2 Th toegepast, wat duidelijk hielp bij het afleiden van eiwitverhoudingen uit de intensiteiten van de fragment-ionen.

De in **Hoofdstuk 3** beschreven Ac-IPG-tag⁴ verhoogde de multiplex-capaciteit van de fragment-ion gebaseerde kwantificering verder en verminderde de complexiteit van de MS2-spectra met de helft, aangezien alleen de ionen uit de y-serie verschillend zijn tussen de labelingskanalen, terwijl de ionen uit de b-serie identiek zijn. Daarnaast maakt deze aanpak gebruik van de IPTL-eigenschap voor het kwantificeren van peptides op basis van peptide-specifieke fragment-ionen, waardoor het probleem van ratio-vertorming, dat vaak wordt waargenomen voor kwantificeringsmethoden gebaseerd op reporter-ionen zoals TMT en iTRAQ, wordt vermeden.

In **Hoofdstuk 4** introduceerden we de isobare Ac-AG tag⁵ die peptides kwantificeert

op basis van peptide-gekoppelde reporter-ionen en compatibel is met zowel DDA- als DIA-modi. Deze methode is een waardevol alternatief voor de kwantificering van peptide-gekoppelde reporter-ionen met behulp van TMTc⁶- en EASI-tags⁷. Het is van belang om op te merken dat de Ac-AG tag intense peptide-gekoppelde reporter-ionen genereert. Ongeveer 90% van de PSM's in een complex eiwitmonster vertoonden peptide-gekoppelde reporter-ionen met een complete isotoop-distributie die voor kwantificering kon worden gebruikt. De verhoudingen tussen labelingskanalen worden simpelweg afgeleid uit de intensiteiten van de peptide-gekoppelde reporter-ionen in DDA-modus, maar de verhoudingen kunnen ook worden afgeleid uit de isotoop-distributie van de peptide-gekoppelde reporter-ionen in DIA-modus door middel van deconvolutie. De Ac-AG-tag is dus een veelzijdige isobarische tag die kan worden toegepast voor multiplex-kwantificering in zowel DDA- als DIA-modi zonder afhankelijk te zijn van extreem hoge MS2-resolutie.

De in **Hoofdstuk 5** beschreven Ac-IP-tag resulteert in differentieel gelabelde peptides met verschillende precursor-ionen die de kwantitatieve informatie bevatten, maar die identieke MS2-spectra opleveren, aangezien het isotopisch gelabelde Ac-Ile-deel na CID als een ongeladen fragment verloren gaat. In de survey-scan verandert de isotoop-distributie van de precursor-ionen met verschillende mengverhoudingen, en de kwantitatieve informatie kan worden afgeleid na deconvolutie. De multiplex-capaciteit kan worden uitgebreid tot 10-plex en maakt multiplexing voor DIA mogelijk met bijna dezelfde capaciteit als die voor DDA gebruikelijk is, terwijl de voordelen van reproduceerbaarheid en onbevooroordeelde selectie van precursor-ionen behouden blijven.

Concluderend verliep het in dit proefschrift beschreven werk langs twee onderzoekslijnen: (1) het elimineren van het probleem van ratio-verneming veroorzaakt door cofragmentatie van peptides en het verhogen van de multiplex-capaciteit in de data-afhankelijke acquisitie (DDA)-modus; (2) het bereiken van multiplex kwantificering in data-onafhankelijke acquisitie (DIA)-modus met algemene DIA LC-MS-instellingen. Opgemerkt dient te worden dat identificatie van peptides met

de vier methodes (SMD-IPTL, Ac-IPG-tag, Ac-AG-tag en Ac-IP-tag) werd uitgevoerd met algemeen beschikbare proteomics-software (bijv. PEAKS en SearchGUI met Andromeda) met standaard-instellingen. Kwantificering werd uitgevoerd met zelfgeschreven Python-scripts die op GitHub zijn gedeeld. Al deze methodes bevorderen de multiplex kwantificering in zowel DDA- als DIA-modi en maken het toegankelijk voor een bredere groep van proteoomonderzoekers.

Referenties:

- (1) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology* **1999**, *17*, 994-999.
- (2) Thompson, A.; Wölmer, N.; Koncarevic, S.; Selzer, S.; Böhm, G.; Legner, H.; Schmid, P.; Kienle, S.; Penning, P.; Höhle, C.; Berfelde, A.; Martinez-Pinna, R.; Farztdinov, V.; Jung, S.; Kuhn, K.; Pike, I. TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-Based Isobaric 16-Plex Tandem Mass Tag Reagent Set. *Analytical Chemistry* **2019**, *91*, 15941-15950.
- (3) Tian, X.; de Vries, M. P.; Visscher, S. W. J.; Permentier, H. P.; Bischoff, R. Selective Maleylation-Directed Isobaric Peptide Termini Labeling for Accurate Proteome Quantification. *Analytical Chemistry* **2020**, *92*, 7836-7844.
- (4) Tian, X.; de Vries, M. P.; Permentier, H. P.; Bischoff, R. A Collision-Induced Dissociation Cleavable Isobaric Tag for Peptide Fragment Ion-Based Quantification in Proteomics. *Journal of Proteome Research* **2020**, *19*, 3817-3824.
- (5) Tian, X.; de Vries, M. P.; Permentier, H. P.; Bischoff, R. A Versatile Isobaric Tag Enables Proteome Quantification in Data-Dependent and Data-Independent Acquisition Modes. *Analytical Chemistry* **2020**, *92*, 16149-16157.
- (6) Sonnett, M.; Yeung, E.; Wuhr, M. Accurate, Sensitive, and Precise Multiplexed Proteomics Using the Complement Reporter Ion Cluster. *Analytical Chemistry* **2018**, *90*, 5032-5039.
- (7) Winter, S. V.; Meier, F.; Wichmann, C.; Cox, J.; Mann, M.; Meissner, F. EASI-tag enables accurate multiplexed and interference-free MS2-based proteome quantification. *Nature Methods* **2018**, *15*, 527-530.

Acknowledgments

Time flies, the end of my PhD stage is around the corner. I still clearly remember the first day when I came to Groningen. Closing eyes and recalling the last four years, many scenes came into my mind, working with mass spectrometers, doing sample preparation, struggling with data analysis, presenting results in different meetings and of course many memories about entertainment, fencing, boxing, climbing, and laser game in the Lab Day and a variety of tasty food in the Christmas Dinners. These colorful experiences make up the whole enjoyable PhD journey. To be honest, work and life are not always smooth and I have had some difficult times. Because of the culture gap, language barrier, and changing research direction from protein chemistry synthesis to proteomics, I found it difficult at the beginning of my PhD. Lucky thing is that I am surrounded by supportive supervisors and accommodating colleagues. Herein, I would like to express appreciation to those people who supported and helped me in this unforgettable journey.

First of all, I would like to appreciate my supervisors **Prof. R.P.H. Bischoff** and **Dr. H. P. Permentier** for taking me into the Analytical Biochemistry group. It is because of your recognition and choice that I can start my PhD journey.

My primary supervisor **Prof. R.P.H. Bischoff**, Dear **Rainer**, it is my honor to be your student. In my impression, you are easy-going and gentle in life, serious and rigorous at work. I really appreciate the encouragement, trust, and freedom you gave in the work. Meanwhile, you reminded me of the progress at the right time and adjusted the research directions. The beginning of my PhD was not easy and I struggled for a long time. I am greatly indebted to you for encouraging me when I wanted to work on isotope labeling and giving me time to learn Python, which turned out to be very useful for the projects. In addition, thank you much for modifying manuscripts in details and timely, I learn a lot from it. The scene of the first time I came to the lab still fresh that you led me to visit the laboratory and introduced me to the members of the group, the instruments, and the surroundings. Time flies too fast, and you are about to retire. I wish you will

enjoy life after retirement and all the best to your family.

My co-supervisor **Dr. H. P. Permentier**, Dear **Hjalmar**, you are a very kind and affable person, I am more inclined to regard you as a friend than a supervisor. I can't forget the moment that you said to me "Don't worry, that is not a big deal" when I was about to be assessed in the six-month evaluation, that was the first formal evaluation in PhD and I was anxious at that time. You really comforted me. Besides, you are a very organized person, you can work efficiently as schedule until getting off work and make full use of all kinds of "fragmented time". This is the quality I really admire and need to continue learning. Thank you for all supports and practical suggestions in LC-MS and data processing! The templet of Python scripts you shared with me really helped the data analysis in my projects. Moreover, you also taught me how to deal with some nonscientific problems and how to speak out. I really learned a lot from you in and outside of academia. I wish you everything goes well at work and a happy life with your family.

Dear **Marcel**, my big brother, you are the man with a cold appearance but a warm heart and the person who looks old but actually mentally young. You are a super professional and experienced Mass Spectrometer specialist and can always give straightforward solutions to my kinds of questions. You are like a "baby keeper" and take good care of all the instruments and make them run smoothly. You are my source of joy in work and I really enjoy the time working with you. I like talking with you for work and also for jokes. I can't finish the projects so smoothly without your support and help. Hi **Ydwine**, my big sister, your laugh is really distinctive. I really didn't expect that you would leave the group before me. I will miss the time we working and talking together. I won't forget your dancing in the lab with rhythmic music.

Hi **Jos**, you are always so kind and gentle. I still remember that you guided me to read the manual of TSQ at the very beginning of PhD and explained every technical term to

me in detail. Besides, you also gave me very practical suggestions for LC-MS, such as the automatic quantification in Xcalibur, which saves me lots of time. It is very nice to work together with you. Hi **Karin**, you are so energetic, kind and, always willing to help others. You are the person who bridges the gap between clinic biology and proteomics. I am glad to have you on my PhD journey. Hi **Peter**, thanks for your suggestions in data processing and other IT-related affairs. I was very happy when I knew that you were also interested in the concept of multiplexing in DIA and wish you good results in this direction.

Dear **Ali**, my friend, you are the one I have the most memories of among students. We together went to Southampton, London, Minster, Eindhoven for meetings and also entertainment. It was an interesting experience to have pizza together after a formal banquet. Ugh, pizza fits us better. In addition, what a wonderful and unforgettable experience at the Ferris wheel in London. Hi **Ola**, a quiet and polite person, it was nice to talk about anime and manga with you after intensive experiments, *Attack on Titan* has our common memory. I will also think of you when I think of this animation in my future life. Hi **Bas**, I am so glad that I met you. You are the one who always had lunch with me. I like talking with you about the difference between Dutch and Chinese culture. Hi **Baubek**, a very hard-working person, you are the one I always met in the lab during the weekends. You are an impressive person who balances well family life and work. **Janine**, **Alienke**, **Sara**, **Rik**, and **Alex**, thank you very much for all of you. You create a very inspiring and friendly atmosphere to work. Wish all students will have a successful PhD.

Larry (Tao Zhang), **Jiaying**, **Wenxuan**, and **Yang**, thanks a lot to all of you for helping me in work and life, even before I came to Groningen. Tips from you made my life and work easier. I still remember the scene of cycling on the island, having hot pot, and playing PUBG together, although it was rare to be the final winner, there was always a lot of joy. Wish all of you a successful career and a happy life with your family.

I also would like to thank all my Chinese friends in Groningen, **Xiaodong Feng**, **Zhangping Xiao**, **Shanshan Song**, **Bidong Zhang**, **Li Gao**, **Bohuan Lin**, **Xiaocui Wang**, **Weijia Yao**, **Yuchen Tian**, **Chongnan Ye**, **Yi Yu**, **Xinghong Zhao**, **Lu Zhou**, and **Huaiying Zhang**, all of you make me feel warm and not too far away from home.

Furthermore, I would like to express my gratitude to the China Scholarship Council (CSC) and the University of Groningen (RUG) for the PhD fellowship.

Last but not least, I would like to appreciate my families. My dear wife, **Chenyao**, who continuously accompanies, supports and understands me. Thank you for encouraging me when I am depressed, accompanying me when I am lonely, and supporting and understanding me when I facing indecisive choices. Thank you for your efforts to bring us closer and your persistence in our feelings, even though we have been in a long-distance relationship for a long time. I really look forward to living together every day in the near future.

亲爱的公、婆、爸、妈，感谢你们的养育之恩，感谢你们一如既往的支持与理解。虽然你们没有很多文化，但你们教会了我勤奋、自强、吃苦耐劳。这些品质一直、也将会继续在我人生中助我成长。

Xiaobo Tian
Groningen
June 2021

List of publications

1. **Xiaobo Tian**, Marcel P. de Vries, Susan W. J. Visscher, Hjalmar P. Permentier, and Rainer Bischoff. Selective maleylation-directed isobaric peptide termini labeling for accurate proteome quantification. *Analytical Chemistry* **2020**, 92, 7836–7844.
2. **Xiaobo Tian**, Marcel P. de Vries, Hjalmar P. Permentier, and Rainer Bischoff. A collision-induced dissociation cleavable isobaric tag for peptide fragment ion-based quantification in proteomics. *Journal of Proteome Research* **2020**, 19, 3817–3824.
3. **Xiaobo Tian**, Marcel P. de Vries, Hjalmar P. Permentier, and Rainer Bischoff. A versatile isobaric tag enables proteome quantification in data-dependent and data-independent acquisition modes. *Analytical Chemistry* **2020**, 92, 24, 16149–16157.
4. **Xiaobo Tian**, Marcel P. de Vries, Hjalmar P. Permentier, and Rainer Bischoff. The Ac-IP tag enables multiplexed proteome quantification in data-independent acquisition mode. *Analytical Chemistry* **2021**, 93, 23, 8196–8202.
5. **Xiaobo Tian**, Marcel P. de Vries, Hjalmar P. Permentier, and Rainer Bischoff. Chemical isotope labeling for quantitative proteomics. *Mass Spectrometry Reviews* **2021**, 1–31.