

MAXIMIZING PROTEOME RECOVERY AND DIGESTION EFFICIENCY
FOR HIGH-THROUGHPUT BOTTOM-UP MASS SPECTROMETRY

by

Jessica Lynn Nickerson

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University

Halifax, Nova Scotia

December 2022

© Copyright by Jessica Lynn Nickerson, 2022

Dedication

To Dad

*The lessons and values that you instilled
inspired every page*

Table of Contents

Dedication.....	ii
<i>Table of Contents</i>	<i>iii</i>
List of Tables	vi
List of Figures	vii
Abstract	x
List of Abbreviations Used.....	xi
Acknowledgements	xiv
1. Introduction	1
1.1 An Introduction to Proteomics	1
1.1.1 -Omics.....	2
1.1.2 MS-based Proteomics.....	5
1.1.3 Multi-omics.....	8
1.2 The Proteomics Workflow	9
1.3 Protein Solubility and Precipitation Mechanisms	10
1.4 Proteolytic Digestion	18
1.5 Automation and Miniaturization	29
1.6 Summary of Thesis Goals	31
2. Rapid and Quantitative Protein Precipitation for Proteome Analysis by Mass Spectrometry	33
2.1 Introduction	33
2.2 Methods	35
2.2.1 Materials.....	35
2.2.2 Growth and Extraction of Yeast Proteome.....	35
2.2.3 Acetone Precipitation: Varying Incubation Time and Temperature.....	35
2.2.4 BCA Quantitation.....	36
2.2.5 SDS PAGE Analysis	37
2.2.6 Trypsin Digestion.....	37
2.2.7 In-Gel Digestion.....	37
2.2.8 LC-MS/MS Analysis.....	38
2.2.9 Data Analysis	38
2.3 Results and Discussion	39
2.3.1 Rapid Precipitation at Room Temperature: Salt Trends.....	39
2.3.2 Higher Salt Increases Precipitation Kinetics in Acetone.....	41
2.3.3 Rapid Recovery of Dilute Samples	43
2.3.4 Temperature is Key to Rapid Recovery	44
2.3.5 Characterization of the Pellet and Supernatant	46
2.4 Conclusions	51

3. Critical Evaluation on the Repeatability of Sample Coverage and Quantitation in Bottom-Up Sample Preparation Strategies	52
3.1 Introduction	52
3.2 Methods	54
3.2.1 Growth and Extraction of Yeast.....	54
3.2.2 Solution Digestion.....	55
3.2.3 In-gel Digestion.....	55
3.2.4 Sample Preparation in the ProTrap XG.....	55
3.2.5 Bottom-up LC-MS/MS Acquisition with Label-Free Quantitation.....	56
3.2.6 DIA-MS Data Analysis.....	56
3.3 Results.....	57
3.3.1 Repeatability of Sample Coverage is Optimized in ProTrap XG Preparations.....	57
3.3.2 Quantitative Precision is Optimized by Sample Preparation in the ProTrap XG	64
3.4 Discussion	69
3.5 Conclusions.....	71
4. Multi-Omics in the ProTrap XG: Optimizing a Selective Precipitation Strategy for Mass Spectrometry Analysis of the Proteome, Peptidome, and Metabolome	72
4.1 Introduction	72
4.2 Methods	74
4.2.1 Growth, Culture, and Lysis of Adherent Cell Line	74
4.2.2 Precipitation-based Metabolome Purification	75
4.2.3 Targeted HILIC-MS Metabolomics Analysis	75
4.2.4 Metabolomics Data Analysis	76
4.2.5 Protein Precipitation for Proteome Recovery in the ProTrap XG	76
4.2.6 Sequential Peptidome Recovery and Preparation.....	76
4.2.7 Bottom-up Proteome Sample Preparation.....	76
4.2.8 LC-MS/MS of Sequentially Precipitated Pellets: DIA Acquisition with Label-Free Quantitation.....	77
4.2.9 Data Analysis of Bottom-up Protein Preparations and Intact Peptide Fractions.....	77
4.3 Results.....	78
4.3.1 Metabolomics Coverage and Quantitation is Optimized in the ProTrap XG	78
4.3.2 Proteomics Reveals Equivalent Coverage from Acetone and Methanol Precipitation.....	82
4.3.3 Peptidomics Analysis is Enhanced by Protein Depletion with Acetone	85
4.4 Discussion	87
4.5 Conclusions.....	88
5. Effects of Common Denaturing Additives on Trypsin Activity and Stability	89
5.1 Introduction	89
5.2 Methods	91
5.2.1 Initial Activity Assays.....	91
5.2.2 Time Course Activity Assays.....	91

5.2.3. Modeling Trypsin Deactivation Kinetics	92
5.3 Results.....	92
5.3.1 Effects of Surfactants on Trypsin Stability	93
5.3.2 Effects of Chaotropes on Trypsin Stability	98
5.3.3 Effects of Organic Solvents on Trypsin Stability.....	100
5.4 Discussion	102
5.5 Conclusions and Future Work	104
6. Maximizing Cumulative Trypsin Activity with Calcium at Elevated Temperature for Enhanced Bottom-Up Proteome Analysis	105
6.1 Introduction	105
6.2 Materials and Methods.....	107
6.2.1. Trypsin Activity and Stability Assays.....	107
6.2.2. Modeling Trypsin Deactivation Kinetics	107
6.2.3. Bottom-Up Proteome Sample Preparation	108
6.2.4. Bottom-Up LC-MS/MS Data Acquisition	108
6.2.5. LC-MS/MS Data Analysis	109
6.2.6. Data Availability	109
6.3. Results.....	109
6.3.1. Cumulative Trypsin Activity is Maximized at 47 °C with 10 mM Calcium Ions	109
6.3.2. Proteome Identifications Show High Similarity between Digestion Conditions	114
6.3.3. Relative Quantitation Reveals Trypsin Cleavage Is Accelerated at Elevated Temperature with Added Calcium Ions.....	118
6.4. Discussion	122
6.5. Conclusions.....	126
7. Future Work & Conclusions	128
7.1 Thesis Summary	128
7.2 Future Work	131
7.2.1 Rapid and quantitative organic solvent precipitation	131
7.2.2 Evaluating the Repeatability of Sample Preparation in the ProTrap XG	132
7.2.3 Precipitation-Based Multi-omics.....	132
7.2.4 Effect of Denaturants on Trypsin Activity	133
7.2.5 Calcium-Assisted Rapid Digestion in the ProTrap XG.....	133
7.3 Conclusions.....	133
<i>Bibliography</i>	<i>135</i>
<i>Appendix A</i>	<i>172</i>
Required Copyright Reproduction Agreements.....	172

List of Tables

Table 1.1 Precipitation conditions used in Cohn Fractionation [123].....	15
Table 2.1 Rate constant determination for room-temperature precipitation in the low-salt condition	43
Table 5.1 Summary of trypsin deactivation kinetics and cumulative activity over time in the presence of select surfactants, organic solvents and chaotropic agents.....	93
Table 6.1 Summary of trypsin deactivation rates at various temperatures in the presence or absence of calcium ions.	112

List of Figures

Figure 1.1 Arrow-pushing mechanism of peptide bond hydrolysis by trypsin.....	19
Figure 2.1 Workflow diagram of precipitation time course.	36
Figure 2.2 Influence of salt on protein precipitation.....	39
Figure 2.3 Influence of salt types on precipitation efficiency	40
Figure 2.4 Influence of salt concentration on precipitation recovery..	42
Figure 2.6 Precipitation recovery of dilute samples.	44
Figure 2.7 Bottom-up protein identifications from low- and room-temperature precipitation.....	45
Figure 2.8 Box-and-Whisker plots of structural properties of rapid precipitation pellets.	46
Figure 2.9 Venn diagrams comparing protein identifications from bottom-up LC-MS/MS analysis of pellets and supernatants.	46
Figure 2.10 Venn diagrams of protein identifications made from duplicate MS injections.....	47
Figure 2.11 Venn diagrams comparing proteins identified from replicate MS injections	48
Figure 2.12 Box-and-Whisker plots of protein properties following rapid precipitation.	49
Figure 2.13 Histogram of proteins identified from rapid precipitation.	50
Figure 2.14 Box-and-Whisker plots of protein properties following room-temperature precipitation.	51
Figure 3.1 Venn diagrams of bottom-up peptide and protein identifications from replicate preparations.....	58
Figure 3.2 Summary of bottom-up peptide and protein identifications.....	59
Figure 3.3 Miscalculation analysis of peptides identified from each preparation strategy.	61
Figure 3.4 Correlation of qualitative repeatability with peptide properties.....	63
Figure 3.5 Assessment of quantitative repeatability for peptides with no missing values..	65
Figure 3.6 Violin plots of peptide coefficients of variation.....	67
Figure 3.7 Correlation plots of peptide intensities across replicate pairs.	68
Figure 4.1 Metabolomics coverage following various precipitation strategies.	79

Figure 4.2. Heatmap of average metabolite intensities.....	82
Figure 4.3 Bottom-up protein identifications from triplicate preparations.....	83
Figure 4.4 Label-free bottom-up protein quantitation analysis..	85
Figure 4.5 Endogenous peptidome coverage from various protein depletion strategies.....	87
Figure 5.1 Effect of sodium dodecyl sulfate on trypsin activity and stability.....	94
Figure 5.2 Effect of sodium deoxycholate on trypsin activity and stability.....	97
Figure 5.3 Effect of sodium laurate on trypsin activity and stability.....	98
Figure 5.4 Effect of chaotropes on trypsin activity and stability.....	99
Figure 5.5 Effect of organic solvents on trypsin activity and stability.....	102
Figure 5.6 Summary of cumulative trypsin activity in the presence of denaturants.....	105
Figure 6.1 Effect of calcium ions and temperature on initial trypsin activity.....	110
Figure 6.2 Time course assessment of trypsin activity.....	111
Figure 6.3 Effect of calcium concentration of trypsin stability.....	111
Figure 6.4 Second-order kinetics models of trypsin de-activation.....	112
Figure 6.5 Cumulative trypsin activity across a 16 h time course.....	113
Figure 6.6 Venn diagram comparing <i>in-silico</i> digestion with identified peptides.....	115
Figure 6.7 Qualitative bottom-up proteome identifications from experimental digests.....	115
Figure 6.8 Miscleavage analysis of bottom-up proteome identifications.....	116
Figure 6.9 Venn diagrams of bottom-up peptide identifications between experimental and control digests.....	117
Figure 6.10 Venn diagrams of bottom-up protein identifications between experimental and control digests.....	117
Figure 6.11 Venn diagrams of bottom-up protein identifications contributed by fully-cleaved peptides.....	118
Figure 6.12 Relative quantitation of fully-cleaved, singly miscleaved, and doubly miscleaved peptides.....	119
Figure 6.13 Tukey Box-and-Whisker plots of relative peptide quantitation.....	120

Figure 6.14 Evaluation of cleavage specificity in rapid digest.	121
Figure 6.15 Characterization of miscleaved peptides.	122
Figure 6.16 Histogram of the protein identification rates as a function of abundance.	125

Abstract

With current mass spectrometry and bioinformatics platforms, proteome analysis is at the forefront of characterizing complex biological systems. Through deep qualitative coverage combined with precise quantitation, proteomics represents a powerful tool in the pursuit of understanding disease-driving mechanisms and elucidating precise therapeutic approaches. However, the stringency of proteomics output is limited by the coverage and precision afforded by front-end preparation strategies. Much of the current proteomics literature relies on the maximum potential of state-of-the-art MS acquisition technologies without leveraging optimal front-end processing. Furthermore, many of the existing sample preparation strategies (reviewed in Chapter 1 of this thesis) impart a trade-off between recovery, digestion efficiency, and precision.

The present thesis aims to evaluate the factors limiting front-end workflows and propose practical alternatives that maximize coverage, quantitative precision, and throughput. Organic solvent-based precipitation as a means of proteome purification has often been overlooked based on conflicting reports of efficiency. Following previous work from this group, Chapter 2 of this thesis assesses the rate-limiting variables associated with protein precipitation and demonstrates a rapid and robust approach to precipitation-based proteome recovery. Chapter 3 provides an evaluation of the repeatability of a precipitation-based bottom-up proteome workflow on the basis of sample coverage and the precision of peptide quantitation. Chapter 4 evaluates the potential of the enhanced precipitation approach towards multi-omics preparations.

Bottom-up proteome strategies rely on robust enzymatic digestion with trypsin. Many common proteomics additives, however, impede the enzyme's stability. Chapter 5 of this thesis characterizes the effects of several denaturing additives, demonstrating that these solubilizing agents are included at the expense of proteolytic efficiency. A wide variety of alternative digestion approaches have been described towards improved throughput over the conventional overnight incubation, although the limited validation reduces their potential for precise quantitation. Chapter 6 of this thesis characterizes the effects of elevated temperature in combination with the stabilizing effects of calcium ions towards a rapid approach to complete digestion while demonstrating the implications for bottom-up proteome analysis. Future studies, summarized in Chapter 7, suggest the application of the described rapid precipitation and enzymatic digestion to the development of targeted assays in large-scale clinical settings.

List of Abbreviations Used

Abbreviation	Description
μFASP	micro filter-aided sample preparation
2D	two-dimensional
2D GE	two-dimensional gel electrophoresis
Arg	arginine
Asp	aspartic acid
BAEE	N α -Benzoyl L-arginine ethyl ester
BCA	bicinchoninic acid
BRCA	BReast CAncer
BSA	bovine serum albumin
CE-MS/MS	capillary electrophoresis-tandem mass spectrometry
CMC	critical micelle concentration
CMW	chloroform/ methanol/ water
CV	coefficient of variation
Cys	cysteine
DDA	data-dependent acquisition
DIA	data-independent acquisition
DIA-NN	data-independent acquisition neural network
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
eFASP	enhanced filter-aided sample preparation
ESI	electrospray ionization
FA	formic acid
FASP	Filter-Aided Sample Preparation
FDR	false discovery rate
FFPE	formalin-fixed, paraffin-embedded
Gdn HCl	guanidine hydrochloride
GE	gel electrophoresis
Glu	glutamic acid
GRAVY	grand average of hydropathicity
HeLa cells	Henrietta Lacks immortalized cell line
HIFU	high-intensity focused ultrasound
His	histidine
HLA	human leukocyte antigen
HPP	Human Proteome Project
IAA	iodoacetamide

Abbreviation	Description
ID	identification
i.d.	internal diameter
IDA	information dependent acquisition
ILIS	isotopically labeled internal standard
IMER	immobilized enzyme reactor
ISD	in solution digestion
iST	In Stage Tip
iTRAQ	Isobaric Tags for Relative and Absolute Quantification
K	lysine
LC	liquid chromatography
LC-MS/MS	liquid chromatography-tandem mass spectrometry
LOD	limit of detection
LOQ	limit of quantitation
Lys	lysine
<i>m/z</i>	mass to charge ratio
MALDI	matrix-assisted laser desorption/ionization
MAR	missing at random
MaSDeS	MS-compatible degradable surfactant
MCAR	missing completely at random
MHC	major histocompatibility complex
MNAR	missing not at random
MRM	multiple reaction monitoring
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MV	missing value
MWCO	molecular weight cut-off
Nano3	nanoparticle-packed n+C40 nanoreactor for nano-proteomics
nanoPOTS	nanodroplet processing in one-pot for trace samples
NK	natural killer
OD600	optical density at 600 nm
PBS	phosphate-buffered saline
PC	pancreatic cancer
PCA	principal component analysis
PCR	polymerase chain reaction
Phe	phenylalanine
pI	isoelectric point
PSM	peptide spectral match
PTM	post translational modification
QconCAT	Quantification conCATamer

Abbreviation	Description
R	arginine
RNA	ribonucleic acid
RPLC	reversed phase liquid chromatography
RSD	relative standard deviation
RT	room temperature
SARS CoV-2	severe acute respiratory syndrome coronavirus 2
SC ₅₀	drug concentration eliciting 50 % of the maximum stimulation
SDC	sodium deoxycholate
SDS	sodium dodecyl sulfate
SDS PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SILAC	Stable Isotope Labelling of Amino Acids in Cell Culture
SiTrap	Simultaneous Trapping
SNP	single nucleotide polymorphism
SP3	single pot solid phase sample preparation
SPE	solid-phase extraction
SPME	solid phase microextraction
S-Trap	Suspension Trapping
SWATH-MS	Sequential Window Acquisition of All Theoretical Mass Spectra
TCA	trichloroacetic acid
TEAB	triethylammonium bicarbonate buffer
TFA	trifluoroacetic acid
TMT	tandem mass tag
TPCK	L-tosylamido-2-phenylethyl chloromethyl ketone
Trp	tryptophan
Tyr	tyrosine
UV	ultraviolet
vol	volume
YPD	yeast extract-peptone-dextrose

Acknowledgements

I would first like to thank my supervisor, Dr. Alan Doucette. Throughout these years, you have shown me unwavering support and greater confidence in me than I sometimes had in myself. Your infectious passion and energy made what could have been “hard work” an absolute joy and will continue to inspire my future endeavors.

To my immediate family, I thank my husband, Brayden MacQuarrie; my mother, Janet Nickerson; and my brother, Daniel Nickerson for their love and belief in me through every stage of this journey. I would also like to acknowledge my late father, Greg Nickerson. Although he only saw the first year of my degree, the amount of excitement he shared for my scientific career will last a lifetime. I additionally thank my extended family, especially Kristina Fudge and my MacQuarrie family. This degree was, in many respects, a team effort—I could not have done it without all of your support, patience, grounding, and encouragement.

I would like to thank a couple of friends, who are truly more like family. Specifically, I thank Nicole Unterlander for many critically-decompressing wine nights (both in-person and virtual), and for “showing me the ropes” as I got my start in the lab. I would also like to thank Paul Bissonnette for being a constant source of inspiration and much-needed laughter since the early days of our academic journeys.

I would like to thank my committee members Dr. Alejandro Cohen, Dr. Jan Rainey, and Prof. Amares Chatt for your insight and support across my degree. I extend my thanks to Dr. Dajana Vuckovic and Dr. James Fawcett for reviewing my final thesis. An additional thanks goes to Dr. Cohen for his contribution to much of my MS data acquisition at Dalhousie’s BMS Core Facility. I also thank Dr. Hugo Gagnon and his team at the PhenoSwitch Biosciences facility for their MS and data analysis services. Additionally, I thank Dr. Jeanette Boudreau and Stacey Lee (PhD candidate) for the opportunity to gain training in their lab and to collaborate on an impactful application of our novel sample preparation strategies.

1. Introduction

1.1 An Introduction to Proteomics

The diverse field of proteomics provides a powerful means of investigating biological systems, relying heavily on sophisticated separations, mass spectrometry platforms, and bioinformatics technologies to characterize protein expression. Proteomics has grown rapidly over the last several decades, stemming from protein mapping by way of two-dimensional gel electrophoresis and evolving with improved capabilities of mass spectrometry [1]. Since John Fenn's adaptation of electrospray ionization (ESI) towards the ionization of biological macromolecules [2], liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) and subsequently capillary electrophoresis with tandem mass spectrometry (CE-MS/MS) platforms have been the gold standard for proteome analysis [3]. MS-based protein identification was conventionally performed by sequencing smaller protein segments generated by enzymatic digestion in a process known as bottom-up proteomics. With improvements in MS sensitivity and resolution, the capacity for direct MS characterization of intact proteins (i.e., top-down proteomics) has become increasingly popular.

To maximize the potential of modern MS instrumentation to characterize complex protein systems, fast, robust, and reliable front-end sample preparation strategies are also needed. The structural heterogeneity across a proteome mixture is large, with molecular weights spanning 5-500 kDa [4,5], a range of solubility differences between cytosolic and membrane proteins, charge states ranging from highly negative to highly positive, and varying degrees of intrinsic disorder. This diversity contributes biases in sample recovery when subjecting the sample to various manipulations associated with front-end proteome workflows. This is especially concerning when considering the vast dynamic range of a proteome, spanning ~ 10 orders of magnitude. Often, the most interesting analytes are those present in low abundance, and so it is critical that preparative workflows conserve the entire proteome, especially as sample sizes diminish to as low as a single cell.

1.1.1 -Omics

Proteomics is just one of several “-omics” initiatives, referring to fields of study to comprehensively examine the diverse classes of molecules from biological systems (e.g., genomics: DNA, transcriptomics: mRNA, proteomics: proteins/proteoforms, metabolomics: metabolites). Genomics, for example, grew from simple DNA sequencing into a highly ambitious goal of sequencing the full set of DNA expressed in a given organism. The Human Genome Project was formerly established and undertaken between 1990-2003, being far quicker than anyone had anticipated [6]. Proteomics emerged shortly after the Human Genome Project was initiated to characterize the biologically functional products encoded by the genome. Similarly, transcriptomics characterizes the complete set of RNA transcripts produced by the genome. Omics fields of study have since expanded to include glycomics, lipidomics, metabolomics, and others, with growing interest in multi-omics analyses (e.g., metabolomics + proteomics) enabling increased coverage of complex cellular processes.

Genomics encompasses the physical and functional mapping, study, and altering of the genes expressed by an organism. The full DNA sequence of an individual encodes a diverse array of biological functions, affording insight into disease-driving mechanisms. The field of genomics had its earliest origins in the 1960’s as Robert Holley elucidated the first nucleic acid sequence of transfer RNA [7]. This was followed by Walter Fiers who determined the first sequence of a gene in the 1970’s [8,9]. In 1977, Frederick Sanger reported the chain-termination sequencing method (also known as Sanger sequencing), relying on chemical chain-termination followed by both gel electrophoretic analysis [10]. While Next Generation DNA sequencing techniques emerged in the early 2000’s (and continue to evolve) [11,12], the Sanger method is still widely used today, including for the analysis of the SARS-CoV-2 virus [13]. In 1985-1986, The Human Genome Project was proposed, with the aim of sequencing the complete human genome (>3 billion base pairs) [14–16]. This huge initiative, often referred to as *Biology’s Moonshot* represented one of the largest scientific collaborations, relying on the efforts of hundreds of researchers from 20 international sites [17,18]. The first draft was released in 2001 [19], comprising 90 % of the total genome, followed by what was deemed an “essentially complete” version in April of 2003 [20]. Additional refinements were reported recently in May 2021, with a final gapless assembly most recently released in January 2022 [21].

The published human genome, combined with the efficiency and accessibility of Next Generation genome sequencing techniques have been invaluable in medicine, from deconvoluting the genetic basis of diseases to predicting drug response. Genetic sequencing is nearly a standard of care for several malignancies, enabling precision and personalized medicine approaches for the targeted treatment of heterogeneous diseases. Mutations in the BRCA (BRCA1 and BRCA2) gene, for example, are highly informative markers in breast, ovarian and pancreatic cancer [22–24]. Unfortunately, the genome does not completely encode the diverse array of biological functions. While functional genomics technologies are available to explore the relation between gene, RNA and protein products, the very nature of transcription and translation results in a set of proteins that are more complex than the genome that encodes them. The additional heterogeneity of the proteome arises in part from alternative splicing events, open reading frames in mRNA translation, and single nucleotide polymorphisms (SNP). This is further compounded by post translational modification (PTM) of the protein. While the human genome encompasses roughly 20,000 genes, the number of distinct forms of proteins (i.e., proteoforms) could potentially be 2 orders of magnitude higher [25]. The study of proteins comes with the added challenge of differential expression. While genomics analysis benefits from polymerase chain reaction (PCR) amplification to maximize sensitivity [26], the differential expression of proteins in human plasma is estimated to span 10 orders of magnitude [27], complicating the detection of low-abundance analytes.

The term “proteome” was coined by Wilkins in 1994 to describe the total complement of proteins encoded by a given genome. While the term has only been formally defined for less than thirty years, the earliest proteomics experiments date back to the 1970’s whereby two-dimensional (2D) gel electrophoresis (GE) was employed by O’Farrell to map *E. coli* proteins [28]. The method of 2D GE separates proteins in the first dimension by isoelectric focusing and subsequently according to molecular weight, resulting in a 2D map of spots, each representing a distinct protein. Protein identification/characterization relies largely on the determination of the primary amino acid sequence, which first relied on chemical approaches such as Edman degradation [29] or alternatively, subjecting the protein spots to immunoassays, such as Western blotting [30] wherein the identity of a protein is confirmed by selective interaction with an antibody. Eventually, these techniques gave way to MS-based strategies for protein sequencing. While gel-based methods for protein identification continue to be exploited today [31,32], their limited throughput motivated alternative strategies for larger scale goals of deep proteome profiling.

Protein detection by mass spectrometry was not widely utilized until the advent of soft ionization techniques, namely, matrix-assisted laser desorption/ionization (MALDI) [33] and ESI [2]. In 1987, Tanaka demonstrated ionization of proteins using MALDI with metal particles suspended in a glycerol matrix. Subsequent efforts showed that higher molecular weight proteins were more effectively ionized from nicotinic and cinnamic acid matrices [34]. In addition to desorptive methods for ionization of biological molecules, in 1989 John Fenn effectively coupled electrospray ionization (first described by Dole et al., [35]) to a mass spectrometer for effective detection of intact proteins. Tanaka and Fenn shared the Nobel Prize in 2002. The ability to ionize such large, non-volatile biological molecules from the liquid phase enabled the invaluable direct coupling with liquid phase separations.

While Fenn's work paved the way for modern proteomics methodologies, it was two decades earlier that Hunt employed tandem MS to determine the amino acid sequence of trace proteins following enzyme and/or chemical digestion, with liquid chromatography (LC) fractionation to simplify the resulting mixture [36]. It was stated that the total time for this process "rarely exceeds 4 or 5 days". Indeed, MS-based proteome sequencing has come a long way. With electrospray ionization (ESI) extending MS toward larger biomolecules, direct sequencing of intact proteins quickly followed, forging the "top-down" approach [37]. Given the limitations of MS resolution and high sample complexity, detection of digested peptides was more efficient. The modern era of bottom-up proteome analysis [38] was therefore established, coupling MS with separation platforms, together with bioinformatic approaches [39] that leveraged the information compiled in expanding protein sequence databases [40]. Rounding the toolbox for bottom-up proteome analysis were front-end strategies to facilitate protein extraction, digestion, and purification [41]. The current section focuses on MS acquisition strategies; sections 1.3 & 1.4 provide descriptions of the variables influencing protein solubility and enzymatic digestion, respectively.

1.1.2 MS-based Proteomics*

Currently, bottom-up MS can rapidly profile proteomic mixtures, with ~100-200 unique protein identifications per minute [42–44]. With such capacity, the Human Proteome Project, proposed by Williams in 1996, is now approaching completion. At present, over 90 % of the 19,873 protein entries have been catalogued with high stringent MS data representing a minimum of two non-nested peptides reported at a false discovery rate (FDR) <1 % [45]. However stringent, the detection of a protein based on surrogate peptides inherently leaves a void in the capacity of the HPP to fully disclose expressed gene products.

Several MS acquisition strategies have been developed to accommodate the diverse goals of proteome experiments. For example, discovery-based studies and whole-sample profiling rely on untargeted/shotgun approaches. These experiments have conventionally been achieved by data-dependent acquisition (DDA), which subjects the most intense precursors detected by MS1 to fragmentation and MS2 detection. These tandem MS spectra are searched against a spectral library generated from an *in-silico* digest of the indicated proteome sample [46,47]. Matched spectra are filtered globally based on a false positive rate (usually 1 %) and individually based on a scoring system [48,49].

Advances in the scan speed and sensitivity of mass spectrometry instrumentation enable a broader approach, data-independent acquisition (DIA). In DIA workflows, every precursor ion from the MS1 scan is subject to fragmentation. Contrasting with DDA, this provides tandem MS detection of all ions within a selected *m/z* range, theoretically enabling deeper qualitative characterization and more accurate quantitation [50,51]. However, the simultaneous fragmentation of several precursors produces complex tandem MS spectra, posing challenges with data analysis. The sequential fragmentation employed in DDA facilitates MS2 spectra containing fragment ions from a single precursor, making peptide identification easier. DIA spectra are processed by one of two approaches—spectrum- or peptide-centric. As reviewed by Halder et al. [52], in a spectrum-centric approach, DIA-MS spectra are converted to pseudo-DDA MS/MS spectra, which can subsequently be searched against a spectral library of an *in silico* digest [53,54]. The alternative,

* Section 1.1.2 is based on the published article Nickerson J.L., Baghalabadi V., Rajendran S.R.C.K., Jakubec P.J., Said H., McMillen T.S., Dang Z., Doucette A.A. Recent advances in top-down proteome sample processing ahead of MS analysis. *Mass Spectrom. Rev.* 2021 doi: 10.1002/mas.21706.

peptide-centric approach simulates a spectral library based on the input protein sequence. These predictive strategies are now benefitting from Deep Learning approaches based on large DDA datasets and are facilitated by various algorithms such as Pecan [55] and DIA-NN [56] analysis platforms. As reviewed by Krasny et al., DIA offers significant enhancements in identification repeatability compared to DDA, which is more prone to missing a given peptide from one run to the next based on variances in the elution profile [57]. This enhanced qualitative identification rate inherently improves quantitative results, whereby the reduced frequency of missing values increases the precision of peptide and protein quantitation. A 2017 update on the Human Proteome Project reported organ-specific protein databases to facilitate the application of DIA-MS strategies for disease investigations [58].

The enhancements in sensitivity and scan speed that are enabling the use of DIA-MS approaches are also facilitating broader quantitative strategies. In a targeted quantitation approach, complex samples may be surveyed for a particular protein, e.g., a diagnostic biomarker, by way of multiple reaction monitoring (MRM) [59]. In these strategies, only the precursors of interest are subject to fragmentation, which optimizes the sensitivity for a small fraction of the complex sample. Such MRM assays are often employed in biomarker detection and for accurate quantitation [60–62]. By contrast, untargeted quantitation has traditionally relied on isotopic labelling strategies. The repeatability of MS acquisition impedes the comparison of a precursor's intensity from one run to the next, so comparative samples may be differentially labeled, pooled, and analyzed simultaneously for a direct comparison of MS1 peak areas. Such labeling strategies include Stable Isotope Labelling of Amino Acids in Cell Culture (SILAC) [63], reductive dimethyl labelling [64], and tandem mass tagging (TMT) [65] for relative quantitation as well as Isobaric Tags for Relative and Absolute Quantification (iTRAQ) [66], Quantification conCATamer (QconCAT) [67], and internal standard approaches for absolute quantitation [68]. These labeled relative quantitation approaches provide benefits with multiplexing (up to 18 samples with the TMT strategy); however, they are limited by variances in labeling efficiency, throughput, and cost. The enhanced repeatability of DIA-MS approaches provides high agreement in identifications as well as precursor ion intensities across LC-MS/MS runs (with reported coefficients of variation of <10 % [69,70]). By this, label-free quantitation facilitates a low-cost approach to precise quantitation, with particular benefits for large scale analysis, or challenging systems such as single cell analyses. However, the run-to-run instrumental variance should be assessed with respect to

labelling efficiencies to determine whether a labeled or label-free approach would provide the greatest quantitative accuracy for a given platform.

Thus far, the presented descriptions of MS acquisition have assumed a bottom-up approach, where intact proteins are digested into lower molecular weight peptides for easier detection and data analysis. The alternative approach to MS-based proteome characterization, namely top-down proteomics, eliminates the step of enzymatic digestion, instead performing tandem MS of the intact protein. Such large molecules pose challenges for tandem MS analysis by resisting fragmentation, which has led to low sequence coverage. Several advances in fragmentation strategies have been reported recently, including sequential ETD and HCD, collisionally activated dissociation, and mesh fragmentation [71–73]. The associated enhanced sequence coverage offers advantages in confidence of protein identification and detecting correlated post translational modifications. Deconvolution of MS1 and MS2 spectra of intact proteins is challenging owing to wide charge state envelopes and variable fragmentation efficiency. Despite these challenges, several strategies for protein identification based on top-down spectra have been developed, relying on shotgun annotated databases [74,75], fragment ion mapping [76], long peptide sequence tagging [77], and spectral alignment [78,79]. Given the numerous post translational events that can decorate a protein, including phosphorylation, glycosylation, acetylation, sumoylation, top-down proteomics allows a wholistic view of the distinct form of the protein. A protein exhibiting all of its final post-translational modifications was coined *proteoform* by Smith and Kelleher in 2013 [80]. With all possible degradation products, sequence variants, and permutations of PTMs, the complexity of the human proteome is orders of magnitude greater than the encoding genome. These modifications have high relevance in cell-to-cell communication, proliferation, and thereby influence a protein's role in driving functional mechanisms such as disease states. Given the functional discrepancy between a protein and its complement of proteoforms, a Human Proteoform Project has launched, proposed by Smith et al. in 2021 [81], with the goal of characterizing the drivers of human health and disease at the proteoform level. To date, 61,770 proteoforms have been characterized from 5705 proteins coded by 4689 genes [82], reflecting the increase in heterogeneity between the genome and the functional proteome.

The peptidome represents an interesting fraction of the proteome, comprising the products of endogenous enzymatic reactions. There has been an interest in characterizing these analytes since proteomics emerged, particularly in neurobiology [83] and the immune system [84]. The

inherent low abundance and diversity of endogenous peptides poses challenges in sample preparation, detection, and data analysis, however, advances in LC-MS and bioinformatics platforms are enabling peptide characterization based on MS/MS spectra despite low concentrations and without the constraint of trypsin specificity [85–88]. The Human Immunopeptidome Project was launched in 2015 with the objective of mapping the entire complement of peptides produced by human leukocyte antigen (HLA) molecules [89,90].

1.1.3 Multi-omics

Following the successes of the human genome project and the progress in the field of proteomics, scientists continue to probe deeper into the complexity of human biology by way of many other omics fields. Transcriptomics, for example, represents the study of RNA transcripts, enabling an understanding of differential gene expression [91] and emerged around the same time as proteomics. While the genome, transcriptome and proteome reflect the drivers of many biological processes, the metabolome provides an opposite perspective, representing the end products of completed cellular functions. By this, metabolomics—first described in 1999 by Nicholson [91]—most accurately evidences the current physiology of the cells under investigation. Metabolomics strategies have often employed GC-MS, benefitting from its high resolving power. However, LC-MS approaches are favored for the enhanced sensitivity, throughput, and mitigated derivatization. Advances in metabolome technologies are enabling its application in disease diagnostics [92], precision medicine [93] and drug interaction studies [94].

Each of the omics fields that have been described exhibit different advantages and limitations while enabling unique perspectives into the complex workings of a biological system. Typically, any of these -omics workflows target the recovery and detection of a particular fraction of the sample. However, multi-omics approaches are gaining popularity due to the rich information obtained by combining the profiles of the genome, transcriptome, proteome, and/or metabolome. As reviewed by Huang et al., recent advances in data integration have been vital in the progress of multi-omics [95]. With sample quantities often being limited, single sample preparation workflows are being developed to recover and prepare multiple “omes” simultaneously, however throughput seems to be a remaining limitation. A 2016 study by Quinn et al. described an integrated genome/metabolome pipeline that enabled multi-omics analysis in <48 h [96]. In 2021, Bechmann et al. claimed to develop a single sample preparation workflow to recover the proteome and

metabolome from mouse adrenal tissue, however, the sample was aliquoted for differential treatment following cell lysis [97]. A 2016 review by Bock et al. highlights recent preparative approaches to multi-omics analyses on single cells [98]. These diverse integrative approaches clarify the connections between the genome and final phenotype and have great potential in understanding disease mechanisms and drug interactions [99,100].

1.2 The Proteomics Workflow

Proteome analysis generally relies on rigorous front-end workflows to quantitatively recover all proteins from a complex biological matrix with sufficient purity for effective ionization and detection by mass spectrometry. The structural heterogeneity across a proteome demands robust catch-all approaches to solubilization, often relying on MS-incompatible denaturants (e.g., surfactants, chaotropes), which are subsequently depleted before MS analysis. Given the complexity and dynamic range of proteome samples, detection efficiency is optimized by fractionating the sample. This is achieved through a variety of approaches spanning chromatography, electrophoresis, or membrane filters. Many workflows benefit from multi-dimensional combinations of these separations to maximize peak capacity and thereby the depth of sample coverage. In the case of a bottom-up workflow, fractionation steps may occur before or after enzymatic digestion (offline) and are often directly coupled to the MS source (online). Depending on the experiment and the exploited instrumentation, these workflows can vary greatly—some being optimized for a broad approach aiming for a snapshot of the whole sample (untargeted, shotgun), while others select for specific analytes of interest (targeted).

Proteome samples can include tissue specimens, body fluids (e.g., urine, blood, saliva) or cultured cell lines. These complex matrices necessitate robust front-end sample preparation to selectively and quantitatively recover the proteome, deplete interferences, digest (for bottom-up workflows), and fractionate for optimal MS characterization. The structural/chemical heterogeneity across a complex proteome poses challenges associated with solubility and biased manipulation based on size or hydrophobic character. MS instrumentation is affording increasingly robust identification and quantitation accuracy, demanding equivalent stringency from sample preparation while the scale of current proteome initiatives favors increased throughput and automation. The challenges and recent advances in front-end sample processing are described here, with a focus on the controlling variables of solubility and digestion.

1.3 Protein Solubility and Precipitation Mechanisms

Proteome workflows largely take place in solution, which necessitates careful consideration of the heterogeneous solubility properties across a complex proteome sample. A native protein adopts a secondary and tertiary folded state that minimizes the Gibb's free energy of its primary amino acid sequence in water [101]. With the H-bonding network of polar residues and the peptide backbone being stabilized in both the folded and unfolded state, folding towards the minimal free energy is largely entropically driven. The *Hydrophobic Effect* describes the tendency of hydrophobic amino acid residues to be buried at the interior of the folded structure, while more polar residues interact favorably with the aqueous solvent. When hydrophobic residues are exposed to the polar aqueous solvent, the surrounding water molecules become highly ordered. By this, when these residues become buried within the folded structure, the disorder of the solvent increases. The combination of increased entropy and stabilizing van der Waals interactions [102] affords a thermodynamically favorable energy of reaction, optimizing the intermolecular interactions between the protein and solvent. Complete protein solvation relies on the formation of stabilizing intermolecular interactions between surface of the protein and surrounding water molecules, comprising the solvation layer. These interactions include hydrogen bonding (e.g., at asparagine and glutamine residues) and ion-dipole interactions at charged residues (e.g., aspartic acid, glutamic acid). This layer of solvating water has been reported to contribute a radius of several angstroms between the protein and bulk solvent. The influence of complete solvation on protein function has also been discussed [103,104].

Protein solvation is inherently complex, with recent works describing an influence on water dynamics extending into the bulk solvent. Thus, it is unsurprising that co-solutes have a direct influence on protein stability in solution. The influence of ionic strength (μ , given by Equation 1.1, where c_i is the concentration of an ion, and z_i is its charge) on solvating protein-water interactions has been researched since the 1950's and can be understood through the Debye-Hückel theory. In 1923, Debye and Hückel characterized the ordering of dilute electrolyte whereby ions in solution are surrounded by a stabilizing cloud of oppositely charged ions through Coulombic interactions [105]. The result is an overall shielding effect, which decreases the activity of the electrolyte in solution. Similarly, Coulombic ion-ion interactions will occur between dissolved electrolyte and charged residues on a protein. This ordering results in an overall shielding effect, where the

effective concentration of the charged species is most accurately expressed through the activity coefficient. The activity coefficient (γ_x) is given by Equation 1.2, where α_x is the diameter of the hydrated ion in picometers. The product of the activity coefficient with the ion's concentration (Equation 1.3) gives the activity of the ion.

$$\mu = \frac{1}{2} \sum c_i z_i^2 \quad (1.1)$$

$$\log(\gamma_x) = \frac{-0.509 z_x^2 \sqrt{\mu}}{1 + \frac{\alpha_x \sqrt{\mu}}{305}} \quad (1.2)$$

$$A_i = \gamma_i c_i \quad (1.3)$$

By reducing the activity (i.e., effective concentration) of charged residues on the protein's surface, the electrostatic free energy is minimized, promoting the protein's solvation. This process is commonly known as *salting in*. By contrast, at high ionic strength, electrolyte in solution imposes the opposite effect, favoring protein-protein interactions by way of *salting out*, which can be instrumental in inducing protein precipitation (discussed in further detail shortly).

While salting in enhances protein solubility by stabilizing charged (polar) residues, other proteins, specifically membrane-associated proteins, are folded to optimize hydrophobic interactions with the lipid bilayer. This hydrophobic character makes membrane proteins particularly challenging to solubilize in aqueous buffers, considering the entropic costs of solvating non-polar residues. However, membrane proteins make up approximately one third of the human proteome and are of high biological interest due to their roles in intercellular signaling, drug interactions, and immune responses.

Various additives have been employed to optimize the solubility across complex proteome samples, the most common being surfactants, such as sodium dodecyl sulfate (SDS). Above the critical micelle concentration, surfactants effectively denature proteins by disrupting the tertiary and secondary structure in favor of ionic interactions between charged residues and the polar head group as well as hydrophobic interactions between non-polar residues and the twelve-carbon chain

on the surfactant [106]. The resulting protein-SDS complex continues to be the subject of molecular dynamics studies, showing differences based on concentration of both the protein and the surfactant as well as across protein type, temperature, and ionic strength [107]. In 1980, Tanford et al. proposed a “beads on a string” model wherein the unfolded protein became wound around SDS micelles through ionic interactions between charged residues and the surfactant’s sulfate group [108]. Ten years later, Lundahl suggested that all protein-bound SDS molecules formed a cylindrical micelle around which the unfolded protein would form a helical structure [109]. Also in 1990, Guo employed small angle neutron scattering to investigate protein-SDS binding interactions, which lead to their proposed correlated necklace model [110,111]. More recently, a molecular dynamics simulation by Winogradoff et al. characterized two mechanisms of secondary structure disruption [112], both of which were initiated by a single SDS molecule disrupting inter-beta strand Hydrogen bonding. Both initial pathways were shown to precede the formation of a non-native helical structure around SDS micelles at sufficient concentrations, supporting the long standing “beads on a string” and “necklace” descriptions.

Ionic surfactants contribute invaluablely to proteome sample preparation, whether it be for cell lysis, protein solubilization/denaturation, or SDS PAGE (polyacrylamide gel electrophoresis) separation. However, the amphiphilic character that makes them amenable to protein stabilization also makes them incompatible with the favored reversed phase chromatography and of even greater consequence, electrospray ionization [113]. Electrospray works by de-solvating ions within an aerosol at high potential through sequential Coulombic fission events [114]. By this, the solutes occupying the outer surface of the solvent droplet enter the gas phase allowing subsequent detection by mass spectrometry. The solvation of hydrophobic species is entropically unfavorable, eliciting highly ordered/rigid structure among surrounding water molecules. Thus, it becomes favorable for the hydrophobic tail of surfactants such as SDS to occupy the air-water interface of an electrospray droplet, significantly reducing the ionization efficiency of peptide/protein analytes [113]. In combination with ion suppression, surfactants also deactivate the enzymes exploited for proteolysis in bottom-up workflows and disrupt chromatographic separation of the proteome sample, causing major challenges in later steps of a proteome workflow.

The reduced efficiency of downstream preparative steps and ultimately, MS detection imposed by ionic surfactants necessitates near-quantitative depletion. Previous work from the Doucette group has demonstrated that MS signal intensity is optimized when SDS is depleted

below ~10 ppm [115]. Conventional strategies include sample dilution to below the SC_{50} [116], dialysis [117], precipitation with KCl [118], and chromatographic strategies [119,120]. However, these pose major throughput limitations (with dialysis taking multiple days) and largely fail to remove protein-bound surfactant. In-gel digestion was widely adopted but is limited by low throughput, variable sample recovery, and inherent incompatibility with top-down approaches. Some of the most common approaches to SDS removal rely on protein precipitation—a century-old approach—which is a central theme of this thesis. Effective protein precipitation has been induced by a variety of strategies, driven by excessive ionic strength (i.e., salting out) or the addition of organic solvents. The present section discusses the thermodynamics controlling protein precipitation and the recent contributions presented by the Doucette group to optimizing the technique for high recovery and high throughput in bottom-up or top-down proteome workflows.

At sufficient ionic strength, electrolytes are associated with reducing protein solubility in a process known as salting out. In addition to the salt concentration, the type of electrolyte plays a significant role in determining its effect on protein solubility. In 1888, Franz Hofmeister described the effects of various salt species on the solubility of proteins, elucidating what is now known as the Hofmeister series [121]. It has long been explained that the most hydrated anions (those with a high charge density—kosmotropes, e.g., CO_3^{2-} , SO_4^{2-} , etc.) promote the ordering of water, decreasing the amount of water available to form solvating interactions with the protein. This loss in entropy promotes protein aggregation through the hydrophobic effect and Coulombic interactions. By contrast, chaotropic anions are characterized by a lower charge density (e.g., ClO_4^- , SCN^- , etc.), which favors the disruption of the water's hydrogen bonding network, increasing the free energy of the solvent, and thereby contributing a net solubilizing effect.

The mechanism of salting in/out has long been debated, and was recently reviewed by Hyde et al. [122]. They describe that the inclusion of salt reduces the entropic penalty associated with the ordered structure of water surrounding the protein in favor of ion-dipole interactions between water and dissolved electrolyte. By this, the protein surface is de-solvated and stabilized through protein-protein ionic interactions and the hydrophobic effect, which leads to large-scale aggregation.

The pH of the solvent system can also be exploited to manipulate protein solubility. The diverse acid/base character of the twenty amino acids results in different net charges across proteins depending on the pH of the solution. The thermodynamics of protein folding favor the

exposure of charged residues at the outer surface of the folded structure, enabling ion-dipole interactions with solvating water molecules. Charged residues also enhance protein solubility by contributing electrostatic repulsions. This is why proteins are least soluble at or near their isoelectric point. By this, a protein's solubility has been observed to reach a minimum at its isoelectric point (pI) [123]. Isoelectric precipitation has been exploited for commercial preparations (e.g., cheese, pharmaceuticals) and historically contributes to the method of Cohn fractionation (described in greater detail below). However, it poses practical limitations with the recovery of complex proteome mixtures, lending favor to alternative means of precipitation for MS-based proteome workflows.

Almost 20 years after Hofmeister's characterization of the salting out effects exhibited across salt types, Mellanby et al. studied the efficiency of precipitation and described the earliest use of organic solvent (ethanol) to induce protein precipitation [124]. The addition of organic solvent reduces the dielectric constant of the solvent system compared to that of pure water [125]. Coulomb's Law describes the force (F) between two point charges in a particular medium and is given by Equation 1.4, where ϵ_0 is the permittivity of a vacuum, ϵ_r is the permittivity of the medium, q_1 and q_2 are the charges on two particles, and r is the distance between the two charged species. From Coulomb's Law, a reduced dielectric constant imparts stronger Coulombic forces (F) (either attractive or repulsive) between charged particles in solution [126]. The strengthened electric field promotes aggregation between oppositely charged proteins through ionic interactions. Furthermore, in the presence of organic solvent, water-solvent interactions (e.g., hydrogen bonding and dipole-dipole) become more favorable than solvating water-protein interactions. The preferential interaction of water with the organic solvent is thought to promote dehydration of the protein's solvation layer. With less water available to solvate the protein, precipitation is favored entropically through protein-protein hydrophobic interactions.

$$F = \frac{1}{4\pi\epsilon_0\epsilon_r} \frac{q_1q_2}{r^2} \quad (1.4)$$

In a classic 1907 study, Mellanby et al. provide extensive discussion on the precipitation conditions that showed a controlling influence on protein recovery [124]. They found that precipitation efficiency is optimized at cold temperatures (tested down to 0 °C) using high concentrations of ethanol. They also described a dependence on protein concentration, where

samples with greater protein content precipitated more readily and under milder conditions. Their investigation on precipitation time was of practical interest, as they show the effect of precipitation time on protein recovery, but also the negative impact that long incubations have on the pellet's solubility. Interestingly, Mellanby's analysis of the effect of salt (ammonium sulfate) concentration on precipitation efficiency revealed that different types of proteins required varying concentrations of salt to achieve similar precipitation recoveries. The observation that precipitation efficiency differs based on sample type inspires the question of which differential properties between two samples are limiting precipitation rate (i.e., molecular weight, isoelectric point, hydrophobicity).

While Mellanby et al. showed that the type of protein was an influential factor on the sample's precipitation efficiency, it was a 1946 study by Cohn et al. that exploited this phenomenon towards a technique whereby precipitation could be adopted as a practical fractionation methodology [123]. Cohn's precipitation-based approach to proteome (specifically, plasma) fractionation combined the principles of salting out, isoelectric and organic solvent precipitation. The specific conditions for the sequence of precipitated fractions are outlined in Table 1.1. Their work ultimately laid the foundation for organic solvent precipitation as it is implemented today.

Table 1.1 Precipitation conditions used in Cohn Fractionation [123]

Fraction	1	2	3	4	5
% Ethanol	8	25	18	40	40
pH	7.2	6.9	5.2	5.8	4.8
Temperature / °C	-3	-5	-5	-5	-5

Based on the reported dependencies of other processes (sedimentation, diffusion, solubility) on a protein's molecular weight and charge, these two properties were suspected by Cohn et al. to also have a controlling influence on precipitation. Cohn et al. were the first to systematically exploit the pH and solvent solubility of proteins as a tool for protein fractionation. Through their understanding of minimized solubility at a protein's pI, they developed a method by which plasma proteins could be separated into five fractions by ramping organic solvent, pH and ionic strength while also controlling precipitation temperature. For example, the fifth fraction

showed selective precipitation of albumin, which provided particularly noteworthy clinical advantages by replacing whole blood transfusions for wounded soldiers during World War II.

Despite the long history of applications for protein precipitation, several works throughout the proteomics era point to limitations of the organic solvent-based approach, including variable and biased recovery. In 1976, Barritault assessed the efficiency of acetone precipitation of ribosomal proteins, noting the gains in throughput over the conventional dialysis approach [127]. The authors of this study highlight the dependence of initial protein concentration on the rate of precipitation, showing that a rapid (5 minute) precipitation at room temperature afforded quantitative recovery for samples containing 5-20 mg/mL total protein, while samples 10× more dilute benefited from overnight incubation at -20 °C. Almost thirty years later, Thongboonkerd et al. evaluated the efficiency of acetone precipitation compared to cartridge-based ultrafiltration for the analysis of urinary proteins [128]. Contrasting with Barritault's robust recovery, the authors report biases in the efficiency of acetone precipitation, favoring proteins with acidic and hydrophilic character. In 2002, Fountoulakis' group compared the precipitation efficiency of salting out (ammonium sulfate), trichloroacetic acid (TCA), and acetone for preparation of plasma samples [129]. Assessment of spot intensity in a 2D gel revealed that salting out effectively depleted albumin, but whole proteome recovery was optimized by precipitation with TCA or acetone. A 2010 study by Fic et al. compared rat brain proteome coverage resulting from chloroform/methanol, TCA, TCA-acetone, and acetone precipitation, demonstrating the highest and least biased recovery from acetone [130]. Even so, recovery was highly variable ranging from 35 to 90%. In 2016, Santa evaluated TCA-acetone and acetone precipitation for the recovery of "very dilute" (0.1 mg/mL) proteome samples [131]. Their quantitative SWATH-MS (Sequential Window Acquisition of All Theoretical Mass Spectra) results showed greater reproducibility from acetone precipitation with respect to peptide and protein identifications as well as quantitation with 74% of proteins showing coefficients of variation <20%. They even note better quantitation precision from the acetone precipitated sample than a non-precipitated control. The varying results across sample types and lab groups reported over fifty years suggested that there was perhaps another variable not being controlled for.

A 2013 study by Crowell et al. demonstrated that organic solvent does not independently induce protein precipitation, but rather, recovery is dependent on the sample's ionic strength [132]. All proteome samples would inherently contain ionic matrix components, but perhaps at varying

activities. Previously speculated mechanisms of organic solvent precipitation suggest that the reduction of dielectric constant upon addition of organic solvent strengthened the Coulombic attraction between oppositely charged proteins. Crowell's work supported a greater contribution from Coulombic repulsion, conserving proteins in solution in the presence of organic solvent, while recovery was maximized by adding 10-30 mM ionic strength. It was thus speculated by this group that with the added electrolyte, protein charges are neutralized, enabling effective precipitation entropically driven by the hydrophobic effect. In 2015, Kachuk evaluated this newly optimized acetone precipitation approach against several other SDS depletion strategies (e.g., in-gel digestion, Pierce kit, FASP II (Filter-Aided Sample Preparation), TCA precipitation, SDS-KCl precipitation) [133]. Acetone precipitation was shown to be the most robust alternative to the low-throughput gel-based method, offering quantitative recovery and depletion of >99 % of the initial SDS. FASP II is credited with unparalleled SDS removal, however at the expense of 50 % recovery. In 2020, Pérez-Rodríguez et al. assessed this group's acetone/salt precipitation protocol for the comprehensive proteome recovery from Chinese hamster ovary cells compared to CMW (chloroform/ methanol/ water) and TCA-acetone [134]. The authors report quantitative recovery from the revised acetone precipitation approach and CMW precipitation, and attributed the lower recovery observed from TCA-acetone to decreased pellet solubility. The authors note, however, that recovery of low molecular weight proteins (<15 kDa) was low compared to the rest of the proteome.

With current throughput requirements in large-scale proteomics settings, it was of interest to determine the influence of ionic strength on the rate of acetone precipitation. Barritault's 1976 study pointed to a dependence on sample concentration and the presence of SDS, but extrapolated conclusions on the required temperature without isolating it among other variables [127]. Chapter 2 provides an update to Crowell's optimal precipitation protocol, where it is demonstrated that quantitative recovery of proteome samples is possible following 2-5 min precipitation at room temperature, starting from concentrations as low as 0.01 mg/mL, resulting in optimal proteome coverage from bottom-up LC-MS/MS analysis [135]. This study shows agreement with Pérez-Rodríguez's study, demonstrating a slight bias towards the rapid recovery of high-molecular weight proteins. The slightly reduced precipitation efficiency of low-MW proteins was subsequently addressed by Baghalabadi, where it was shown that effective precipitation of low-MW proteins and even peptides is achieved by increasing the acetone concentration from the

conventional 80% to 95-97%, in combination with optimized ion effects from 100 mM zinc sulfate [136].

1.4 Proteolytic Digestion

Proteolytic digestion is a critical step in bottom-up proteome workflows, enabling protein level characterization on the basis of MS identification of lower molecular weight proteolytic peptides. As reviewed by Zhang (Yates) et al., several proteases have been exploited towards complete proteome digestion [137], but trypsin is the gold standard in the field owing to its stringent specificity and the amenability of its products to fragmentation for peptide mass fingerprinting [138]. Trypsin is classified as a serine protease (meaning its active site depends on a nucleophilic serine residue). Its globular structure is characterized by a double beta-barrel with the catalytic region housed between the two six-stranded barrels. The enzyme selectively hydrolyzes the amide bond at the C-terminus of lysine and arginine residues, as defined by the “Keil Rules” [139]. This selectivity derives from the conformation and modulated charge state of the active site [140]. Under controlled pH (~8), the negatively charged aspartic acid residue in the well of the catalytic pocket invites Coulombic interactions with positively charged side chains on a substrate protein, as hypothesized by Mares-Guiaï in 1965 [141]. The depth and constricted width of the pocket further selects for the long narrow side chains of lysine and arginine.

Proteolysis with trypsin proceeds via a hydrolysis reaction along a substrate’s peptide backbone involving a charge-relay system between a serine, histidine, and aspartic acid residue comprising the enzyme’s catalytic triad. The exact mechanism has been a subject of debate [142–144] with recent computational simulations describing the role of enzyme and substrate dynamics in the catalysis [145–147], however, the generally accepted mechanism is described here. As illustrated in Figure 1.1, a cleavage reaction is initiated by the nucleophilic Ser donating a proton to the His—its nitrogen group exhibits elevated electronegativity due to an H-bonding bridge formed with the Asp residue. This initial step generates an acyl-enzyme intermediate, which is subject to subsequent attack by an activated water molecule. The C-terminus of the substrate’s peptide bond is freed when the proton on the activated His is returned to Ser, marking the re-establishment of the initial conformation.

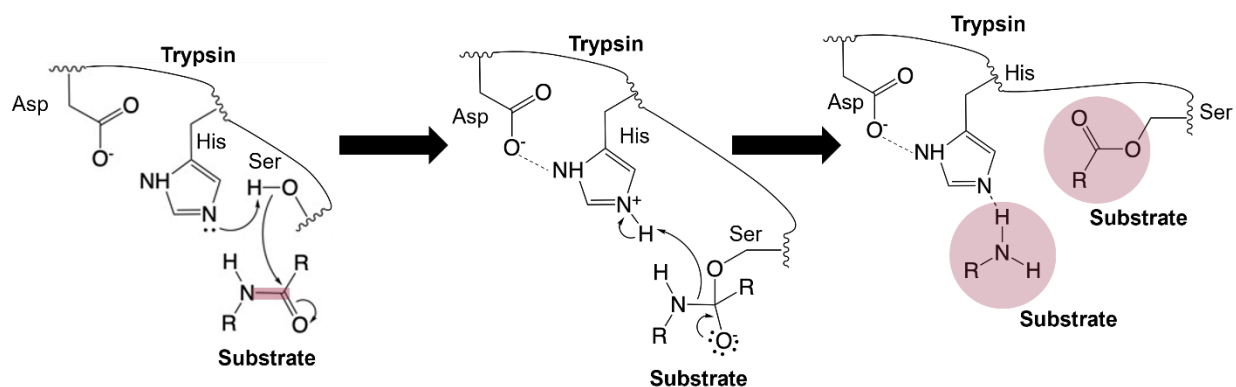


Figure 1.1 Arrow-pushing mechanism of peptide bond hydrolysis by trypsin, adapted from the work of Vandermarliere, Mueller, and Martens [138].

The selective specificity trypsin exhibits for cleavage at lysine and arginine residues produces peptides that are amenable for bottom-up proteome analysis by mass spectrometry in several regards. The average yeast protein produces tryptic peptides with lengths of 7-35 residues [148], providing molecular weights that are easily resolved by mass spectrometry. Furthermore, cleavage at these basic residues results in multiply charged peptides, which are readily fragmented with both product ions potentially being detected, which ultimately facilitates high confidence peptide and protein identifications based on MS/MS spectra. Even so, trypsin specificity has its limitations, which continue to be characterized by recent work.

Commercial preparations of trypsin suitable for proteomics analysis workflows need to be purified (i.e., depleted of chymotrypsin). Chymotrypsin is another serine protease, with high similarity to trypsin, but it cleaves C-terminal to aromatic residues, contaminating MS spectra with un-identifiable peptides, while reducing the intensity of specific peptides. Depletion of chymotrypsin has been achieved through chromatographic approaches [149,150], and as an additional step, commercial trypsin is often treated with a chymotrypsin inhibitor, L-tosylamido-2-phenylethyl chloromethyl ketone (TPCK), minimizing any residual non-specific activity [151,152].

Assuming a purified enzyme, an abundance of work has been done to characterize the cleavage patterns of trypsin towards optimal proteome database searching. Misleaved peptides are an inherent product of all digests [153], and even offer the advantage of improved sequence coverage. However, it is of interest to understand trypsin specificity in order to accurately assign

miscleavage sites and optimize the selection of proteotypic precursors in targeted and quantitative workflows [154]. In 2000, Thiede et al. demonstrated that 90 % of miscleavage sites exhibit an adjacent tryptic residue, an Asp or Glu, or a C-terminal Pro [155]. In 2004, Olsen et al. described strict specificity, suggesting that peptides exhibiting apparent non-tryptic cleavage were degraded products of fully-tryptic peptides [156]. A 2008 study by Rodriguez et al. aimed to re-evaluate the long accepted Keil rules for trypsin proteolysis—trypsin cuts C-terminal to lysine and arginine, but not before proline—exploiting 14.5 million MS/MS spectra to more precisely characterize cleavage frequency across the twenty proteinogenic amino acids [157]. Their analysis demonstrated several deviations from the accepted Keil rule, starting with significant cleavage frequencies at basic sites followed by a proline residue. Additionally, they show reduced cleavage efficiency at dibasic sites, or when the Lys or Arg is adjacent to Cys, Trp, Asp, or Glu. Arginine residues were cleaved more frequently than lysine (as previously reported by Yen et al. [158]), unless followed by Tyr, Phe, or His. In 2012, Lawless and Hubbard reported an algorithm to predict miscleavage sites based on their observations that Lys or Arg inhibit cleavage within five residues N-terminal to a scissile site, while Asp and Glu are most inhibitory within 2 residues C-terminal to the tryptic site [159]. Similar findings were noted by Walmsley the following year in a comparison of bovine and porcine trypsin cleavage patterns [160]. In 2015, Šlechtová et al. reported a systematic analysis of the hydrolysis kinetics of synthetic peptides with varying motifs surrounding the scissile site [161]. The authors quantify rate constants of enzymatic cleavage for fifteen acidic and dibasic model cut sites, showing a 3-fold faster cleavage of arginine sites compared to lysine, as well as a greater inhibition by acidic residues preceding the scissile site as opposed to following it—contrasting with Lawless’ results. Dibasic residues are cleaved 2-3-fold slower than a simple R or K, with the hydrolysis of the miscleaved peptide to the fully-cleaved peptide being 15-fold slower. With rate constants of potential cleavage motifs varying by four orders of magnitude (from Šlechtová’s work), optimizing the digestion conditions to sustain the enzyme’s activity is critical. Moreover, proteome workflows continue to demand increasingly high throughput—to this end, it is unsurprising that there is at least as much literature on trypsin activity and potential mechanisms to accelerate trypsin cleavage.

Trypsin has been characterized as a highly unstable enzyme, requiring precise temperature and pH conditions to sustain activity. Additionally, it is prone to self-digestion (autolysis), reducing its effective concentration over time under optimal digestion conditions. The factors influencing

trypsin activity and stability have been the subject of discussion for over a century with reports of enhanced catalysis at elevated temperature, varying ratios of enzyme/substrate, with a range of additives (e.g., surfactants, organic solvents, chaotropes), and physical treatments such as microwave, ultrasound and elevated pressure. Among the diverse approaches to optimal trypsin digestion, the most widely adopted development across the proteomics community is the use of chemically modified trypsin, commonly referred to today as MS- or sequencing-grade enzyme. This section will review the evidence supporting (and dismissing) each strategy towards optimizing catalytic potential while minimizing autolytic activity and thermal denaturation.

It has long been accepted that trypsin activity and stability are optimized at the physiological temperature, 37 °C, and pH 8.0 (optimizing the charge state of trypsin's active site and tryptic side chain residues for effective Coulombic interaction), which is reflected by conventional proteome digestion protocols, recommending overnight incubation under these conditions. Nonetheless, several reports claim that proteolysis with trypsin is accelerated by increasing the digestion temperature. A 2003 study by Havliš et al. reported optimal catalytic activity of modified trypsin at 50 °C, proposing a 30 min in-gel digest [162]. Finehout et al. evaluated the kinetics of digestion with modified trypsin in a 2005 study, reporting ~9% faster cleavage at 48 °C compared to 37 °C [163]. In 2019, Heissel et al. described a procedure for methylating trypsin, claiming the modified enzyme's enhanced tolerance to 58 °C digestion [164]. In 2021, Takemori et al. described a rapid approach to in-gel digestion, employing a <1 h incubation at 70 °C [165]. While cleavage rates are accelerated at higher temperatures, trypsin activity is reported to drop in excess of 50 °C, likely owing to thermal denaturation and aggregation, reducing the stability of the enzyme [166].

In addition to elevated temperature, several other physical treatments have been reported to accelerate proteolysis with trypsin, such as pressure, microwave and ultrasound irradiation. A 2008 study by López-Ferrer et al. reported a 1 min digest using pressure cycling at 5-35 kpsi, validated based on sequence coverage of a standard protein [167]. In 2010, Yang et al. employed a significantly lower pressure (6 atm = 88 psi), by conducting trypsin digestion for 30 min in a syringe [168]. The degree of digestion was evaluated based on MS sequence coverage of two standard proteins compared to a control that employed a 35% organic solvent digestion buffer. While these works claim enhanced digestion efficiency based on high sequence coverage, the findings are likely contingent on simple protein mixtures, with limited use for the methods in the case of complex proteome samples since sequence coverage benefits from some degree of

miscleavage. Additionally, López-Ferrer et al. speculate that digestion efficiency is enhanced by promoting substrate denaturation at the high pressure condition, but trypsin is also subject to unfolding towards an inactive state at elevated pressure as described in 1997 by Ruan et al. [169]. In 2011, Lee et al. reported another 1 min digest using pressure cycling at 25 kpsi, however the trypsin was immobilized on nanoparticles [170], which offers the advantage of a high enzyme-to-substrate ratio while slowing autolysis and denaturation. In 2005, López-Ferrer et al. reported another 1 min digestion (with solution and in-gel compatibility), this time employing high-intensity focused ultrasound (HIFU) [171], which they describe as providing localized pulses of high temperature and pressure as well as enhanced mixing. The same group followed up in 2009 with a 15-second ultrasound-assisted digest of BSA (bovine serum albumin), this time immobilizing trypsin on beads, benefitting from the associated increase in enzyme-to-substrate ratio [172]. The digestion method intended for ¹⁸O-labeled quantitation was validated based on comparable peptide and protein identification rates to control digests; however, quantitative workflows demand the utmost stringency in digestion completion, which can not be deduced solely on protein identification rates. A 2011 study by Carrera et al. applied López-Ferrer's HIFU method for the authentication of fish species, demonstrating the suitability of such rapid digests for protein identification despite the likely low degree of digestion completion [173]. In a 2012 evaluation of rapid digestion approaches, Dycka noted a 20 % increase in miscleavages from the ultrasound treatment compared to a conventional overnight approach [174].

Microwave treatment has been another major topic in the interest of rapid trypsin digestion, with several studies using it in solution [175,176] while others combine it with immobilized trypsin [177–180], FASP [181], in-tip [182], and in-gel [176,183] digestion approaches. A 2010 report by Reddy et al. recommends careful consideration before applying microwave-assisted digestion strategies, especially if complete digestion is a concern [184]. Similarly, in 2012, Damm et al. criticized the proposed benefits of microwave-assisted digestion compared to conventional heating, demonstrating the consequences of temperatures exceeding 50 °C on trypsin activity, correlating with incomplete digestion [185]. They show that just a few seconds of microwave treatment imparts temperatures around 100 °C, while trypsin activity is demonstrated to plummet at 60 °C.

Digestion efficiency relies on not only an active enzyme, but also the conservation of sample solubility across the digestion period. While membrane proteins represent some of the most

biologically relevant analytes, their hydrophobic character makes them some of the most challenging proteins to manipulate during front-end sample preparation. Several studies have described the advantages of surfactants towards augmented digestion of membrane proteome samples. In 2003, Tian et al. demonstrated the use of the surfactant, Tween 80 and a sugar additive (mannitol) to protect trypsin from the deactivating effects of ultrasonic irradiation [186]. In 2004, Zhang et al. evaluated several common surfactants employed in proteome workflows, concluding optimal sequence coverage of bacteriorhodopsin—an integral membrane protein—by digesting it with 0.1% SDS [187]. In 2007, the same group compared the digestion efficiency of a membrane enriched fraction of *E. coli* using 1% SDS and 60% methanol, demonstrating greater membrane coverage using the methanol-assisted digest [188]. While the inclusion of SDS undoubtedly enhances the solubility of hydrophobic membrane proteins, the denaturant is a detriment to trypsin activity. In 2003, Yu et al. used N α -benzoyl L-arginine ethyl ester (BAEE) assays to report a 5-fold drop in the initial trypsin activity in the presence of 0.1% SDS [189]. A 2015 review by Takeda et al. described the kinetics of SDS-protein binding, showing that in 0.2% SDS (or greater), the interaction is optimized, promoting an irreversible conversion to an inactive state [190]. In 2020, Ma et al. conducted a molecular dynamics study on trypsin-SDS interactions, showing the favorable interaction between SDS and hydrophobic residues in trypsin's catalytic pocket—inhibiting substrate interactions with the active site [191].

Sodium deoxycholate (SDC), commonly known as bile salt, and naturally present in the pancreas where trypsin is produced, has been exploited for trypsin digestion. SDC is associated with improved enzyme and LC-MS compatibility, warranting many studies towards its use for membrane proteome sample preparation. A 2006 study by Zhou et al. demonstrated trypsin's higher tolerance to SDC than SDS, whereby they digested rat hippocampal membrane fractions in 0.1% of each detergent, with the SDC digest identifying more peptides and more membrane proteins [192]. In 2008, Masuda et al. evaluated trypsin activity spectroscopically in the presence of twenty-seven common denaturants [193], reporting a 5-fold enhancement in residual activity following overnight incubation in 0.01-1% SDC, with increased activity observed in up to 10% SDC. In the same year, Lin et al. evaluated trypsin activity across a range of SDC concentrations, reporting negligible loss in 0.1% SDC, with 77% relative activity measured in 10% SDC [194]. While these measurements of the enzyme's initial activity show promise towards conserving trypsin's catalytic potential while enhancing the sample's solubility, the rate of enzyme

deactivation remains in question. From Lin's results, a >10-fold drop in initial activity is noted across the 5-minute spectroscopic assay. A few years later, in 2013, Lin et al. reported a bottom-up workflow that optimized their observed advantages of both SDS and SDC [195]. SDS was used for cell lysis and proteome solubilization, depleted by acetone precipitation, after which the pellet was re-solubilized in SDC prior to digestion at a final concentration of 1%. In 2015, they published a follow-up noting enhanced re-solubilization efficiency with 5% SDC followed by dilution to 1% for "optimal" digestion [196]. In 2020, Shahinuzzaman et al. described an optimized preparation of methanol-chloroform precipitated proteins by digesting in 2% SDC, reporting a 27% increase in protein identifications [197]. A 2016 study by Moore et al. demonstrated 26% increase in membrane protein identifications with overall greater intensity by digesting in the presence of 1% SDC compared to the conventional approach [198]. Several of the mentioned studies show that trypsin exhibits greater resistance to deactivation in SDC compared to SDS with limited discussion of its stability over a digestion period. The work described in Chapter 5 evaluates the effect of various surfactants on trypsin stability. It is speculated that SDC does in fact "assist" digestion when sample solubility is the limiting variable, however, from the perspective of trypsin activity, proteolysis occurs sufficiently *despite* the SDC, as opposed to there being an associated enhancement factor.

SDS and SDC are likely the most common and widely adopted surfactants, however, several others have shown promise based on enhanced MS compatibility by way of acid-labile or photocleavable character. RapiGest (3-[(2-methyl-2-undecyl-1,3-dioxolan-4-yl)methoxy]-1-propanesulfonate) is an acid-labile anionic surfactant that has become widely adopted to enhance proteome digestion with recent reports noting particular benefit for glycosylated proteins [199]. A 2009 study by Norrgran et al. optimized RapiGest-assisted digestion parameters for quantitation strategies [200]. The authors report optimal digestion is achieved following 2 h incubation in 0.04% RapiGest using excess trypsin (2.5:1 enzyme to substrate ratio), at conventional temperatures. By contrast, a 2014 report by Pop et al. demonstrated reduced membrane protein representation when using RapiGest compared to a no-additive control [201]. MaSDeS is another acid-labile surfactant that was shown by Chang et al. in 2015 to optimize proteome coverage from tissue samples [202], but has not gained much traction since.

In recent years, Ying Ge's group has recently developed a photocleavable surfactant, Azo (4-hexylphenylazosulfonate) which promises compatibility with both, bottom-up and top-down

workflows owing to its efficient degradation by ultraviolet (UV) irradiation [203]. In the first report of the surfactant in 2019, the authors demonstrate comparable extraction efficiency to that achieved with SDS and MS signal intensities similar to a no-surfactant control. The following year, Brown et al. reported an optimized bottom-up membrane proteome workflow [204], while Knott et al. demonstrated an Azo-assisted workflow for bottom-up characterization of extracellular matrix proteins [205]. The utility of the photocleavable surfactant has since been shown for exosome [206], tissue [207], and plant proteomics [208], offering a promising robust alternative to SDS.

Chaotropic agents are common alternatives to aid sample solubility while being presumably milder towards enzyme deactivation. Masuda's 2008 comparative study on digestion additives reported optimized activity in the absence of guanidine hydrochloride and a 26% enhancement in initial activity with 3 M urea [193]. In 2010, Proc et al. employed an MRM (multiple reaction monitoring) assay of several plasma peptides to evaluate the digestion rate in a diverse range of denaturants spanning organic solvent, surfactants, and chaotropic agents [209]. The assessed chaotropes (urea and guanidine hydrochloride, Gdn HCl) demonstrated the slowest rate of digestion product formation; however, this result confounds cleavage efficiency with undesired covalent modifications which alter the mass of the digestion product. They found that 0.1% SDS and 1% SDC provided equivalent digestion efficiencies, ultimately recommending a 9 h digest with 1% SDC due to the ESI interferences associated with SDS. In 2012, Poulsen et al. evaluated digestion efficiency (with endoprotease Lys-C) in the presence of 8 M Gdn HCl, which they used to elute their sample from an affinity purification cartridge [210]. The authors found that 30 min digestion in 8 M Gdn HCl offered greater peptide and protein coverage than digestion in the presence of urea, while affording the use of elevated temperatures without concern for variable carbamylation. A 2018 study by Betancourt et al. reported enhanced digestion efficiency in 1 M urea at room temperature compared to the conventional overnight incubation at 37 °C, likely due to a reduction in carbamylation rates [211]. Similar to the perspective on the advantages/disadvantages of surfactant-aided digests, it is speculated that chaotropes offer the benefit of enhanced sample solubility, but at the expense of trypsin activity to varying degrees. A 2022 study by Dušeková et al. demonstrated that trypsin stability is optimized in increasingly *kosmotropic* environments [212], showing the influence of various anions on the enzyme's rigidity, thermal stability, and ligand binding.

Rounding off the denaturants that have been employed for trypsin digestion, organic solvents spanning acetonitrile, methanol, ethanol, 2-propanol, and others have been reported to enhance proteome digestion efficiency at concentrations of 10-90%, despite the overlap with common precipitation conditions [132,213,214]. As reviewed by Mattos and Ringe in 2001, enzymes are speculated to conserve their activity in organic-aqueous buffers from strengthened intra-protein Hydrogen bonding, promoting a more rigid folded structure [215]. In 2001, Russell et al. reported an optimal rate of digestion in 80% acetonitrile based on high sequence coverage of myoglobin after just 5 min [216]. Similarly, in 2006, Strader et al. reported digestion efficiency advantages in 80% acetonitrile, showing increased peptide identifications from a standard protein mixture [217]. It is noted that variably cleaved digestion products increase the number of unique peptides identified, so this metric does not evidence better digestion efficiency. A 2007 study by Hervey et al. compared the effects of several surfactants, chaotropes and organic solvents on the digestion efficiency of ribosomal and microtubule-associated proteins [218]. The authors report optimized ribosomal coverage by digesting for 1 h in 80 % acetonitrile. A recent study by Liu et al. employed a more modest 10 % acetonitrile buffer for the digestion of ricin, reporting optimized yield of tryptic peptides from a ricin digest in 4 h at 45 °C [219]. These results contrast with earlier work by Östin et al. wherein they suggested optimal digestion of the same protein in 50% methanol using a 1:2 enzyme-to-substrate ratio at room temperature [220]. While low concentrations of organic solvent may help optimize the enzyme's conformation and sample solubility, concentrations exceeding 30-40% organic risks the precipitation of yet-to-be-digested intact proteins, with probable bias towards the loss of water-insoluble proteins. In 2011, the Doucette group evaluated digestion efficiency in the presence of organic solvents, demonstrating reduced digestion completion using 80 % acetonitrile compared to an aqueous buffer [221]. In 2020, Sproß et al. employed ion mobility analysis to characterize enzyme conformation following exposure to various organic solvent conditions [222]. The authors describe the influence that buffers such as ammonium acetate have on trypsin's tolerance to acetonitrile, noting the conservation of activity in up to 20-25% acetonitrile when ammonium acetate is included. Despite the stabilization of a supporting buffer, a 2021 study by Espinosa et al. employed a buffer-free digestion, including 20% acetonitrile for the characterization of the binding domain of the SARS-CoV-2 spike protein [223]. The apparent consensus that trypsin activity is largely conserved in modest concentrations of organic solvent supports the findings of Casteneda-Agullo et al., reported in 1958 [224].

Many strategies that apparently achieve enhanced digestion primarily do so by aiding the solubility of the substrate, while compromising the enzyme's activity and stability. A simple solution to moderate reductions in activity is to just use more enzyme, which is a major advantage of immobilized enzyme cartridges. By digesting the sample within a column of immobilized enzyme, the effective enzyme-to-substrate ratio is drastically increased compared to the conventional 1:50. Enzyme-substrate interactions are further enhanced by modulating the flow rate through the reactor. An additional benefit is the reduction in autolysis rates, improving the enzyme stability, which in turn affords rapid digestion [225]. As reviewed by Kecskemeti et al., the enzyme is bound to a stationary phase via either covalent linkages, adsorption, bioaffinity, or enzyme entrapment or crosslinking [226]. These versatile systems have been reported since the 1990's with a wide range of applications and adaptation into online [227] and microfluidic platforms [226,228]. Although immobilized enzyme reactors show high digestion efficiency, reusability, and compatibility with online platforms, they are not cost effective and may jeopardize sample recovery. These limitations lend favor to further optimizing solution digests.

Conventional digestion conditions—including buffer composition, temperature, and enzyme-to-substrate ratio—are the most widely exploited among the diverse reported strategies, however, chemically modified enzyme is one development that has been widely adopted. Several different covalent modifications have been investigated [229–232], although the most widely exploited modification is acetylation of the free amino groups. In 1954, Ram, Terminiello and Nord reported the influence of acetylation on trypsin stability [233]. They describe the reduction in the rate of autolysis (owing to the modification acting on the enzyme's own tryptic residues) as well as a decrease in the rate of dissociation—both of which contribute overall enhanced stability. Acetylated trypsin is now sold commercially as “sequencing grade” or “MS grade” enzyme due to its enhanced stability [163] and specificity [234]. In 2005, Finehout et al. characterized the kinetics of modified trypsin, demonstrating its thermostability at temperatures above that of conventional digestions (37 °C) [163]. A 2012 study by Burkhart et al. compared the digestion efficiency from several trypsin products, ranging in TPCK treatment and acetyl modification [234]. Comparison of the bottom-up peptide identifications from each preparation showed enhanced cleavage efficiency and specificity from the modified enzyme (Promega sequencing grade), resulting in maximized bottom-up peptide and protein identifications. Given the enhanced digestion efficiency and specificity of modified trypsin, it has become a standard for bottom-up proteomics applications

[235–237]. However, despite having shown greater thermostability, overnight incubation at the conventional 37 °C is the most common approach.

The assessment of initial enzyme activity is an indication of trypsin digestion efficiency. Equally important is the enzyme stability over time. For example, while 37 °C is frequently employed for tryptic digestion, multiple researchers have noted enhanced enzyme activity at elevated temperatures [238–240]. However, these gains are counteracted by increased enzyme deactivation, greater thermal aggregation of proteins [166], the potential for accelerated side chain modifications [241–244] and enhanced chymotrypsin activity [245,246]. This has deterred widespread adoption of high temperature digestion [163,165,184,242,244,247,248]. The influence of calcium ions on enhancing trypsin activity as well as preserving thermal stability has been known since the early 1900's, though surprisingly, the inclusion of calcium ions is not widely exploited across the proteomics community. The mechanism of calcium's stabilizing effects [249] was introduced by Gorini in 1951 [250] and expanded by Green and Neurath, whereby it was demonstrated that calcium decreases the rate of autolysis in both active and inactive forms of trypsin [251]. A series of follow up studies explored the effects of calcium on autolysis rates, conformational changes, and temperature dependence of trypsin-calcium interactions [252–254]. The 1954 study by Ram, Terminiello and Nord indicates equivalent stabilization effects are imparted by calcium ions on the unmodified and acetylated enzyme [233]. Given the extensive body of evidence to support a more active, stable and selective enzyme, it was among the present objectives to incorporate calcium-stabilized trypsin in an accelerated digestion workflow. It is a goal of this work to achieve maximal, sustained tryptic activity, resulting in consistent, specific, and (ideally) complete proteome digestion. In Chapter 6, the kinetics of trypsin deactivation are characterized as a function of temperature and inclusion of calcium. Based on the cumulative activity estimated from the enzyme's stability, the digestion efficiency is assessed by quantifying the relative intensities of fully-cleaved, fully-tryptic, and unmodified peptides, compared to those produced by a conventional digestion.

In summary, trypsin digestion relies on several factors, including the enzyme's folded structure, charge state, temperature, and substrate solubility. In the interest of accelerating the cleavage reaction to increase bottom-up throughput, temperature is often increased (either directly or by way of ultrasound or microwave treatment); however, this can compromise the folded structure, and thereby the enzyme's specificity. Likewise, substrate solubility is enhanced by the

addition of denaturing additives such as surfactants, organic solvents, and chaotropes, but presumably at the cost of the enzyme's integrity. Furthermore, "accelerated" digestion approaches are often only partially complete, limiting their application in quantitative objectives. Chapter 5 and 6 provide discussions on the effects of digestion additives and the unique combination of elevated temperature and calcium ions to stabilize the enzyme towards accelerated digestion.

1.5 Automation and Miniaturization

Current proteome initiatives are constantly demanding higher throughput and robust processing of increasingly small sample quantities. To accommodate these trends, many semi- and fully-automated sample preparation platforms have been developed, promising better repeatability and faster processing than manual workflows.

Filter-aided sample preparation (FASP) was described by Wiśniewski et al. in 2009 [41], offering quantitative SDS depletion followed by a solution digestion with peptide recovery from elution through a MWCO membrane. In 2014, Erde et al. reported enhancements to the original digestion strategy (eFASP), substituting 8 M urea for 0.2% SDC [255]. They report enhanced digestion efficiency, however the surfactant needs to be subsequently depleted. In the same year, Yu et al. described a 96-well plate format (96FASP) [256], offering complete automation for quantitation of a urine proteome. In 2022, Sandbaumhüter et al. optimized well-plate μ FASP for processing of limited sample quantities [257]. While FASP technologies have been widely applied, they are notorious for lengthy processing and low recovery, especially for low sample loadings.

Suspension trapping (S-Trap) followed in 2014 [258], based on a methanol precipitation step to deplete SDS (or other compatible lysis buffers [259]) prior to reduction, alkylation, and digestion on a quartz mesh. Similar to FASP, S-Trap methods have been adapted for low sample quantities [260], cysteine capture [261], and automation in a 96-well plate format [262]. In 2018, HaileMariam et al. evaluated S-Trap, showing a reduction in processing time compared to FASP from 3 h to 15 min, increased protein identification rates, and improved quantitation [263]. Another comparison by Ludwig et al. demonstrated a more complete digestion than FASP or solution digests with an increase in protein identifications [264]. In 2020, Zougman et al. reported Simultaneous Trapping (SiTrap)—an adapted S-Trap workflow that facilitates simultaneous proteome and metabolome preparation for multi-omics analysis [265].

The In Stage Tip (iST) technology was also developed in 2014 by Matthias Mann's group [266]. The sample preparation strategy employs a pipet tip packed with a C₁₈ stationary phase for isolation and digestion of purified protein. The iST strategy offers greater amenability to low sample quantities [267], high proteome coverage, quantitation accuracy, and reproducibility, however, its major limitation is its incompatibility with surfactants such as SDS.

Single pot solid phase sample preparation (SP3) was developed by Christopher Hughes' group in 2019 [268], wherein protein is retained on affinity-coated beads for buffer exchange and digestion. The SP3 approach offers compatibility with SDS-containing samples, low starting quantities [269], and top-down workflows. In 2017, Sielaff et al. compared the quantitation accuracy and reproducibility of FASP, iST, and SP3 workflows [270]. iST and SP3 both showed good compatibility with low sample loadings (~1 µg) and high quantitative precision, while FASP provided fewer identifications and showed reduced quantitative precision when sample loading was <10 µg. In 2021, Suparsi et al. compared the performance of SP3 with S-Trap and FASP [271]. SP3 showed more efficient detergent depletion, digestion, and MS identification. A recent study by Johnston et al. demonstrated a solvent precipitation-aided SP3 workflow (SP4), where proteins were precipitated onto glass beads and recovered by centrifugation rather than a magnetic interaction [213]. The authors demonstrate improved recovery, particularly of membrane proteins compared to the conventional SP3 strategy, and better scalability using the lower-cost beads. In 2021, Yang et al. employed a *nanoparticle*-packed *nanoreactor* for *nanoproteomics* (Nano3), which demonstrated enhanced recovery and identification rates compared to SP3 [272]. The Nano3 cartridge employs the same paramagnetic beads that are used in SP3 workflows, packed in a 30 nL cartridge for purification and digestion of SDS-containing samples.

Chip-based formats are gaining popularity as the field moves towards single-cell preparations in volumes down to the nanoliter scale [226,273–278]. In 2018, Yu et al. developed NanoPOTS, enabling lysis, reduction, alkylation and digestion within the chip's nanowells [279]. The authors report the greatest proteome coverage from a single HeLa cell to date. In 2020, Zhou et al. extended the chip's applications to top-down approaches [280]. In 2021, Woo et al. developed a modified nested nanoPOTS chip (N2). Trace samples are prepared in a dense array of nested nanowells prior to TMT labeling for quantitative LC-MS/MS [275]. The reduction in sample volume showed increased digestion rates and overall processing throughput by >10-fold. In 2022, Ctortocka et al. described proteoCHIP, featuring single cell isolation and nanoliter sample

volumes, contributing to maximized sensitivity [281]. These miniaturized platforms are affording low-cost and online multiplexed preparation of single-cell samples, which have broad applications, for example, in studying diseases that exhibit cellular heterogeneity and circulating tumor cells.

In 2013, the Doucette group developed a two-stage spin cartridge, the ProTrap XG [282], which facilitates the optimized precipitation-based workflow by automating the isolation of the recovered protein pellet from the contaminating supernatant. The precipitation cartridge enables quantitative recovery and SDS depletion from a 2 min precipitation step [135,283], followed by optional digestion for bottom-up approaches or re-solubilization for top-down analysis. Having validated a ProTrap XG-based workflow based on precipitation recovery, Chapter 3 describes a recent evaluation of the repeatability of identifications and quantitation for bottom-up preparations in the ProTrap XG compared to conventional solution and in-gel digests. Future objectives aim to optimize the ProTrap XG for trace (e.g., single-cell) samples and adapt the cartridge for automation in a well-plate format.

1.6 Summary of Thesis Goals

The -omics discipline is expanding with proteomics at a particular pinnacle and being central to most multi-omics approaches. However, with recent enhancements in LC-MS platforms, front-end sample preparation is currently a limiting factor in throughput and sample coverage. With solubilization and digestion arguably being the most controlling variables in proteome characterization, the projects in this thesis aim to elucidate a high throughput bottom-up workflow based on fundamental investigations of protein solubility (both towards robust extraction and to optimize precipitation efficiency) and proteolytic enzyme activity. Current sample preparation strategies appear to be unnecessarily complex—employing diamonds [284], high pressure [167], ultrasonic irradiation [173], etc. to recover and digest a proteome. The approaches presented in this thesis are based on century-old techniques, which guide robust precipitation-based sample preparation strategies, by optimizing the rate and recovery of protein precipitation. Furthermore, the use surfactants is permitted to maximize extraction efficiency. Subsequently, the effects of common digestion additives and digestion conditions on trypsin activity are characterized in this thesis, together with the kinetics of enzyme deactivation, enabling the elucidation of a rapid digestion approach that exploits modest elevations in temperature in combination with the long-described stabilizing effects of calcium. It is further demonstrated that these approaches are

amenable to multi-omics and targeted assays, promising a simple and robust approach to high throughput and high impact omics analyses.

2. Rapid and Quantitative Protein Precipitation for Proteome Analysis by Mass Spectrometry[†]

2.1 Introduction

With over 100 years of application [124], solvent-based precipitation is widely employed to isolate proteins in proteomic and metabolomic workflows [37,285,286], as well as for preparative (commercial) processing [287–289]. As a specific example, solvent precipitation is considered the best practice approach to eliminate mass spectrometry (MS) interferences (salts, SDS) ahead of top-down proteomic analysis [116]. Unfortunately, as reviewed by Vuckovic [290], low and/or inconsistent protein recoveries are persistently reflected in the many non-standardized approaches to precipitation (ranging in solvent composition, temperature, time, etc.). Alternative protein purification strategies, including cartridge-based platforms, are often favored over precipitation in proteomic workflows, particularly with respect to the potential for higher throughput [41,291]. For precipitation to be considered a favored sample preparation ahead of MS analysis, the approach should consistently provide high (quantitative) recovery, in minimal time, facilitating optimal sample throughput.

Several comparative studies examining the efficiency of protein precipitation have concluded that acetone precipitation provides the highest and most consistent protein recovery over alternative organic solvents [129–131]. Despite this, current protocols involving acetone precipitation appear unable to recover all proteins with equal (unbiased) efficiency [128,130]. Working with 50% acetone (overnight incubation, –20 °C), Thongboonkerd et al. reported a biased loss of low-molecular-weight, basic, as well as hydrophobic proteins. In 80% acetone (1 h, –20 °C), Fic et al. observed a significant decline in precipitation efficiency (to ~50 % yield) when precipitating proteins isolated from the frontal cortex of rat brains. This lower recovery was associated with the increased hydrophobicity of proteins from this region of the brain. By contrast, in 80 % acetone, this group demonstrated an equivalent recovery of cytosolic vs. membrane proteins isolated from yeast [132,292]. It is noted that the latter study incorporated previous findings of the essential role of ionic strength in maximizing protein recovery through acetone

[†] This chapter is based on the published article: Nickerson, Jessica L, and Doucette, Alan A “Rapid and Quantitative Protein Precipitation for Proteome Analysis by Mass Spectrometry.” *Journal of Proteome Research* 19.5 (2020): 2035–2042. DOI: 10.1021/acs.jproteome.9b00867

precipitation. The necessity of including charged species, such as simple salts but also ionic buffers or detergents, together with organic solvents has provided new insights into the mechanistic understanding of precipitation, being directed by solvent-enhanced ion pairing, which neutralizes protein charge, promoting aggregation [293].

Aside from solvent composition, protein concentration is also known to influence precipitation efficiency [124,127,131]. In a 1976 study by Barritault, while >98% recovery was achieved in only 5 min when precipitating concentrated proteins (5 vol acetone, room temperature), protein recovery dropped below 30 % when dilute samples (0.2 g/L) were subject to the same rapid precipitation protocol [127]. However, recovery was restored through overnight incubation at $-20\text{ }^{\circ}\text{C}$. By contrast, Santa reported a high protein recovery of “extremely dilute” proteins (0.1 g/L), through precipitation from an SDS buffer with 6 vol cold acetone ($-20\text{ }^{\circ}\text{C}$, 20 min) [131]. In working with SDS, the latter study satisfies the requirement of maintaining a high ionic strength to improve precipitation. However, it is unsure if proteins can be effectively precipitated below a starting concentration of 0.1 g/L. With respect to incubation temperature, Cohn, Askonas, and others have discussed the importance of conducting precipitation at low temperature for preservation of protein structure and activity [123,294], though its influence on recovery is poorly understood [127,131]. While prior studies from this group employed conventional overnight, $-20\text{ }^{\circ}\text{C}$ acetone precipitation [132,292], here, the assumptions that lengthy, low-temperature incubations are necessary for optimal protein precipitation efficiency are challenged.

The present study provides standardized guidelines to maximize the efficiency of solvent precipitation in acetone, with respect to obtaining rapid and high protein recovery, demonstrating the robustness of this protocol as a sample preparation technique ahead of bottom-up proteomic analyses. It is shown that a unique combination of high salt and elevated temperature significantly improves the precipitation efficiency of proteins in acetone. Quantitative recovery is achieved by rapid (2 min) precipitation of a complex proteome mixture at room temperature and the strategy is also applied to dilute protein systems. Precipitation recovery is assessed by bicinchoninic acid (BCA) assay and SDS PAGE. Detailed characterization of the proteins recovered in the precipitation pellet of a yeast total protein extract is provided by bottom-up LC-MS analysis, which confirms the near-homogeneous recovery of all proteins, irrespective of molecular weight,

hydrophobicity, and isoelectric point. The presented study is therefore anticipated to streamline proteomic sample preparation by simultaneously maximizing recovery and throughput.

2.2 Methods

2.2.1 Materials

Bovine serum albumin and cytochrome *c* were purchased from Millipore Sigma (Oakville, Canada). *Saccharomyces cerevisiae* was obtained as active dry pellets from a local grocery store and cultured in yeast extract-peptone-dextrose (YPD) broth (Millipore Sigma). The Pierce BCA assay kit, sodium chloride, as well as HPLC-grade acetone, acetonitrile, and formic acid were from Thermo Fisher Scientific (Ottawa, Canada). MilliQ-grade water was purified to 18.2 MΩ cm. Tris, iodoacetamide (IAA), dithiothreitol (DTT), SDS, and other SDS PAGE materials were from Bio-Rad (Mississauga, ON, Canada).

2.2.2 Growth and Extraction of Yeast Proteome

S. cerevisiae was cultured overnight (30 °C) in YPD broth to an OD₆₀₀ of 1, according to standard procedures [295]. The cell pellet was twice washed with water and ground under liquid nitrogen in a mortar and pestle to extract the protein into water. Protein in the clarified supernatant was diluted to 1.0 g/L following a BCA assay.

2.2.3 Acetone Precipitation: Varying Incubation Time and Temperature

Aqueous samples of protein or of yeast lysate (100 μL, 0.001-1 g/L) were prepared with varying concentrations of salt or other ionic species, namely, NaCl, Tris-HCl (pH 8.0), or SDS, each ranging from 1 to 100 mM, and gently mixed with 4 volumes acetone and then incubated at the specified temperature for times ranging from 2 min to 24 h. In this work, room temperature refers to 22 ± 2 °C. A low-temperature precipitation employed prechilled acetone (−20 °C) combined with refrigerated (4 °C) protein. A specified volume of supernatant was transferred to a clean vial for further analysis, while the remaining supernatant was removed, leaving <5 μL residual supernatant. The pellet was air-dried.

A precipitation “time course” experiment was also conducted, whereby 1 g/L yeast was initially precipitated in the presence of 20 mM NaCl for 2 min at $-20\text{ }^{\circ}\text{C}$. The recovered protein pellet was washed with 400 μL acetone, centrifuged (9460 g, 5 min), the wash solution was discarded, and the pellet was allowed to air-dry. The collected supernatant was incubated an additional 58 min, prior to a second centrifugation (9460 g, 5 min) to isolate newly precipitated proteins (60 min time point). The 60 min pellet fraction was collected as described above, and the recovered supernatant was incubated 23 h (24 h in total) followed by centrifugation to isolate the final pellet fraction. This precipitation time course workflow is summarized in Figure 2.1 below.

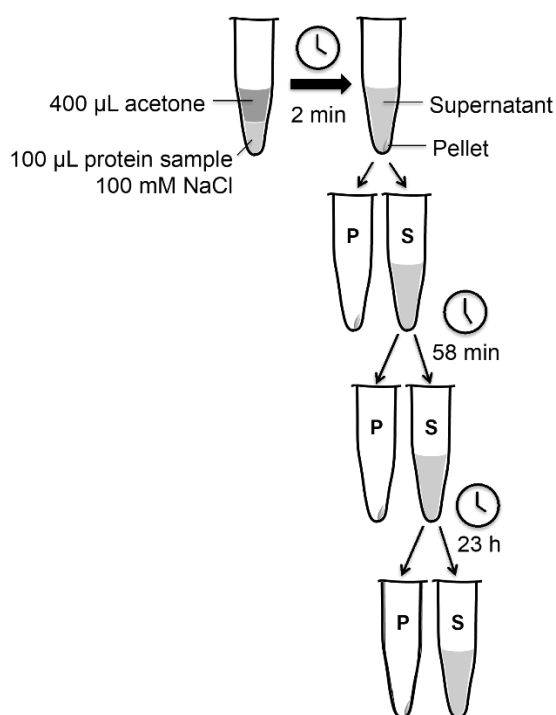


Figure 2.1 Precipitation time course. Workflow diagram depicting precipitation time course employed for cold precipitation ($-20\text{ }^{\circ}\text{C}$). Supernatant fractions were isolated following 2 min and continued incubating for 58 min; newly precipitated proteins were isolated, and the remaining supernatant fraction continued incubating for 23 h (for a total of 24 h precipitation).

2.2.4 BCA Quantitation

Solvent from the collected supernatant (300 μL , of 500 μL total) was evaporated by SpeedVac. Residual protein from the supernatant or pellet fractions was dissolved in 50 μL of 1% SDS, with 1 h benchtop incubation followed by 30 min sonication. The resolubilized protein was quantified

by BCA assay, using a non-precipitated control of the same sample to construct a calibration curve over 1 order of magnitude.

2.2.5 SDS PAGE Analysis

Following precipitation of 1 g/L yeast in the presence or absence of 20 mM NaCl, the dried supernatant (495 μ L, of 500 μ L total) and pellet fractions were resolubilized in 50 μ L of 1 \times Laemmli buffer [296]. Re-solubilization was aided by 20 min of sonication and heating at 95 $^{\circ}$ C for 5 min. Samples were loaded (25 μ L of 50 μ L total) into 1 mm of 12% T SDS PAGE gels and resolved at 100 V until the dye front reached the bottom of the gel. Control lanes consisted of 50 μ g of a non-precipitated yeast proteome extract. Protein bands were visualized by Coomassie staining [297] and imaged with a digital camera.

2.2.6 Trypsin Digestion

Based on the known protein recovery, determined by BCA assay, all precipitated yeast protein fractions were digested at a final concentration of 0.1 g/L. This was achieved by solubilizing the sample in varying volumes of 8 M urea (20% of final volume), then diluting with 0.1 M Tris (pH 8.0), followed by reduction/alkylation with DTT/IAA. Samples were digested overnight at 37 $^{\circ}$ C with 2% (w/w) trypsin, as described previously [115]. The acidified fractions were desalted, and peptides were quantified by LC–UV on a self-packed Poros 20 R2 column, as described previously [298].

2.2.7 In-Gel Digestion

Four replicate samples of a 1 g/L yeast proteome extract (100 μ L) were precipitated with 80 % acetone and 20 mM NaCl (2 min, room temperature) and then resolubilized in 100 μ L of 1 \times Laemmli buffer, with sonication and boiling (100 $^{\circ}$ C, 5 min). A total of 25 μ L of each replicate were loaded onto independent lanes of a 1-mm 12% T SDS PAGE gel, alongside equivalent amounts of a non-precipitated control yeast sample. The gel was resolved at 100 V until the protein had just entered the resolving gel and then subject to Coomassie staining. For each lane, the entire protein band (\sim 5 mm) was resected as a single fraction and subject to tryptic digestion as per

conventional protocols [299]. As above, the extracted peptide fractions were desalted and quantified by LC–UV.

2.2.8 LC-MS/MS Analysis

Proteome analysis of the digested yeast involved duplicate injection of 1 μ L of sample (consisting of 0.5 μ g total peptides) onto a 30-cm self-packed monolithic C18 column, coupled to a 10 μ m New Objective PicoTip noncoated Emitter Tip (Woburn, MA). A Dionex Ultimate 3000 LC nanosystem (Bannockburn, IL) delivered a 2 h linear gradient from 0.1 % formic acid in water to 35 % acetonitrile. The LTQ Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific) operated in top 5 data-dependent mode at a resolution of 30 000 full width at half-maximum (FWHM) for MS1, scanning in rapid mode for MS2 (66,666 Da/s) with a resolution of <0.6 Da FWHM.

2.2.9 Data Analysis

The raw MS data were searched using the Proteome Discoverer software version 1.4 (Thermo Fisher Scientific), searching the UniProt *S. cerevisiae* filtered database (9931 entries, downloaded May 29, 2019) with a false discovery rate of 1%, while also requiring that the protein was identified in each duplicate injection. Carbamidomethylation of cysteine residues was included as a fixed modification and oxidation of methionine residues was considered a variable modification. A precursor mass tolerance and fragment mass tolerance of 20 and 5 ppm were used, respectively. Proteins were profiled using the online software, Venny 2.1. The Proteome Discoverer software reports the protein's molecular weight and pI. Grand average of hydropathicity (GRAVY) scores were extracted from an online calculator, <http://www.gravy-calculator.de/>. The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE28 partner repository [300] with the data set identifiers PXD015674 and 10.6019/PXD015674.

2.3 Results and Discussion

2.3.1 Rapid Precipitation at Room Temperature: Salt Trends

Consistent with prior work on protein precipitation [132], Figure 2.2A illustrates the controlling influence of ionic strength on maximizing protein recovery. Following overnight incubation at -20 °C with 80% acetone and minimal salt (<1 mM NaCl), recovery is effectively negligible ($3 \pm 2\%$). However, upon the addition of >10 mM NaCl, the 1 g/L sample of bovine serum albumin (BSA) was quantitatively recovered ($98 \pm 4\%$). Extended incubations at low temperature are generally recommended to optimize recovery through precipitation while simultaneously preserving protein activity [123,127,294]. However, as seen, in Figure 2.2A, with sufficient ionic strength, room-temperature precipitation ($+20$ °C) will provide equivalent high recovery of BSA ($99 \pm 2\%$), and with similar levels of NaCl required to maximize recovery (5-10 mM or higher). These results therefore suggest that solvent precipitation, with inclusion of sufficient salt concentrations, does not require a reduction in temperature to maintain high protein recovery.

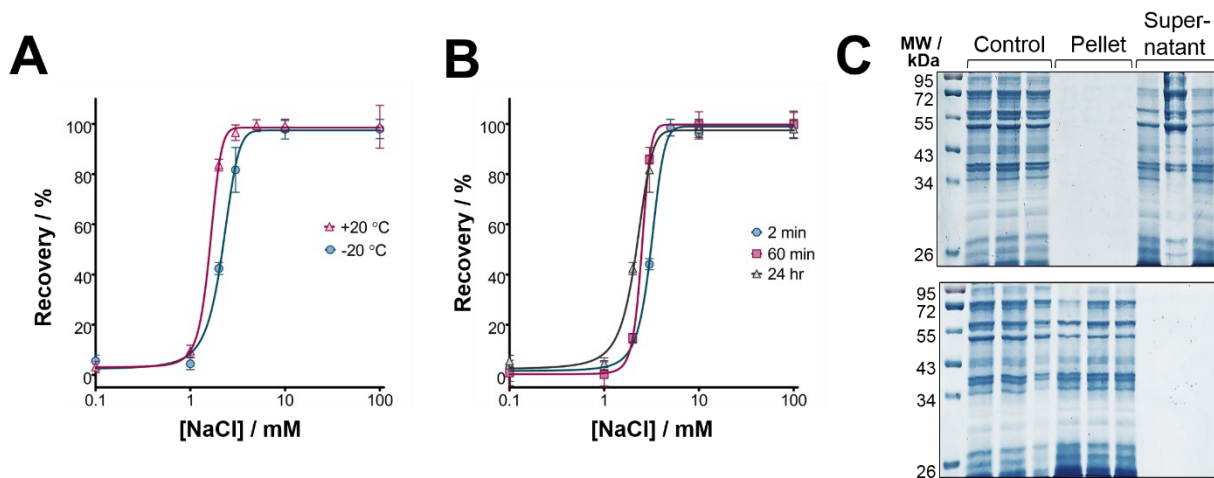


Figure 2.2 Influence of salt on precipitation of BSA (80% acetone). Protein recoveries were determined via BCA assay following precipitation at (A) differing incubation temperatures and (B) varying incubation times (at $+20$ °C). Samples of 1.0 g/L BSA were combined with 80% acetone and a range of NaCl concentrations (0.1-100 mM) for times ranging from 2 min to 24 h. Error bars represent the standard deviation from five replicates. (C) SDS PAGE analysis following rapid precipitation (2 min, room temperature) of a yeast proteome reveals the absence of protein in the pellet when no salt is added (top), and quantitative recovery of proteins in the pellet with the addition of 20 mM NaCl (bottom).

While rapid precipitation protocols are well established, such as by TCA/acetone [135], or chloroform/ methanol/ water [214], prior reports on the required incubation period to maximize

protein recovery through acetone precipitation are conflicting. However, as the ionic strength of the solution was not controlled in these previous studies involving acetone, the present work explores how the NaCl concentration influences protein recovery in 80% acetone when incubated for short (2 min), intermediate (1 h), or extended (24 h) periods. As Figure 2.2B demonstrates, quantitative recoveries (>98%) were consistently obtained for precipitation of 1 g/L BSA, despite employing a minimal incubation period. The establishment of a rapid, high recovery precipitation protocol offers the potential to enhance sample throughput for proteome processing. From Figure 2.2B, minor differences in salt concentration are required to maximize recovery. It is also realized that some proteins require greater salt concentrations than others to maximize yield [132]. It is therefore recommended to include between 20-100 mM NaCl, along with 80 % acetone. This level of salt ensures maximal recovery, as well as precipitation kinetics (discussed below). The rapid precipitation protocol does not specifically require the addition of “sodium chloride”, but rather that the ionic strength of the solution be maintained at a sufficiently high level to induce protein aggregation. As supporting evidence, similar sigmoidal protein recovery curves can be obtained using varying concentrations of Tris-HCl (pH 8.0), or SDS, alongside 80 % acetone, in place of NaCl, as shown in Figure 2.3. It is also noted that the BCA assay has varying compatibilities with SDS and Tris-HCl, potentially contributing to an apparent reduction in the maximized recovery. Evidently, different cationic or anionic species may contribute minor differences in the required concentration to maximize recovery, as seen in the supporting figure. Unless otherwise stated, all subsequent precipitations were conducted with inclusion of NaCl (i.e., “salt”).

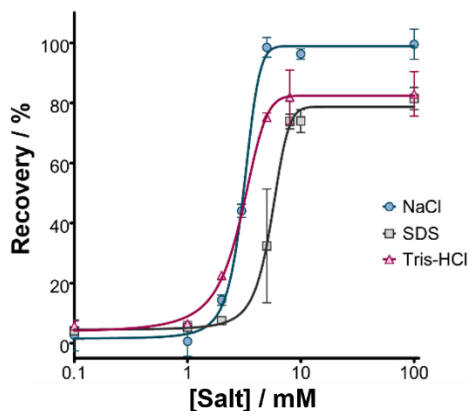


Figure 2.3 Influence of salt types on precipitation efficiency in 80 % acetone and 100 mM of either NaCl, SDS, or Tris-HCl buffer. Protein recoveries were determined via BCA assay following 2 min precipitation at room temperature.

Next it is examined whether the rapid precipitation protocol was equally applicable to the recovery of a complex proteome mixture. A yeast extract was chosen to reflect a range of molecular weights, pI, and protein hydrophobicity values. SDS PAGE analysis (Figure 2.2C) displays the contrasting precipitation efficiencies achieved following 2 min incubation of an *S. cerevisiae* lysate in 80% acetone without the addition of salt (top) or with inclusion of 20 mM NaCl (bottom). In the absence of salt, all proteins from the yeast extract remain fully soluble through the 2 min incubation and are observed in the supernatant fraction. Conversely, in the presence of 20 mM NaCl, all proteins were exclusively observed in the pellet fraction. Residual proteins may be present in the supernatant but would be below the detection limit of Coomassie staining. From Figure 2.2C, the quantitative recovery of proteins appears without bias toward protein molecular weight, extending over the range of 10-100 kDa. The potential for complete proteome characterization by mass spectrometry, including quantitative analysis, relies on avoiding front-end loss of select proteins. Thus, rapid precipitation observed here establishes a standardized approach for subsequent MS analysis (see section 2.3.5).

2.3.2 Higher Salt Increases Precipitation Kinetics in Acetone

Prior studies have investigated the rate of organic-solvent-based precipitation, demonstrating a proportionality between recovery and incubation time, often extending for hours. Such extended incubations fail to achieve the desired throughput required for efficient sample preparation. While it has been established that a minimum ionic strength is required to maximize recovery, here it is demonstrated for the first time that the rate of solvent-based protein precipitation is also dependent on the salt concentration. Test solutions consisting of single protein standards (BSA, cytochrome *c*), as well as a complex yeast proteome extract, were each made to 1.0 g/L and subject to precipitation over a range of incubation times. A high level of salt was selected, (100 mM, Figure 2.2), as well as a lower concentration previously determined as the minimal level required for maximal recovery (0.1-2 mM, depending on the protein sample; shown in Figure 2.4). As Figure 2.5A shows, the rate of precipitation significantly increases in the high-salt condition. With low salt, the observed recovery trends are reminiscent of those reported by others, showing that increased recovery is obtained through longer incubations. In the low-salt condition, the recovery trend can be modeled based on the anticipated first-order kinetics (Table 2.1).

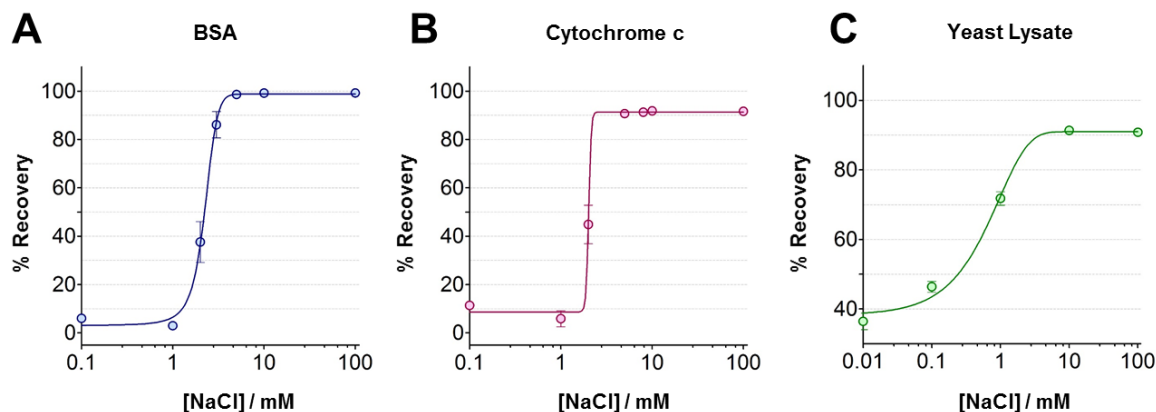


Figure 2.4 Influence of salt concentration on precipitation recovery. Recoveries were calculated based on a BCA assay of the dried supernatant collected after 60 min precipitation for 1 g/L samples of (A) BSA, (B) cytochrome c, and (C) yeast lysate.

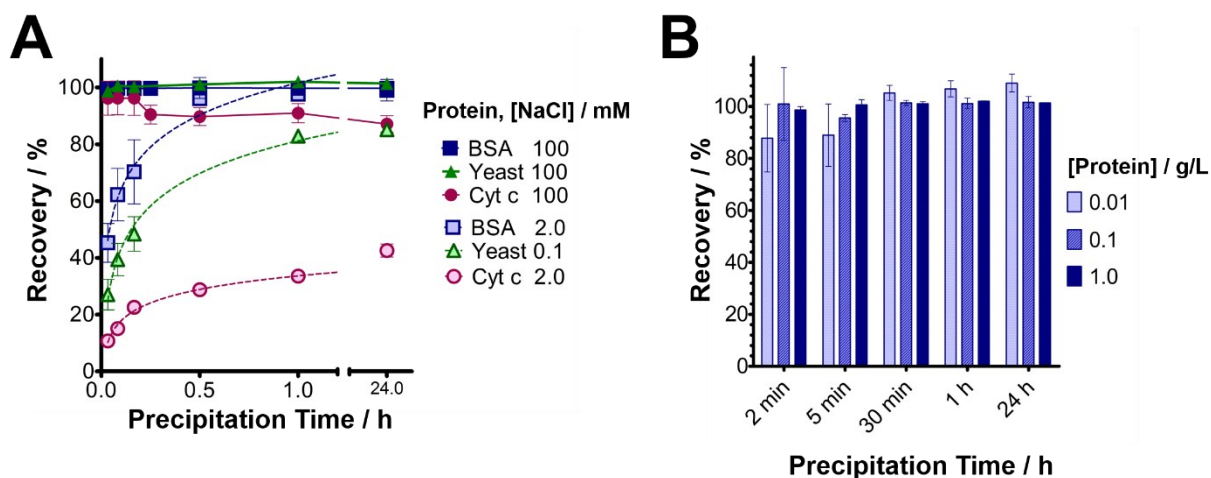


Figure 2.5 Protein recoveries as a function of time following room-temperature precipitation of (A) BSA, cytochrome c, and yeast (1 g/L) with high and low salts, and (B) yeast proteome at varying concentrations with 20 mM NaCl. The second-order rate constants determined for the low-salt condition are summarized in Table 2.1. Recoveries were determined by BCA assay, with error bars representing the standard deviation from five replicates.

Table 2.1 Rate constant determination for room-temperature precipitation in the low-salt condition

Sample	Rate Equation $\ln[A] = \ln[A]_0 - kt^a$	Rate Constant, k (s^{-1})	$t_{1/2}$ (min)	pI
BSA, 2 mM NaCl	$\ln[A] = 103226 + 207.3(t)$	207 ± 38	12 ± 2	4.8
Cyt c, 2 mM NaCl	$\ln[A] = 13276 + 4.437(t)$	4.4 ± 0.2	39 ± 2	9.6
Yeast, 0.1 mM NaCl	$\ln[A] = 62516 + 57.45(t)$	57 ± 8	13 ± 2	6.8

^a[A] = concentration of protein (mol/L); t = time

Calculated half-lives for BSA, cytochrome *c*, and yeast ranged from 12 to 39 min. By contrast, with sufficient salt included, protein recovery for all three test samples exceeds 98% after only 2 min incubation, exhibiting half-lives of less than 30 s. To simultaneously achieve both quantitative recovery with minimal incubation for acetone precipitation, the ionic strength of the sample must therefore be higher than the minimum value previously reported to maximize recovery [132]. With consideration to certain proteins requiring higher salt to maximize recovery (e.g., those with extremes of pI) [132], it is recommended to add 100 mM salt together with 4 volumes of acetone, providing a standardized approach to precipitation with high recovery and throughput, which greatly enhances the value of precipitation as a comprehensive sample preparation technique.

2.3.3 Rapid Recovery of Dilute Samples

As precipitation closely follows second-order kinetics, the initial protein concentration of the sample should be a controlling factor in the rate of recovery. It was of interest to determine whether this rapid precipitation could be extended to more dilute protein samples. When the yeast sample was diluted to 0.1 g/L, essentially quantitative recovery was maintained after 2 min incubation (Figure 2.5B). A slight reduction in recovery was seen for the 0.01 g/L sample following only 2 min incubation (just under 90% yield), with recovery restored to quantitative levels ($105 \pm 3\%$) at 30 min incubation. A 0.001 g/L sample of yeast exceeded the limits of the BCA assay, though recovery at an equivalent dilution was assessed, beginning with BSA, as shown in Figure 2.6. Again, near-quantitative recovery was attainable with 30 min incubation. When precipitating samples at extreme dilutions, the addition of 0.1% SDS to the protein facilitates the formation of a tight pellet that is more easily recovered during protein re-solubilization. Based on these results, it can be concluded that the optimized high-salt/room-temperature precipitation protocol is applicable to samples as dilute as 0.01 g/L, with lower protein concentrations requiring somewhat

lengthier incubations (30 min) in keeping with the expected kinetics of precipitation being dependent on protein concentration. The current depiction of high recovery and throughput for dilute samples described in this chapter exceeds the best performance of acetone precipitation reported to date [131].

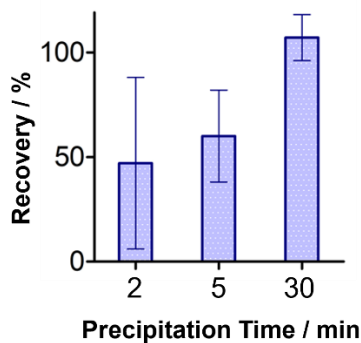


Figure 2.6 Comparing precipitation recovery following 2, 5, and 30 min precipitation of a dilute (0.001 g/L) BSA sample with 80% acetone and 100 mM ionic strength.

2.3.4 Temperature is Key to Rapid Recovery

To this point, all kinetic investigations were conducted using room-temperature precipitations since it was shown that high recovery was not dependent on lowering the sample temperature to $-20\text{ }^{\circ}\text{C}$ during incubation (Figure 2.2A). Here, the use of elevated temperature is determined to be a controlling factor for rapid acetone precipitation in the presence of high salt concentrations. Precipitation efficiency was assessed by bottom-up LC-MS/MS analysis of the proteins identified in the pellet fractions, collected at varying times from elevated ($+20\text{ }^{\circ}\text{C}$) and low ($-20\text{ }^{\circ}\text{C}$)-temperature precipitations. Interestingly, precipitation at $-20\text{ }^{\circ}\text{C}$ provided effectively negligible recovery through the short incubation. Only when the sample had been incubated in the cold for 24 h was high recovery obtained. From Figure 2.7A, only 46 proteins were identified by MS in the 2 min pellet, while 518 proteins were identified in the 24 h pellet. While cold incubations may be employed in the interest of preserving protein activity, the approach is detrimental to a fast precipitation protocol, enforcing the need to precipitate for longer periods of time under the (conventional) cold conditions. The 46 proteins recovered in the 2 min pellet were

examined, but they displayed no distinguishing features in terms of molecular weight, pI, and hydrophobicity (GRAVY score) (Figure 2.8), suggesting they may simply represent the highest abundance proteins. By contrast to the results from cold precipitation, when precipitation was conducted at room temperature, 538 proteins were identified by MS in the 2 min pellet. Considering the three time points examined, 92% of all protein identifications were recorded in the 2 min pellet. Based on total PSM counts from the 2 min precipitation at elevated vs. reduced temperatures, the recovery of proteins was a factor of 50 times greater when incubated at room temperature. Figure 2.9 provides an overview of the proteins identified in the pellet vs supernatant from room-temperature precipitation (2 min, 1 h, 24 h). The full list of identified proteins through the various precipitation protocols is provided as an appendix in the form of Excel spreadsheets (Tables A2.1–A2.20). It is therefore concluded that elevated temperature is an essential factor for obtaining high recovery with rapid precipitation, together with the inclusion of high-salt concentrations.

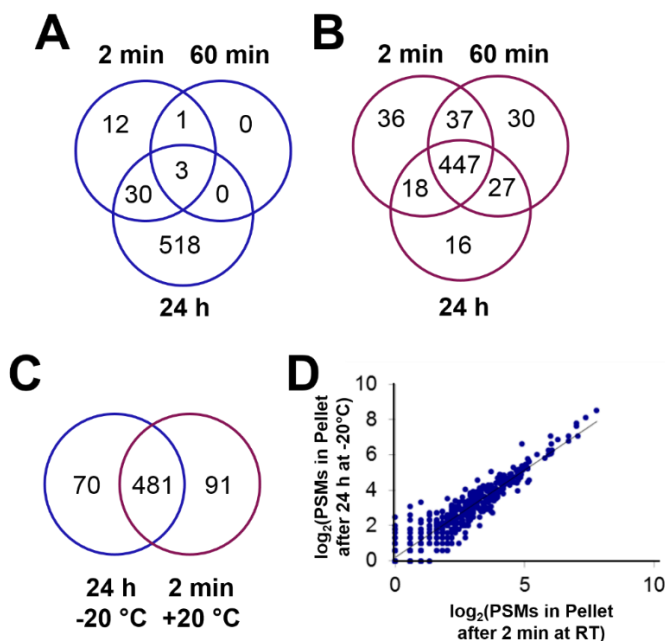


Figure 2.7 (A–C) Venn diagrams of proteins identified in the pellet through (A) cold precipitation (-20°C) and (B) room-temperature precipitation. Shown are the number of yeast proteins identified through bottom-up LC-MS/MS analysis following a time course precipitation conducted with 80% acetone and 100 mM NaCl at the two temperatures. In the cold condition, the majority of proteins were solely identified in the 24 h pellet, while the 2 min RT pellet contained 92% of all identifications achieved by the 24 h point. (C) Comparison of proteins identified following 24 h precipitation at -20°C vs. rapid (2 min) precipitation at room temperature shows equivalent identifications in the two precipitation protocols. (D) Analysis of peptide spectral match (PSM) counts for proteins identified in the pellet collected after 2 min precipitation at room temperature (RT) vs. 24 h at -20°C . The trendline has a slope of 0.99 and a correlation coefficient of 0.9.

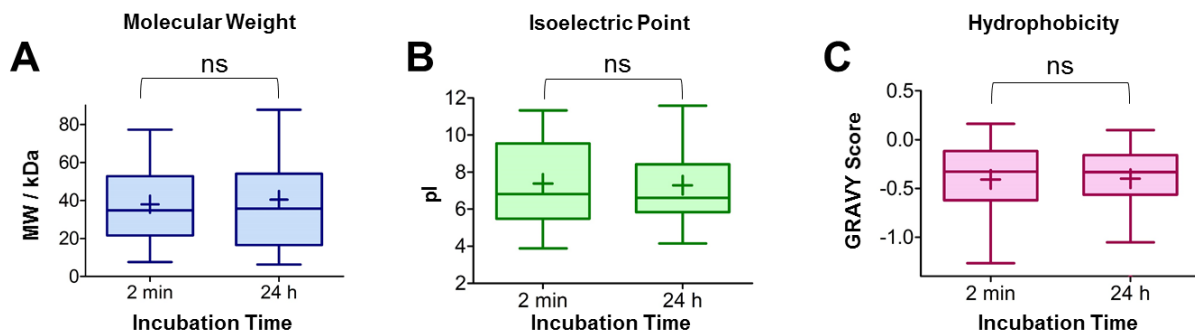


Figure 2.8 Tukey Box-and-Whisker plots comparing the distributions of (A) molecular weight, (B) isoelectric point, and (C) hydrophobicity between the 46 proteins identified in the pellet collected after 2 min and 24 h precipitation at -20°C .

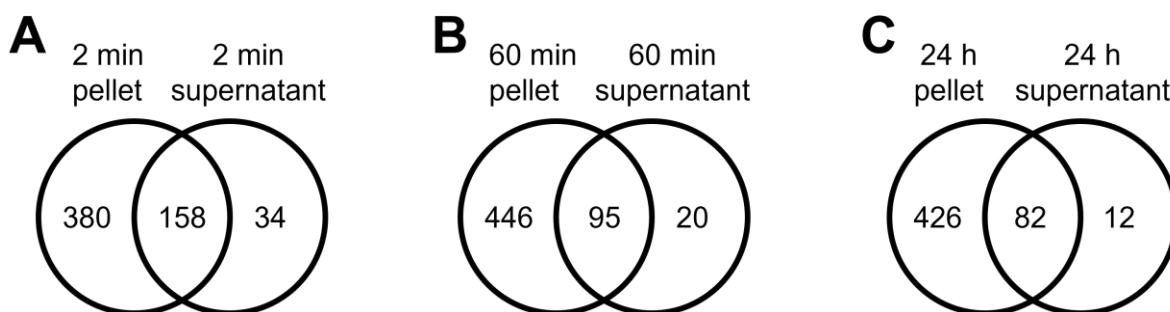


Figure 2.9 Venn diagrams comparing protein identifications from bottom-up LC-MS/MS analysis of pellets and supernatants collected after (A) 2 min room temperature precipitation, (B) 60 min room temperature precipitation, and (C) 24 h room temperature precipitation.

2.3.5 Characterization of the Pellet and Supernatant

A comparison is made between the proteins recovered from rapid (2 min) precipitation at room temperature to a conventional 24 h precipitation at -20°C (both using 100 mM NaCl) (Figure 2.7). While 551 proteins were identified following the conventional overnight protocol (2.7A), 572 proteins were detected from the 2 min precipitation at room temperature (2.7B). A total of 75% of the identified proteins were in common between the two precipitation protocols (Figure 2.7C). Given that replicate MS analysis of identical samples tends to only provide a 70% overlap (Figure 2.10), these results imply a considerable similarity between the two precipitation protocols. A

strong correlation between protein PSM counts across the two precipitation protocols (Figure 2.7D) suggests a similar recovery for the distinctly identified proteins obtained from the two sample preparation methods. The high overlap in identifications and correlation of PSM counts confirm that the fast, room-temperature precipitation significantly improves throughput for sample preparation, without compromising recovery and subsequent analysis of a proteome.

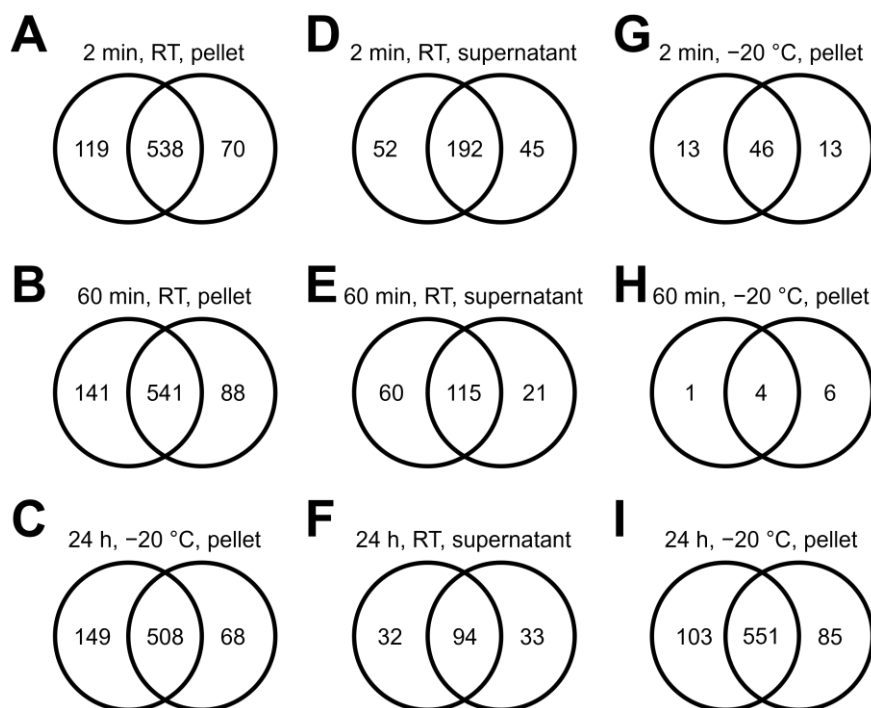


Figure 2.10 Venn diagrams comparing yeast protein identifications made from duplicate MS injections (left = run 1, right = run 2) of (A-C) pellets collected after room temperature incubations, (D-F) supernatants collected after room temperature precipitations, and (G-I) pellets collected after precipitation at $-20\text{ }^{\circ}\text{C}$. (RT = room temperature, $+20\text{ }^{\circ}\text{C}$).

A more detailed analysis of the protein profiles obtained from the two precipitation protocols is presented in Figure 2.12. The distribution of identified proteins is plotted with respect to the molecular weight (Figure 2.12A), isoelectric point (2.12B), and hydrophobicity (2.12C). Pairwise t-testing revealed no statistical bias in the proteins recovered after 2 min precipitation compared to the 24 h, $-20\text{ }^{\circ}\text{C}$ protocol. A comparative analysis of proteins isolated from a rapid, room-temperature precipitation of the yeast proteome extract is performed relative to that of a non-precipitated control. In this instance, precipitated proteins were resolubilized with gel buffer and loaded into an SDS PAGE gel alongside the controls, to ensure that there was no bias in

resolubilizing the resulting pellet. A full listing of identified proteins and peptides is provided as supplementary Excel tables, provided as part of the appendix, and is summarized in Figure 2.11. As shown in Figure 2.12D-F, the resulting proteins again display no bias in terms of the molecular weight, pI, or hydrophobicity profiles. This confirms the effectiveness of the rapid precipitation protocol for proteome profiling. While precipitating under optimal conditions (including 20 mM salt and at room temperature), protein recovery is exceptionally high (~99%), though it is acknowledged that even a 1% loss can still present a certain degree of bias. The question is asked whether the protein content of the supernatant was enriched in specific classes of proteins, which would indicate biased sample loss. MS analysis of the supernatant fraction results in the detection of a significant number of proteins (Figure 2.9), though it is essential to point out by comparison to the analysis of proteins in the pellet, the supernatant fraction was concentrated 100-fold prior to MS. Thus, assuming ~99% of all proteins were recovered without bias in the pellet, one would therefore expect to detect an equivalent set of proteins in the supernatant, and with equivalent intensity, vs. those identified in the pellet fraction. From the bottom-up MS data, the number of proteins in the supernatant constitutes only 19-36% of the total unique proteins observed in the pellet. By way of comparison with total PSM counts, the ratio of protein abundance (supernatant/pellet) is further diminished, ranging from 6 to 15% of the total PSMs observed in the pellet. A smaller subset of proteins (by type and by abundance) is also observed in the supernatant as the precipitation incubation time is increased. However, even when considering the 2 min incubation period, the observed PSM ratios indicate that the fraction of proteins in the supernatant constitutes even less than the predicted 1% by mass of the initial protein.

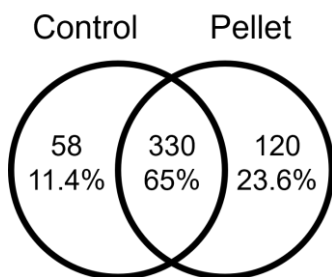


Figure 2.11 Venn diagrams comparing yeast protein identifications made from replicate MS injections ($n = 4$) of a sample that was subject to rapid (2 min, room temperature) acetone precipitation, relative to that of a non-precipitated control. Although unique proteins were identified across each of the resulting samples, the overall distribution of proteins in terms of their molecular weights, pI and hydrophobicity are indistinguishable.

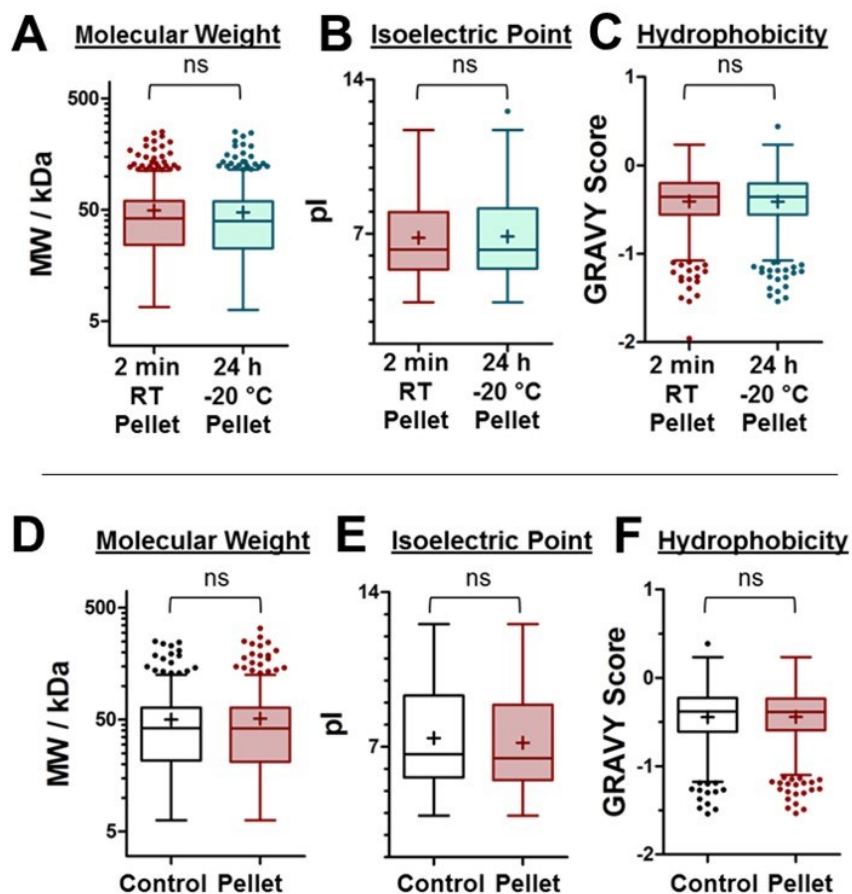


Figure 2.12 Tukey Box-and-Whisker plots comparing the molecular weight (A), isoelectric point (B), and GRAVY score profiles (C) of proteins identified following MS analysis of a precipitation pellet collected after 2 min incubation at room temperature or after 24 h precipitation at $-20\text{ }^{\circ}\text{C}$. Pairwise t-tests reveal no statistically significant difference in the protein profiles. (D–F) display the equivalent comparison between a non-precipitated control yeast extract to that of the pellet collected after 2 min precipitation at room temperature. Again, pairwise t-testing reveals no statistical differences.

While a small percentage (by mass) of proteins partition to the supernatant following precipitation, it was of interest to determine if specific proteins preferentially partition to the supernatant. From Figure 2.13, the detection of a given protein in the supernatant is highly correlated ($R^2 = 0.96$) to its respective abundance (PSM) in the pellet. In particular, 80% of those proteins that were detected in the pellet with relatively high abundance (>30 PSMs) were also detected in the supernatant, while only 20% of the proteins with low abundance (<4 PSMs) were identified in the supernatant. This dependence shows that the protein content of the supernatant fraction is largely composed of the most abundant proteins in the sample by mass.

Such an observation supports that the proteins are not preferentially partitioned to the supernatant according to other distinct properties, such as size, isoelectric point, or hydrophobicity.

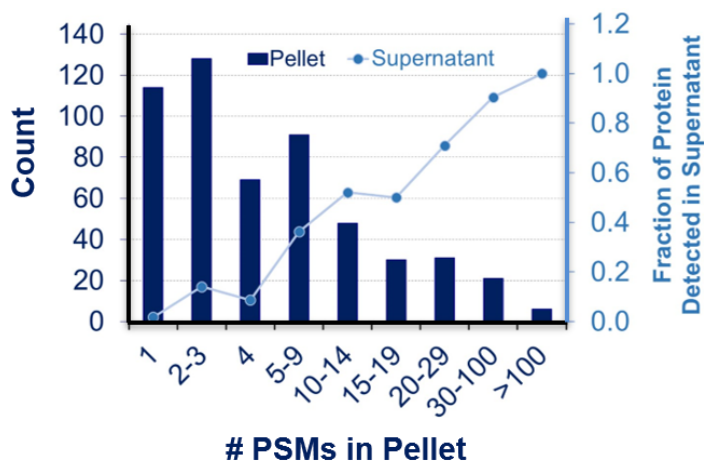


Figure 2.13 Histogram of proteins identified in the 2 min, room-temperature pellet. Proteins are ranked on the x-axis according to abundance as interpreted by their PSM count (peptide spectral match). The fraction of proteins identified in the supernatant is quantified on the secondary y-axis, showing that proteins present in high abundance in the pellet have a correspondingly high probability of being detected in the supernatant.

Previous reports have indicated that solvent precipitation favors the recovery of proteins of high molecular weight, acidic pI, and/or hydrophilic character. It is therefore examined whether these biases are present in the reported optimized rapid precipitation, comparing the pellet and supernatant fractions collected after 2 min, 1 h, or 24 h. Figure 2.14 plots the distribution of proteins in the pellet and supernatant fractions through Box-and-Whisker plots, comparing proteins based on their molecular weight, isoelectric point, and hydrophobicity (GRAVY score). The molecular weight was the only characteristic contributing a statistically significant difference between the pellet and supernatant. The decreasing p-value from pairwise t-tests across the three time points indicates that the low-molecular-weight proteins are being increasingly recovered with additional incubation time. This does not imply that low mass proteins are not recovered in the pellet. Recalling that the supernatant was concentrated 100-fold relative to the pellet, a total of 192 proteins were detected in the supernatant after 2 min precipitation. Of these, 158 (94%) were also identified in the corresponding pellet. Of the remaining 34 proteins uniquely identified in the supernatant (6% of total proteins detected), the molecular weight distribution is similar to that of the total sample. The inability to detect these specific proteins in the pellet may also be attributed

to variations in the capacity of MS to detect low-abundance proteins in the sample. There were no significant differences between the isoelectric points nor the GRAVY scores of any of the pellet/supernatant pairs (Figure 2.14). These results show that rapid precipitation recovers all proteins in a sample, with minimal bias, validating the robustness of this optimized approach as it pertains to proteomic sample preparation.

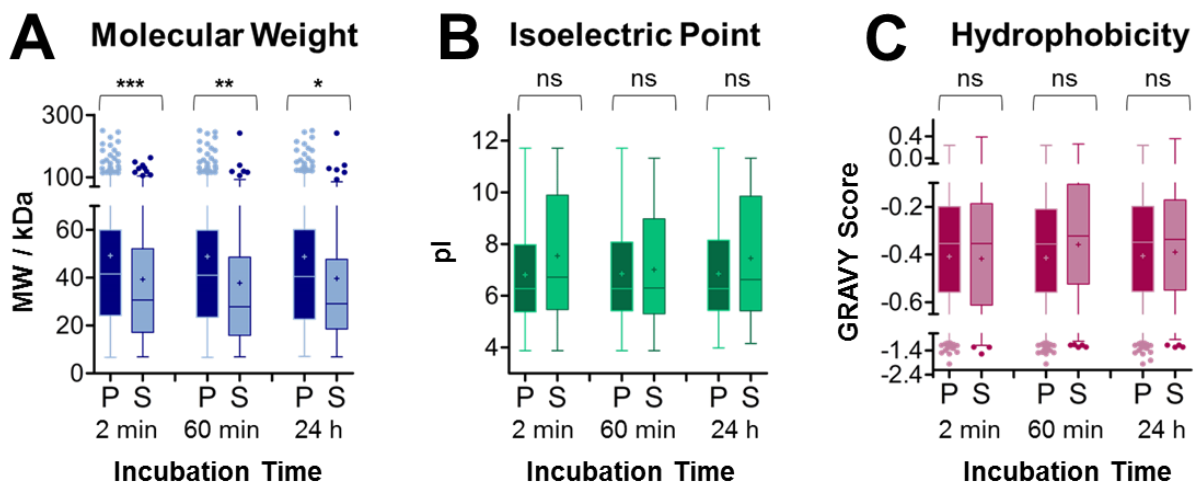


Figure 2.14 Tukey Box-and-Whisker plots of protein properties following room-temperature precipitation. Shown are the (A) molecular weight (MW), (B) isoelectric point (pI), and (C) hydrophobicity (GRAVY, grand average of hydropathicity) score of proteins collected in the pellet (P) and supernatant (S) fractions after 2 min, 60 min, and 24 h precipitation. Pairwise t-testing was used to compare the distributions; asterisks indicate significance: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; ns = not significant.

2.4 Conclusions

The presented work provides an optimized protocol for near-quantitative precipitation of all proteins from simple and complex mixtures. Precipitation at room temperature with inclusion of salt presents a high-throughput approach to isolate a complete proteome with high efficiency. Based on the robustness of the presented approach, rapid, room-temperature precipitation with salt is proposed as a standardized protocol for isolating protein in proteomic workflows, prior to MS analysis. The recoveries obtained here are the highest of any protein precipitation protocol, with minimal bias among the collected proteins, facilitating the comprehensive characterization of a proteome. Therefore, precipitation offers a simple, rapid, and affordable approach to quantitatively isolate proteins. It is anticipated that the high precipitation efficiency of this protocol will grant new opportunities for proteomics, metabolomics, and commercial protein processing.

3. Critical Evaluation on the Repeatability of Sample Coverage and Quantitation in Bottom-Up Sample Preparation Strategies

3.1 Introduction

Proteome initiatives continue to dive deeper as they explore complex biological systems, demanding confident identification with precise quantitation of increasingly low-abundance analytes [301] or minimal sample quantities [302,303]. High quality data not only relies on state-of-the-art MS instrumentation, but on increasingly robust workflows that minimize loss and bias throughout the process from sample collection to MS detection. From prior studies, the consistency of proteomic data acquisition by modern LC-MS/MS platforms [304–309], as well as subsequent analysis software has been well characterized [310–312]. With increasing scan speed, sensitivity, acquisition and fragmentation modes, one can now expect highly repeatable proteomic data, affording rich qualitative and quantitative characterization of proteome samples [308,312–314]. Unfortunately, real sample acquisition continues to be challenged by the limitations of front-end sample processing, questioning the stringency of analytical data.

Comprehensive bottom-up characterization relies on robust sample preparation that affords complete solubilization and enzymatic digestion, especially with the growing interest in bottom-up protein quantitation. Whether quantitation is being conducted by a relative or absolute approach, the accuracy and precision of measured peptide and protein abundances are subject to variances introduced during sample preparation. Total protein solubilization benefits from the inclusion of surfactants, such as SDS or sodium deoxycholate (SDC) [116,315], facilitating unbiased recovery of the proteome irrespective of structural heterogeneity. However, these solubilizing additives compromise digestion efficiency and peptide detection by LC-MS, necessitating effective depletion of these MS interferences. Although many strategies effectively deplete surfactants from proteome samples, they often achieve this purification at the cost of variable recovery [128,131,264,316]. Such trade-offs pose challenges in determining an optimal workflow.

The reported low and variable recovery associated with traditional in-gel and precipitation-based workflows [316] lends favor to the adoption of various cartridge and bead-based strategies [41,258,266,268,269,317]. Several comparative studies evaluate the evolving sample preparation technologies based on sample and sequence coverage as well as quantitative reproducibility. The

first widely adopted cartridge-based platform, FASP, has been shown to offer greater quantitative precision than solution digests, but often at the expense of peptide and protein identifications. A 2014 study by Tanca et al. evaluated direct trypsinization, in solution digestion and FASP based on several metrics including reproducibility [318]. Equivalent overlap in protein identifications was observed between solution and FASP preparations while FASP improved peptide overlap across triplicates by 12%. However, peptide quantitation was, on average, 9% more agreeable across solution preparations. In 2018, Ludwig et al. showed 9% greater protein overlap across two FASP preparations compared to two solution digests, which was met with a 36% greater correlation between protein intensities [264]. A similar trend was observed for peptide overlap, however quantitative precision at the peptide level was equivalent between solution and FASP approaches (R^2 of 0.71 for solution replicates vs. 0.68 for FASP replicates). In 2021, Davalieva et al. compared a RapiGest-based preparation to the FASP approach for fresh and formalin-fixed, paraffin-embedded (FFPE) preserved tissue proteomics. They found that RapiGest provided on average 7 % more repeatable extraction of FFPE samples than FASP and 50% more repeatable extraction of fresh frozen tissues [319].

Subsequent cartridge and bead-based formats have demonstrated improvements in efficiency (e.g., processing time, cost) as well as tolerance to low starting quantities, digestion completion, identification rates, and quantitative precision over the original FASP cartridge. In 2017, Sielaff et al. evaluated FASP, SP3, and iST approaches on the basis of quantitative reproducibility at low sample loadings [270]. FASP demonstrated reduced quantitative precision when processing <10 μ g, while SP3 and iST conserved peptide intensity correlation coefficients of 0.97 and 0.93 respectively. A 2022 study by Wu et al. evaluated the reproducibility of bottom-up identifications following preparation in the S-Trap compared to FASP and solution digests [260]. S-Trap was demonstrated to provide marginal enhancements in protein overlap (82% across two samples compared to 78 % and 79 % for in solution digestion, ISD and FASP respectively) as well as the greatest peptide overlap of the three approaches (69% vs. 60% from ISD). The quantitative precision of the FASP approach was recently improved by Loroch et al. by adapting the cartridge to an automated 96-well plate format [320]. While these comparative studies demonstrate the relative precision across a selection of semi-automated sample preparation technologies, there remains little to no discussion on the factors that influence a strategy's repeatability at the identification and quantitative level. This limits the scope of the conclusions to

the chosen sample type [319,321] or the particular targeted assay [322–327]. A 2022 evaluation of 16 common preparative strategies by Varnavides et al. evaluated 16 of the most common sample preparation approaches. High quantitative precision was demonstrated for all cartridge and bead-assisted platforms, as well as acetone precipitation-based strategies. Lower precision was noted when denaturing additives were included during digestion or removed by precipitation with chloroform/methanol [328]. Considering the trade-off between high recovery and quantitative reproducibility among current strategies, there remains a need for a high-throughput approach to sample preparation that offers quantitative, repeatable and reproducible sample recovery and digestion, enabling high confidence characterization by mass spectrometry.

Recent efforts from this group have elucidated an optimized organic solvent-based precipitation approach, relying on controlled ionic strength in combination with the conventional 80% acetone [132,135]. The optimized precipitation approach is further facilitated by a commercialized ProTrap XG precipitation cartridge—a two-stage filtration cartridge that automates the quantitative isolation of a precipitated pellet from the contaminating supernatant [282]. The present study characterizes the repeatability of sample coverage and quantitative precision provided by the ProTrap XG compared to in-solution and in-gel controls, all benchmarked against inherent instrumental variation. It is demonstrated that identifications and repeatability are optimized by including the surfactant followed by a precipitation-based workflow in the ProTrap XG, offering significant gains in sample coverage and quantitative precision over the conventional in-solution and in-gel approaches. The high repeatability afforded by the semi-automated optimized precipitation strategy shifts the primary source of variation back to the MS detection and subsequent peak integration. The mechanisms behind the biases contributed by each step are further discussed, characterizing the proteins and peptides that contribute variability during extraction/solubilization, precipitation and re-solubilization, digestion, and recovery from a reverse phase desalting cartridge.

3.2 Methods

3.2.1 Growth and Extraction of Yeast

S. cerevisiae was cultured overnight (37 °C) in YPD broth to an OD₆₀₀ of 1, according to standard procedures [295]. The cell pellet was twice washed with water and ground under liquid nitrogen in a mortar and pestle to extract the protein into water or 5% SDS. Protein in the clarified

supernatant was diluted to 2.0 g/L following total protein quantitation by an LC-UV assay with absorbance at 214 nm.

3.2.2 Solution Digestion

A single aliquot of yeast extracted into Tris-HCl (pH 8.0) was combined with 5 mM DTT. The sample was incubated at 56 °C for 30 min, combined with 11 mM IAA and incubated at room temperature in the dark for 1 h. Following reduction and alkylation, the sample was split into four equal aliquots. TPCK-treated trypsin was added at a substrate to enzyme ratio of 50:1. Samples were digested overnight at 37 °C followed by acidification with 0.1% TFA (trifluoroacetic acid). Digested samples were subject to reversed phase clean-up by LC-UV with peptide quantitation at $A_{214\text{nm}}$ and fraction collection.

3.2.3 In-gel Digestion

A single aliquot of yeast extracted into SDS was combined with 5× Laemmli buffer, mixed by vortex, and incubated at 95 °C for 5 min. 50 µg aliquots of the protein sample were loaded into a 12% T SDS PAGE gel, electrophoresed at 100 V across approximately 2 cm of resolving gel, and stained with Coomassie blue. Each lane was excised and sliced into ~1 mm³ portions and transferred to four respective clean microcentrifuge tubes. Conventional in-gel digestion with trypsin was performed as previously described [299] using a 50:1 substrate-to-enzyme ratio. Extracted peptides were subject to additional desalting, quantitation, and fraction collection by reversed phase liquid chromatography (RPLC) as described above.

3.2.4 Sample Preparation in the ProTrap XG

Two aliquots of yeast (one extracted into water, one extracted in SDS) were combined with 50 mM Tris-HCl (pH 8.0) and 5 mM DTT. The samples were incubated at 56 °C for 30 min, combined with 11 mM IAA and incubated at room temperature in the dark for 1 h. Following reduction and alkylation, eight equal aliquots of the SDS extract and one aliquot of the aqueous extract, each comprising 100 µg total protein, were transferred to ProTrap XG filtration cartridges. Samples were combined with 100 mM sodium chloride and four volumes of acetone and subject to rapid (2-min) precipitation as described previously [283]. Following precipitation, precipitated protein was isolated on the membrane of the filtration cartridge by centrifugation and the supernatant was

discarded. Pellets were washed with a fresh aliquot of acetone and immediately centrifuged. Pellets were briefly dried in a fume hood followed by re-solubilization in 8 M urea. Pellets were disrupted by repeat pipetting and periodic mixing by vortex followed by overnight re-solubilization. Samples were diluted to 1.5 M urea, combined with 50 mM Tris-HCl (pH 8.0), and digested at 37 °C overnight with 50:1 trypsin as described above. Digests were quenched with 0.1% TFA. Four of the SDS-containing preparations as well as the precipitated and digested no-SDS extract were immediately subject to RPLC-UV clean-up, quantitation, and fraction collection as described above. The remaining four SDS-containing preparations were subject to additional desalting by the ProTrap XG's associated solid-phase extraction (SPE) cartridge. Samples were loaded onto the spin cartridge in 5% acetonitrile/ 0.1 % TFA and eluted in 50% acetonitrile/ 0.1% TFA. Desalted peptide samples were then subject to RPLC-UV clean-up, quantitation, and fraction collection as described.

3.2.5 Bottom-up LC-MS/MS Acquisition with Label-Free Quantitation

LC-MS/MS analysis of proteome digests was conducted by Phenswitch Biosciences (Sherbrooke, Quebec) according to their standard protocols. Samples were subject to online reversed phase separation using a Kinetex XB C18 column 0.3 mm i.d., 2.6 µm particles, 150 mm (Phenomenex), maintained at 60 °C across a 60-min ethanol gradient containing 3% DMSO/ 0.1% FA at a flow rate of 3 µL/min. Samples were injected by overfilling a 5 µL loop.

MS acquisition was performed on an ABSciex TripleTOF 6600 (ABSciex, Foster City, CA, USA) equipped with an electrospray interface with a 25 µm inner diameter capillary coupled to an Eksigent µUHPLC (Eksigent, Redwood City, CA, USA). Analyst TF 1.8 software was used to control the instrument and for data processing and acquisition. The acquisition was performed in Information Dependant Acquisition (IDA) mode to establish the spectral library. The samples were analyzed in SWATH acquisition mode. The source voltage was set to 5.35 kV and maintained at 325 °C; curtain gas was set at 50 psi, gas 1 at 40 psi, and gas 2 at 35 psi.

3.2.6 DIA-MS Data Analysis

Peptides and proteins comprising the spectral library were identified using Protein Pilot (Sciex). Peptide and protein SWATH quantification were done using DIA-NN (version 1.8.1). Bottom-up tandem MS spectra were searched against the publicly available spectral library of the human

proteome (SWATHAtlas). Spectra from intact peptide fractions were searched against the information dependent spectral library (section 4.2.8). Appendix Tables A3.1-A3.5 contain all bottom-up MS/MS identifications and associated label-free intensities.

Peptide quantitation was evaluated by comparing the coefficient of variation across all identifications. Violin plots and heat maps were generated using the online graphing software, Plotly Chart Studio (Plotly Technologies Inc. 2015) [329].

3.3 Results

3.3.1 Repeatability of Sample Coverage is Optimized in ProTrap XG Preparations

The reproducibility of bottom-up peptide and protein identifications are in part contributed by biases in the recovery and/or digestion efficiency imparted during sample preparation. Here a conventional solution-based digestion is compared to the classic in-gel digestion workup, with the precipitation-based workflow using the ProTrap XG, described previously in Chapter 2. Figure 3.1 shows Venn diagrams of peptide and protein identifications across four replicate solution digests, ProTrap XG preparations, and five replicate LC-MS/MS injections. The precipitation-based spin cartridge provided a 30% enhancement in peptide overlap and a 28% gain in protein overlap compared to the solution digest.

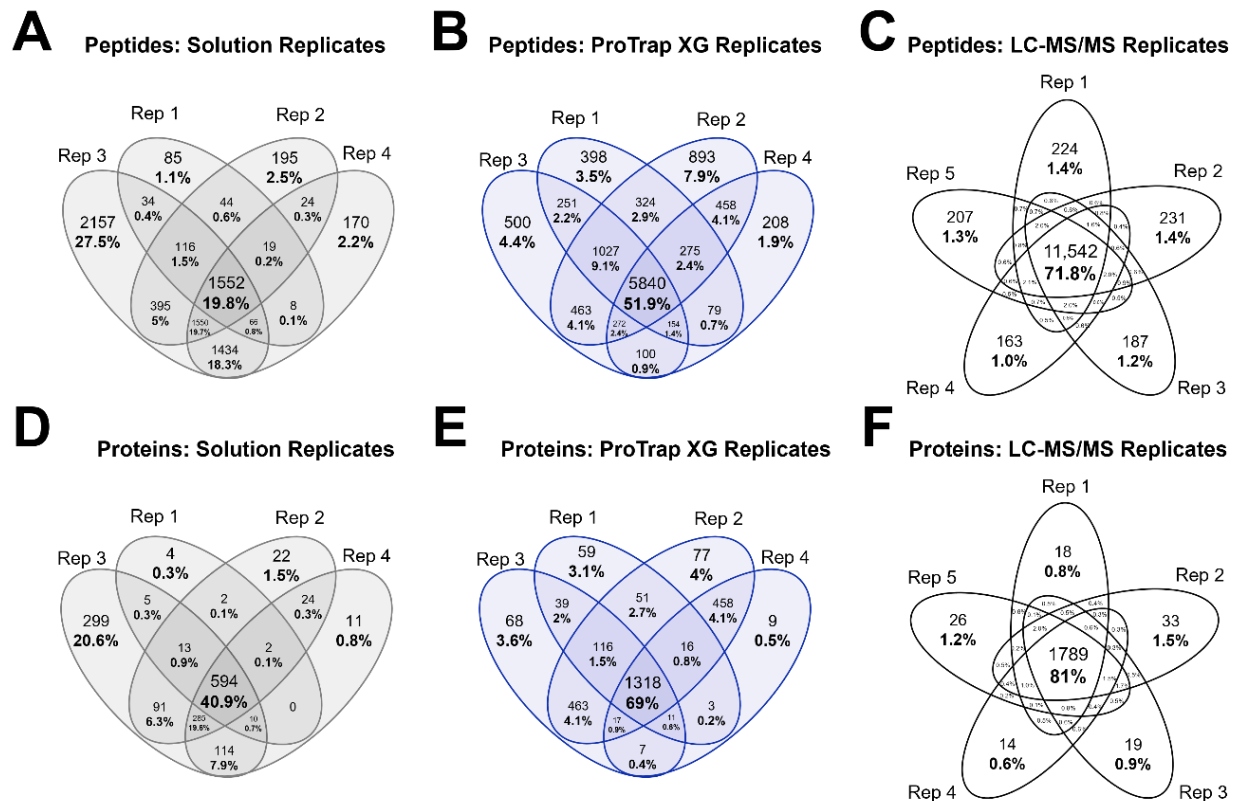


Figure 3.1 Venn diagrams of bottom-up (A-C) peptide and (D-F) protein identifications from solution and ProTrap XG digests compared to replicate LC-MS/MS injections.

Figure 3.2A summarizes the total identifications and overlap for all four evaluated sample preparation strategies. It is shown here that the number of non-redundant peptide and protein identifications was maximized in the ProTrap XG-based preparations. Moreover, the overlap across replicate preparations was optimized when samples were processed using the disposable spin cartridges. At the protein level, the ProTrap XG and ProTrap XG +SPE samples showed 85% and 82% overlap, respectively, which is equivalent to the overlap seen in replicate LC-MS/MS injections (83%) of a single sample. This is reflective of high repeatability in the precipitation, detergent removal, and protein re-solubilization efficiency achieved in the spin cartridge. This represents >13% greater overlap than replicate in-gel digests and >40% greater overlap than solution preparations. As expected, peptide overlap was lower than the intact protein level, however, peptides recovered from ProTrap XG-based strategies showed significantly greater overlap than in-gel and solution strategies. Replicate LC-MS/MS injections of ProTrap XG-based preparations demonstrated 72% overlap across five runs. The other two precipitation-based

preparations demonstrated 60% agreement across four preparation replicates. By contrast, replicate solution digests showed just 19% agreement, with 33% of peptides being unique to one sample. This further supports significantly enhanced repeatability relative to the 51% overlap from in-gel and 21% overlap from solution preparations. While proteins have many chances to be detected, the reduced repeatability of peptide identifications confounds the effects of variable digestion, (recovery from gel), ionization, and identification efficiency.

As demonstrated by the Venn diagrams in Figure 3.1, not all peptides are identified in 4 out of 4 replicate preparations. For example, if a peptide is detected in 3 of 4 samples, it is described as having one missing value. The frequency of missing values poses challenges in practical quantitative investigations, so it is of common interest to minimize the source of these differences in identification profiles. Peptide identifications were sorted based on the number of missing values and plotted in Figure 3.2B. From this, it is shown that peptides identified from ProTrap XG-based preparations exhibit the lowest frequency of missing values, with <25% of identifications showing >1 missing value across 4 replicates. By contrast, 33% of in-gel and 57% of solution identifications are found in $\leq 2/4$ replicates, reflecting low reproducibility in digestion and/or recovery from the gel.

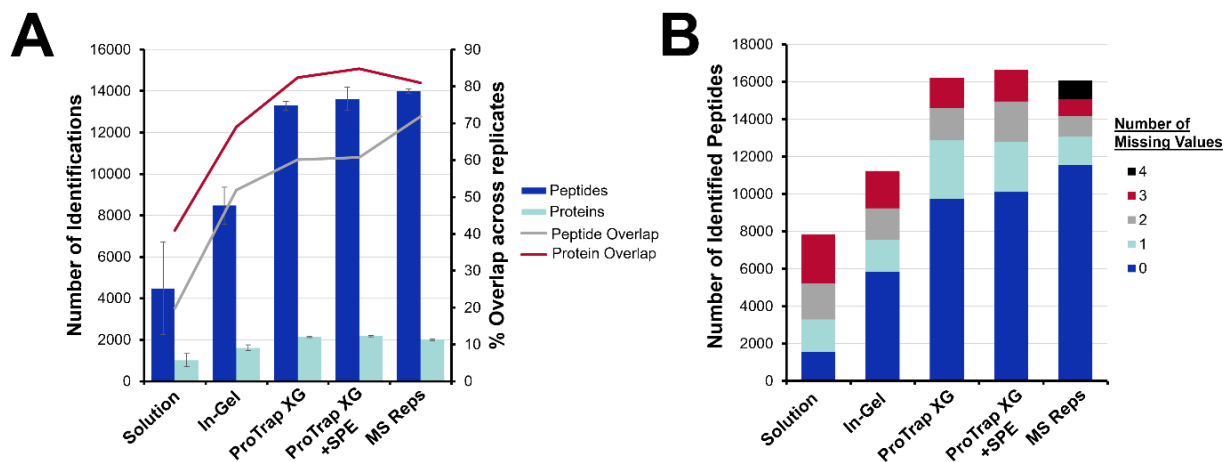


Figure 3.2 Summary of bottom-up peptide and protein identifications (A) total counts and percent overlap across replicates and (B) sorted based on number of missing values. Error bars represent the standard deviation from replicate sample preparations.

The degree of digestion completion directly influences the precision of bottom-up quantitation by way of splitting the signal for a particular sequence across variably miscleaved

products. Figure 3.3 shows the number and frequency of fully-cleaved digestion products resulting from each preparation strategy. From this, ProTrap XG-based preparations demonstrate maximized digestion efficiency along with the least variance across replicate digests. Sample preparation in the spin cartridge resulted in the identification of $10,000 \pm 200$ fully-cleaved peptides without additional clean-up in the SPE. When the SPE was employed, an average of 10,000 fully-cleaved peptides were still identified however the standard deviation increased to 500 across the four preparations. This (slight) reduction in repeatability reflects the variance introduced by recovery from the SPE cartridge. By strong contrast, the in-gel digestion identified 50 % fewer fully-cleaved peptides with a >3-fold increase in the RSD of the frequency of fully-cleaved products. The apparent reduction in digestion efficiency seen from the in-gel approach is consistent with results reported by Yang et al. in 2018, where they compared the efficacy of precipitation-based, cartridge-assisted and in-gel strategies [330]. The authors reported a 30% increase in the frequency of miscleavages from the in-gel approach compared to an acetone precipitation workflow (as is used in the ProTrap XG preparations). The solution digest resulted in the greatest variance in the number of identified fully-cleaved peptides with an RSD of 0.5 across four preparations. However, the fraction of identifications representing fully-cleaved products was highly repeatable, with an RSD of 0.009. In combination with the reduced protein overlap shown in Figure 3.2A, this suggests that the variance in the solution digest is largely contributed by substrate recovery rather than digestion efficiency.

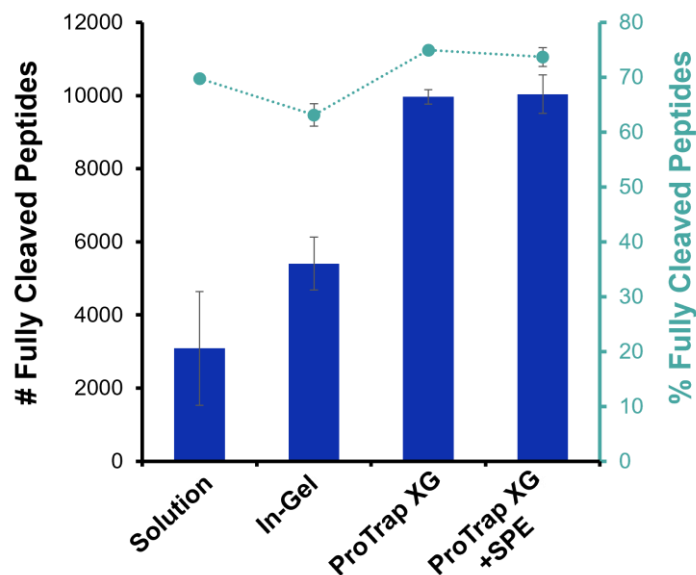


Figure 3.3 Miscleavage analysis of peptides identified from each preparation strategy. The number and frequency of fully-cleaved peptides are optimized in ProTrap XG-based approaches. Error bars represent the standard deviation from replicate sample preparations.

Peptides identified from each preparation were sorted based on their number of missing values (MVs) and each group was compared on the frequency of miscleavages, frequency of modifications, and median intensity; the resulting correlations are shown in Figure 3.4. As reviewed by Webb-Robertson et al. missing values are categorized as *missing completely at random* (MCAR), *missing at random* (MAR), or *missing not at random* (MNAR) [331]. Values are classified as MCAR if their missingness does not correlate with low intensity or any other characteristic of the peptide. Values are considered MAR if there is a correlation to a particular peptide characteristic. And values are classified as MNAR if their missingness correlates with low intensity that may be below the limit of detection. Considering that the total peptide mass was normalized for each injection, differential intensity for a given peptide from one run to the next indicates a differential abundance between the respective preparations. From Figure 3.3, it is shown that the number of missing values for a given peptide from any of the evaluated preparation strategies correlates with low intensity, suggesting that many MVs could be classified as MNAR.

From Figure 3.4(A,F,K), it is shown that for solution digests, a peptide's missing values do not correlate significantly with modification or miscleavage frequency, suggesting that partially digested products are generated with equivalent repeatability as completely digested products. Together, these results support the previous speculation that the variance in solution digests is most

attributable to variable recovery. It was also found that 24% of proteins identified in the set of solution digests were unique to 1 of 4 preparations, which suggests that the variance in recovery is at the protein level.

From Figure 3.4G, the in-gel digests show an interesting trend where repeatability of peptide identifications correlate with degree of miscleavage. In combination with the low digestion efficiency shown in Figure 3.3, this suggests that partial digestion products are generated with the greatest frequency by in-gel digests. Contrasting with the solution digest, it was found that 80% of proteins associated with peptides that were identified in 1 of 4 preparations were also identified by more repeatable peptides. This shows that protein recovery is highly repeatable (likely owing to the inclusion of SDS) while peptide recovery from the gel matrix is a source of variation across replicates.

Repeatability as a function of miscleaved peptides shows the opposite trend in the ProTrap XG. From Figure 3.4H, it is seen that the most repeatable peptides are more completely digested than those with more missing values. This suggests that partially digested peptides are not as consistently produced and identified, which is reflective of the enhanced digestion efficiency depicted in Figure 3.3. Given that there is a similar correlation across replicate LC-MS/MS injections (Figure 3.4J,O), the reduced repeatability of miscleaved and modified peptides may be attributed to detection and identification efficiency rather than variances in the preparation. It was also found that 85 % of protein groups associated with peptides that were unique to 1 of 4 preparations were identified by other peptides with greater repeatability. This suggests that a peptide's variability is largely attributable to its cleavage and ionization efficiency as opposed to the recovery of its associated protein. This is further supported by the high solubilization efficiency expected from SDS-containing preparations as well as the quantitative precipitation recovery associated with the ProTrap XG [135,283].

Figure 3.4M shows an additional correlation between the number of missing values and frequency of modifications. Spectra were searched with carbamidomethylation of cysteine residues as a fixed modification and oxidation of methionine as a variable modification. It was found that the frequency of cysteine-containing peptides as a function of missing values was constant (8-9%); therefore, the relationship between repeatability and modification frequency is a function of variable oxidation at methionine residues. Methionine-containing peptides can be observed in both the oxidized and non-oxidized form; it was found that 52% of oxidized peptides

were identified without the modification with greater frequency. This suggests that the oxidation of methionine residues occurs with relatively low frequency, making modified peptides appear less repeatable. Taken together, missing values from ProTrap XG preparations include MNAR (missing not at random) and MAR (missing at random) values.

Samples prepared in the ProTrap XG with additional clean-up in the associated SPE cartridge demonstrate no correlation between a peptide's missing values and miscleavage or modification rates. This suggests that peptide retention and elution from the reversed phase SPE cartridge is introducing some degree of random variance across replicates, which is to be expected by adding any additional preparative step. Potential biases in SPE recovery rates are discussed further in section 3.2.2.

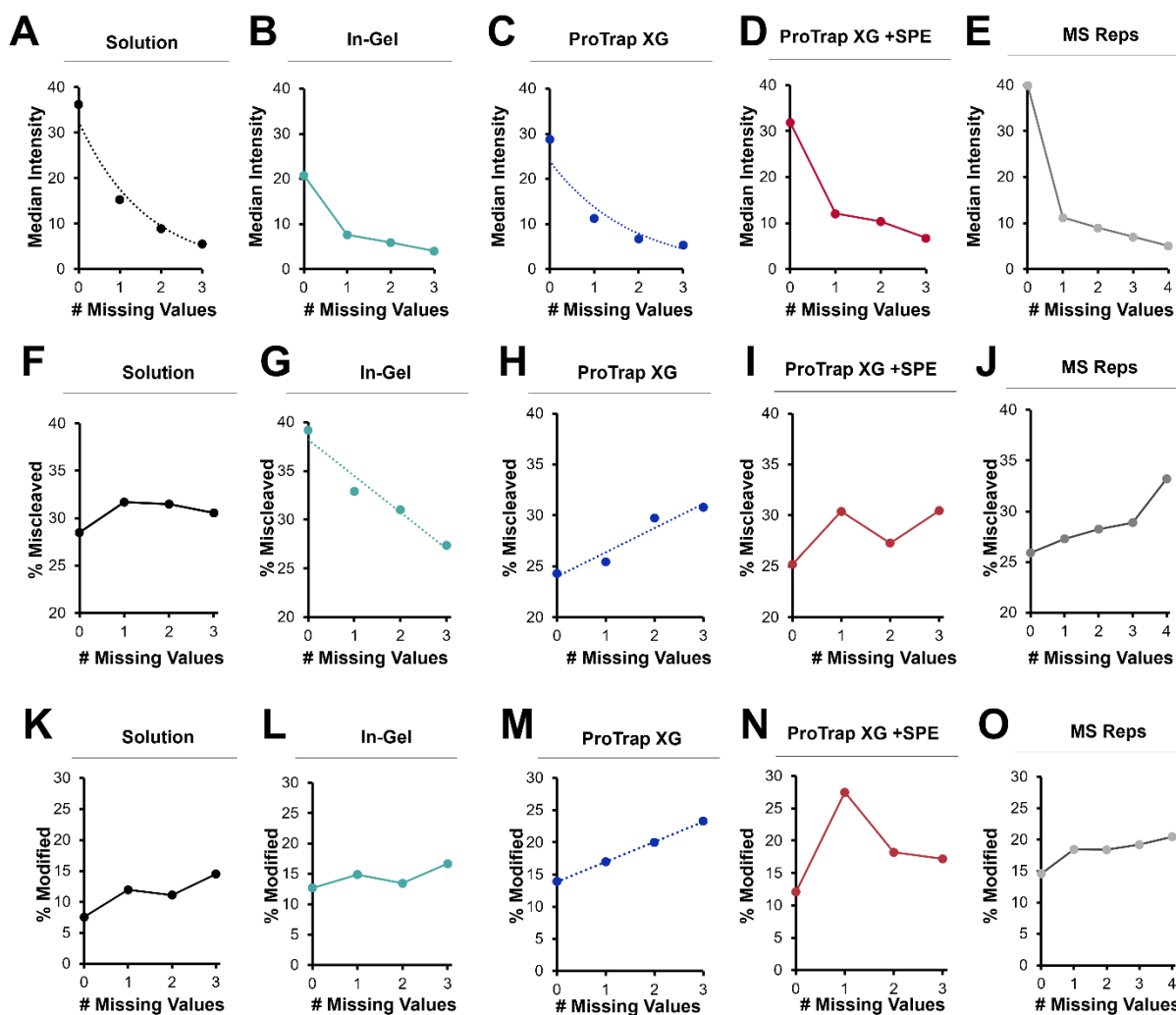


Figure 3.4 Correlating number of missing values with (A-E) peptide intensity, (F-J) miscleavage, and (K-O) modification rates. All preparation strategies demonstrate correlated identification repeatability with peptide intensity.

3.3.2 Quantitative Precision is Optimized by Sample Preparation in the ProTrap XG

While the reproducibility of protein identifications is of high importance for shotgun proteomics, precision at the peptide level has even greater influence for quantitative initiatives. SWATH quantification was used to compare peptide intensities. The variance in LC-MS/MS replicates was modeled as a function of peptide intensity. Based on the error model, the LOD and LOQ were determined to be 8.95 and 140.54, respectively. Figure 3.5A shows the fraction of peptides from each set that exceed the LOD and LOQ, demonstrating that sample preparation in the ProTrap XG enables confident detection and quantitation of the greatest number of peptides and by consequence, proteins. The heatmap in Figure 3.5B shows the high variance in peptide intensity across solution digests. In-gel preparations show improvement over the solution samples, while the ProTrap XG and ProTrap XG +SPE samples show the greatest agreement across replicates among all the strategies. Principal Component Analysis was used to deconvolute the large datasets and evaluate the precision of replicates. From Figure 3.5B, replicate samples from the ProTrap XG and ProTrap XG +SPE show the highest degree of consistency second to the replicate LC-MS/MS injections. In-gel digests show slightly greater variance, while solution samples account for the majority of the variance across all twenty-one samples with significant spread carrying over to the second principal component. Based on these comparisons, preparation strategies can be ranked from most repeatable to least repeatable as follows: ProTrap XG, ProTrap XG +SPE, in-gel, solution, which shows agreement with the trends observed across the repeatability of identifications.

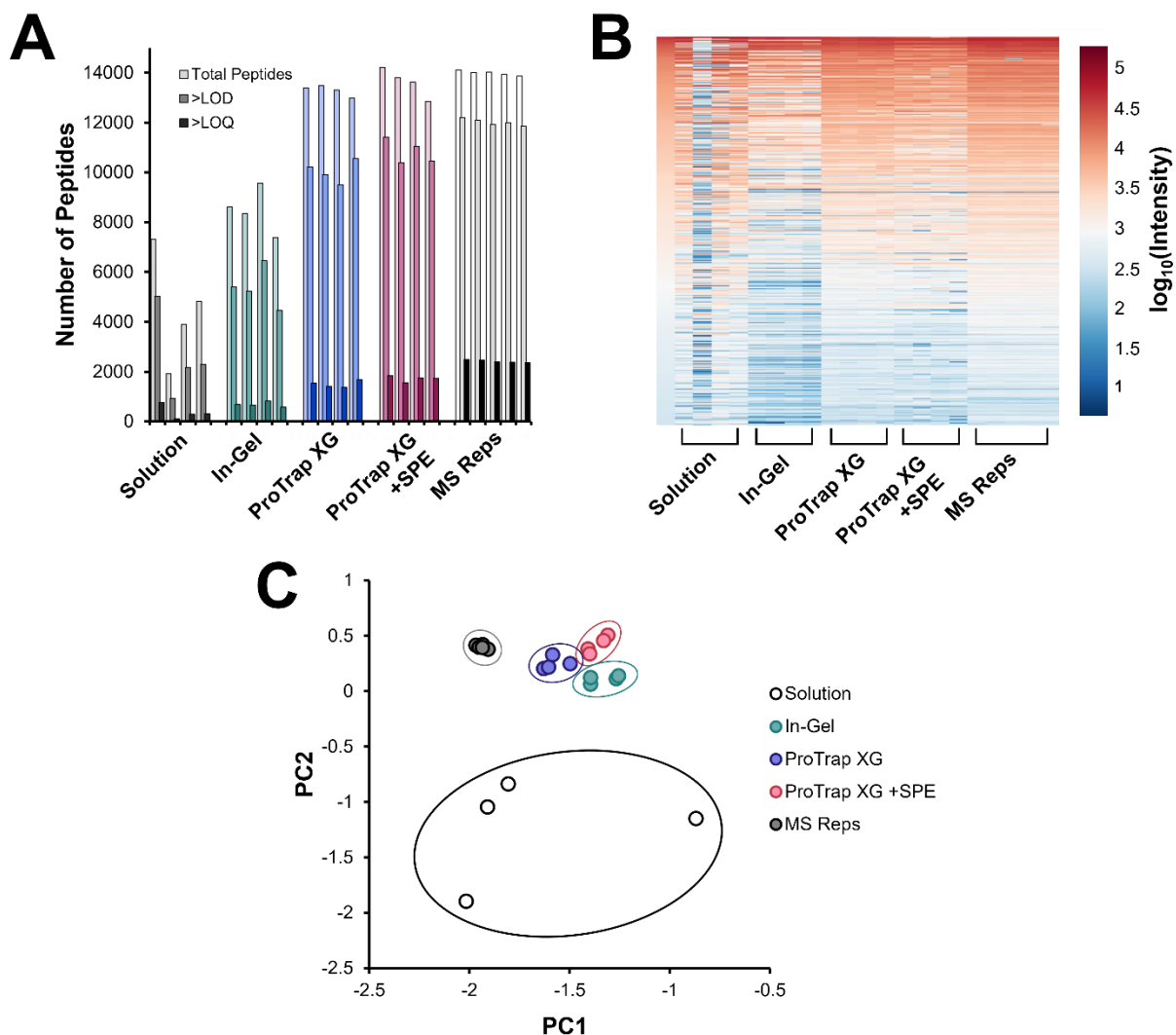


Figure 3.5 Assessment of quantitative repeatability for peptides with no missing values. (A) Peptide identifications classified based on LOD and LOQ. (B) Heatmap of peptide intensities reflecting maximized precision across five replicate LC-MS/MS injections, followed by ProTrap XG-based preparations. In-gel and solution digests show high variance across replicates over the range of plotted intensities. (C) Principal component analysis further reflects high precision across ProTrap XG samples compared to in-gel and solution digestions.

Bottom-up quantitative repeatability was further evaluated by comparing the distribution of coefficients of variance for peptides commonly identified across all replicates of each preparation strategy. From Figure 3.6A, it is shown that replicate LC-MS/MS injections have a mean CV of 10% with a standard deviation of 7%. Samples prepared in the ProTrap XG with and without additional clean-up in the SPE demonstrate comparable mean CVs of $25 \pm 19\%$ and $18 \pm$

15%, respectively. This shows that approximately 40% of quantitative variance is contributed by the LC-MS/MS detection, while precipitation and digestion in the ProTrap XG contributes 32% and the additional SPE clean-up adds 28% of the total variance. In-gel digests exhibit a broader distribution but contribute a similar average variance to the SPE workflow with a mean CV of 26%. Once again, the solution preparations demonstrate the weakest repeatability among the evaluated strategies with an average CV of $48 \pm 26\%$. Contrasting with the precipitation-based workflows, this shows that 80 % of quantitative variance is contributed by the solution digest preparation. This continues to support the observation that SDS-containing workflows are significantly more repeatable with precipitation-based approaches in the ProTrap XG offering greater precision than in-gel digests.

Variance in a peptide's measured intensity was speculated to correlate with signal intensity since shot noise becomes a greater contributor. Peptides commonly identified across all twenty-one samples were binned into 10 fractions according to their normalized intensity. Mean CV was calculated for each bin and plotted against the median intensity (Figure 3.6C). Replicate LC-MS/MS injections demonstrate the greatest correlation between peptide intensity and quantitative repeatability, especially in the lowest 50%. The ProTrap XG samples demonstrate a similar trend, albeit with greater CVs at high intensities compared to LC-MS/MS replicates. This shows that sample preparation contributes variance that is independent of peptide intensity (as described previously with respect to miscleavage and modification rates). ProTrap XG +SPE and in-gel digests continue to exhibit some dependence between repeatability and peptide intensity, but with a lesser slope than the other ProTrap XG preparations. This weaker correlation suggests there are more significant sources of variation from the SPE recovery and in-gel preparation than a peptide's intensity when compared with the other precipitation-based strategies. Solution digests show an opposite relationship whereby high intensity peptides exhibit lower quantitative precision than lower intensity peptides, however at any intensity, quantitation is at least 15% less precise from a solution digestion than any of the other evaluated preparations. Tanca's 2014 study demonstrated a comparable trend where the absolute concentration of a protein digest recovered from RapiGest and FASP preparations correlated with its quantitative precision [318]. Although a peptide's intensity is a product of both, its abundance and its ionization efficiency, their results support that the stronger a protein's total signal, the more repeatable its bottom-up quantitation will be. From Figure 3.6C, it is shown that for ProTrap XG-based preparations, quantitative precision is greater

for fully-cleaved peptides than miscleaved peptides. This is attributed to there being >5-fold more total signal intensity being contributed by complete digestion products compared to partial digestion products in these two approaches. In-gel and solution preparations show a similar bias towards greater repeatability of fully-cleaved peptides although with a lower degree of confidence, which correlates with the greater signal contribution from miscleaved products in these two sets. Chiva et al. reported a similar observation where they showed the distribution of peptide standard deviation was independent of its degree of miscleavage for solution and FASP preparations [332].

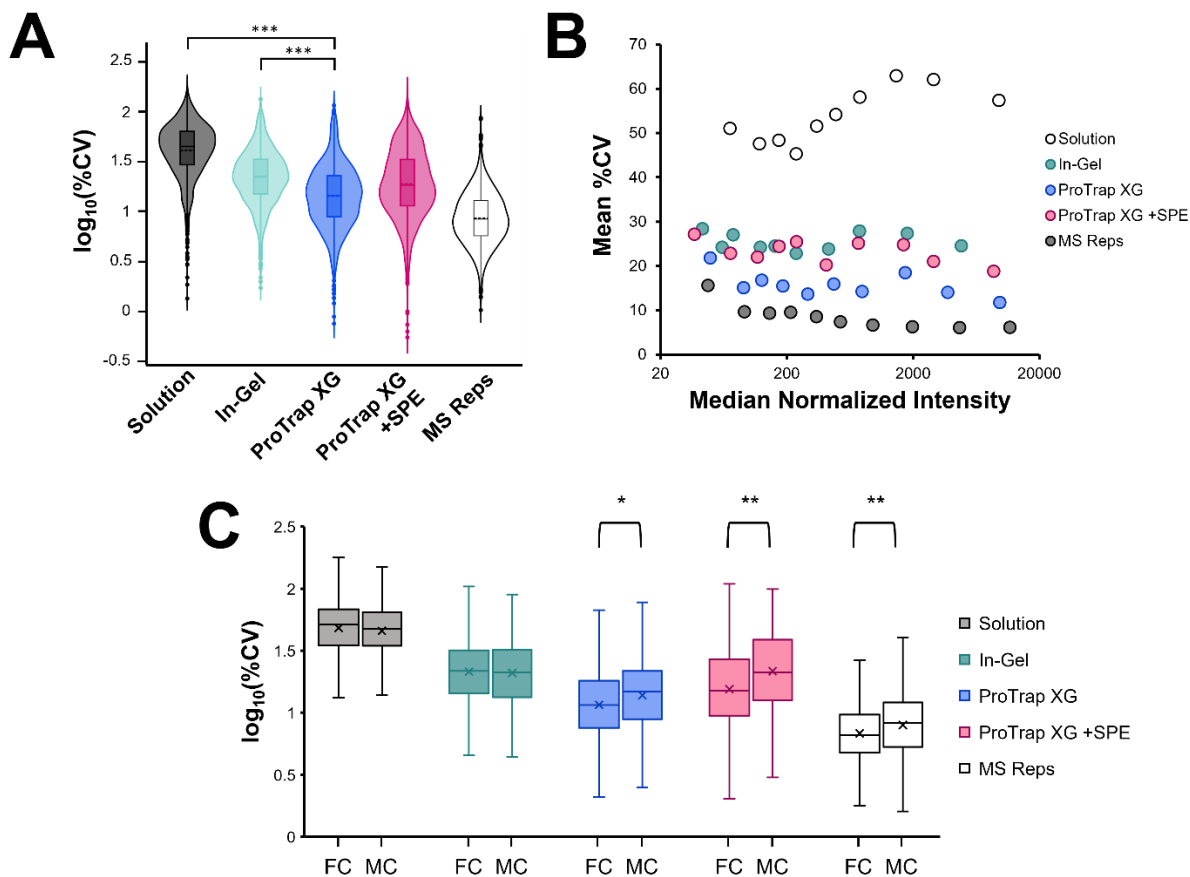


Figure 3.6 Coefficient of variation (CV) plotted (A) as a violin plot representing the distribution for each preparation strategy, and (B) as a function of median intensity. (C) Tukey Box-and-Whisker plot of CV distributions for fully-cleaved peptides (FC) compared to miscleaved peptides (MC) for each preparation strategy.

Quantitation agreement as a function of intensity is further observed in Figure 3.7, where a peptide's intensity from one preparation is plotted against its intensity in another replicate. For

all preparation approaches, variance between samples is greatest at low intensities with increasing precision (and linearity) as intensity increases. Outliers are noted in the correlation of SPE-cleaned samples (Figure 3.6D), while solution digests (Figure 3.6A) show low precision irrespective of peptide intensity. Comparison of correlation coefficients (Figure 3.6F) demonstrates the similarity between the ProTrap XG preparations and LC-MS/MS replicates. In-gel digests show a greater spread in the precision of replicate pairs and solution digests continue to demonstrate the least agreement across replicate preparations with an average correlation coefficient 10% lower than any other approach.

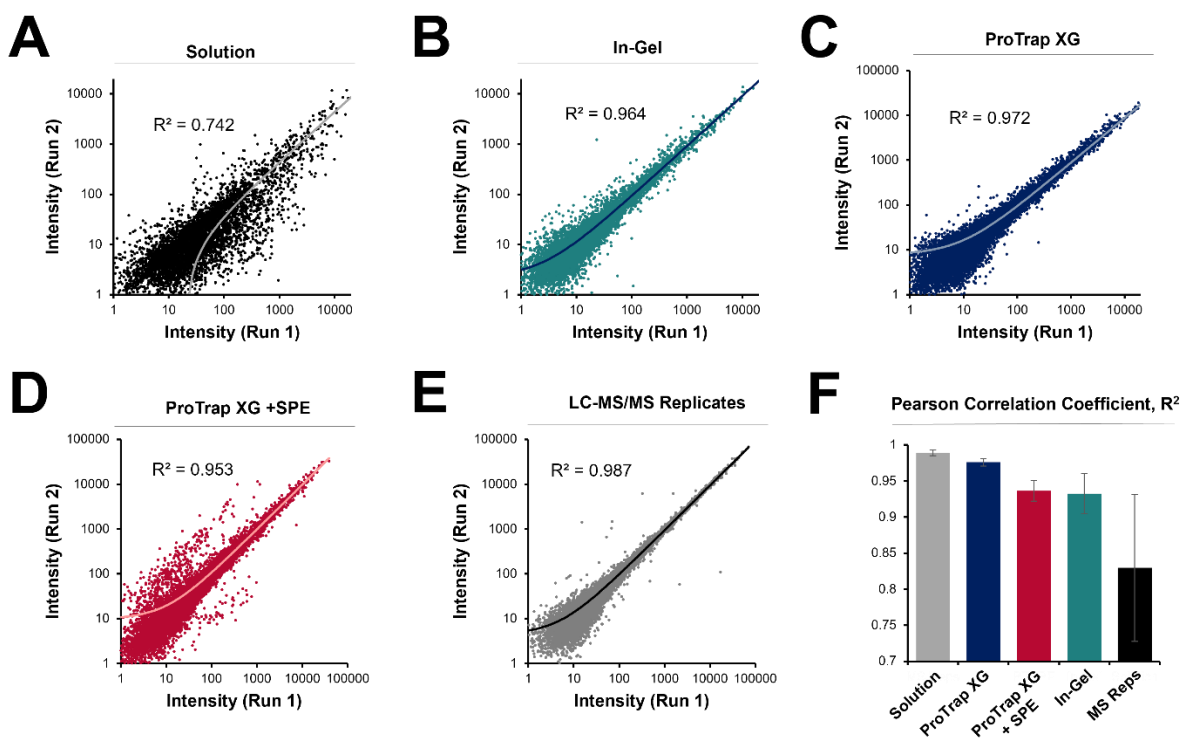


Figure 3.7 Correlation plots of peptide intensities across a selected pair of replicate preparations for (A) solution, (B) in-gel, (C) ProTrap XG, (D) ProTrap XG +SPE, (E) replicate LC-MS/MS injections. (F) Average Pearson correlation coefficients, R^2 , error bars represent the standard deviation across all possible pairs of samples.

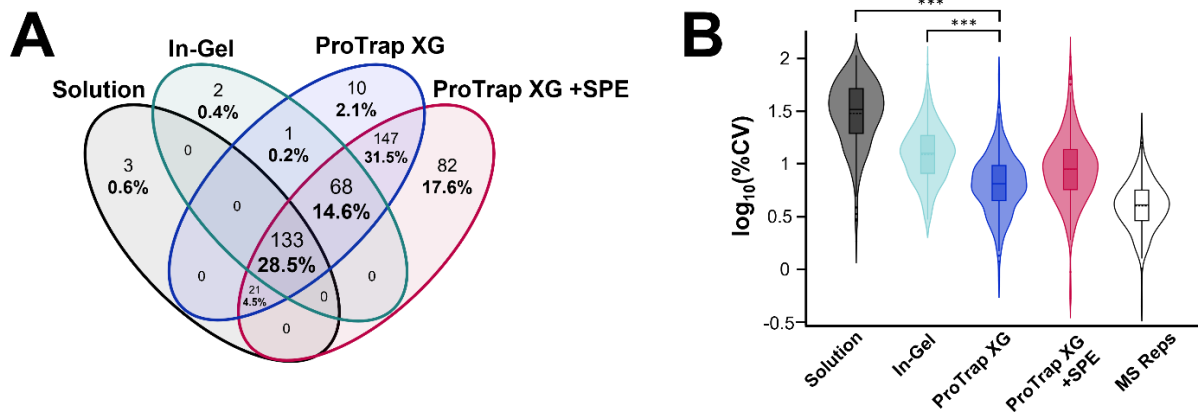


Figure 3.8 Characterizing the precision of bottom-up protein quantitation. (A) Venn diagram of protein groups with associated peptides whose intensity exceeded the limit of quantitation. (B) Violin plot of the distribution of coefficients of variation for proteins quantified in all sample preparation strategies.

3.4 Discussion

Bottom-up protein quantitation has impactful applications including clinical assays for biomarker detection [333]. The combined need for high throughput and precise measurements in these settings demands robust sample preparation that minimizes front-end variances. The present study evaluated the repeatability of peptide and protein identifications as well as bottom-up quantitative precision for solution, in-gel, and two precipitation-based workflows in the ProTrap XG. It is herein demonstrated that both the repeatability of identifications and quantitative precision are optimized by exploiting an SDS solubilization step followed by a robust precipitation approach in the semi-automated spin cartridge and subsequent pellet digestion.

Maximized repeatability of peptide and protein identifications across sample preparations is the first step in achieving precise quantitation, which motivates the adoption of DIA-MS acquisition strategies over DDA. However, missing values are not completely eliminated in DIA, leaving a point of discussion on how to accurately and precisely quantify peptides that are not identified in all cases [331,334]. A 2022 report by Li and Smyth proposes that all missing values in DIA should be classified as MNAR, speculating that intensity is the only factor influencing a peptide's detection [333]. It was shown in Figure 3.4 that the repeatability of peptide identification correlated with signal intensity across all evaluated preparation strategies. The repeatability of identifications and quantitative precision were both greater for fully-cleaved digestion products compared to miscleaved peptides, however these properties were subsequently shown to correlate

with the peptide's intensity, which supports the idea that missing values are “not at random” but rather, below the limit of detection.

The differential abundance of a peptide is controlled by its protein's recovery as well as its cleavage efficiency. Of the evaluated sample preparation strategies, the solution digests demonstrate the lowest identification agreement and quantitative precision by a large margin. However, the frequency of miscleaved digestion products is only 5% greater than is seen in more precise (ProTrap XG-based) approaches. This suggests that the recovery of the intact protein is limiting the robustness of solution digests, as opposed to digestion efficiency. Considering the lack of denaturing additives, initial solubilization efficiency is reduced. Additionally, thermal aggregation is accelerated at the temperature used for reduction by DTT (56 °C) [166] in the absence of chaotropes such as urea. By contrast, in-gel and ProTrap XG preparations benefit from the protein solubilization afforded by an SDS extraction. Subsequent reduction and alkylation proved to be a source of variation in the ProTrap XG samples that underwent clean-up by SPE. A series of outlier points seen in Figure 3.7D were analyzed and it was found that this group is enriched in cysteine-containing peptides, with these analytes comprising 80.5% of the outlier group compared to 9 % of the total peptide population from this set of samples. A 2013 study by Jiang et al. discusses the influence of different alkylation modifications on a peptide's retention in reversed phase separations [335]. The authors show that with TFA as an ion-pairing modifier, carbamidomethylated peptides from IAA alkylation demonstrate weaker retention compared to peptides modified by other alkylating reagents. Future investigations will evaluate the repeatability of peptide recovery from the ProTrap XG's associated SPE cartridge using a range of cysteine alkylating reagents and sample loading conditions.

Enzymatic digestion has been described for its controlling influence in bottom-up quantitation precision [332]. Digestion efficiency was shown to be optimized in ProTrap XG preparations, benefiting from effective pellet re-solubilization in 8 M urea. The high digestion efficiency is thought to enhance quantitative precision by increasing the intensity of fully-cleaved peptides rather than splitting a sequence's signal across variably miscleaved products. In-gel digestion demonstrated reduced completion by way of increased miscleaved peptides, which correlated with reduced quantitative precision. It is speculated that variances in digestion completion could arise from trypsin's ability to permeate the gel, followed by a reduction in peptide signal intensity associated with variable extraction efficiency from the gel.

Evaluating the front-end workflow as a whole, the results presented here suggest that protein and peptide recovery are highest and most reproducible in the SDS-ProTrap XG preparations. Additional clean-up in the SPE cartridge contributes some variance especially in the recovery of cysteine-containing peptides, although the total distribution of variance resulting from SPE preparations is statistically similar to that without the SPE clean-up. In-gel preparations show improved repeatability compared to solution digests, likely owing to enhanced protein solubilization efficiency, however the reduced digestion efficiency correlated with lower quantitative precision than is observed for the ProTrap XG.

3.5 Conclusions

The present study demonstrated that identification repeatability and the precision of bottom-up quantitation benefit from complete proteome solubilization enabled by the inclusion of surfactants followed by robust surfactant depletion and proteome digestion. It was shown that the low repeatability of solution digests severely limits the application of the simplistic approach when stringent analytical results are required. In-gel digestion showed improved repeatability in peptide and protein identifications as well as quantitative precision but was limited by the high frequency of partially cleaved peptides. By contrast, precipitation-based workflows conducted in the ProTrap XG afforded complete solubilization with SDS in combination with robust digestion, which correlated with optimized repeatability of identifications and quantitative precision, second only to replicate LC-MS/MS injections. Based on this, precipitation-based strategies facilitated by the ProTrap XG are recommended for future bottom-up quantitation applications.

4. Multi-Omics in the ProTrap XG: Optimizing a Selective Precipitation Strategy for Mass Spectrometry Analysis of the Proteome, Peptidome, and Metabolome

4.1 Introduction

Multi-omics approaches are gaining popularity in clinical [100,336] and food science [337,338] settings owing to the depth of information afforded by integrating multiple -omics analyses. In particular, different combinations of genomics, transcriptomics, proteomics, and metabolomics are often integrated. The combined data provide greater insight on disease pathology [339–341], for example, than an individual -omics approach. Despite the sensitivity of modern mass spectrometry, these workflows are often limited by low sample quantity with each class of analytes being expressed with a unique abundance [97,342]. This poses challenges in developing a single sample preparation approach that provides optimal coverage of several -omics fractions. Some of the most common multi-omics approaches combine genomics/ transcriptomics/ proteomics, or proteomics/metabolomics. However, the present study will focus on the integration of proteomics, peptidomics, and metabolomics through a sequential precipitation approach.

Metabolome extraction is often achieved by solid phase microextraction (SPME) strategies or organic solvent-based protein precipitation. A 2019 review by Reyes-Garcés describes the advantages of SPME for metabolomics, offering a simple high-throughput approach with good selectivity between the desired analytes and interferences of complex matrices [343]. The main disadvantage of SPME is low recovery, which confounds with the limitations of frequently small sample quantities. Diverse precipitation and liquid-liquid extraction (LLE) strategies have been employed and are individualized based on the metabolites of interest. The standard approach to metabolome preparation involves a $-20\text{ }^{\circ}\text{C}$ protein precipitation with 80% methanol for varying precipitation times. With respect to the LLE approach, a 2020 review by Roca discussed the structural and solubility diversity across a metabolome resulting in the differential recovery of each class of metabolites with different extraction solvent systems [344,345]. Polar analytes such as amino acids, nucleotides, and carbohydrates are recovered from extraction with various compositions of methanol /ethanol/ acetonitrile/ water, while non-polar species such as glycerolipids and steroids show better extraction efficiency in isopropanol, methyl tertbutyl ether and ethyl acetate [346]. Bechmann et al. demonstrated a five-step metabolome workflow in 2021,

where six aliquots of an adrenal tissue lysate were differentially extracted to maximize metabolome coverage [97]. A 2018 study by Wawrzyniak et al. showed that hydrophobic metabolites often co-precipitate during the protein depletion step [347]. The authors showed that pre-incubation with proteinase K improved non-polar metabolome coverage. The enzyme induced a partial digestion, reducing the folded structure of native proteins, which released trapped metabolites. This demonstrates that proteome/ metabolome multi-omics approaches will benefit from further efforts to maximize the selectivity between the two sets of analytes. In addition to the diverse solvent systems used for extraction, the incubation conditions are critical in order to optimize recovery without compromising the native sample. Longer extractions and increased temperature offer improved recovery, while protein precipitation is often conducted at $-20\text{ }^{\circ}\text{C}$ for several hours or overnight in order to conserve the un-modified metabolome [96,348]. The two strategies have associated advantages and drawbacks—precipitation poses the risk of co-precipitating certain metabolites alongside the protein pellet, while a liquid extraction may favor the recovery of some metabolites over others based on the chosen solvent system. LLE recovery would also be improved by extracting with a greater volume, reducing the total analyte concentration. Recent efforts from the Doucette group showed quantitative protein precipitation is achieved in 2-5 min under optimized conditions [132,135], which was hypothesized to have dual application for proteome and metabolome preparations.

Peptidomics is emerging as a complement to proteome analysis, offering particular value in immune response characterization. As reviewed by Dallas in 2014 and Fan in 2022, peptidome sample preparation often relies on complete recovery of the proteome and peptidome followed by a separation of the high- and low-molecular weight fractions [349,350]. Protein/peptide fractionation has been achieved by molecular weight cut-off strategies, low-temperature protein precipitation, size exclusion chromatography (SEC), and solid phase extraction (SPE). As described for metabolome sample preparation, each strategy has unique advantages and disadvantages, however, the potential for integrating peptidomics analyses in clinical settings [351] demands a preparative approach that prioritizes high recovery, throughput, repeatability, and accuracy (i.e., does not modify the native peptidome). Conventional precipitation approaches involving overnight incubation at $-20\text{ }^{\circ}\text{C}$ as well as SPE purification benefit from reduced modification/degradation rates, however recovery and peptide/protein selectivity are variable [352]. Meanwhile, the low throughput of MWCO and SEC approaches risk compromising the

native peptides while exhibiting variable selectivity and reproducibility. A 2021 review by Foreman recommended the combination of protein depletion by organic solvent precipitation followed by metabolite removal by SPE [351] for optimal peptidome coverage. A recent study by Baghalabadi and Doucette demonstrated high-efficiency precipitation of low molecular weight proteins and peptides using the unique combination of 95% acetone, 100 mM ZnSO₄ and rapid benchtop incubation [136].

While sample preparation strategies vary among proteome, peptidome, and metabolome initiatives, organic solvent-based protein precipitation is a common strategy that has been exploited for optimal coverage of all three -omics analyses. Given this group's recent development of protein and peptide precipitation approaches that simultaneously maximize recovery and throughput, the present study aimed to integrate the two precipitation strategies towards a rapid approach for the recovery of the proteome, peptidome and metabolome of a single sample. Evaluation of metabolome, proteome, and peptidome coverage and quantitative precision following a range of precipitation strategies revealed the utility of acetone- or methanol-based sequential precipitation strategies for a proteome/ peptidome/ metabolome multi-omics workflow. Qualitative and quantitative results are herein described for each of the three fractions. It was found that protein depletion with 80 % acetone followed by peptide precipitation of the supernatant optimized peptidome coverage, while initial protein precipitation with 80% methanol improved metabolome analysis. The two precipitation strategies demonstrated equivalent efficacy for proteome recovery. The mechanisms behind the apparent differences in selectivity of the two solvent systems are discussed.

4.2 Methods

4.2.1 Growth, Culture, and Lysis of Adherent Cell Line

Human cell lines, AsPC-1 and Panc-1, were cultured at 37 °C in 5 % CO₂ in RPMI-1640 media with 10 % heat inactivated fetal bovine serum, 100 U/ml penicillin, and 100 mg/mL streptomycin sulfate until confluence. Cells were rinsed with phosphate buffered saline at pH 8 (PBS), trypsinized using Trypsin-EDTA (ethylenediaminetetraacetic acid) (0.05%), harvested by centrifugation, and stored at -20 °C prior to lysis. Frozen cell pellets were suspended in 100 mM PBS buffer and lysed by grinding under liquid nitrogen. The lysate was sonicated on ice for 1 h and cell debris was pelleted by centrifugation. The clarified supernatant was subject to total protein

quantitation by BCA assay, diluted to a final concentration of 2 g/L and brought to an ionic strength of 50 mM with PBS. Solubilized samples were stored at $-20\text{ }^{\circ}\text{C}$ prior to precipitation.

4.2.2 Precipitation-based Metabolome Purification

Aliquots of the solubilized cell lysate containing 100 μg total protein were added to twelve ProTrap XG filtration cartridges and three conventional vials. The three vial samples were combined with four volumes methanol and precipitated for 5 min at room temperature with gentle mixing followed by centrifugation at $4\text{ }^{\circ}\text{C}$. Supernatants were decanted to clean vials and stored at $-20\text{ }^{\circ}\text{C}$ prior to metabolomics analysis by LC-MS.

Triplicates of the ProTrap XG samples were combined with varying organic solvent systems at final concentrations of 70% acetone, 80% acetone, 95% acetone, and 80% methanol. Samples were precipitated for 5 min at room temperature followed by isolation of the supernatant by centrifugation at 3000 rpm at $4\text{ }^{\circ}\text{C}$.

Equal volumes of all recovered metabolome extracts were combined as a control for relative quantitation. This pooled reference sample as well as all recovered metabolome extracts were stored at $-20\text{ }^{\circ}\text{C}$ prior to LC-MS analysis.

4.2.3 Targeted HILIC-MS Metabolomics Analysis

Metabolome analysis was conducted at Dalhousie University's Biological Mass Spectrometry CORE Facility according to standard protocols. Data were acquired on an Agilent 1290 Infinity II liquid chromatograph coupled in-line to a QTRAP 5500 (Sciex, Framingham, MA, USA) triple-quadrupole linear ion trap tandem mass spectrometer operated in selected reaction monitoring (SRM) mode. Samples were subject to online chromatographic separation by injecting 50 μl on a hydrophilic interaction chromatography (HILIC) XBridge® Amide 3.5 μm particle, $1.0 \times 50\text{ mm}$ column (PN 186004871; Waters, Milford, MA, USA) at $37\text{ }^{\circ}\text{C}$ across an 8-minute gradient starting at 5% 20 mM ammonium carbonate (pH 9.8) and 95% acetonitrile, ending at 98% 20 mM ammonium carbonate pH 9.8. The LC flow rate was set at 200 $\mu\text{l}/\text{min}$. A $475\text{ }^{\circ}\text{C}$ heat-assisted electrospray source (TurboIonSpray; Sciex) was used for ionization; curtain gas was kept at 20 units and gas 1 and 2 at 33 units (arbitrary units in Analyst, version 1.6.2; Sciex). Data were acquired in both negative and positive modes using ionization spray voltages of -4.5 and 5.5 kV , respectively. The combined injections contained a total of 565 selected reaction monitoring (SRM)

target transitions for the analysis of main central metabolites. The SRM data was acquired using Analyst 1.6.2 software (Sciex) with peak integration obtained from Skyline Version 19.1, developed by the MacCoss Lab [353].

4.2.4 Metabolomics Data Analysis

Tandem MS spectra were searched against a panel of 565 target metabolites in Skyline Version 19.1 [354]. Precursor ion areas were normalized to the pooled control sample. Appendix Table A4.1 contains all targeted metabolomics output.

4.2.5 Protein Precipitation for Proteome Recovery in the ProTrap XG

Six aliquots of the solubilized cell lysate containing 100 µg total protein were added to ProTrap XG filtration cartridges. Three were combined with four volumes acetone and three with four volumes methanol. Samples were precipitated for 5 min at room temperature with gentle mixing followed by pellet isolation by centrifugation at 3000 rpm at 4 °C. Pellets were briefly air dried in a fume hood to evaporate any residual organic solvent.

4.2.6 Sequential Peptidome Recovery and Preparation

Supernatants were recovered following protein precipitation (described in section 4.2.5) and combined with additional acetone at a final concentration of 95 % organic solvent and combined with ZnSO₄ (100 mM) [136]. Samples were precipitated for 5 min at room temperature followed by centrifugation at 4 °C. Supernatants were decanted as described previously [283] and pellets were re-solubilized in 80 % formic acid at -20 °C for 1 h with periodic vortex mixing [292]. Solubilized pellets were subject to reversed phase clean-up with quantitation at A₂₁₄ and fraction collection in 45% acetonitrile. Collected fractions were dried in a SpeedVac and stored at -20 °C prior to LC-MS/MS analysis.

4.2.7 Bottom-up Proteome Sample Preparation

Dried protein pellets (Section 4.2.5) were combined with 50 µL 8 M urea and re-solubilized for 1 h with periodic vortex mixing and repeat pipetting. Samples were diluted to a final concentration of 1.5 M urea and combined with 50 mM Tris-HCl (pH 8.0). Samples were subject to thiol reduction with 5 mM DTT for 1 h at 56 °C followed by alkylation with 11 mM IAA for 1 h at

room temperature in the dark. TPCK-treated trypsin was added at a substrate-to-enzyme ratio of 50:1 and samples were digested overnight at 37 °C and subsequently quenched with 0.1% TFA. Acidified digests were subject to a reversed phase LC peptide clean-up with quantitation at A₂₁₄ and fraction collection in 45% acetonitrile. Collected fractions were dried in a SpeedVac and stored at –20 °C prior to bottom-up LC-MS/MS analysis.

4.2.8 LC-MS/MS of Sequentially Precipitated Pellets: DIA Acquisition with Label-Free Quantitation

Proteome digests were subject to online reversed phase separation using a Kinetex XB C18 column 0.3 mm i.d., 2.6 µm particles, 150mm (Phenomenex), maintained at 60 °C across a 60-min ethanol gradient containing 3% DMSO/ 0.1% FA at a flow rate of 3 µL/min. Samples were injected by overfilling a 5 µL loop.

MS acquisition was performed on an ABSciex TripleTOF 6600 (ABSciex, Foster City, CA, USA) equipped with an electrospray interface with a 25 µm inner diameter capillary coupled to an Eksigent µUHPLC (Eksigent, Redwood City, CA, USA). Analyst TF 1.8 software was used to control the instrument and for data processing and acquisition. The acquisition was performed in Information Dependant Acquisition (IDA) mode to establish the ion library. The samples were analyzed in SWATH acquisition mode. The source voltage was set to 5.35 kV and maintained at 325 °C; curtain gas was set at 50 psi, gas one at 40 psi, and gas two at 35 psi.

4.2.9 Data Analysis of Bottom-up Protein Preparations and Intact Peptide Fractions

Peptides and proteins comprising the ion library were identified using Protein Pilot (Sciex). Peptide and protein SWATH quantification were done using DIA-NN (version 1.8.1). Bottom-up tandem MS spectra were searched against the publicly available ion library of the human proteome (SWATHAtlas). Spectra from intact peptide fractions were searched against the information dependent ion library (section 4.2.8) at an FDR of 1%. Appendix Table A4.2 contains all bottom-up proteomics output. Appendix Table A4.3 contains all peptidomics identifications and intensities.

Bottom-up protein quantitation was evaluated by summing all peptide intensities associated with each protein identification and comparing the average coefficient of variation

across all associated peptides. Violin plots and heat maps were generated using the online graphing software, Plotly Chart Studio (Plotly Technologies Inc. 2015) [329].

4.3 Results

4.3.1 Metabolomics Coverage and Quantitation is Optimized in the ProTrap XG

Metabolome coverage was compared against a panel of 600 targets following a variety of protein precipitation strategies. Shown in Figure 4.1A, metabolite identifications were optimized by precipitation with 80% methanol in the ProTrap XG. From Figure 4.1B, 63% of identifications were common between all precipitation strategies. The efficiency of pellet isolation in the spin cartridge afforded 10% more metabolite identifications than using the same procedure in a conventional vial. The increased identification rates from precipitation in the ProTrap XG (highlighted by the Venn diagram in Figure 4.1C) enabled enhanced coverage of additional metabolic pathways including fatty acid and lipid biosynthesis, secondary metabolite biosynthesis, and amino acid biosynthesis and degradation. Compared to the methanol approach in the ProTrap XG, 70-80% acetone was the next most effective, showing only a 7% drop in the total number of metabolite identifications. The Venn diagram in Figure 4.1D compares the metabolite overlap between 80% acetone and 80% methanol precipitation in the ProTrap XG. The identifications unique to the methanol precipitation enabled enhanced coverage of pathways including nucleoside/nucleotide biosynthesis, and amino acid biosynthesis/degradation. Increasing the acetone concentration to 95% showed a greater loss in metabolome coverage, providing 8% fewer identifications than 70-80% acetone and 16% fewer than were observed from 80% methanol. Following a comparison of metabolite identifications from 95% acetone with 80% methanol in the ProTrap XG, it was noted that the analytes unique to the acetone approach were enriched in non-polar compounds, including ceramide, cholesterol, and Vitamin K2, while those unique to the methanol approach largely consisted of polar analytes such as sugars and amino acids.

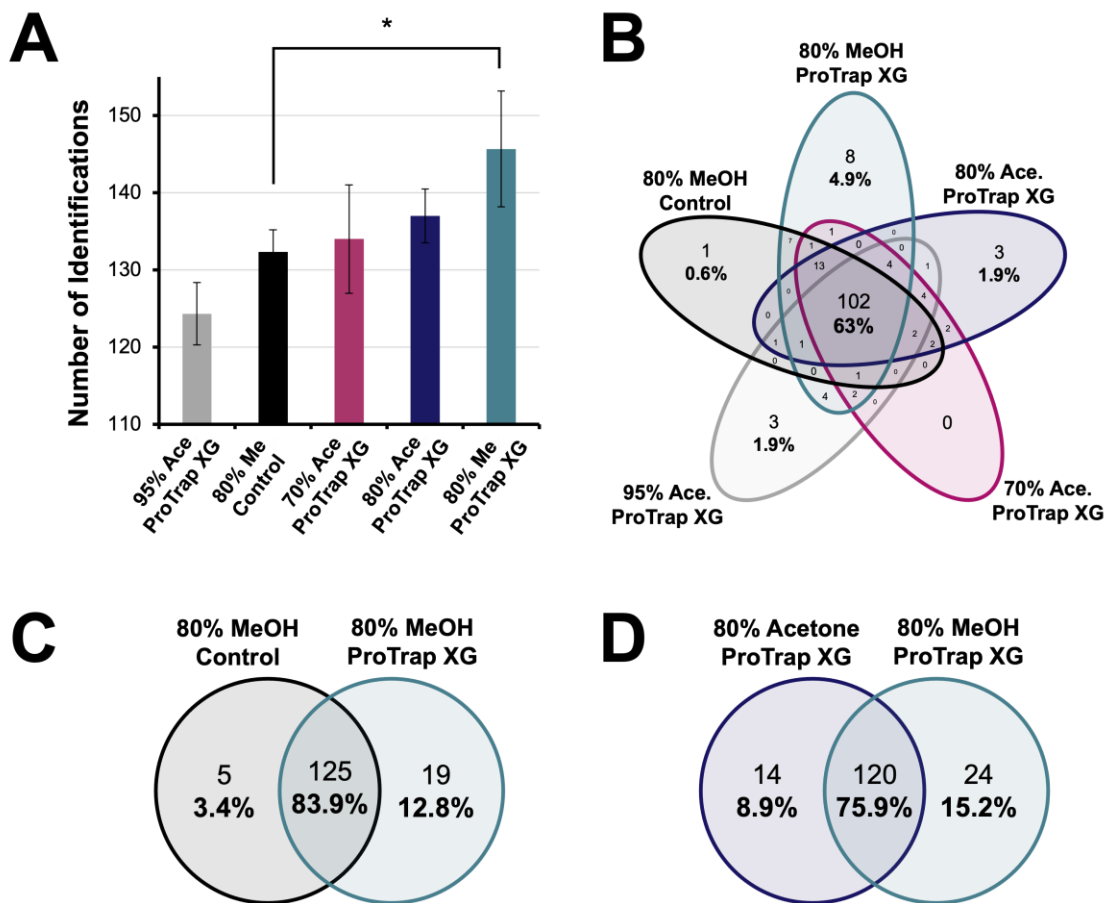
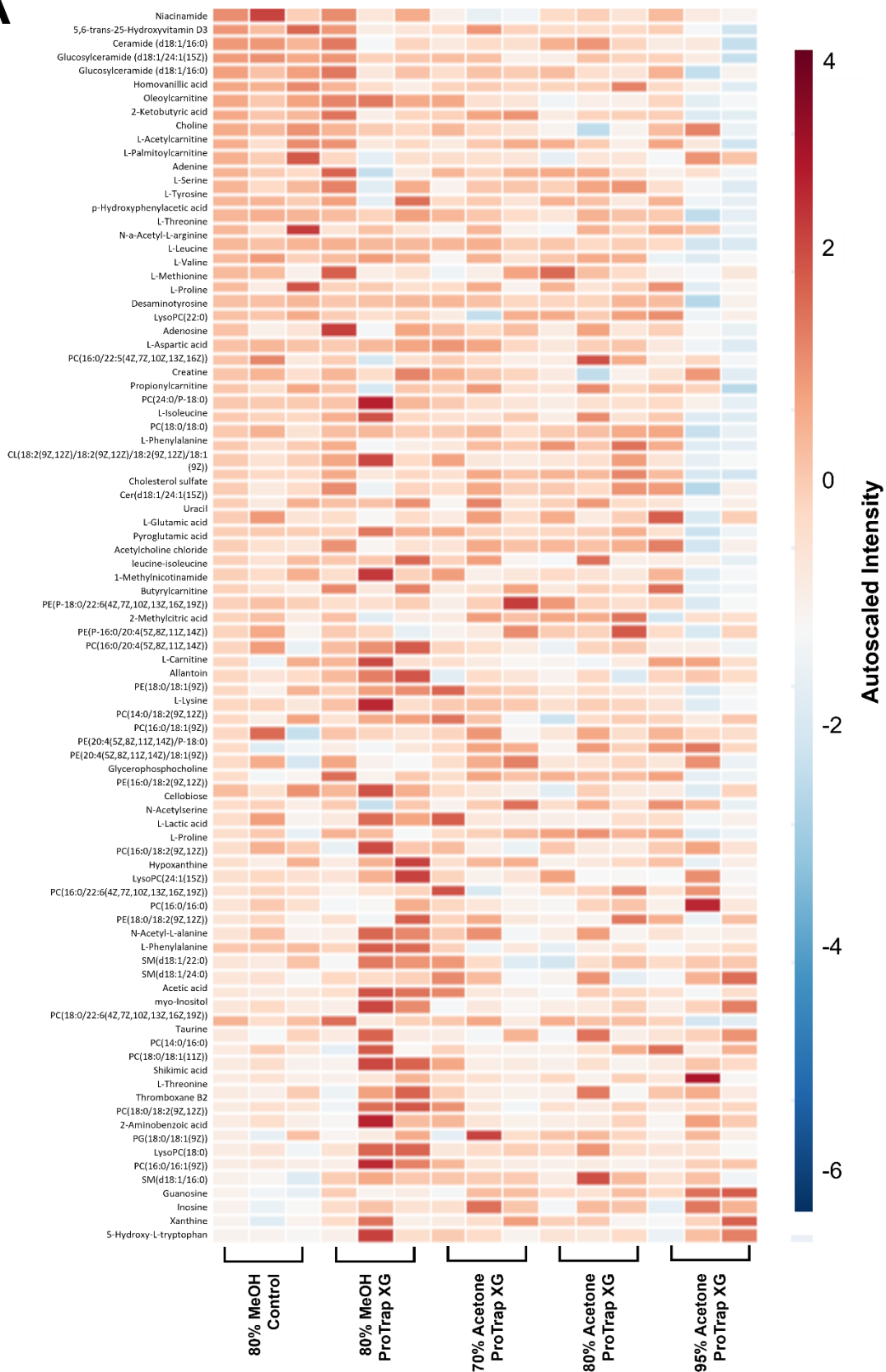


Figure 4.1 Metabolomics identifications (A) total counts from each protein depletion strategy. * $p < 0.05$. Venn diagrams comparing metabolites identified (B) by each protein depletion strategy, (C) using 80 % methanol precipitation in a conventional vial vs. the ProTrap XG, and (D) following protein precipitation with 80 % acetone vs. 80 % methanol.

The quantitative reproducibility from the metabolome fraction was subsequently evaluated across the precipitation strategies. The heatmap in Figure 4.2 demonstrates the differential metabolite intensities following various preparations in the ProTrap XG cartridge compared to the conventional approach of precipitation with 80% methanol in a standard vial. From this, it is seen that 60% of all commonly identified metabolites are detected with a greater average intensity following precipitation with 80% methanol in the ProTrap XG compared to a conventional vial. In particular, medium-polarity organic acids exhibit enhanced intensity from the ProTrap XG-based preparation. When comparing 80% acetone to 80% methanol (both in the ProTrap XG), several non-polar metabolites demonstrate increased signal from the acetone preparation, including ceramide, phosphatidylethanolamine, and phosphatidylcholines. This trend is supported

by the respective dielectric constants of acetone (20.7) and methanol (32.6). The higher dielectric constant of methanol is more likely to cause co-precipitation of highly organic metabolites compared to the less polar solvent. It is further noted that many of the metabolites that show a greater intensity following acetone precipitation compared to methanol precipitation also gain intensity when conducting methanol-based protein depletion in a conventional vial format compared to the ProTrap XG. It is speculated that some metabolites may partially co-precipitate with the protein fraction. Thus, these metabolites may be more easily recovered by directly pipetting the supernatant fraction, as opposed to passing the solution through a MWCO membrane of the ProTrap XG to isolate the supernatant. The limitations of filtration-based metabolome preparation has been discussed previously [355], namely showing a reduced recovery of hydrophobic metabolites. It is additionally noted that although acetone precipitation showed a marginal reduction in total metabolome coverage, the intensity of recovered metabolites is more homogeneous across structural characteristics, suggesting less of a bias towards the recovery of the most polar analytes.

A



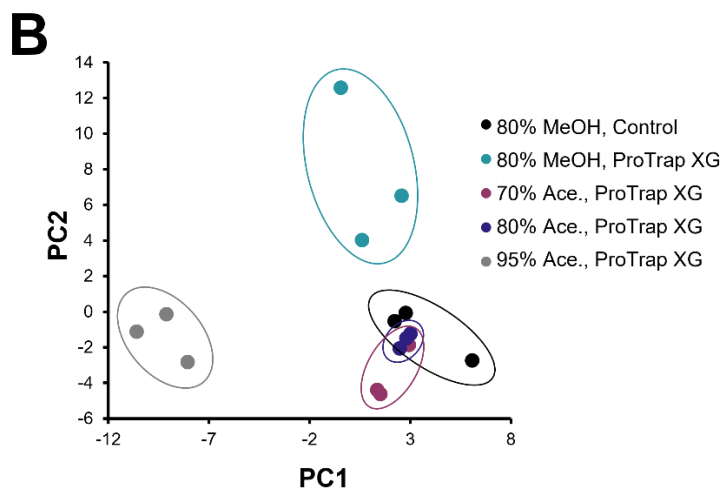


Figure 4.2 (A) Heatmap of metabolite intensities following precipitation with 80% methanol or 70-95% acetone in the ProTrap XG compared to a conventional methanol-based preparation. (B) PCA analysis of metabolite intensities across triplicate preparations.

4.3.2 Proteomics Reveals Equivalent Coverage from Acetone and Methanol Precipitation

Precipitation-based proteome sample preparation is often achieved with acetone, chloroform/ methanol/ water, or TCA solvent systems. In order to determine a single precipitation-based workflow to optimize metabolome and proteome recovery, protein pellets recovered from 80% acetone and 80% methanol in the ProTrap XG were compared. Total precipitation recovery was evaluated by an LC-UV assay, resulting in $99 \pm 7\%$ recovery from the acetone pellets and $97 \pm 6\%$ recovery from methanol, demonstrating quantitative recovery from both solvent systems. Subsequent bottom-up LC-MS/MS analysis of the digested pellets revealed equivalent total protein identifications from the acetone and methanol preparations, with an average of 5674 and 5705 respectively (Figure 4.3A). 28% of tryptic peptides were unique to the methanol preparation, however, these included 100% of the identified cysteine-containing peptides, suggesting the bias from ineffective thiol alkylation in the acetone preparations. Acetone and methanol precipitation strategies demonstrated equivalent repeatability of identifications across triplicate preparations (Figure 4.3B,C), with 93 and 94% of protein identifications in common across the respective sets of replicates. As anticipated from the similarity in protein counts, Figure 4.3D demonstrates high overlap in the proteins identified with high frequency (i.e., identified in 3 of 3 samples) between the two precipitation strategies. Gene ontology analysis was conducted on the proteins unique to acetone and methanol preparation strategies. It was found that the 136 proteins unique to the

acetone precipitation approach showed a 30% enrichment in membrane proteins, while those unique to the methanol pellet showed a 7-fold enrichment in nuclear proteins and a 70% enrichment in cellular membrane proteins, both of which are associated with more hydrophobic character than cytosolic proteins. This reflects a lower repeatability in the identification of hydrophobic proteins in both precipitation strategies, with greater coverage of these proteome fractions following precipitation with 80% methanol compared to 80% acetone.

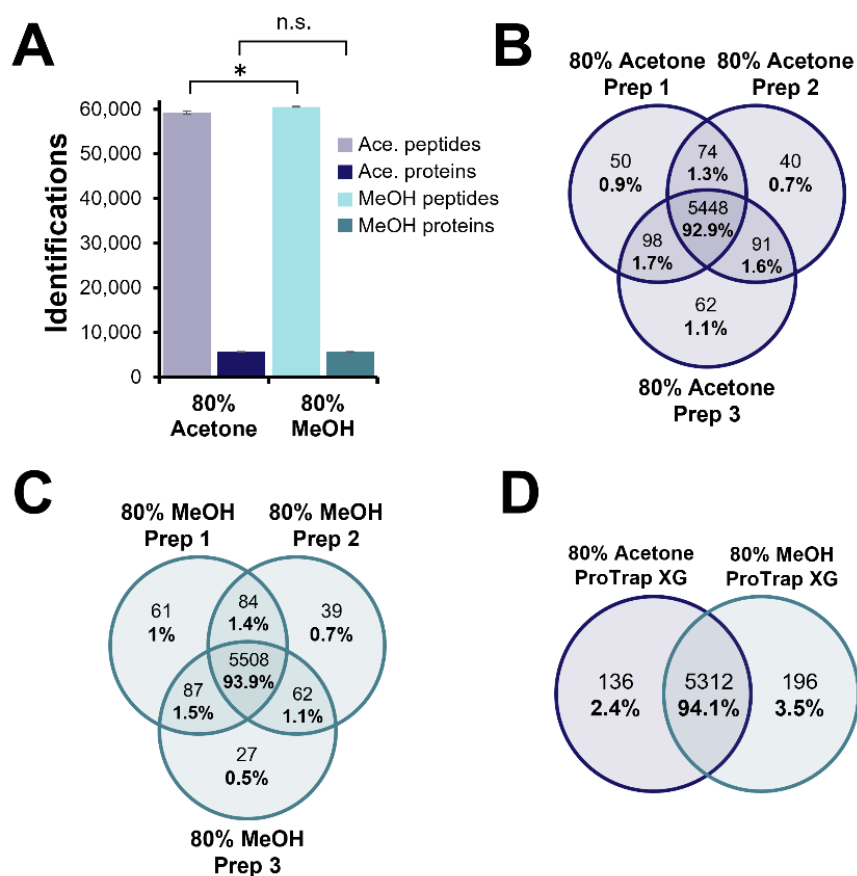


Figure 4.3 Bottom-up protein identifications from triplicate preparations using precipitation with 80% (A) acetone in the ProTrap XG, (B) methanol in the ProTrap XG and (C) comparing the most repeatable proteins between the two precipitation approaches. (D) Summary of total bottom-up peptide and protein identifications from precipitation with 80 % acetone and 80 % methanol in the ProTrap XG. * $p < 0.01$.

Comprehensive characterization of a protein fraction often includes quantitative information. The precision of bottom-up protein quantitation is limited by variable recovery as well as digestion efficiency, as discussed in Chapter 3. The strong linearity between a protein's intensity following acetone precipitation and methanol precipitation (Figure 4.4A) demonstrates

equivalent recovery of all proteins on a global scale. The distributions of coefficients of variance (CV) from label-free bottom-up protein quantitation are presented in Figure 4.4B. It is shown that both precipitation strategies result in similar quantitative precision with mean CVs of $13.5 \pm 6\%$ for acetone and $12.5 \pm 7\%$ for methanol pellets. This difference was shown to be statistically significant at the 95% confidence level, however, it does not likely have significant practical implications. It was found that 52% of proteins exhibit enhanced precision following methanol precipitation. Gene ontology analysis of these proteins demonstrated a >2-fold enrichment in mitochondrial membrane proteins and approximately 80% enrichment of proteins from the membrane of membrane-bound organelles such as exosomes. Figure 4.4C compares the quantitative repeatability of protein intensities measured from acetone vs. methanol precipitated pellets between the top 10% and bottom 10% based on intensity. As described for peptide precision in Chapter 3, the repeatability of protein quantitation across replicate preparations is influenced by the intensity of the protein. The top 500 proteins show a high degree of quantitative precision across all 6 preparations. The proteins with lower intensity exhibit reduced quantitative precision, however, this is consistent across the two preparation strategies. Likewise, Figure 4.4A shows greater agreement between the two precipitation approaches at higher intensities, with greater variance seen for less intense proteins.

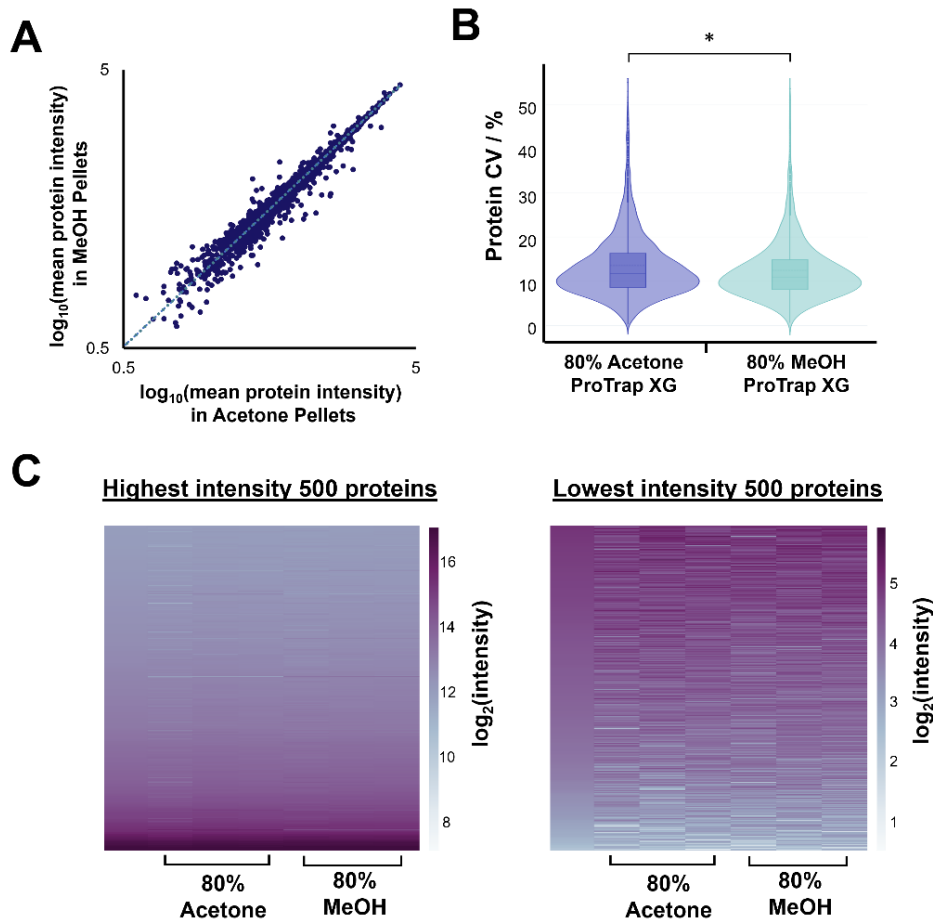


Figure 4.4 Label-free bottom-up protein quantitation analysis. (A) Linear correlation between the proteins' mean intensity measured in the acetone pellet and methanol pellet. Linear regression analysis revealed a slope of 1.035 ± 0.002 and a Pearson correlation coefficient of 0.97. (B) Violin plot comparing the distribution of average protein coefficients of variation between acetone and methanol precipitated pellets * $p < 0.05$. The most precise peptide (with an intensity $> \text{LOQ}$) associated with each protein was included. (C) Heat maps of average protein intensity across triplicate preparations using acetone and methanol precipitation for the top 500 most intense proteins compared to the bottom 500.

4.3.3 Peptidomics Analysis is Enhanced by Protein Depletion with Acetone

Following the isolation of the protein pellet from acetone and methanol precipitations, the supernatants were transferred to a clean vial for subsequent precipitation of a low-molecular weight peptide-rich fraction. The recovered pellets were analyzed without enzymatic digestion and, thus, spectra were searched with no-enzyme specificity. From Figure 4.5A, peptidomics identifications were optimized following initial protein depletion with acetone compared to methanol. Taken together, the three peptide fractions following the acetone-based preparation identified a total of 1466 peptides, >2.3 times more than were identified following initial protein depletion with methanol. The Venn diagram in Figure 4.5B shows that only 0.3% of peptides are

unique to the methanol preparations, highlighting enhanced peptidome coverage from an acetone-precipitated sample. The greater peptide identification rate in acetone preparations suggests an enhanced selectivity for molecular weight on stepping from 80% to 95% acetone compared to 80% methanol to 95% acetone. This shows a similar trend to metabolome recovery, suggesting co-precipitation of low molecular weight peptides in the initial 80% methanol solvent owing to its higher dielectric constant. Identification repeatability across acetone preparations is also enhanced over methanol, whereby 76% of peptides were identified in at least 2 of 3 preparations contrasting with 35% in the methanol precipitated samples. Figure 4.5C compared the molecular weight distribution between peptides identified in the acetone vs. methanol preparations, showing no statistical difference. Similarly, peptides were compared based on GRAVY score (Figure 4.5D), showing no bias in recovery across the two precipitation approaches based on hydrophobicity. This suggests that the difference in peptide recovery observed between acetone and methanol precipitation strategies is not biasing these structural properties, but rather representing a greater partitioning of all low molecular weight species to the supernatant following precipitation with 80% acetone, enabling their recovery in the following precipitation step. The peptides commonly identified between the two protein depletion strategies demonstrate high quantitative agreement, shown in Figure 4.5E. The correlation of average peptide intensities was found to have a slope of 0.81 ± 0.02 , showing an average of 20% greater intensity in the acetone preparation. However, >90% of the data is correlated with a slope of 1.003 ± 0.005 with a correlation coefficient of 0.98.

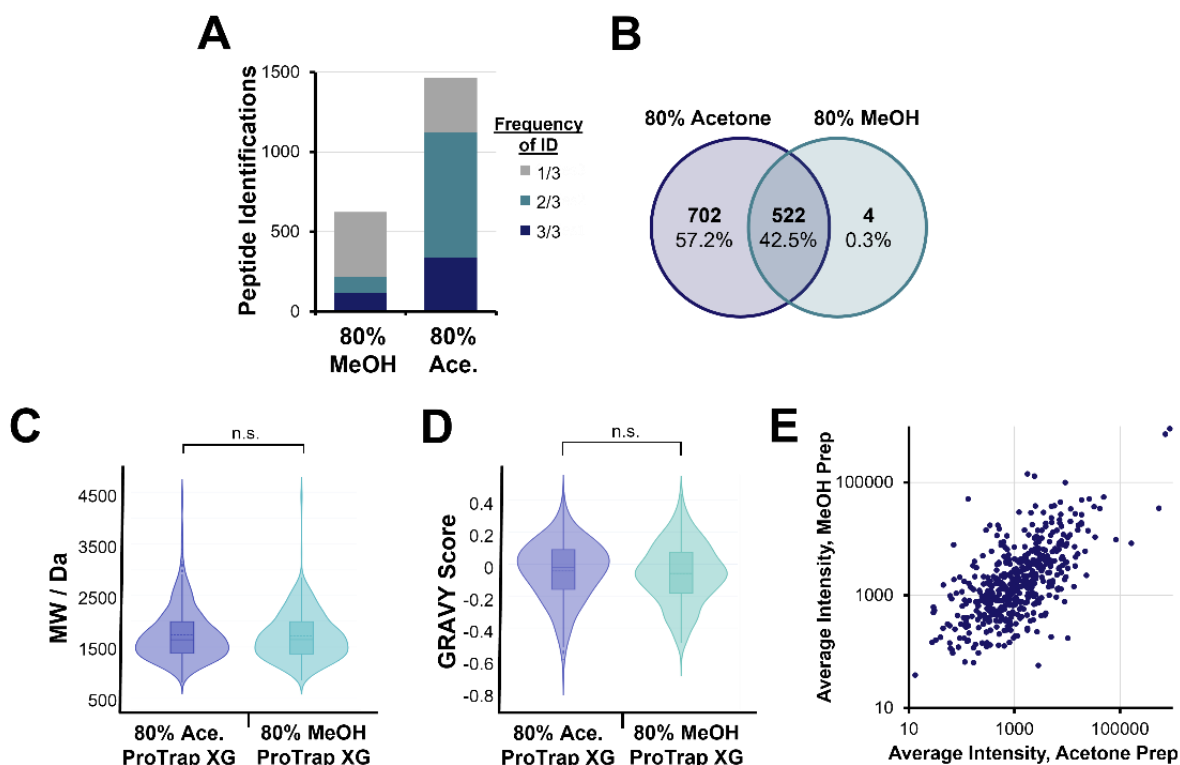


Figure 4.5 Endogenous peptidome coverage and intensity analysis of sequentially precipitated pellets following initial protein depletion with 80% acetone vs. 80% methanol. (A) Total peptide identifications sorted based on identification frequency. (B) Venn diagram of sequences identified across the two protein depletion approaches, showing 2.3-fold more identifications from the acetone preparation. (C, D) Comparison of (C) molecular weight and (D) hydrophobicity properties of peptides identified from acetone and methanol protein depletion strategies (E) Correlation of average peptide intensities between acetone and methanol-based preparations.

4.4 Discussion

Precipitation-based multi-omics workflows were evaluated starting with protein depletion with 80% acetone or 80% methanol followed by peptide precipitation at 95% organic solvent with added $ZnSO_4$. The two strategies provided equivalent proteome coverage and quantitation precision following bottom-up mass spectrometry. However, they exhibited differences in their respective coverage and quantitation of metabolome and peptidome fractions. Methanol-based metabolomics showed enhanced detection of sugar phosphates and organic acids, while the acetone precipitation approach demonstrated better recovery of medium polarity compounds such as ceramides. As described by others, it is speculated that the less polar solvent reduces the co-precipitation of less polar metabolites during the protein depletion step [344,346]. This is supported by the enhanced peptidome coverage demonstrated by the acetone-based workflow as

well. The lower identification rate of endogenous peptides following protein depletion with methanol suggests that these low-MW analytes were recovered in the proteome pellet.

In summary, proteome coverage and quantitation were each equivalent between the two strategies, while peptidomics was optimized by the acetone approach and metabolomics showed enhanced coverage from methanol precipitation. Considering the 7.5% differential in metabolome coverage compared to the 230% difference in peptidome coverage between acetone and methanol strategies, the total sample coverage is optimized by conducting initial protein precipitation with 80% acetone. However, if maximized metabolome coverage and quantitation is of greater priority for a particular application, the proteome fraction should be recovered with the more conventional approach of 80% methanol.

4.5 Conclusions

Rapid and high-efficiency precipitation strategies were applied and optimized for multi-omics analysis of the proteome, peptidome, and metabolome of a single sample. It was demonstrated that total sample coverage is optimized by a sequential precipitation strategy employing 80% acetone, followed by 95% organic solvent for recovery of endogenous low-molecular-weight peptides and retention of some of the initial supernatant for metabolome analysis. Metabolome quantitative precision was optimized by initially depleting the protein fraction with 80% acetone, which also afforded the benefit of more diverse metabolite identifications (including more non-polar analytes). The presented discussion on selectivity in acetone and methanol systems should guide the choice of precipitation strategies based on the objectives of an experiment. This high throughput approach to multi-omics preparation enables deep profiling of the proteome, endogenous peptidome, and metabolome, which shows potential for the elucidation and detection of multi-omic biomarker panels.

5. Effects of Common Denaturing Additives on Trypsin Activity and Stability

5.1 Introduction

Enzymes are highly valued biomolecules in preparative workflows, owing to their selective and catalytic reactivity. Optimizing their activity and stability is of interest across a wide range of applications including the growing field of proteomics and many industrial processes (e.g., food processing, detergent production). Michaelis Menten kinetics describe an enzyme's turnover rate as a function of substrate concentration. However, many other factors will influence an enzyme's activity and stability in solution, such as temperature, solvent composition and co-solutes. Unfortunately, the buffer composition that optimizes substrate solubility often accelerates enzyme deactivation. For example, surfactants such as SDS are typically employed to improve protein solubility, but also contribute to denaturation of enzymes including trypsin. Prior studies have examined the influence of surfactants and other additives on initial trypsin activity, while others indirectly evaluate enzyme efficiency based on the detection of select hydrolytic products. However, to date no study has quantified the differential rates of trypsin deactivation as a function of solvent additives.

Surfactants have previously been shown to enhance enzymatic reactions, offering particular benefit for enzymes that work at air/liquid or liquid/liquid interfaces [356–359]. Similarly, at appropriately low concentrations, surfactants are considered proteolytic digestion enhancers in the proteomics community based on improved sample coverage by LC-MS/MS-based proteomics [194,360,361]. However, beyond a certain concentration, enzymes are often hindered by the presence of surfactants. As reviewed by Savelli et al. in 2000, ionic surfactants can inhibit enzyme activity by either competing with the substrate-enzyme interaction, altering the chemistry of the buffer, or reducing the free substrate concentration [362]. In 2008, Ghosh et al. evaluated the effects of SDS on trypsin activity [363], showing that above the CMC, SDS micelles compromise the alpha helicity, beta sheets and random coils including those that comprise trypsin's active site. More recently, in 2020 Ma et al. conducted a molecular dynamics simulation of the interactions between trypsin and SDS [191]. It was shown that even below the CMC, SDS monomers interfere with trypsin's activity by preferentially interacting with its active site and denaturing the catalytic pocket. These molecular interactions suggest that there is no tolerable

concentration of SDS to maintain optimal enzyme activity. Nonetheless, reports of optimal (initial) trypsin activity and digestion have been described at SDS concentrations ranging from 0.01 % to 0.1 % SDS, indicating that there is no clear consensus in the field [193,364,365].

As SDS is strongly denaturing, and also interferes with LC-MS analysis, alternative surfactants have been evaluated on the basis of conservation of enzyme activity while maximizing substrate solubility. Both sodium deoxycholate [196,198] and sodium laurate [366] have been explored, offering enhanced versatility over SDS due to their phase transferability. In 1973, Blinkhorn et al. described the controlling influence of a surfactant's head group chemistry on its denaturing effects [367], which supports the adoption of sodium deoxycholate (SDC) as a more favorable additive in enzymatic applications compared to SDS. Several studies have claimed enhanced enzyme activity in the presence of SDC compared to SDS [193,194,368–370]. The surfactant is widely applied in the proteomics community to enhance bottom-up preparations of membrane proteins [192,196,209,371]. However, despite its milder denaturing effects, the surfactant has long been shown to ultimately accelerate enzyme deactivation [372,373].

Contrasting with the deactivating ionic and hydrophobic interactions between surfactants and proteins, organic solvents affect enzyme integrity by way of dehydrating the active site. It has been described that a protein's solvation shell is crucial for its biological function. By this, it is unsurprising that addition of dehydrating solvent has been shown to reduce enzyme activity [374,375]. Nonetheless, organic solvents are commonly included in bottom-up proteome preparations, being favored for their compatibility with reversed phase LC-ESI MS analysis. A 2006 study by Strader et al. claimed to optimize the digestion of low sample quantities in the presence of 80% acetonitrile [376]. However, the conclusion of a more complete digestion can not be drawn based on elevated identification rates since the products of a partial digest are more heterogeneous than complete proteolysis. Similarly, Blonder et al. reported a digestion approach to identify mammalian membrane proteins using 60% methanol [377]. This method, too, was validated based on membrane protein identifications [378,379], which does not independently reflect enzyme performance.

The present study characterizes the activity and stability of TPCK-treated trypsin in the presence of common classes of protein solution additives. Surfactants, organic solvents, and chaotropes are commonly included to optimize substrate solubility, however there are conflicting opinions on their respective implications on enzyme activity and digestion efficiency. It is shown

here that trypsin deactivation follows second-order kinetics and is variably accelerated by different surfactants, organic solvents and chaotropes. Cumulative activity is estimated across conventional trypsin digestion periods to elucidate the overall effect of solubilizing additives on the enzyme's function.

5.2 Methods

5.2.1 Initial Activity Assays

The activity of TPCK-treated trypsin (Millipore Sigma, Oakville, CA) was measured using spectroscopic assays that monitored the rate of hydrolysis of N- α -Benzoyl-L-Arginine Ethyl Ester (BAEE) (Millipore Sigma, Oakville, CA) [380] in the presence of a range of additives. The control BAEE substrate solution was prepared at 0.086 mg/mL containing 50 mM Tris-HCl (Fisher Scientific, Ottawa, CA) at pH 8. Experimental substrate solutions contained a range of concentrations of surfactant, chaotropic, and organic solvent additives. SDS was purchased from Fisher Scientific (Ottawa, CA). Sodium deoxycholate (SDC) and sodium laurate (SL) were purchased from Millipore Sigma (Oakville, CA). Urea was purchased from Bio-Rad (Hercules, USA). Guanidine hydrochloride (Gdn HCl), methanol and acetonitrile were purchased from Fisher Scientific (Ottawa, CA). In a temperature-controlled quartz cuvette, 1.5 mL of the substrate solution was combined with 1.5 μ g trypsin at 37 °C. Absorbance was monitored at 253 nm for 2-5 min with measurements taken at 15-20 s intervals. Activity was taken to be proportional to the slope of Δ absorbance/ Δ time. Relative activity was determined by the ratio of slopes between experimental conditions and the no-additive control.

5.2.2 Time Course Activity Assays

To determine enzyme stability over time, bulk trypsin samples were buffered (pH 8) with the indicated denaturants and aged at 37 °C for times ranging from 5 min to 16 h prior to combining with fresh BAEE substrate. At the indicated aging time, 1.5 μ g of enzyme was sampled for BAEE assay, and the resulting slope was normalized to that of the control (37 °C, no aging, no additives).

5.2.3. Modeling Trypsin Deactivation Kinetics

The normalized residual activity was fitted using a second-order kinetic model. Deactivation rate constants were extrapolated from the slopes of the linear correlation plot (inverse activity vs. time), and trypsin activity half-lives were calculated from Equation 5.1, where k is the deactivation rate constant, and A_o is the initial trypsin activity. The second-order models were also integrated to determine the cumulative enzyme activity across the measured time course.

$$t_{1/2} = \frac{1}{kA_o} \quad (5.1)$$

5.3 Results

Common protein additives are shown to have a strong influence on both the enzyme's initial activity as well as its stability over time. Table 5.1 summarizes the effects of select surfactants, organic solvents, and chaotropic agents on the kinetics of trypsin deactivation. Initial activity measurements were determined by BAEE assay while the deactivation rate constants were extracted as described in the previous section. Among the tested additives, it is noted that initial activity is optimized in the presence of 0.2% deoxycholate, 10% acetonitrile, or 20% methanol—all showing approximately 40 % enhancement. This is consistent with findings from a 2011 study by Wall et al. where they demonstrated optimal activity in 10 % acetonitrile [221]. Considering the influence of these additives on the stability of trypsin, it is seen that the gains in initial activity afforded by SDC are quickly lost, leading to a 10-fold drop in cumulative activity relative to the no-additive control. By contrast, optimal cumulative activity is observed with 10 mM guanidine hydrochloride, showing a 22% enhancement relative to the conventional no-additive control. These results enforce that initial enzyme activity measurements do not adequately describe the impact of solvent additives on trypsin digestion efficiency.

Table 5.1 Summary of trypsin deactivation kinetics and cumulative activity over time in the presence of select surfactants, organic solvents and chaotropic agents

Condition	Relative Initial Trypsin Activity / %	Deactivation Rate Constant, k / h^{-1}	Enzyme Half-Life / min	Relative Cumulative Activity in 16 h / %
No-Additive Control	100 ± 14	0.42 ± 0.01	140	100
0.005 % SDS	93 ± 5	1.5 ± 0.6	43	44
0.01 % SDS	114 ± 4	4.1 ± 0.2	13	21
0.02 % SDS	34 ± 4	645 ± 16	0.3	0.3
0.1 % SDC	137 ± 2	8.3 ± 0.7	5.4	13
0.2 % SDC	142 ± 0.1	12 ± 1	3.6	10
0.5 % SDC	115 ± 3	22 ± 2	1.2	6
1 % SDC	111 ± 5	36 ± 3	1.2	4
0.05 % SL	100 ± 11	4.8 ± 0.4	13	19
10 % acetonitrile	140 ± 10	0.78 ± 0.05	55	77
20 % methanol	145 ± 22	0.49 ± 0.07	84	105
1.5 M urea	104 ± 16	0.6 ± 0.2	96	86
10 mM Gdn HCl	110 ± 10	0.29 ± 0.03	190	122

5.3.1 Effects of Surfactants on Trypsin Stability

The effects of SDS and SDC on trypsin activity and stability were evaluated across a range of concentrations. From Figure 5.1.A, trypsin demonstrated optimal initial activity in the presence of 0.01% SDS, showing a $14 \pm 4\%$ enhancement relative to the control. As the SDS concentration approached the CMC, trypsin activity dropped to essentially zero. These results are in sharp contrast to previous work by Masuda et al., in which they correlate 0.01% SDS with a 3.5-fold enhancement in activity. However, their results were subject to significant misinterpretation, whereby activity was inferred from an absolute absorbance measurement, rather than the rate of substrate hydrolysis. The modest gains in initial enzyme activity at low concentrations of SDS may appear promising to enhance protein digestion. However, Figure 5.1B demonstrates the effect of the surfactant on enzyme deactivation over time. This same condition that optimized initial enzyme activity caused a 10-fold increase in deactivation rate. Trypsin deactivation was shown to obey second-order kinetics, and as seen in Figure 5.1C, appears to correlate exponentially as a function of SDS concentration. At only 0.02% SDS, trypsin is completely deactivated in just 5 min. The second-order trends were integrated to estimate the cumulative activity across a potential digestion period, shown in Figure 5.1D. The sustained activity of the control sample (no SDS) is reflected by a continuous increase in the cumulative activity across 16 h. By contrast, the cumulative activity of an overnight digest in only 0.005% SDS is equivalent to 3.5 h under conventional conditions. At 0.02%, an overnight digestion would be equivalent to only 30 s in the absence of the surfactant.

To summarize, considering the enhanced deactivation of trypsin in the presence of SDS, even trace levels of the surfactant have a detrimental effect on the enzyme, and should not be considered beneficial to enhance activity. A 1973 study by Jones et al. demonstrated complete deactivation of ribonuclease A in 10 times less surfactant (0.001%) [381], which contrasts with the modest deactivation observed here in 0.005% SDS. This suggests an individualized effect of the surfactant on different enzymes, which is speculated to be influenced by the secondary and tertiary structures that form a given enzyme's active site. The influence of the surfactant on improving protein solubility and substrate unfolding, to afford greater access of the enzyme to cleavage sites, may still provide a benefit. Such effects are proposed to be explored in future work (described in Chapter 7).

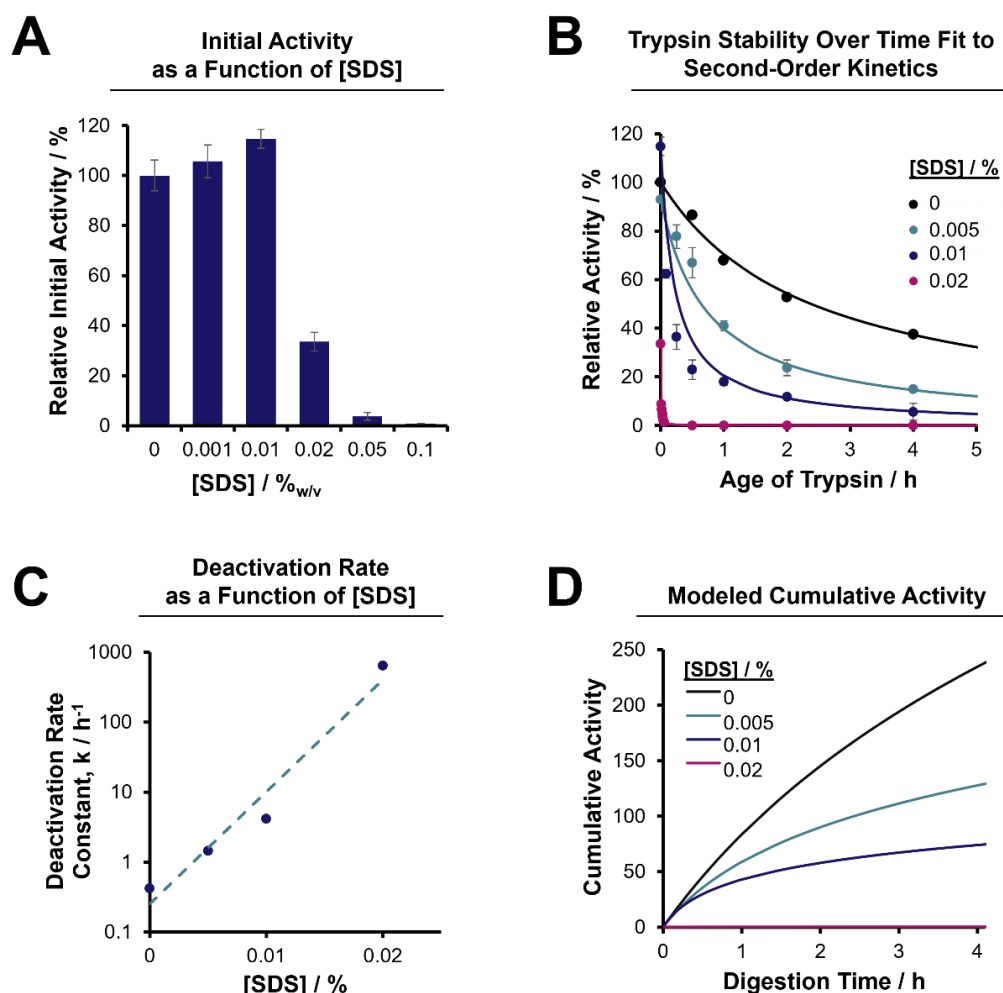


Figure 5.1 Trypsin (A) initial activity and (B) stability over time measured by BAEE assays of variably aged enzyme. (C) Exponential correlation between deactivation rate constant and SDS concentration. (D) Estimated cumulative activity as a function of SDS concentration extracted from the integral of stability curves in (B).

Sodium deoxycholate (SDC) has proven to be advantageous over SDS for protein processing due to its phase transferability and enhanced enzyme compatibility [382]. Initial trypsin activity was evaluated in the presence of 0.1-2.0% SDC, shown in Figure 5.2.A. Compared to SDS, trypsin demonstrates a tolerance for higher concentrations of SDC, with 0.2% SDC offering $50 \pm 10\%$ increase in the initial activity. This result supports the literature observations that SDC is a favorable surfactant for enhancing enzyme activity. Furthermore, maximizing activity at or near the surfactant's CMC (200× more concentrated vs. 0.01% SDS) provides greater opportunity for the surfactant to solubilize the substrate protein. Even at 1-2% SDC, initial activity remains higher than the no-additive control. The current evaluation of initial trypsin activity in SDC differs from Masuda's results where they report 0.01% SDC to optimally enhance initial trypsin activity 5-fold relative to the absence of surfactant [193]. By comparison, a 2008 study by Lin et al. reported 77% relative activity in the presence of 10% SDC [194], however, like Masuda, their conclusions are not fully supported by their activity assays. The rate of hydrolysis was shown to decrease throughout the course of their 5 min assay, which they attributed to the reduced substrate concentration. However, this conclusion goes against the principles of the kinetic assay. It is more likely that the reduction in hydrolysis rate over time is evidence of reduced enzyme activity. In 1973, Jones et al. characterized the influence of a surfactant's head group chemistry on its denaturing effects [367]. They describe that surfactants with bulkier head groups show reduced interaction with an enzyme's active site due to steric hindrance. By this, the bulkier structure of SDC may not directly compete with enzyme-substrate interactions as is shown for SDS [191].

Considering the influence of SDC on enzyme deactivation over time, trends similar to SDS were observed. Figure 5.2B shows that trypsin deactivation in the presence of SDC follows second-order kinetics, with <10% relative activity remaining after 2 h with added SDC. However, while only 0.02% SDS is shown to deactivate the enzyme in seconds, trypsin sustains measurable activity for 15 min in the presence of 1% SDC. The deactivation rate constants are plotted in Figure 5.2C, showing a linear correlation with SDC concentration. Compared to the exponential trend in deactivation rates with SDS (Figure 5.1C), this linear relationship further demonstrates an elevated tolerance to SDC relative to SDS. Figure 5.2D illustrates the cumulative activity estimated from integrating the second-order stability curves. In the presence of 0.1-1% SDC, an overnight

incubation offers <10% of the total activity from a conventional digest, which may be sufficient in particular settings where complete digestion is not critical [194,383,384].

Figure 5.2E emphasizes the brevity of initial enhancements from SDC. Here it is shown that the initial enhancements from 0.1-0.2% SDC are only sustained up to approximately 5 and 4 min, respectively. Following longer incubations, the greater stability in the absence of surfactant affords the control sample greater cumulative activity. While the increased stability of trypsin in SDC compared to SDS seems promising, Bogdanova et al. demonstrated that its milder amphiphilicity and rigid structure is associated with up to a 10-fold reduction in solubilization capacity [385], which mitigates the benefits of including a surfactant for substrate solubility.

In summary, these two common surfactants are shown to optimize initial activity at a critical concentration. SDC offers the greatest enhancement factor compared to a control, and at a much higher concentration than optimal SDS-containing conditions. Further analysis of the enzyme's stability in either surfactant revealed that the apparent initial benefits are quickly lost due to accelerated deactivation. SDS was proven to be a much more potent inhibitor than SDC, which is anticipated to elicit greater consequences in practical applications.

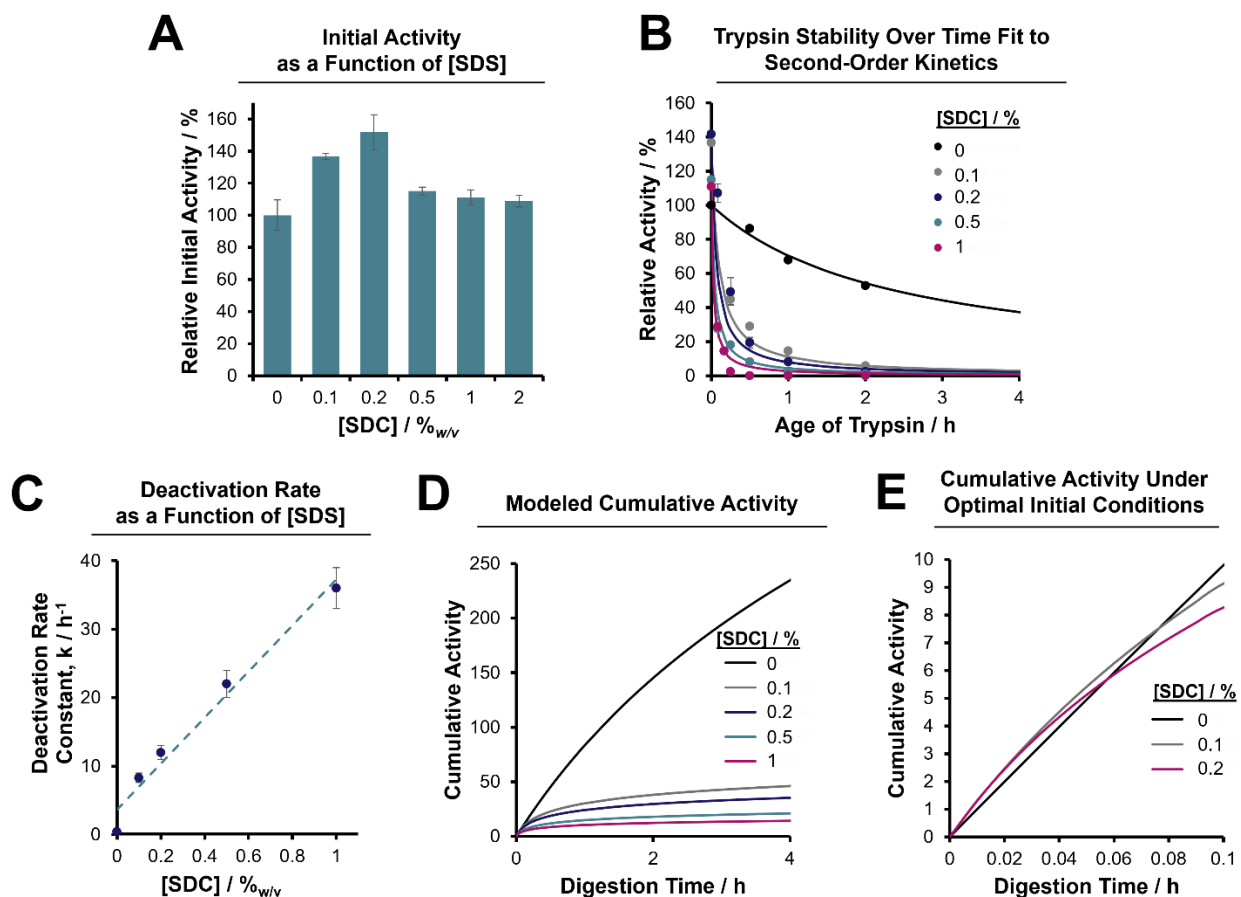


Figure 5.2 Trypsin (A) initial activity and (B) stability over time measured by BAEE assays of variably aged enzyme. (C) Linear correlation between deactivation rate constant and SDC concentration. (D) Estimated cumulative activity as a function of SDS concentration extracted from the integral of stability curves in (B). (E) Comparing the short-term cumulative activity under optimal initial conditions.

Sodium laurate was reported by Lin et al. in 2013 as a viable alternative for proteomics preparations [366]. Similar to SDC, sodium laurate can be removed by phase transfer, facilitating subsequent LC-MS analysis. The authors reported superior trypsin compatibility and bottom-up membrane protein characterization using SL over SDS and SDC, which they attribute to its structure, which seemingly optimizes the solubilization capacity of a 12-carbon chain tail of SDS with the less-denaturing carboxylic acid head group of SDC. Their initial activity assays demonstrate a 30% enhancement in activity in the presence of 0.1% SL compared to the control, while 1% SL reduced activity by about 24%. These results conflict with theirs somewhat, whereby statistically similar initial activities were observed in 0-0.05% SL, with greater concentrations showing deactivation. It was subsequently demonstrated that the equivalent initial activity seen at

0.05% surfactant was associated with accelerated deactivation by more than 10-fold. Nonetheless, this represents enhanced enzyme compatibility compared to SDS, with comparable deactivation rates to SDC.

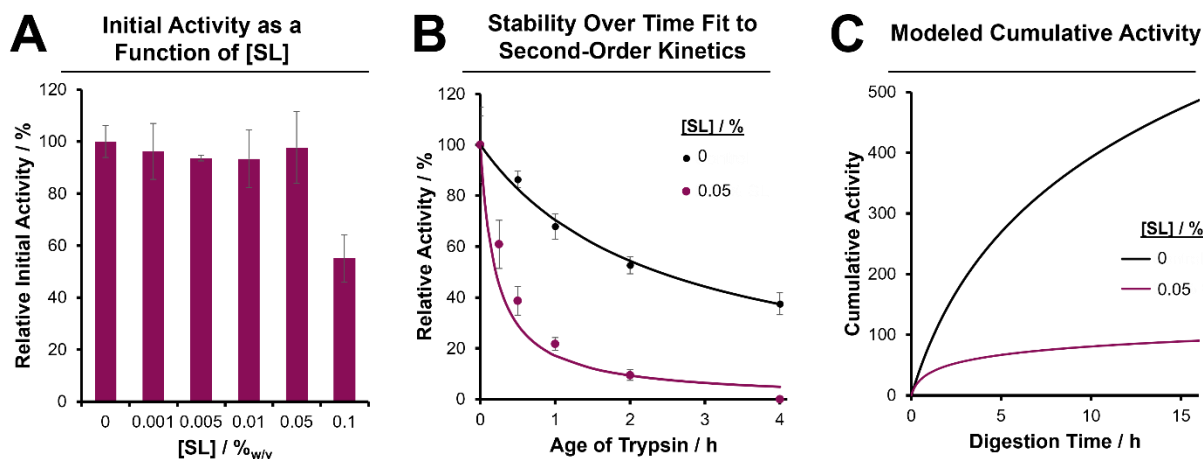


Figure 5.3 Trypsin activity measured by BAEE assays in the presence of sodium laurate (SL), (A) initially, and (B) across a 4 h time course fit to second-order kinetics trends. (C) Cumulative activity was estimated by integrating second-order stability trends across an overnight digestion period.

5.3.2 Effects of Chaotropes on Trypsin Stability

Chaotropes are valued in protein processing for their salting-in properties, thus they too have been extensively studied with respect to their influence on enzyme behaviour. Urea induces denaturation by disrupting critical H-bonding within an enzyme's active site. It was shown by Rajagopalan et al. in 1961 that urea-induced denaturation of some enzymes is reversible upon dilution [386], enabling protein solubilization at high chaotrope concentrations with subsequent enzymatic digestion under milder conditions. Gdn HCl, by contrast, reduces enzyme activity at much lower concentrations (e.g., 20 mM) [142,387]. In 1993, Ghatge et al. showed that Gdn HCl causes denaturation by preferentially interacting with carboxy groups [388]. This specific interaction suggests that the influence of Gdn HCl on a given enzyme's activity will depend on the role of carboxy groups in the enzyme's catalytic mechanism. The authors also discuss that Gdn HCl does not readily induce conformational changes, which may better conserve selectivity compared to other additives. Figure 5.4.A compares initial trypsin activity across 0.01-8 M urea and guanidine HCl, showing consistent drops in activity with 0.1-2 M guanidine HCl and an apparent tolerance up to 1.5 M urea. 0.01 M Gdn HCl provides $10 \pm 9\%$ enhancement in initial

activity, though it is shown to be statistically insignificant by an unpaired t-test. 0.5-1.5 M urea allows equivalent initial activity relative to the control, with a subsequent drop to 10% relative activity in the presence of 8 M urea. Figure 5.4B shows the enzyme stability in the presence of the maximum concentration that conserved 100% initial activity. In sharp contrast to the results seen with each of the surfactants, trypsin demonstrates marginally enhanced cumulative activity in the presence of 10 mM Gdn HCl. For example, at 4 h incubation, 10 mM Gdn HCl exhibits 15% more residual activity than the control. On the other hand, the addition of 1.5 M urea will accelerate enzyme deactivation, albeit modestly relative to the rates seen with added surfactants. Following a 30 min incubation, the urea-containing sample demonstrated 40% less residual activity than the control, however, the two samples exhibit <10% differential activity by the 4 h point.

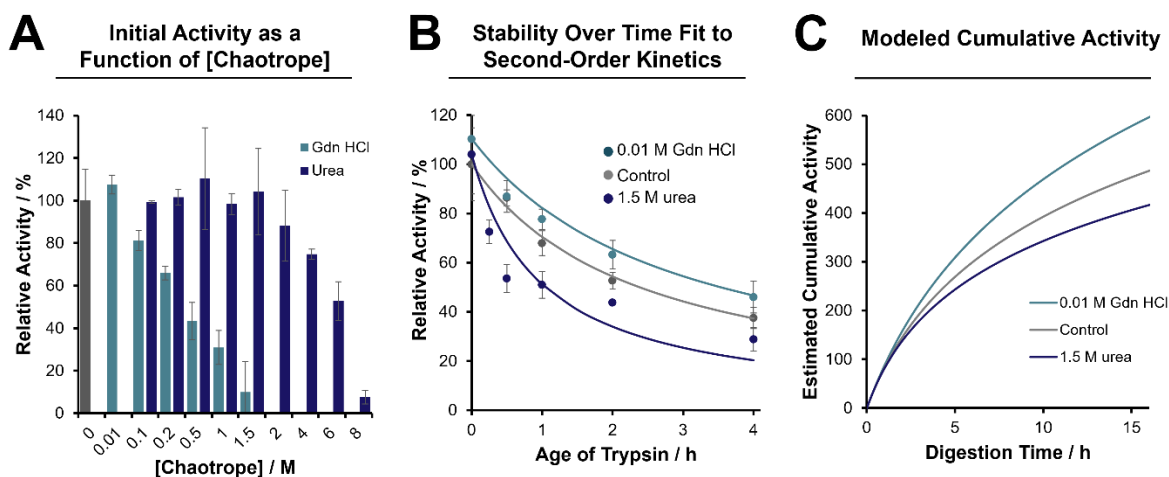


Figure 5.4 Trypsin activity measured by BAEE assays in the presence of chaotropes, urea and GdnHCl (A) initially, and (B) across a 4 h time course. (C) Cumulative activity was estimated by integrating second-order stability trends across an overnight digestion period.

The evaluated chaotropes (urea and guanidine hydrochloride) demonstrated significantly different enzyme compatibilities, which is supported by the different structural moieties that they respectively disrupt (i.e., urea reversibly disrupts Hydrogen bonding, guanidine interacts with polar residues). The conclusion that 100% initial relative activity is conserved in up to 2 M urea is supported by the longstanding recommendation in proteomics to dilute an 8 M sample at least 5-fold prior to adding trypsin. In 1.5 M urea, trypsin deactivation was shown to accelerate by approximately 40%, with the fastest deactivation observed in the first hour. Residual activity measured after 4 h was just 10 % lower than that of the control. However, the earlier deactivation

is reflected by an overall reduction in cumulative activity across the time course—albeit by only 14% at the 16 h point, which does not likely have practical implications. Similar to these findings, a 1956 study by Perlmann et al. evaluated the effects of urea on pepsin stability. They demonstrated equivalent deactivation to the control in 2 M urea, while 4-8 M conditions were increasingly deactivating. Also in 1956, Léonis et al. evaluated the reversibility of urea denaturation of xanthine oxidase. It was shown that 1-2 M urea induces approximately 40% initial inhibition, however, the denaturing effects of a 60 min incubation in up to 6 M urea is quantitatively reversible upon dilution [389]. A 1977 study by Gomez et al. further demonstrated that renaturation following urea treatment is enhanced by added calcium ions [390]. Calcium has been extensively characterized for its enhancement of trypsin activity and stability [252,253,391,392]. It has been shown to increase the rigidity at the active site, which may reduce unfolding that could otherwise be imparted by urea's disruption of H-bonding that also contributes to conservation of the active site conformation.

Compared to urea, guanidine hydrochloride induced deactivation at much lower concentrations, which is in agreement with Léonis' quantification of a 33-fold greater inhibition constant for Gdn HCl than urea for xanthine oxidase [389]. Initial activity was equivalent to the control at 10 mM Gdn HCl but decreased logarithmically as the concentration was increased further. This shows reasonable agreement with a 1964 study by Inagami et al., where they demonstrated a reduction in trypsin activity in 20 mM Gdn HCl [142]. Stability at the low concentration of 10 mM Gdn HCl was within 30% of the control. Estimated cumulative activity is apparently optimized in the presence of the low-concentration chaotrope, however, the kinetic model is amplifying the difference in initial activity which was determined by a t-test to not be significant at the 95% confidence level ($p > 0.3$). Other enzymes may demonstrate a greater tolerance to guanidine salts than trypsin due to the crucial role of polar residues, Asp and Ser in its catalytic triad. A 1996 study by Fan et al. showed that in 2 M Gdn HCl, dihydrofolate reductase was only deactivated after 1.5 h [387].

5.3.3 Effects of Organic Solvents on Trypsin Stability

Figure 5.5.A compares initial trypsin activity in the presence of 5-50% acetonitrile and methanol. From this, the optimal acetonitrile concentration appears to be 10%, offering 30% enhanced initial activity relative to the control. This is consistent with a 1994 study by Batra and

Gupta where they showed a 44% enhancement in activity in the presence of 10% acetonitrile [393]. By contrast, Masuda reported optimal initial activity in 20% acetonitrile, ethanol, and propanol while methanol content was optimized at 40% [193]. Raising acetonitrile content to 20% still shows 25% greater initial activity relative to the control. These results show strong contrast with a 1998 study by Simon et al. which demonstrated conserved initial activity in up to 80 % acetonitrile [394]. The differences may be attributed to their inclusion of 19 mM calcium ion and conducting the activity assays at 25 °C rather than 37 °C which was used in the current study. Compared to any measurements in acetonitrile, a greater enhancement is seen in the presence of 20% methanol, showing a 50% increase relative to conventional conditions. Raising the methanol content to 30% still shows a 40% enhancement, while 30 % acetonitrile reduces activity by 70%. The initial activity in methanol quantified in the current study shows more modest enhancements than the 390% relative activity reported by Masuda (again, from questionable interpretations of their data) [193]. A 1998 study by Park et al. distinguished the effects contributed by methanol concentration and dielectric constant on Michaelis Menten parameters of trypsin-catalyzed hydrolysis [395]. They found that below 30%, the nucleophilicity of the organic solvent raises k_{cat} while the reduced dielectric constant raises K_M . In excess of 30% methanol, they attribute reduced activity to a disruption of enzyme-substrate binding. A 2007 study by Zhang et al. compared the proteome digestion efficiency in 60% methanol to 1% SDS [188]. The authors demonstrate increased identification rates and membrane proteome coverage using 60% methanol over the SDS strategy, which shows agreement with the relative deactivation shown here from these two denaturants.

The solvent compositions that optimized initial activity were evaluated across a time course. As shown in Figure 5.5B, 10 % acetonitrile was shown to reduce the enzyme's stability, with an approximate 2-fold increase in the rate of deactivation compared to the control. 20% methanol also demonstrated accelerated deactivation, however only by 17%, enabling sustained greater activity compared to the control across the 4-hour time course. From Figure 5.5C, cumulative activity is optimized in 20% methanol for up to 48 h. 10% acetonitrile provides enhanced cumulative activity up to 2.5 h, after which the greater stability of the control provides higher total activity. A 2011 study by Wall et al. demonstrated the implications of reduced digestion completion by including highly organic solvent systems [221], while Hervey et al. demonstrated reduced cleavage specificity [218]. Proteins may still be identified by bottom-up

mass spectrometry, however, the abundance fully-cleaved tryptic digestion products is reduced, compromising quantitation accuracy and raising the limit of detection of low-abundance species.

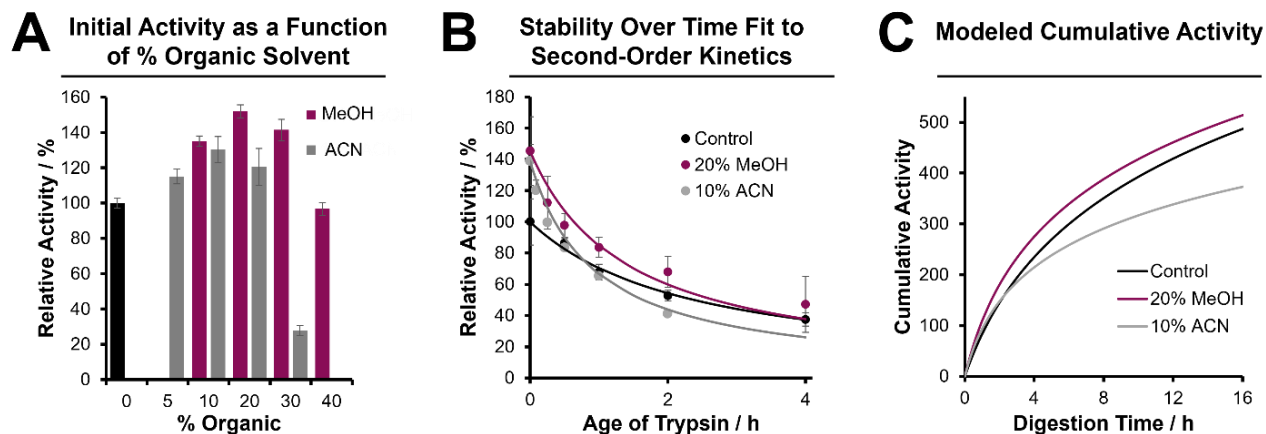


Figure 5.5 Trypsin activity measured by BAEE assays in the presence of organic solvents, methanol and acetonitrile (A) initially, and (B) across a 4 h time course. (C) Cumulative activity was estimated by integrating second-order stability trends across an overnight digestion period.

5.4 Discussion

The rate of trypsin hydrolysis has a controlling influence on throughput of proteomics preparations and industrial enzymatic applications. Many studies propose optimal catalytic conditions either based on results that confound the effects of several variables (e.g., substrate solubility and enzyme activity) [209,369] or are based solely on initial enzyme activity [193,194]. The present study aimed to systematically evaluate a diverse set of common additives for protein processing applications to determine the concentration at which initial activity is optimized and subsequently evaluate the effect of the additive on stability over time. It is anticipated that optimal enzymatic reaction conditions can only be predicted based on the combined effects of initial activity enhancements and deactivation rate.

Figure 5.6 summarizes the relative cumulative activities determined for each of the evaluated denaturing conditions. Taken together, it was found that even trace levels of SDS reduce the overall catalytic potential of the enzyme across a practical digestion period. This further enforces the need for protein purification strategies (such as acetone precipitation in the ProTrap XG) to eliminate SDS ahead of subsequent sample processing. The adoption of SDC for membrane

proteome preparations suggests that trypsin is tolerant at $\geq 0.1\%$, albeit with a significant 10-fold reduction in overnight cumulative activity. This could readily be compensated by undertaking digestion at a higher substrate-to-enzyme ratio (e.g., 10:1 instead of 50:1). It is noted that among all the evaluated additives, trypsin deactivation in the presence of 1.5 M urea demonstrates a weaker correlation with the second-order kinetics model. The slower deactivation supports the inclusion of 1.5 M urea for proteome digestion and may be attributed to conserved cleavage activity despite a disruption of the H-bonding network at the enzyme's active site. In a complex proteome digest, cleavage specificity may be more compromised than what is reflected by the BAAE assay. Of all the evaluated conditions, initial as well as cumulative enzyme activity is conserved in the presence of 10 mM Gdn HCl and marginally enhanced by inclusion of 20% methanol. However, 10 mM Gdn HCl is below the level typically required for this chaotrope to provide a meaningful benefit in terms of protein solubilization. This therefore favors the use of methanol to afford practical gains in solubility as well as digestion efficiency. Contrasting with the purposeful addition of surfactants to enhance bottom-up workflows, it is speculated that any gains in substrate solubility will be proportional to the loss in enzyme activity, meaning the true benefits of surfactant-assisted digestion processes remains to be properly evaluated. For example, in cases where substrate solubility is the limiting factor in an enzymatic process, modest reductions in catalytic rates may have negligible practical implications. Alternatively, one could compensate the reduced enzyme activity through longer incubation or increased enzyme-to-substrate ratio, in cases where surfactants are necessary to solubilize the proteome.

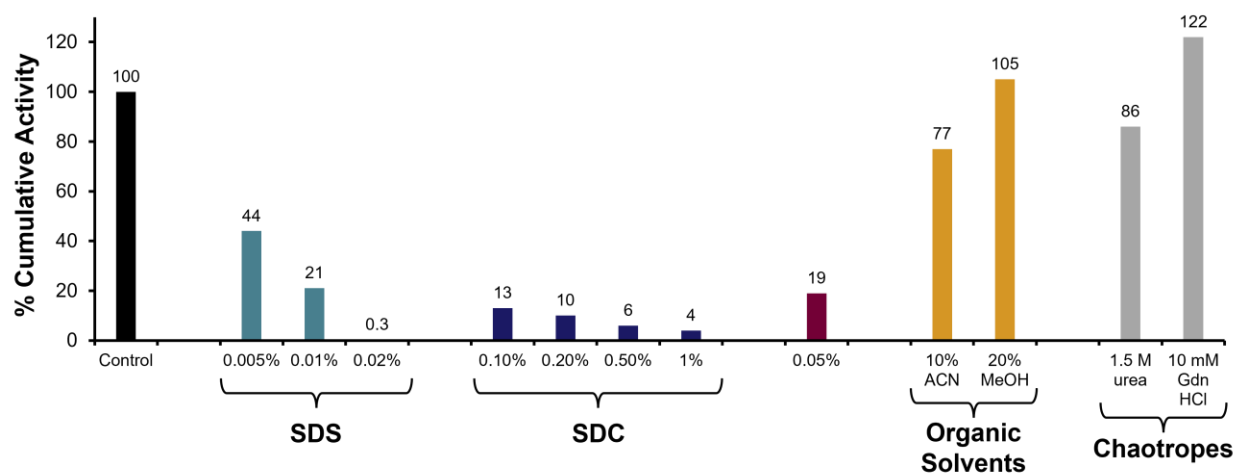


Figure 5.6 Summary of relative cumulative trypsin activities determined for all evaluated denaturing conditions across a 16-h digest.

5.5 Conclusions and Future Work

The effects of a variety of denaturants on trypsin activity and stability were quantified. It was demonstrated that enzyme activity is temporarily enhanced at critical concentrations of each denaturant, however increased rates of deactivation are reflected by reductions in total enzyme function across a digestion period. The evaluated denaturants are often reported to enhance MS characterization of membrane-rich proteome samples, which is attributed to enhanced substrate solubility at the expense of enzyme activity. From the present work, the inclusion of practical concentrations of denaturants is correlated with lower enzyme efficiency. Considering the relevance of membrane protein characterization and accurate quantitation, future investigations will evaluate the implications of denaturing additives for bottom-up proteome preparations.

6. Maximizing Cumulative Trypsin Activity with Calcium at Elevated Temperature for Enhanced Bottom-Up Proteome Analysis[‡]

6.1 Introduction

Improved characterization of complex proteomic systems using enhanced MS instrumentation demands equally robust front-end workflows to fully capitalize on gains in sensitivity and throughput. For bottom-up proteomics, the trypsin digestion step [138] is integral to sample processing, though conventional enzyme digestion (37 °C, overnight, typically at a 50:1 protein-to-enzyme ratio) remains a bottleneck to maximize throughput. While adopting accelerated sample processing techniques is a shared priority in the field, protein digestion time cannot be optimized in isolation of other factors, including the efficiency of trypsin digestion, which directly influences quantitation accuracy, and analytical precision.

Multiple groups have reported enzyme digestion strategies to address sample processing throughput. Among these are the incorporation of chemically modified trypsin, which lowers autolysis [163,230,396], the use of high enzyme-to-substrate ratios including by way of immobilized enzyme reactors (IMER) [397–399], digestion at elevated temperatures or pressure [165,170], imparting energy by microwave radiation or ultrasonication [171,177,181,400,401], or the inclusion of sample additives such as organic solvents or surfactants [216,360,402]. These protocols have been employed to reduce the trypsin digestion time from hours to minutes, or even seconds. Recently, Zare's group disclosed a sub-microsecond digestion strategy, enabled by enhanced trypsin kinetics within micron-sized droplets generated by electrosonic spray ionization [403]. As a metric of the accelerated digestion protocol, Zare's study highlighted the improved protein sequence coverage afforded by the modified digestion strategy. However, it is stressed that the commonly cited result of generating more MS-detectable peptides does not fully reflect the extent of sample digestion, particularly when derived from single or simple protein mixtures. Perhaps counter-intuitively, improved sequence coverage can be obtained through incomplete digestion [184,221]. As shown by Cannon et al. in 2010, a middle-down approach (i.e., partial digestion) provides high sequence coverage of a ribosomal proteome [404]. Viewed from this

[‡] This chapter is based on the published article: Nickerson, Jessica L., & Doucette, Alan A. "Maximizing Cumulative Trypsin Activity with Calcium at Elevated Temperature for Enhanced Bottom-Up Proteome Analysis" *Biology*, 11-10, 2022, 1444. DOI: 10.3390/biology11101444

perspective, elevated temperatures are also known to enhance trypsin deactivation [166,238,253,405–407], while even modest levels of organic solvents can induce protein (and enzyme) precipitation, contributing biased sample loss [408]. While surfactant-assisted digestion protocols exist (e.g., SDS, SDC), these detergents are known to significantly inhibit enzyme activity above threshold levels [194,209], and pose further challenges by their incompatibility with downstream processing [315,409]. Thus, despite the extensive options available, there is limited evidence to suggest these approaches offer a more effective digestion of the sample. Other critical parameters to assess include cleavage efficiency and enzyme specificity, each being relevant to modern proteome investigations.

The spectroscopic assessment of trypsin activity provides a preliminary estimate of digestion efficiency under a controlled set of conditions. A more complete understanding is provided by determining cumulative (integrated) trypsin activity over the full digestion period. To illustrate, while multiple researchers have noted enhanced enzyme activity at elevated temperatures [238–240], these gains will be tempered by accelerated enzyme deactivation [163,231], greater thermal aggregation of proteins [166], the potential for accelerated side modifications [241,242,244,247,410,411], and elevated chymotrypsin activity [245,412]. The ideal temperature for accelerated trypsin digestion presents a balance of enhanced initial activity with sustained trypsin stability over the duration of the incubation.

The influence of calcium ions on enhanced tryptic activity and preserve thermal stability has been known since the early 1900s [250,252,253,413,414], though surprisingly, the inclusion of calcium ions is not widely exploited across the proteomics community. The mechanism of calcium's stabilizing effects was introduced by Gorini in 1951 [250] and expanded by Green and Neurath [251], whereby it was demonstrated that calcium decreases the rate of autolysis in both active and inactive forms of trypsin. A series of follow-up studies explored the effects of calcium on autolysis rates, conformational changes, and temperature dependence of trypsin-calcium interactions [253,391,413,415].

Given the extensive body of evidence to support a more active, stable, and selective enzyme, it was an objective to incorporate calcium-stabilized trypsin in an accelerated, high-temperature digestion workflow. The goal of this work is to establish a routine and robust digestion protocol for accelerated proteolysis, delivering high specificity, and a high degree of cleavage which would in turn deliver high consistency. Through an optimized accelerated digestion

protocol, the cleavage efficiency is assessed for a complex proteome by quantifying the extent of fully-cleaved, fully tryptic and unmodified peptides, relative to those produced through a conventional digest. The merits of this robust, accelerated digestion workflow are discussed.

6.2 Materials and Methods

6.2.1. Trypsin Activity and Stability Assays

Initial trypsin activity was measured using BAEE spectroscopic assays [51]. Briefly, 0.86 g/L BAEE substrate (Millipore Sigma, Oakville, ON, Canada) was combined with Tris (Thermo Fisher Scientific, Whitby, ON, Canada) adjusted to pH 8.0 and 1 μ g TPCK-treated trypsin (cat. num. T1426, Millipore Sigma) in a temperature-controlled cuvette with or without calcium chloride (Millipore Sigma). The hydrolysis of BAEE was monitored by timed absorbance measurements at 253 nm across 2-3 min. The resulting slope (Δ absorbance/ Δ time) was proportional to the enzyme's activity. To determine enzyme stability over time, trypsin samples were buffered (pH 8) with or without calcium and aged at 37-67 $^{\circ}$ C for times ranging from 5 min to 16 h prior to combining with fresh BAEE substrate. At the indicated aging time, 1 μ g of enzyme was sampled for BAEE assay and the resulting slope was normalized to that of the control (37 $^{\circ}$ C, no aging, no added calcium chloride). Control BAEE assays performed in the absence of trypsin confirmed the stability of the substrate at elevated temperatures.

6.2.2. Modeling Trypsin Deactivation Kinetics

The normalized residual activity was fitted using a second-order kinetic model. Deactivation rate constants were extrapolated from the slopes of the linear correlation plot (inverse activity vs. time), and trypsin activity half-lives were calculated from Equation (6.1), where k is the deactivation rate constant, and A_o is the initial trypsin activity. The second-order models were also integrated to determine the cumulative enzyme activity across the measured time course.

$$t_{1/2} = \frac{1}{k A_o} \quad (6.1)$$

6.2.3. Bottom-Up Proteome Sample Preparation

S. cerevisiae was grown overnight at 30 °C in YPD broth (Millipore Sigma) to an OD600 of 1.0, as previously described [295]. Cell pellets were harvested by centrifugation, washed twice with distilled water, and lysed by grinding under liquid nitrogen. The lysate was subject to proteome extraction by boiling in 2% (w/v) SDS (Thermo Fisher Scientific) for 1 h. The resulting extract was isolated from residual cell debris by centrifugation. The total protein concentration was determined by a BCA assay (Thermo Fisher Scientific) and adjusted to 2 mg/mL by dilution with water. Aliquots of yeast proteome extract (100 µg) were subject to protein precipitation as previously described [283], by combining with 50 mM sodium chloride (Millipore Sigma) and 80% (v/v) acetone in the ProTrap XG filtration cartridge (Allumiq, Halifax, NS, Canada). Following 2-5 min incubation with gentle mixing, protein pellets were isolated from the supernatant by centrifugation, washed with 400 µL additional acetone, and re-solubilized overnight in 8 M urea (Bio-Rad, Mississauga, ON, Canada). Solubilized proteins were diluted to 1.5 M urea with triethyl ammonium bicarbonate (TEAB) buffer (pH 8.0) (Millipore Sigma), and subject to reduction and alkylation with 5 mM DTT (Bio-Rad) and 11 mM IAA (Millipore Sigma). Trypsin digestion was performed at a 25:1 mass ratio (protein to trypsin) and with a variety of experimental conditions including the addition or absence of 10 mM calcium chloride, with a range of temperature (37 to 67 °C) and digestion times (15 min to 16 h). The conventional (control) digest consisted of overnight (16 h) incubation at 37 °C without calcium. Following termination of the digests, all samples were subject to reductive dimethylation [54] with deuterated formaldehyde and sodium cyanoborohydride (Cambridge Isotope Laboratories, Tewksbury, MA, USA) for relative quantitation. Each experimental condition exhibited a “light” (+28.03 u) label while the conventional control digest was tagged “heavy” (+36.08 u). Experimental preparations were combined with the control at an equal mass ratio and subject to reversed phase LC-UV clean-up with total peptide quantitation from A₂₁₄ and fraction collection in 45% acetonitrile. Desalted peptides were dried and stored at -20 °C until LC-MS/MS analysis.

6.2.4. Bottom-Up LC-MS/MS Data Acquisition

Bottom-up LC-MS/MS was conducted by injecting 1 µg total peptides onto a self-packed monolithic C18 column, coupled to a 10 µm New Objective PicoTip noncoated Emitter Tip (Woburn, MA, USA). A Dionex Ultimate 3000 LC nanosystem (Bannockburn, IL, USA) delivered

a 2 h linear gradient from 0.1% FA in water to 55% acetonitrile. The Q Exactive mass spectrometer (Thermo Fisher Scientific) operated in data dependent mode scanning the top 10 precursors for MS2, with a resolution of 35,000 full width at half maximum for both full MS and MS2.

6.2.5. LC-MS/MS Data Analysis

MaxQuant proteomics software version 1.6.17.0, developed by the Max Planck Institute of Biochemistry was downloaded from <https://www.maxquant.org> (accessed on 27 July 2022). Raw MS/MS spectra were searched in MaxQuant [416] with relative quantitation based on dimethyl labeling, first with full trypsin specificity and subsequently with semi-tryptic specificity, in both cases using an FDR of 0.01 and allowing up to 2 miscleavages. Search results were filtered at the peptide level, requiring a posterior error probability (PEP) of ≤ 0.01 and at the protein level, requiring ≥ 2 unique peptide IDs per protein identification. Peptide and protein identifications were compared across digestion conditions using Venny 2.1 [417]. Peptide identifications were sorted based on miscleavage frequency. The relative abundance of digestion products was taken as the ratio of peak areas between the experimental condition (dimethyl label: light) and conventional digest (dimethyl label: heavy). Motif diagrams were generated using WebLogo [418].

6.2.6. Data Availability

Raw MS/MS spectra were searched in MaxQuant version 1.6.17.0 [416] with relative quantitation. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org> (accessed on 27 July 2022)) via the PRIDE partner repository [419] with the dataset identifier PXD035682. They are also contained in the appendix – Tables A6.1-A6.6.

6.3. Results

6.3.1. Cumulative Trypsin Activity is Maximized at 47 °C with 10 mM Calcium Ions

Bottom-up proteomics workflows have exploited elevated temperatures for accelerated trypsin digestion (up to 90-100 °C) [162,164,420]. The spectroscopic BAEE assays confirmed an optimal initial trypsin activity at 47 to 57 °C, respectively, contributing a 31 and 28% enhancement relative to the control, 37 °C (Figure 6.1). The addition of calcium not only promotes higher thermal

stability for trypsin but also enhances its initial activity. Inclusion of 10 mM calcium chloride contributes a maximal 86% increase in initial activity at 37 °C. Moreover, 10 mM calcium shifts the optimal temperature for initial activity to 67 °C, whereby trypsin activity was observed to be 340% relative to the control condition. These results confirm the essential role of calcium ions to maximize tryptic activity at elevated temperature. However, such observations do not reflect the expected loss in enzyme activity caused by trypsin autolysis or thermal denaturation, meaning it cannot be concluded that 67 °C is optimal for tryptic digestion.

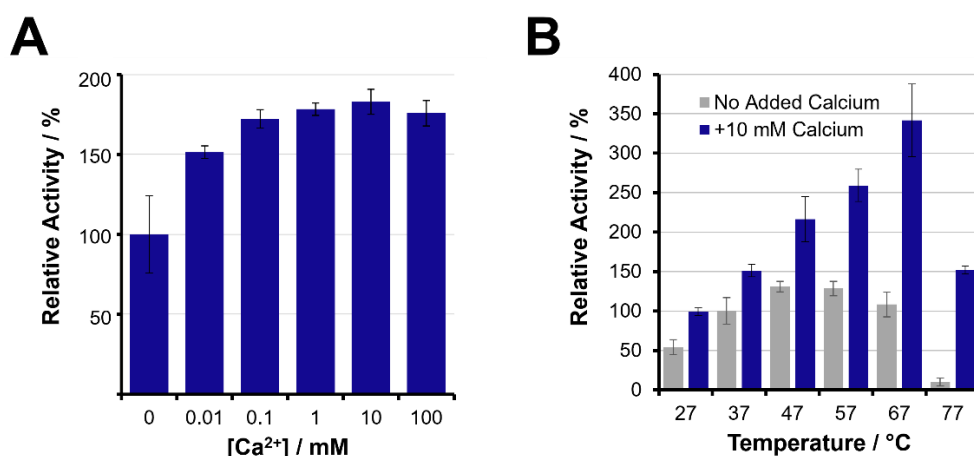


Figure 6.1 Initial trypsin activity of unmodified, TPCK-treated trypsin, as determined by BAEE assay, and normalized to activity at 37 °C in the absence of calcium. (A) Activity increases with addition of CaCl₂; (B) Initial activity at increasing temperatures, showing the benefits of Ca²⁺ for higher temperature incubation.

Enzyme stability studies were conducted by pre-incubating trypsin (pH 8) at various temperatures in varying concentrations of calcium prior to BAEE assays. Without calcium, the enhanced enzyme activity afforded by elevating the temperature is poorly sustained. After 30 min, the residual activity observed at higher temperatures had dropped to levels below that at 37 °C (Figure 6.2A). Figure 6.3 shows similar stabilization benefits from the inclusion of 5-100 mM calcium ions, with 10 mM minimizing the rate of deactivation. From Figure 6.2B, inclusion of 10 mM calcium chloride largely preserves the initial activity gains, though only to a maximum temperature of 57 °C. At 37 °C, enzyme deactivation was immeasurable across the first 4 h, with only a drop in activity after a 16 h incubation (Figure 6.4). The curves presented in Figure 6.2 represent best fit trendlines of the data using a second-order kinetic model (Figure 6.4). These fits allow for quantitative reporting of the rate of trypsin deactivation, as summarized in Table 6.1.

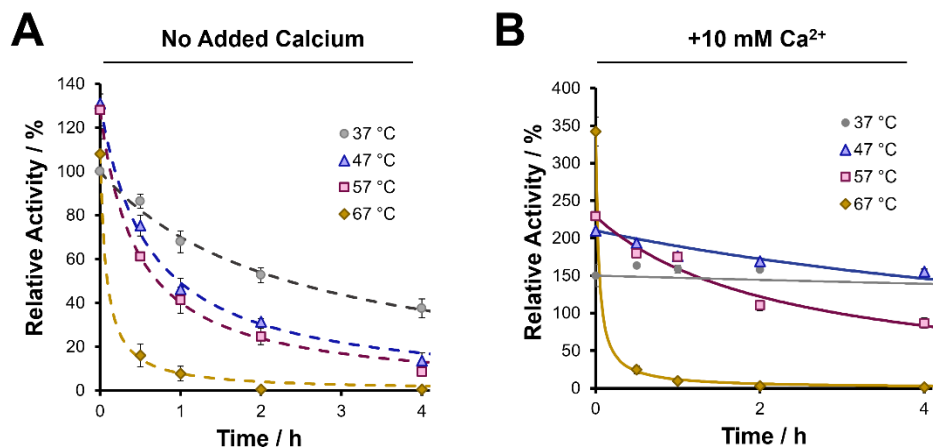


Figure 6.2 Time course assessment of trypsin activity following pre-incubation (pH 8), as determined by BAEE assay. (A) Loss of enzyme activity in the absence of calcium ions is noted at the specified temperatures; (B) Inclusion of 10 mM Ca^{2+} increases initial activity and sustains enzyme stability up to a maximum 57 °C. All values are normalized to the initial activity at 37 °C in the absence of calcium. Trendlines represent a fit to a second order kinetics model of the form $A_T = A_o / \{1 + ktA_o\}$.

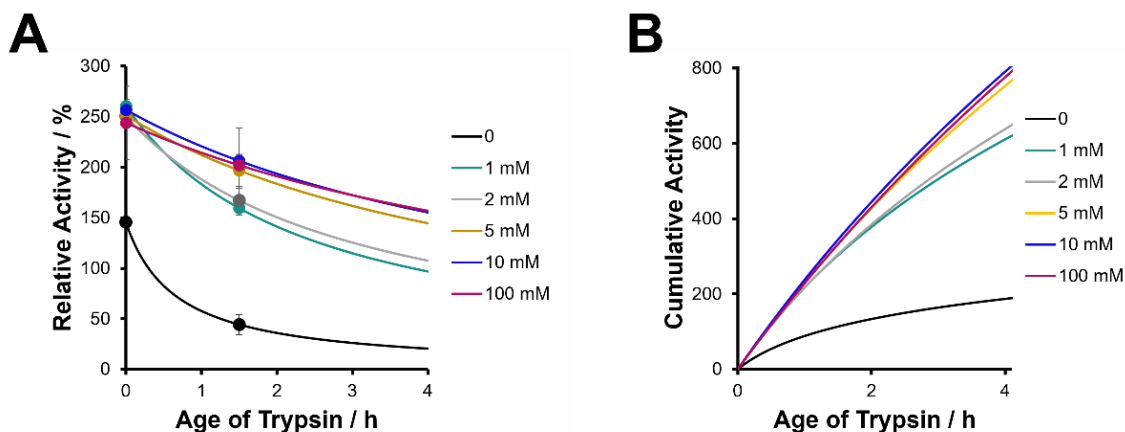


Figure 6.3 (A) Second-order kinetics models of trypsin de-activation at 47 °C with 0-100 mM added calcium chloride. (B) Estimated cumulative activity based on integrated second-order kinetics models.

Table 6.1 Summary of trypsin deactivation rates at various temperatures in the presence or absence of calcium ions.

Temperature	No Added Ca ²⁺			+10 mM Ca ²⁺		
	Initial Activity / %	Rate Constant, k / h ⁻¹	Calculated Half-Life /h	Initial Activity/%	Rate Constant, k / % ⁻¹ h ⁻¹	Calculated Half-Life /h
37 °C	100 ± 15	0.004 ± 1 × 10 ⁻⁴	2.4	150 ± 20	0.0001 ± 2 × 10 ⁻⁵	51.3
47 °C	130 ± 12	0.013 ± 8 × 10 ⁻⁴	0.6	210 ± 17	0.0004 ± 8 × 10 ⁻⁵	11.6
57 °C	128 ± 7	0.017 ± 1 × 10 ⁻⁴	0.5	229 ± 16	0.0018 ± 2 × 10 ⁻⁴	2.4
67 °C	108 ± 20	0.12 ± 8 × 10 ⁻³	0.03	340 ± 20	0.19 ± 0.01	0.03

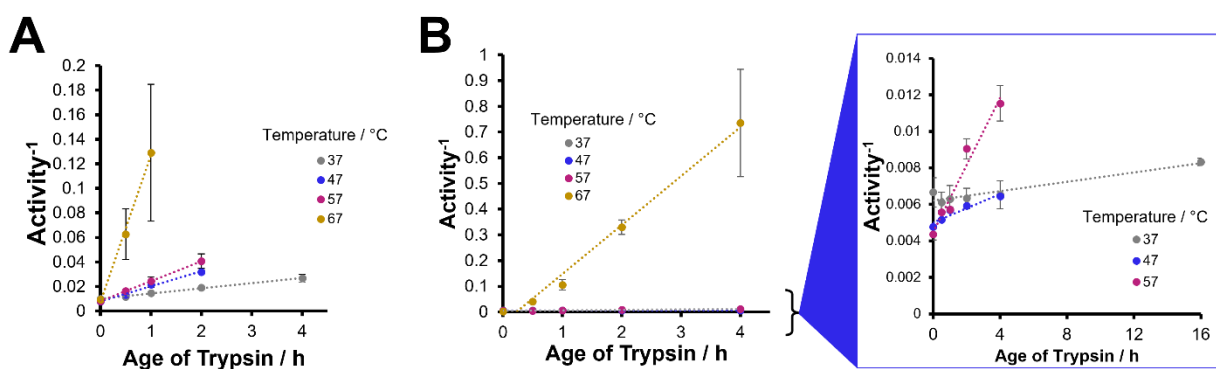


Figure 6.4 Second-order kinetics models of trypsin deactivation at 37-67 °C (A) with no added calcium ions, and (B) with 10 mM added calcium chloride.

From Table 6.1, it is noted a 25-fold improvement in the stability of trypsin at 47 °C with calcium ions compared to the same temperature without calcium, as seen from the drop in the deactivation rate constant. Moreover, the enzyme shows a 5-fold increase in the half-life at 47 °C with calcium ions, at nearly 12 h, compared to only 2.4 h with conventional conditions (37 °C, no calcium). The stability improvement also adds a >2-fold enhancement in initial activity at 47 °C. However, at 67 °C, while the initial activity was considerably higher, the presence of calcium no longer provided a stabilizing effect, resulting in a deactivation rate similar to the no-calcium condition. To reflect trypsin activity over the duration of a digestion, the cumulative (integrated) enzyme activity (A_T) is presented as a function of temperature in Figure 6.5. These plots are obtained through the extrapolated rate constant and integrating the second-order kinetics curves to arrive at Equation 6.2 below:

$$A_T = \frac{1}{k} \ln(1 + A_0 kt) \quad (6.2)$$

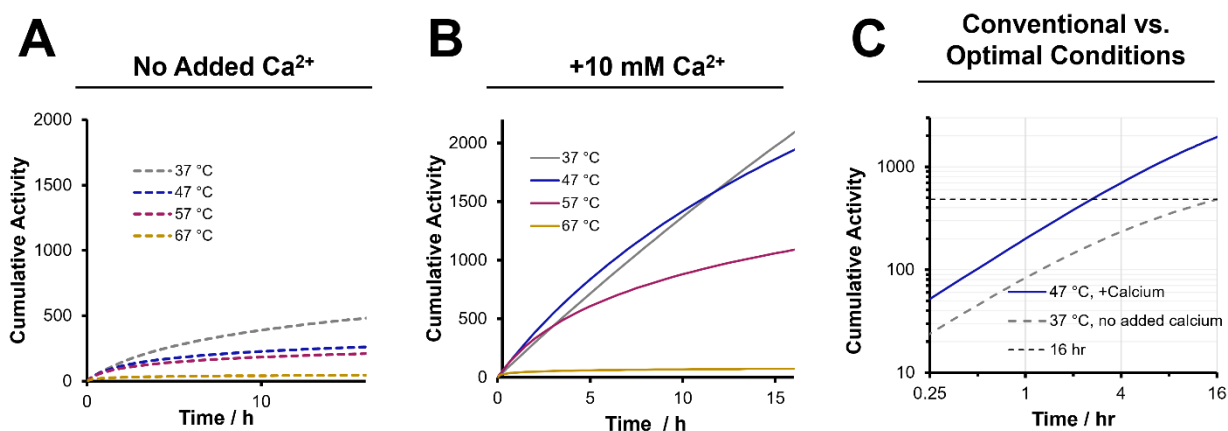


Figure 6.5 Cumulative activity across a 16 h time course estimated from integrated activity over time (A) in the absence of calcium, (B) with 10 mM added calcium ions, and (C) comparing the proposed optimal condition of 47 °C with 10 mM calcium ions to conventional conditions.

In this equation, k is the de-activation rate constant, A_0 is the initial trypsin activity and t is the cumulative digestion time. With inclusion of calcium, the cumulative activity is optimal at 47 °C at all points over a twelve-hour incubation period. It is also evident that higher temperature digests, specifically 67 °C, show a significantly lower cumulative activity, even relative to a 37 °C digestion. A temperature of 47 °C in the presence of 10 mM calcium was chosen as the preferable conditions for the enhanced rapid digestion approach. A direct comparison of the cumulative activity under this condition to that of a conventional digestion (37 °C, no calcium) is provided in Figure 6.5C. An overnight (16 h) digestion is seen to be equivalent to 2.5 h digestion under enhanced conditions. Likewise, a 1 h digest under enhanced conditions provides ~50% of the cumulative activity of a 12 h conventional digest. While these results suggest that accelerated digestion could achieve equivalent cleavage efficiency of a 37 °C digest but in a fraction of the time, it is realized that the BAEE assay does not capture the complexity of a proteomic system undergoing trypsin digestion. To determine the practical influence of these activity and stability enhancements, relative MS-based peptide quantitation was employed following bottom-up preparations ranging in digestion time, temperature, and calcium inclusion.

6.3.2. Proteome Identifications Show High Similarity between Digestion Conditions

Figure 6.6 provides a summary of total peptide and protein identifications from each of the digestion conditions explored. Identifications are optimized in the 1 h digest at 47 °C with calcium, showing a 39% increase in peptide IDs compared to the conventional overnight digest. The Venn diagram in Figure 6.7 shows that miscleaved peptide segments observed in the conventional preparation are often represented as more completely digested peptides in the rapid digest—this therefore increases the overall heterogeneity and non-redundant identifications. Without the addition of calcium, a 1 h digest at 47 °C resulted in 17% fewer peptide IDs and 26% fewer peptides than 1 h at 37 °C. This reflects the consequences of reduced enzyme stability at higher temperature without calcium (4-fold difference in enzyme half-life) and highlights the importance of conserving trypsin activity even across a 1 h digest. Moreover, under enhanced conditions, a 15 min digest produced 12% fewer peptides than the optimal 1 h digest. This suggests that valuable cleavage events are ongoing between the 15 min and 60 min points. The Venn diagrams in Figure 6.6B,C show 61.8% overlap in identified peptides and 94.2% protein overlap when comparing the enhanced 1 h digestion to the conventional overnight protocol. Only 6.8% of peptides (and no proteins) were unique to the conventional digest. Not only does the accelerated digestion protocol capture the proteome profile made visible through conventional digestion, but subsequent analysis of the resulting peptide profile suggests a more complete digestion of the sample using enhanced conditions.

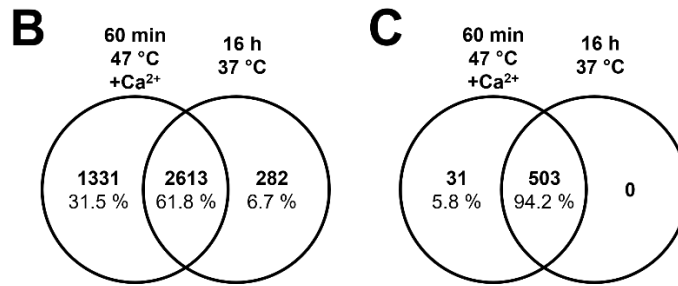
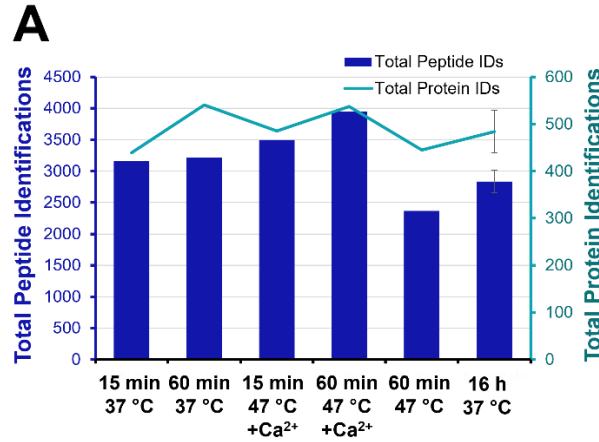


Figure 6.6 Bottom-up proteome identifications. (A) Total peptide and protein identifications following digestions ranging from 15 min to overnight at 37 or 47 °C, with and without 10 mM added calcium chloride. Venn diagrams of (B) peptide and (C) protein identifications in the 1 h digest at 47 °C with 10 mM Ca²⁺ are compared to the conventional overnight digest at 37 °C with no added Ca²⁺.

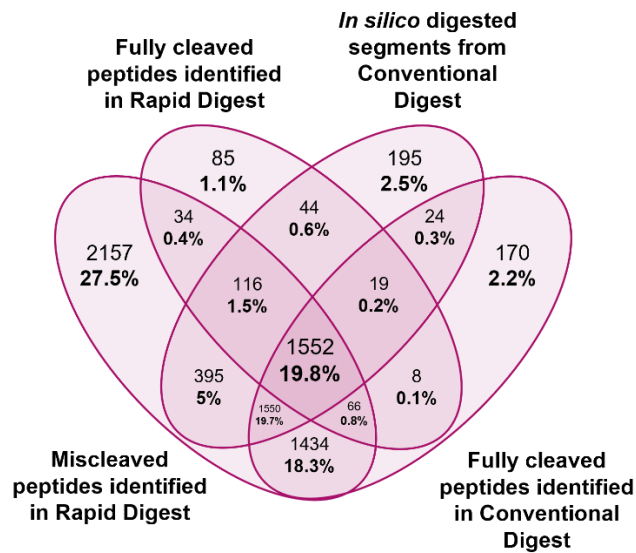


Figure 6.7 Venn diagram comparing *in-silico* digestion of miscleaved peptides identified in the conventional digest with peptides identified in both the rapid (60 min, 47 °C, +10 mM Ca²⁺) and conventional digests.

Proteome identifications were next assessed based on miscleavage frequency to infer the degree of digestion completion (Figure 6.8). As expected, the frequency of fully-cleaved peptides increases with longer digestion under conventional conditions, albeit with only marginal gains beyond 1 h. These diminishing returns are reflective of a decline in enzyme activity from 68% at 1 h to only 6% after 16 h (Figure 6.2). Miscleavage analysis of bottom-up proteome identifications. Absolute count and frequency of fully-cleaved and miscleaved peptides identified across digestion conditions.

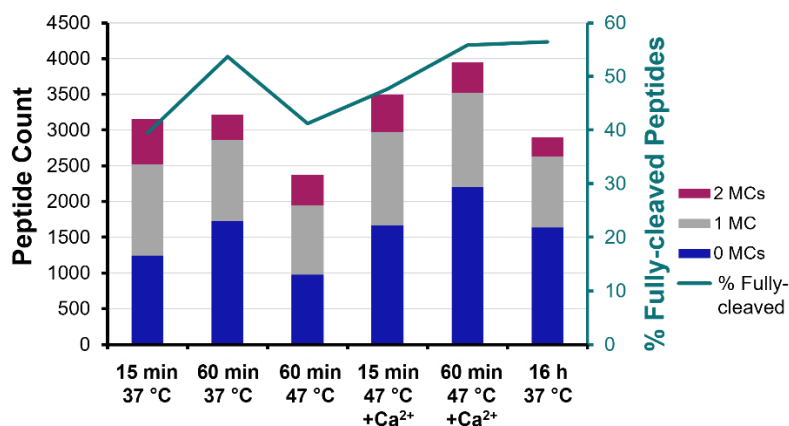


Figure 6.8 Miscleavage analysis of bottom-up proteome identifications. Absolute count and frequency of fully-cleaved (0 MCs) and miscleaved peptides (1-2 MCs) identified across digestion conditions.

Under enhanced conditions (47 °C with calcium), the frequency of fully-cleaved peptides increases from 47.7% at 15 min to 56% at 60 min—equivalent to the overnight digest under conventional conditions (Figure 6.8). Both rapid digests at 47 °C also show 62% overlap with the peptides identified in the conventional sample (Figure 6.9); however, the higher frequency of fully-cleaved peptides lends favor to a 1 h digest over the shorter 15 min protocol, which is attributed to there being four times as much cumulative activity. In the absence of calcium ions, the 1 h digest at elevated temperature had 15% higher miscleavage frequency (Figure 6.8), being comparable to 15 min at 37 °C, despite having more than double the cumulative activity (Figure 6.5). Figure 6.10 demonstrates the consistently high overlap in bottom-up protein identifications from experimental rapid conditions compared to the conventional digest, the exception being a rapid digest at 47 °C in the absence of stabilizing calcium ions (Figure 6.10E). Protein coverage contributed by fully-cleaved vs. miscleaved peptides is compared in Figure 6.11 for the rapid digest (A) and the conventional digest (B). In both cases, >90% of proteins are associated with at least one fully-

cleaved peptide, however the fewer proteins are uniquely identified with miscleaved peptides in the rapid digest suggests a more complete digestion compared to the conventional sample.

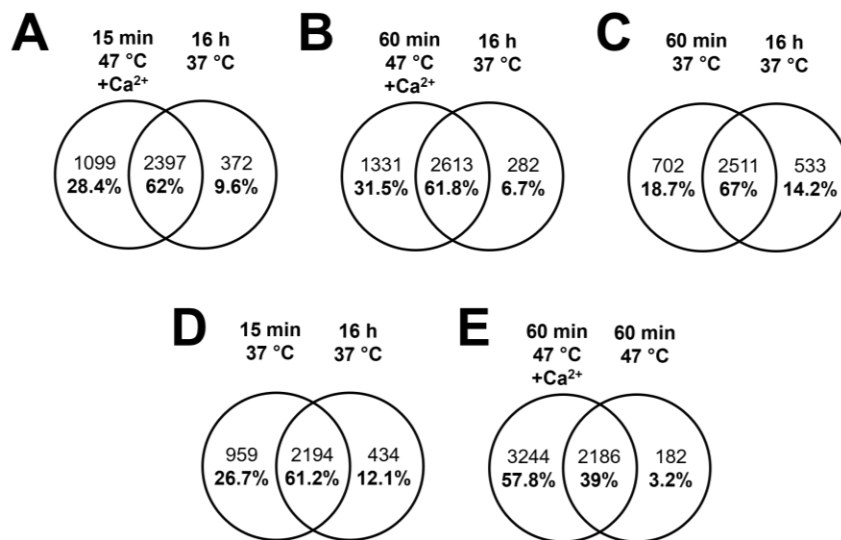


Figure 6.9 Venn diagrams of bottom-up peptide identifications compared across the “light” and “heavy” label, representing the indicated digestion conditions, within each LC-MS/MS injection.

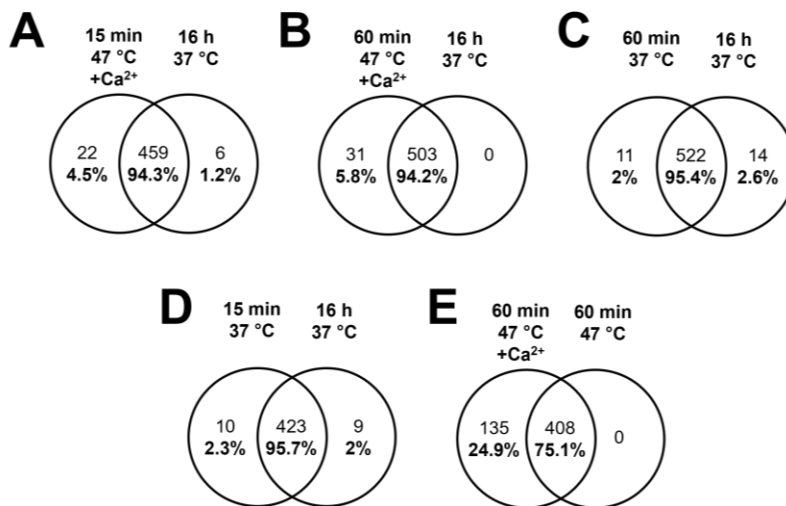


Figure 6.10 Venn diagrams of bottom-up protein identifications compared across the “light” and “heavy” label, representing the indicated digestion conditions, within each LC-MS/MS injection.

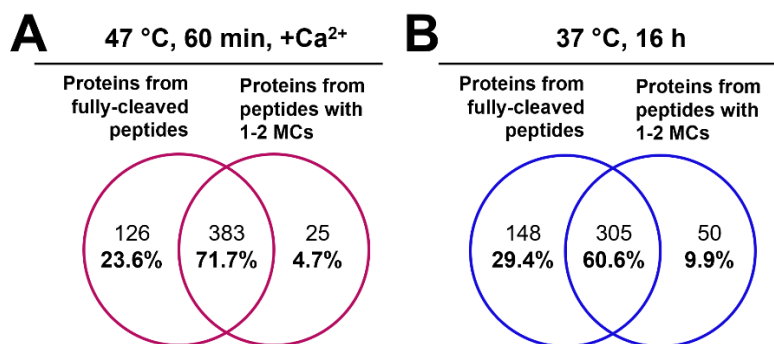


Figure 6.11 Venn diagrams of bottom-up protein identifications contributed by fully-cleaved peptides vs. those with 1-2 miscleavages in (A) the enhanced rapid digest and (B) the conventional/control digest.

6.3.3. Relative Quantitation Reveals Trypsin Cleavage Is Accelerated at Elevated Temperature with Added Calcium Ions

A quantitative examination of digestion products is now provided based on differential dimethyl labeling of various accelerated vs. conventional preparations. Peptides identified from each digestion condition were sorted based on their number of miscleavages and their cumulative signal intensities were plotted over time (Figure 6.12A,B). Under conventional conditions, longer digestions contribute a significant increase in the frequency of fully-cleaved peptides, with a corresponding decrease in the intensity of miscleaved peptides, demonstrating the importance of overnight incubation to maximize cleavage under conventional conditions. At 47 °C with calcium ions present, lengthening the digest from 15 min to 1 h also yields a 10% increase in the intensity of fully-cleaved peptides. However, extending the incubation period beyond 1 h does not yield a statistical gain in digestion efficiency. The intensity ratio of peptides observed from four rapid digestion conditions were compared to those from conventional overnight digestion and grouped according to the degree of cleavage (fully-cleaved, singly, and doubly miscleaved) with results summarized in Figure 6.13. At 37 °C (A&B), shorter digests lead to fully-cleaved peptides with a relative signal intensity below 1, while miscleaved peptides have ratios above 1. This translates to the overnight digestion having maximal signal intensity of fully-cleaved peptides and minimizes signals for peptides with miscleavage sites. At 47 °C with calcium ions, a more complete digest is observed with 1 h incubation vs. 15 min (Figure 6.12B). Interestingly, a 1 h enhanced rapid digest was also seen to increase the signal of all types of peptides compared to the conventional overnight digestion. The intensity ratio of fully-cleaved peptides is well above 1, though so too are the

miscleaved peptides, relative to conventional overnight digestion (Figure 6.13D). From this quantitative assessment, the results show that the 1 h digest at 47 °C with calcium ions improves the overall throughput with enhanced digestion completion, while contributing greater intensity for peptides of all types compared to the conventional approach.

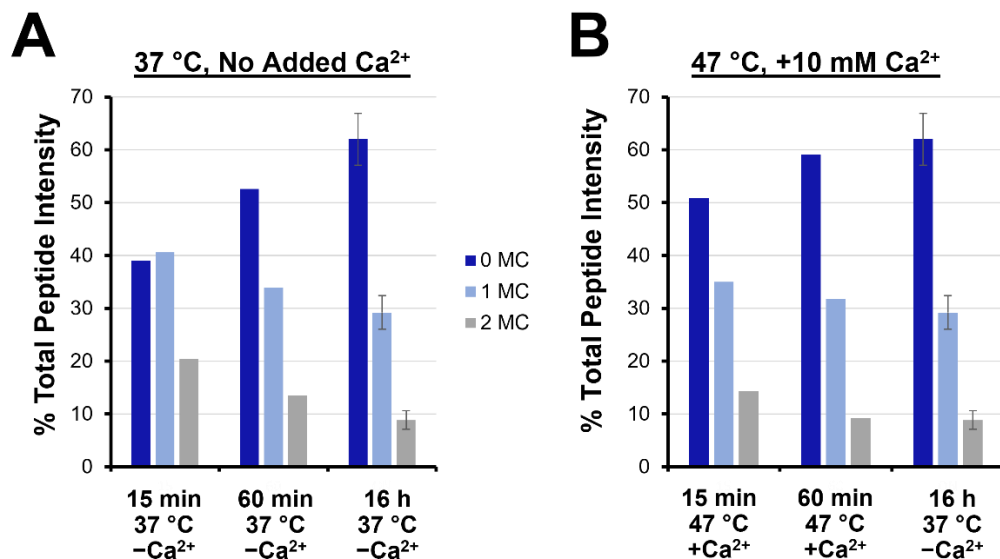


Figure 6.12 Relative quantitation of fully-cleaved (0 MC), singly miscleaved (1 MC), and doubly miscleaved (2 MC) peptides. Time course of digestion completion, inferred by % total peptide intensity contributed by fully-cleaved, singly and doubly miscleaved peptides at (A) 37 °C with no added Ca²⁺ and (B) 47 °C with 10 mM added Ca²⁺ compared to a conventional overnight digest.

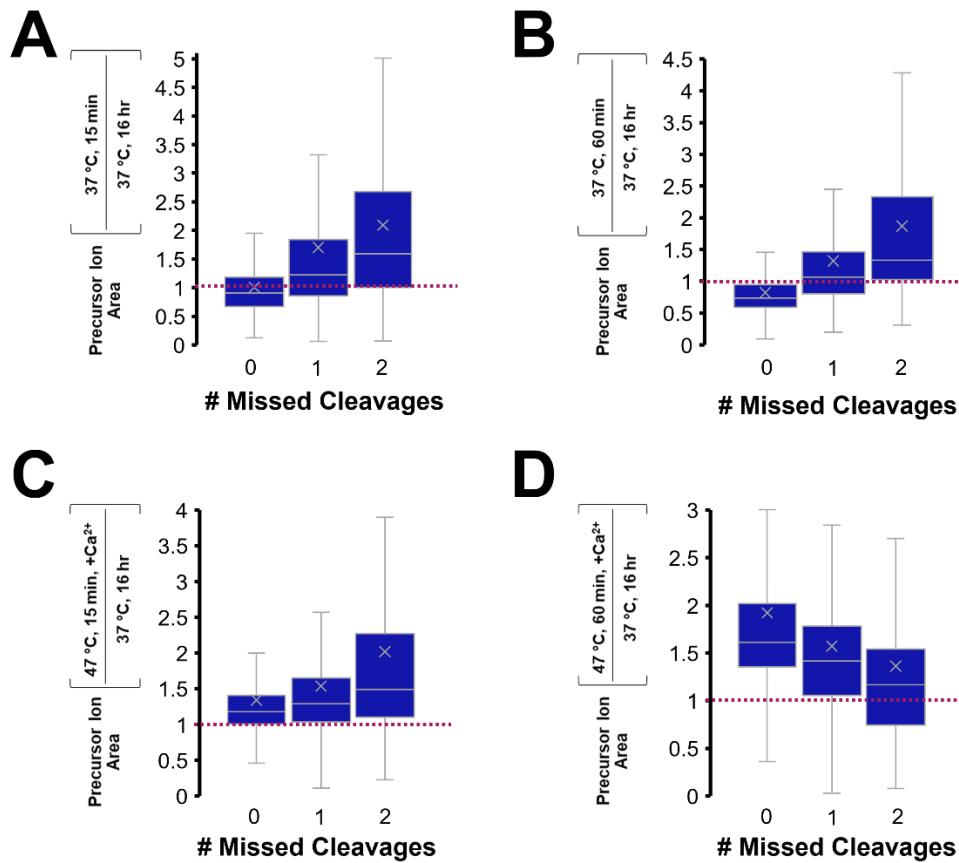


Figure 6.13 Tukey Box-and-Whisker plots of the distribution of relative peptide peak areas following digestion at (A) 15 or (B) 60 min, each at 37 °C with no added calcium or at 47 °C with 10 mM added calcium for (C) 15 or (D) 60 min. The relative peak intensity is compared to a conventional overnight digest at 37 °C.

6.3.4. Trypsin Specificity Is Enhanced in the Rapid Digest at Elevated Temperature with Added Calcium Ions

The normally stringent specificity of trypsin has reportedly been compromised at elevated temperatures [42,43]. To evaluate the enzyme's specificity in the enhanced rapid digest, the MS data were also subject to a search for semi-tryptic peptides. The enhanced and conventional digests demonstrated equivalent frequency of semi-tryptic cleavage, with 81% and 79% of identified peptides having full trypsin specificity, respectively. A Venn diagram comparing the semi-tryptic peptide identifications from the two digests (Figure 6.14A) shows 64% overlap, which is similar to that observed for fully tryptic peptides (Figure 6.7B). However, quantitative analysis of these semi-tryptic peptides shown in Figure 6.14B revealed a lower relative abundance of semi-tryptic peptides for the enhanced digestion compared to fully tryptic peptides, suggesting enhanced specificity is achieved in the rapid digest.

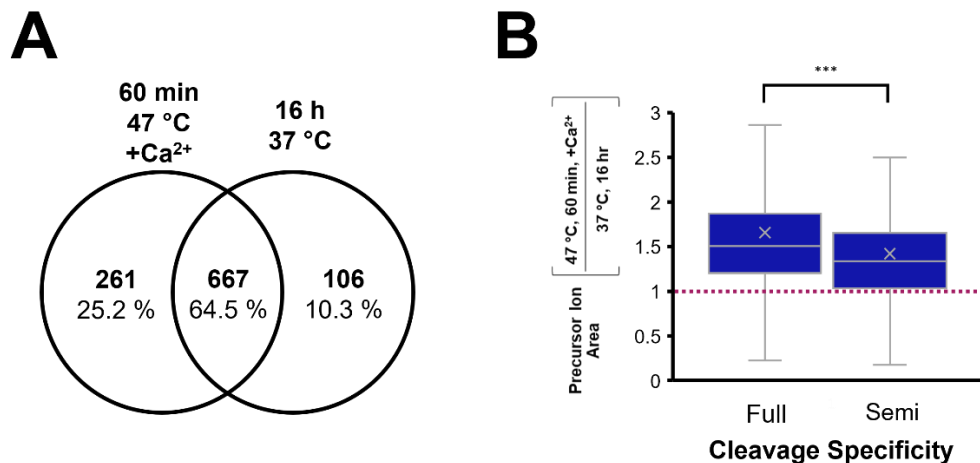


Figure 6.14 Evaluation of cleavage specificity in the 1 h digest at 47 °C with Ca^{2+} compared to a conventional overnight digest at 37 °C. (A) Venn diagram of semi-tryptic peptides identified in each digestion condition. (B) Tukey Box-and-Whisker plot comparing peptide quantitation for fully tryptic and semi-tryptic peptides, showing a greater relative abundance of fully tryptic peptides, *** $p < 0.001$.

The cleavage efficiency of trypsin varies depending on the nature of the residues proximal to the cut site, with reduced digestion expected for dibasic sites and when K and R residues are adjacent to acidic residues [161,421]. The relative cleavage efficiencies of these “hard-to-cut” regions are compared between the enhanced and conventional digests. From Figure 6.15A,B, it was found that 55% of all miscleaved peptides were contributed by acidic or dibasic cut sites. Based solely on the relative frequencies of miscleavage sites, the motif diagrams in Figure 6.15C reflect equivalent efficiency for digestion of acidic and basic amino acids within 2 residues or the C-terminal cleavage site. However, a quantitative assessment of these peptides (Figure 6.15D) shows that acidic and dibasic C-terminal digestion sites were cleaved with significantly greater efficiency in the enhanced digest compared to the conventional ($p < 0.001$). The combination of enhanced cleavage of traditionally difficult sites with a reduction in semi-tryptic peptides lends a conclusion that the accelerated (1 h) incubation at 47 °C with added calcium ions provides an enhanced digestion relative to the conventional overnight (37 °C) approach.

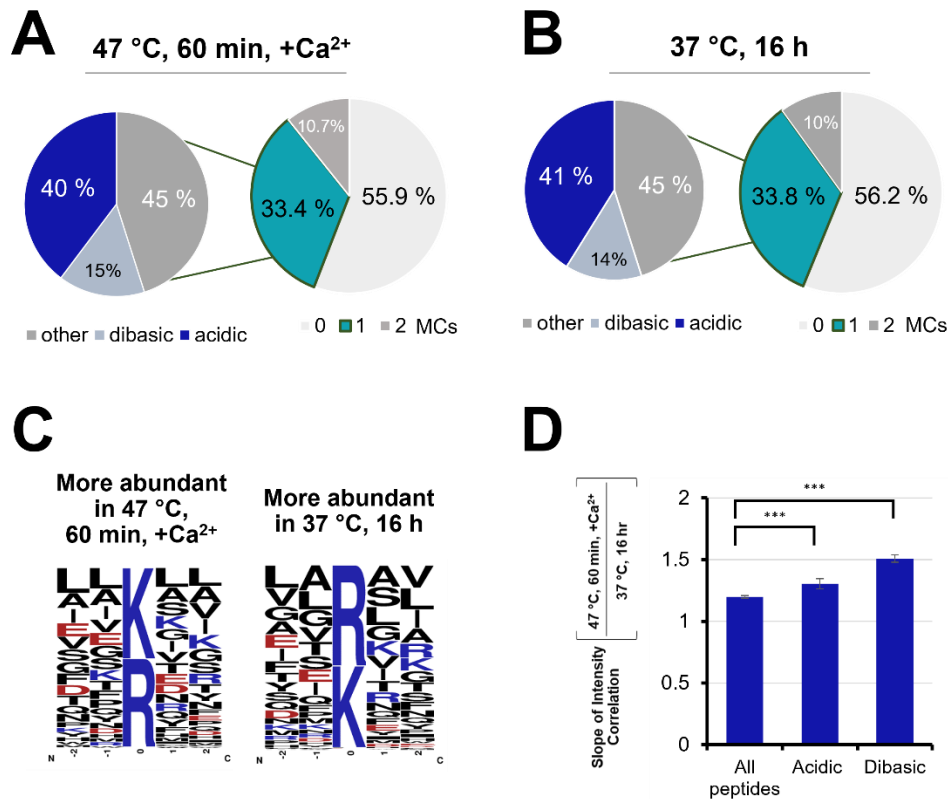


Figure 6.15 Singly miscleaved peptides identified in (A) the rapid digest and (B) conventional digest were further classified based on the nature of the missed cleavage site, with 55 % of singly miscleaved peptides attributed to acidic or basic residues immediately adjacent to the cut site. (C) Motif diagrams highlighting the frequency of acidic and basic residues being within 2 residues of the C-terminal cleavage site. (D) Comparison of the slopes correlating peptide intensity in the rapid vs. conventional digest for all peptides, those whose C-terminal cleavage site is adjacent to an acidic residue and those whose C-terminal cleavage site is adjacent to another lysine or arginine (***) $p < 0.001$).

6.4. Discussion

Proteolysis with trypsin is a critical, yet throughput-limiting component of bottom-up proteome workflows. Many works describe accelerated digests, which are typically validated by achieving high proteome and sequence coverage for simple protein systems. However, the growing interest of quantitative investigations motivates a high-throughput digestion approach that conserves the completion of a conventional overnight digest of a proteomic mixture. The present study first aimed to quantify both the initial activity and deactivation of trypsin over time as a function of temperature as influenced by the presence of calcium ions. Integrating the residual trypsin activity curves over extended periods provides a quantitative evaluation of the cumulative enzyme activity experienced over the duration of a digestion. These results lead to a conclusion

that the cumulative enzyme activity is maximized at 47 °C, justifying the use of this modestly higher temperature digestion but only in the presence of calcium ions. Performing digests at even higher temperatures is not justified as the enhanced initial activity seen in a BAEE assay is not sustained. For example, at 57 °C, the cumulative activity is at most ~50% above that obtained at 37 °C, but only for a short period. Owing to enzyme deactivation, cumulative activity drops below that seen at 37 °C after only 3 min. From the optimal digestion conditions suggested by these activity assays, quantitative bottom-up LC-MS was employed to compare accelerated protocols to a conventional (overnight, 37 °C digest). Not only is the proteome coverage evaluated, but also the quantitative profiling of partially vs. fully-cleaved peptides, as well as the digestion specificity.

Our evaluation of initial trypsin activity as a function of temperature showed strong agreement with the enhancement reported by Sipos et al. in 1973 [253]. They showed a 37% improvement in activity when temperature was increased from 37 °C to 47-57 °C, similar to the $30 \pm 10\%$ increase seen here (Figure 6.1B). Sipos also demonstrated that calcium ions shifted the optimal temperature to 67 °C owing to increased thermostability. Interestingly, though the same temperature shift was observed, a greater enhancement factor was seen (3.4 ± 0.2 -fold compared to their 2.6-fold) (Figure 6.1B). This discrepancy may be attributed to lower enzyme purity in the earlier work. Likewise, the addition of calcium ions provided a 50 % increase in initial activity at 37 °C, which was twice that observed by Sipos (~24%).

The stabilizing effects of calcium can sustain trypsin activity at higher temperatures for longer periods. However, at 67 °C, calcium no longer prevents trypsin deactivation. At elevated temperatures of 47 and 57 °C, 19.3- and 5-fold increases in the enzyme's half-life are observed, respectively. Therefore, the addition of calcium ions provides a drastic improvement in trypsin stability so long as a critical temperature (57 °C) is not exceeded. This suggests that the stabilizing effects of calcium ions are limited by the rate of protein denaturation as opposed to the rate of autolysis at this temperature, which aligns with Trampari's discussion of accelerated unfolding at similar temperatures [166]. The cumulative activity of trypsin calculated by integrating the activity over time justifies the use of only a modest temperature increase, when in the presence of calcium ions. Considering a higher temperature digestion, the enhanced initial activity seen in a BAEE assay is not sustained, meaning cumulative activity suffers. For example, at 57 °C, the cumulative activity difference is maximized at very short times by ~50 % vs. 37 °C. However, cumulative activity drops below that seen at 37 °C after only 3 min.

As reported by Nord and Bier (1956) [252], this study also found that trypsin deactivation initially followed second-order kinetics (Figure 6.4). Sipos reported a similar trend, noting a second-order rate for autolysis, with subsequent deactivation following a pseudo- first order rate [253]. It is speculated that the gradual acceleration of activity loss is due to thermal aggregation, which is reported to occur most readily at temperatures exceeding 54 °C [166]. These trendlines allowed us to integrate the cumulative trypsin activity and show that a 1 h digest at 47 °C with calcium yields a cumulative enzyme activity equivalent to a 6 h digest under conventional conditions. However, applied to a real proteomic system, other benefits of high temperature digestion may be realized.

It is evident that short digestions yield many MS-detectable peptides. In fact, a 15 min incubation at 37 °C generated more peptides than the overnight control digestion, although with fewer total proteins identified. There was also a high degree of overlap (~95%) in the proteome coverage by each of the experimental digestion conditions compared to the control digest (Figure 6.10). Nonetheless, significant differences were observed. For example, shorter digestions translate into more miscleavages, which is evident from the data. Figure 6.16 indeed shows a larger proportion of low-abundance proteins [422] identified using the optimized enhanced digestion protocol. Focusing on the types of peptides, as well as their relative abundances across the various digestion conditions allows a more complete assessment of each digestion condition. In theory, less complete digests yield a greater variety of peptides, which would be desirable for a single protein or simple proteome systems. However, for more complex systems, the variety of peptides generated would further mask lower abundance proteins from being detected. Figure 6.11 shows that miscleaved peptides are not contributing significantly to the total proteome coverage, with 95% of protein identifications having at least one fully-cleaved peptide in both the enhanced rapid and conventional digests. These results demonstrate the relevance of a robust (more complete) digestion for accurate and reproducible bottom-up quantitation.

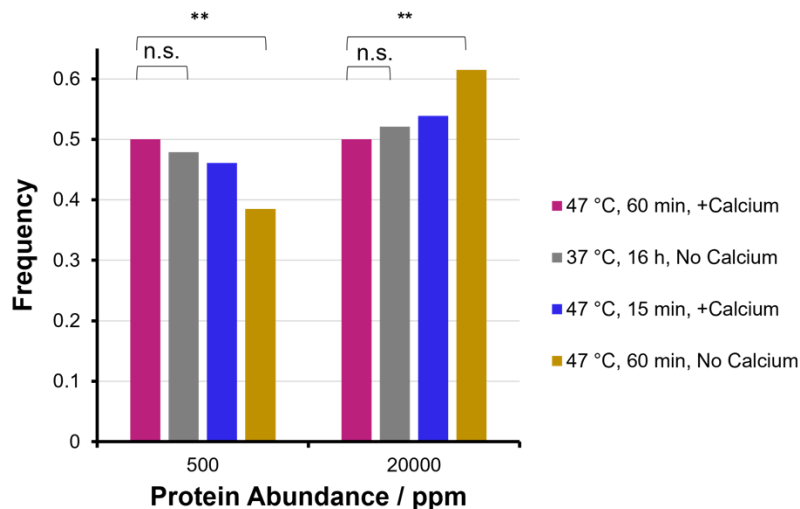


Figure 6.16 Histogram of the frequency of identifying low vs. high abundance proteins in the experimental digestion conditions, showing the greatest frequency of low-abundance proteins in the optimized rapid digest (60 min at 47 °C with 10 mM added calcium ions).

It is assumed that a more complete digest will identify peptides with fewer miscleaved sites than a less complete digest and consequently more homogenous digestion products which lends favor to quantitative and targeted analyses. At 37 °C, fully-cleaved peptides increase in relative intensity as one transitions from 15 min to 1 h to overnight digestion. This reflects a slower digestion rate compared to the enhanced conditions (47 °C with added calcium). Not only does the enhanced digestion produce fully-cleaved peptides with higher relative intensity than the control digestion, but the intensity of peptides with miscleavage sites was also higher than that of the control. In other words, the enhanced digestion protocol provides a more complete digestion, but also yields higher signal intensities than the conventional digestion, meaning a greater number of peptides of all types can be detected. This may reflect a greater degree of protein or peptide loss during extended incubation. It may also be reflective of differences in the degree of peptide modifications over time. The enhanced digestion of the accelerated protocol motivates its application in bottom-up quantitative studies where complete and consistent production of proteotypic peptides is critical.

High temperatures have been reported to compromise the selective structure of the active site of trypsin. However, a qualitative assessment of semi-tryptic peptides showed that non-specific cleavage was less prevalent in the enhanced digestion protocol. Although the proportion of semi-tryptic peptides was equivalent to the control, quantitative analysis revealed a higher

relative abundance of fully tryptic peptides compared to those with semi-specific cleavages. This suggests an enhancement in the enzyme's specificity under the temperature- and calcium-enhanced conditions.

The cleavage efficiency of trypsin is highly dependent on the nature of the amino acid residue adjacent to the lysine or arginine. A 2005 report by Šlechtová et al. described differential cleavage rates depending on the residues surrounding the potential cut site [161]. It was shown there, and by Pan et al. [421], that trypsin cleaves at arginine with greater efficiency than at lysine, owing to the stronger interaction of R with the enzyme's active site [423]. Under conventional digestion conditions, peptides terminating in arginine are generated with greater frequency than those with a C-terminal lysine. Surprisingly, the enhanced digest at high temperature with added calcium demonstrated equivalent cleavage efficiency at arginine *vs.* lysine, suggesting a reduction in the bias towards arginine cleavage under the enhanced conditions. It is speculated that lysine may be able to interact more strongly with the active site of trypsin due to weakening of the intramolecular H-bonding network within the catalytic pocket [140] at the elevated temperature, while the calcium ions may conserve the conformation sufficiently to maintain high selectivity for lysine and arginine [254]. From the relative intensity of miscleaved peptides, one also notes that the enhanced digestion protocol showed a greater degree of cleavage of lysine and arginine when adjacent to acidic or basic residues. These sites are known to electrostatically inhibit the interaction of adjacent tryptic sites with the enzyme's catalytic pocket, suggesting that the elevated temperature, the calcium-assisted protocol increased the cleavage efficiency of 'difficult to digest' cut sites, further supporting enhanced enzyme specificity in the optimized rapid digest.

6.5. Conclusions

The present study characterizes the relationship between trypsin stability and bottom-up MS results in terms of sample coverage and quantitation. Trypsin deactivation follows second-order kinetics until the rate of denaturation exceeds that of autolysis. Further analysis of these kinetics models based on cumulative activity facilitated the estimation of optimal proteome digestion conditions for a bottom-up workflow. Subsequent bottom-up LC-MS/MS comparisons with relative peptide quantitation against a control showed that the number and abundance of fully-cleaved peptides was optimized following a 1 h digest at 47 °C with 10 mM added calcium ions. The corresponding proteome coverage was equivalent to that of a conventional digest. However,

the optimized rapid digest demonstrated enhanced tryptic cleavage specificity by way of reduced abundance of semi-tryptic peptides and greater cleavage efficiency at conventionally inhibited cut sites. In summary, a 1 h digest at 47 °C with 10 mM added calcium ions is shown to provide a more complete and more specific digest than the conventional overnight approach while also increasing throughput. This study employed unmodified, TPCK-treated trypsin. Commercial preparations of modified trypsin may also include calcium in their preparations. While preliminary BAEE assessment of enzyme activity and stability for modified trypsin showed results equivalent to the calcium-enhanced conditions, it is not yet known to what extent the enhanced stability can be attributed to the modified trypsin vs. the addition of calcium. This will be addressed in a future study. It is therefore recommended that proteomics researchers adopt this efficient, economical, and robust approach to optimize reproducibility, quantitation accuracy and turnaround time.

7. Future Work & Conclusions

7.1 Thesis Summary

Proteome characterization by bottom-up mass spectrometry—both qualitative and quantitative—relies on robust sample preparation strategies that simultaneously maximize recovery and digestion efficiency. Additionally, high impact clinical applications of these analytical systems demand maximized depth and quantitative precision along with constant improvements to throughput. The first chapter of this thesis aimed to review the fields that paved the way for modern proteome approaches, including a description of genomics and the expanding multi-omics initiatives. The international objective of completely sequencing the human proteome was described along with a discussion of the progress to date. The unique advantages of DDA and DIA-MS acquisition strategies were compared for qualitative and quantitative approaches.

With the sophistication of modern LC-MS technology, there is a demand for sample preparation strategies that match the potential throughput, sample coverage and quantitative precision allotted by the MS acquisition. A sample preparation strategy must ultimately be highly repeatable to establish reliable MS results. Recovery is discussed as a function of solubilization as well the sustained recovery following subsequent preparative steps including surfactant depletion strategies (e.g., precipitation) and digestion. The effects of surfactants on protein solubility are described, motivating their inclusion for optimal sample coverage including biologically important membrane proteins. The variables influencing protein solubility are further discussed in the interest of optimizing precipitation for surfactant depletion ahead of LC-MS analysis. A review of precipitation strategies is provided including mechanistic discussions, motivating the optimization of precipitation-based bottom-up workflows detailed in Chapters 2, 3, and 4.

Following efficient solubilization and purification, the depth of coverage achieved by bottom-up MS strategies is limited by digestion efficiency with variably miscleaved products increasing the sample's heterogeneity. Moreover, variances in digestion efficiency inhibit the precision of quantitation. Despite the controlling influence of digestion completion on the resulting sample coverage and quantitation “accelerated” digestion approaches often fail to evaluate digestion completion, favoring the result of a longer list of non-redundant peptide identifications and the associated high sequence coverage. A review of rapid/alternative digestion strategies is

provided along with a discussion of current limitations. The mechanisms of trypsin-catalyzed hydrolysis are described in the interest of exploiting optimal enzyme activity and stability for the development of a truly accelerated digestion approach, forming the basis of research projects described in Chapters 5 and 6.

Bottom-up sample preparation workflows have taken on many strategies, varying in complexity and technological demands. Previous work from this group determined that the century-old approach of organic solvent-based precipitation provides quantitative recovery of a proteome sample under specific conditions that control for ionic strength. Additionally, the ProTrap XG cartridge was developed and commercialized, facilitating the isolation of the precipitated protein from the MS-contaminating supernatant. Chapter 2 of this thesis evaluates the role of ionic strength in determining the rate of precipitation. From this work, a rapid approach to precipitation-based bottom-up proteome preparation was proposed, offering a simple and fast strategy of depleting surfactants prior to LC-MS/MS analysis. The reliability of qualitative proteome profiling is influenced by variances/biases in front-end sample preparation. Even more so, quantitative precision relies on highly repeatable workflows that minimize the variance associated with sample recovery and digestion. Chapter 3 describes an evaluation of the repeatability of precipitation-based bottom-up proteome sample preparation in the ProTrap XG with comparisons to conventional solution and in-gel digestion strategies. The inclusion of SDS afforded by the robust precipitation strategy demonstrated enhanced proteome coverage, while the quantitative recovery of the protein pellet and subsequent urea-based digestion enabled optimized quantitative precision. The optimized precipitation-based workflow in the ProTrap XG cartridge was subsequently applied to the development of a multi-omics preparative strategy, described in Chapter 4. The proteome and peptidome fractions were precipitated in sequence for bottom-up and intact LC-MS/MS analysis respectively. The residual supernatant comprising the recovered metabolome was assessed following a range of initial precipitation strategies. The qualitative and quantitative advantages of various precipitation solvent systems for multi-omics analysis were discussed. Briefly, methanol-based precipitation lent favor to increased total metabolome identifications, while the acetone strategy identified fewer total (but more structurally diverse) metabolites and improved peptidome coverage.

The optimization of proteome recovery facilitated by an SDS-rapid precipitation strategy in the ProTrap XG left conventional overnight digestion as the throughput-limiting step in a

bottom-up preparative workflow. Given the importance of optimized throughput as well as robust/complete digestion, the influence of common proteomics additives on trypsin activity and stability over time were evaluated and are described in Chapter 5. It was found that even trace quantities of surfactants reduce the total catalytic activity of the enzyme by several orders of magnitude at varying rates. Contrasting with frequently reported observations by others, it was concluded from this work that surfactants do not enhance trypsin digestion. The conserved activity observed in modest concentrations of chaotropes and organic solvents suggests these are preferable denaturants towards enhanced digestion of hard-to-solubilize membrane proteins.

The historical combination of high temperature and calcium ions for improved trypsin activity was subsequently investigated. Elevated temperatures are frequently exploited towards accelerated enzymatic cleavage; however, the degree of enzyme denaturation and thermal aggregation is often overlooked. The effects of temperature on trypsin's initial activity and stability over time were compared in the presence and absence of stabilizing calcium ions (described in Chapter 6). It was found that 47 °C with 10 mM added calcium ions provided maximized total (cumulative) activity across accelerated digestion periods. A bottom-up relative quantitation approach was employed to analyze comparative digestion products, which demonstrated high digestion completion following a 1 h trypsinization under the temperature- and calcium-enhanced conditions. This leads to a simple and cost-effective approach to accelerated trypsin digestion, offering significant throughput gains compared to the conventional overnight incubation. In combination with the rapid precipitation strategy described in Chapters 2 and 3, this work has reduced a multi-day sample preparation strategy to a couple of hours.

Taken altogether, the results of this thesis point to the benefits of a precipitation-based proteome workflow facilitated by the ProTrap XG filtration cartridge. The rapid precipitation approach presented in Chapter 2 affords the inclusion of surfactants for complete proteome solubilization, which was demonstrated in Chapter 3 to enhance identification repeatability and quantitative precision compared to conventional solution and in-gel strategies. Following precipitation, protein pellets may be re-solubilized in urea and subject to a rapid (60-min) digestion that exploits a dilute chaotrope, increased temperature, and the stabilizing effects of calcium ions, based on the findings from Chapters 5 and 6. This workflow is anticipated to increase the throughput of large-scale bottom-up preparations, affording reproducible proteome profiling and quantitation. Work presented in Chapter 4 demonstrated the potential of the ProTrap XG cartridge

for a multi-omics workflow, highlighting enhanced metabolome coverage and quantitative precision compared to the conventional methanol precipitation approach.

7.2 Future Work

The present thesis described the development and optimization of precipitation-based protein recovery as well as a rapid enzymatic digestion for enhanced throughput and quantitation precision. While both sets of work were validated fundamentally, it is of interest to design and execute future investigations to determine the scope of their application and practical utility in high-impact settings.

7.2.1 Rapid and quantitative organic solvent precipitation

The work described in Chapter 2 demonstrated a rapid approach to organic solvent-based proteome precipitation. The kinetics of precipitation were characterized across temperature and ionic strength, enabling the elucidation of a 2-minute protocol that provides complete recovery of a proteome sample. It was shown, however, that the initial proteome concentration is an influential factor in the rate of precipitation, suggesting 5-30 min incubation would assure robust recovery of dilute samples. Future investigations could further evaluate the efficiency of this rapid precipitation approach for low sample quantities to determine the utility of the approach for single-cell proteome initiatives, for example.

Based on the combined precipitation results presented by this group over the last decade, the role of ionic strength in determining precipitation rate and recovery is critical. An ion-pairing mechanism (strengthened by the reduced dielectric of organic solvents) has been speculated; however, a computational simulation study may further clarify the rate-limiting variables and reveal additional influential factors. Bottom-up LC-MS/MS analysis of variably precipitated pellets and supernatants revealed a slight bias towards the rapid recovery of high-molecular-weight proteins. It has also been shown that increased organic solvent content enhances the recovery of low-molecular-weight species. Thus, a simulation strategy could further characterize the influence of a protein's molecular weight on its rate of aggregation, among other variables (e.g., salt type, pH).

In the presented work, the rapid precipitation strategy was applied to a test sample of cultured *S. cerevisiae*. Further studies should be conducted to validate the scope of this approach

for diverse sample types and research objectives, including targeted analysis of low-abundance proteins. Section 7.3 describes preliminary results following the use of the rapid precipitation strategy in the ProTrap XG towards the detection of potential immunotherapy targets. Additional application studies could confirm the use of the approach for top-down MS with specific PTM characterization.

7.2.2 Evaluating the Repeatability of Sample Preparation in the ProTrap XG

The repeatability of a ProTrap XG-based workflow was characterized against conventional solution and in-gel approaches. However, an abundance of cartridge- and bead-based sample preparation strategies have been reported towards semi-automation and high throughput. Future studies have been designed to evaluate the efficacy and repeatability of a ProTrap XG-based preparation compared to other commercialized cartridges on the basis of sample recovery, qualitative coverage, and quantitation. To complement the results presented on quantitative precision, future experiments are needed to evaluate the accuracy of protein quantitation following the precipitation-based bottom-up preparation. This quantitative assessment could be conducted for a selected target analyte by way of AQUA or iTRAQ approaches.

7.2.3 Precipitation-Based Multi-omics

A precipitation-based approach to simultaneous proteome, peptidome, and metabolome preparation was described. A limitation of the current approach is the avoidance of surfactants in the initial solubilization step in order to conserve the MS-compatibility of the metabolome fraction (i.e., supernatant). Future efforts are needed to evaluate the inclusion of acid-labile or photo-cleavable surfactants, which would afford enhanced coverage of the proteome, peptidome, and perhaps the non-polar fraction of the metabolome.

The current analysis of the peptidome fraction should be considered “preliminary”. Future studies could investigate the use of the sequential precipitation approach for specific characterization of a target hormone or immuno-peptidome characterization. Further application of the whole workflow to the detection of a multi-omic biomarker panel would validate the implantation of this rapid multi-omics preparation approach for clinical assay development.

7.2.4 Effect of Denaturants on Trypsin Activity

Chapter 5 of this thesis characterizes the effects of several denaturing additives on trypsin activity and stability. These surfactants, chaotropes, and aqueous/organic solvent systems lend favor to the solubilization of membrane proteins, which is imperative for whole proteome digestion efficiency. However, the additives assessed here demonstrated significant deactivating effects on the trypsin enzyme. Future studies are currently ongoing in this group to evaluate the implications of including SDS and SDC for trypsin digestion of membrane-enriched proteome samples. Both qualitative sample coverage and quantitative precision of digestion products will be evaluated.

7.2.5 Calcium-Assisted Rapid Digestion in the ProTrap XG

The combined investigations on trypsin activity presented in Chapters 5 and 6 demonstrate the unique benefits of elevated temperature and calcium ions on trypsin's thermostability. Bottom-up LC-MS/MS analysis with relative quantitation, described in Chapter 6, validated the digestion efficiency and cleavage specificity afforded by a 1-hour digest under enhanced conditions. It is of interest to apply this accelerated digestion strategy to the development of MRM assays for clinical samples. Initial experiments are underway towards the application of a rapid ProTrap XG-based sample preparation for the development of an MRM assay for potential immunotherapy drug targets. The motivation, methods, and preliminary results of these efforts are described in Section 7.3. Additional investigations have been proposed to characterize the use of 47 °C and Ca²⁺-enhanced approaches for in-gel and on-bead digestion strategies.

7.3 Conclusions

Modern MS-based proteomics platforms are enabling deep coverage and precise quantitation of complex biological samples, with impactful applications in the understanding, diagnosis, and treatment of diseases. However, the quality of mass spectrometry data is at the mercy of robust front-end sample preparation strategies that enable complete proteome recovery and repeatable digestion while meeting current throughput demands. The present thesis aimed to characterize the variables influencing protein solubility—both in the interest of optimizing solubilization as well as to optimize precipitation efficiency—and proteolytic enzyme activity/stability. Through these diverse investigations, a high-throughput SDS-based preparative

workflow was developed, relying on rapid and quantitative purification by organic solvent precipitation, which is further facilitated by the ProTrap XG filtration cartridge. It was shown that this optimized precipitation approach is also amenable to metabolomics and multi-omics preparations. To complement the throughput of the rapid precipitation strategy, an accelerated digestion approach was elucidated on the basis of maximized cumulative activity through the thermostabilizing effects of added calcium ions. The combination of these high-throughput strategies is currently being applied to the development of an MRM assay to determine immunotherapy targets, exemplifying the potential of this enhanced preparative approach for large-scale and high-impact settings such as the clinic.

Bibliography

1. Aebersold, R.; Mann, M. Mass Spectrometry-Based Proteomics. *Nature*. **2003**, *422*, 198–207.
2. Fenn, J.B.; Mann, M.; Meng, C.K.; Wong, S.F.; Whitehouse, C.M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science*. **1989**, *246*, 64–71.
3. Dakna, M.; He, Z.; Chuan Yu, W.; Mischak, H.; Kolch, W. Technical, Bioinformatical and Statistical Aspects of Liquid Chromatography-Mass Spectrometry (LC-MS) and Capillary Electrophoresis-Mass Spectrometry (CE-MS) Based Clinical Proteomics: A Critical Assessment. *J. Chromatogr. B* **2009**, *877*, 1250–1258, doi:10.1016/j.jchromb.2008.10.048.
4. Das, L.; Murthy, V.; Varma, A.K. Comprehensive Analysis of Low Molecular Weight Serum Proteome Enrichment for Mass Spectrometric Studies. *ACS Omega* **2020**, *5*, 28877–28888, doi:10.1021/acsomega.0c04568.
5. Labeit, S.; Kolmerer, B.; Linke, W.A. The Giant Protein Titin. *Circ. Res.* **1997**, *80*, 290–294, doi:10.1161/01.res.80.2.290.
6. Gibbs, R.A. The Human Genome Project Changed Everything. *Nat. Rev. Genet.* **2020**, *21*, 575–576, doi:10.1038/s41576-020-0275-3.
7. Holley, R.W.; Apgar, J.; Everett, G.A.; Madison, J.T.; Marquisee, M.; Merrill, S.H.; Penswick, J.R.; Zamir, A. Structure of a Ribonucleic Acid. *Science*. **1965**, *147*, 1462–1465, doi:10.1126/SCIENCE.147.3664.1462.
8. Fiers, W.; Contreras, R.; Haegeman, G.; Rogiers, R.; Van de Voorde, A.; Van Heuverswyn, H.; Van Herreweghe, J.; Volckaert, G.; Ysebaert, M. Complete Nucleotide Sequence of SV40 DNA. *Nature* **1978**, *273*, 19–22.
9. Fiers, W.; Contreras, R.; Duerinck, F.; Haegeman, G.; Iserentant, D.; Merregaert, J.; Min Jou, W.; Molemans, F.; Raeymaekers, A.; Van den Berghe, A.; et al. Complete Nucleotide Sequence of Bacteriophage MS2 RNA: Primary and Secondary Structure of the Replicase Gene. *Nature* **1976**, *260*, 500–507.
10. Sanger, F.; Nicklen, S.; Coulson, A.R. DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Natl. Acad. Sci.* **1977**, *74*, 5463–5467, doi:10.1073/PNAS.74.12.5463.
11. Mardis, E.R. Next-Generation DNA Sequencing Methods. *Annu. Rev. Genomics Hum. Genet.* **2008**, *9*, 387–402, doi:10.1146/annurev.genom.9.081307.164359.
12. Slatko, B.E.; Gardner, A.F.; Ausubel, F.M. Overview of Next-Generation Sequencing Technologies. *Curr. Protoc. Mol. Biol.* **2018**, *122*, e59, doi:10.1002/CPMB.59.

13. Lednicky, J.A.; Shankar, S.N.; Elbadry, M.A.; Gibson, J.C.; Alam, M.M.; Stephenson, C.J.; Eiguren-Fernandez, A.; Glenn Morris, J.; Mavian, C.N.; Salemi, M.; et al. Collection of SARS-CoV-2 Virus from the Air of a Clinic within a University Student Health Care Center and Analyses of the Viral Genomic Sequence. *Aerosol Air Qual. Res.* **2020**, *20*, 1167–1171, doi:10.4209/aaqr.2020.05.0202.
14. DeLisi, C. Santa Fe 1986: Human Genome Baby-Steps. *Nature* **1986**, *455*, 876–877.
15. Dulbecco, R. A Turning Point in Cancer Research: Sequencing the Human Genome. *Science*. **1986**, *231*, 1055–1056, doi:10.1126/science.3945817.
16. Sinsheimer, R.L. The Santa Cruz Workshop-May 1985. *Genomics* **1989**, *5*, 954–956, doi:10.1016/0888-7543(89)90142-0.
17. Collins, F.S.; Morgan, M.; Patrinos, A. The Human Genome Project: Lessons from Large-Scale Biology. *Science*. **2003**, *300*, 286–290, doi:10.1126/science.1084564.
18. Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; Fitzhugh, W.; et al. Initial Sequencing and Analysis of the Human Genome. *Nature* **2001**, *409*, 860–921, doi:10.1038/35057062.
19. Craig Venter, J.; Adams, M.D.; Myers, E.W.; Li, P.W.; Mural, R.J.; Sutton, G.G.; Smith, H.O.; Yandell, M.; Evans, C.A.; Holt, R.A.; et al. The Sequence of the Human Genome. *Science*. **2001**, *291*, 1304–1351, doi:10.1126/science.1058040.
20. International Genome Sequencing Consortium Finishing the Euchromatic Sequence of the Human Genome. *Nature* **2004**, *431*, 931–945, doi:10.1038/nature03001.
21. Nurk, S.; Koren, S.; Rhie, A.; Rautiainen, M.; Bzikadze, A. V.; Mikheenko, A.; Vollger, M.R.; Altemose, N.; Uralsky, L.; Gershman, A.; et al. The Complete Sequence of a Human Genome. *Science*. **2022**, *376*, 44–53, doi:10.1126/science.abj6987.
22. Armstrong, N.; Ryder, S.; Forbes, C.; Ross, J.; Gw Quek, R. A Systematic Review of the International Prevalence of BRCA Mutation in Breast Cancer. *Clin. Epidemiol.* **2019**, *11*, 543–561, doi:10.2147/clep.s206949.
23. Swisher, E.M.; Sakai, W.; Karlan, B.Y.; Wurz, K.; Urban, N.; Taniguchi, T. Secondary BRCA1 Mutations in BRCA1-Mutated Ovarian Carcinomas with Platinum Resistance. *Cancer Res.* **2008**, *68*, 2581–2586, doi:10.1158/0008-5472.can-08-0088.
24. Hahn, S.A.; Greenhalf, B.; Ellis, I.; Sina-Frey, M.; Rieder, H.; Korte, B.; Gerdes, B.; Kress, R.; Ziegler, A.; Raeburn, J.A.; et al. BRCA2 Germline Mutations in Familial Pancreatic Carcinoma. *J. Natl. Cancer Inst.* **2003**, *95*, 214–221.
25. Carbonara, K.; Andonovski, M.; Coorssen, J.R. Proteomes Are of Proteoforms: Embracing the Complexity. *Proteomes* **2021**, *9*, doi:10.3390/proteomes9030038.

26. Kramer, M.F.; Coen, D.M. Enzymatic Amplification of DNA by PCR: Standard Procedures and Optimization. *Curr. Protoc. Mol. Biol.* **2001**, *56*, 1–14, doi:10.1002/0471142727.mb1501S56.
27. Hortin, G.L.; Sviridov, D. The Dynamic Range Problem in the Analysis of the Plasma Proteome. *J. Proteomics* **2010**, *73*, 629–636, doi:10.1016/j.jprot.2009.07.001.
28. O'Farrell, P.H. High Resolution Two-Dimensional Electrophoresis of Proteins. *J. Biol. Chem.* **1975**, *250*, 400–4021.
29. Edman, P. Method for Determination of the Amino Acid Sequence of Peptides. *Acta Chim. Scand.* **1950**, *4*, 283–293.
30. Burnette, W.N. “Western Blotting”: Electrophoretic Transfer of Proteins from Sodium Dodecyl Sulfate-Polyacrylamide Gels to Unmodified Nitrocellulose and Radiographic Detection with Antibody and Radioiodinated Protein A. *Anal. Biochem.* **1981**, *112*, 195–203, doi:10.1016/0003-2697(81)90281-5.
31. Arentz, G.; Weiland, F.; Oehler, M.K.; Hoffmann, P. State of the Art of 2D DIGE. *Proteomics - Clin. Appl.* **2015**, *9*, 277–288, doi:10.1002/prca.201400119.
32. Singh, K.K.; Gupta, A.; Bharti, C.; Sharma, H. Emerging Techniques of Western Blotting for Purification and Analysis of Protein. *Futur. J. Pharm. Sci.* **2021**, *7*, 239, doi:10.1186/s43094-021-00386-1.
33. Karas, M.; Hillencamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons. *Anal. Chem.* **1988**, *60*, 2299–2301.
34. Beavis, R.C.; Chait, B.T.; Standing, K.G. Matrix-Assisted Laser-Desorption Mass Spectrometry Using 355 Nm Radiation. *Rapid Commun. Mass Spectrom.* **1989**, *3*, 436–439, doi:10.1002/rcm.1290031208.
35. Dole, M.; Mack, L.L.; Hines, R.L.; Mobley, R.C.; Ferguson, L.D.; Alice, M.B. Molecular Beams of Macroions. *J. Chem. Phys.* **1968**, *49*, 2240–2249, doi:10.1063/1.1670391.
36. Hunt, D.F.; Yates, J.R.; Shabanowitz, J.; Winston, S.; Hauer, C.R. Protein Sequencing by Tandem Mass Spectrometry. *Proc. Natl. Acad. Sci.* **1986**, *83*, 6233–6237, doi:10.1073/PNAS.83.17.6233.
37. Toby, T.K.; Fornelli, L.; Kelleher, N.L. Progress in Top-down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* **2016**, *9*, 499–519, doi:10.1146/annurev-anchem-071015-041550.
38. Kahn, P. From Genome to Proteome: Looking at a Cell's Proteins. *Science (80-.)*. **1995**, *270*, 369–371, doi:10.1126/science.270.5235.369.
39. Mann, M.; Wilm, M. Error-Tolerant Identification of Peptides in Sequence Databases by Peptide Sequence Tags. *Anal. Chem.* **1994**, *66*, 4390–4399, doi:10.1021/ac00096a002.

40. Apweiler, R.; Bairoch, A.; Wu, C.H.; Barker, W.C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; et al. UniProt: The Universal Protein Knowledgebase. *Nucleic Acids Res.* **2004**, *32*, D115–D119, doi:10.1093/nar/gkh131.
41. Wiśniewski, J.R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal Sample Preparation Method for Proteome Analysis. *Nat. Methods* **2009**, *6*, 359–362, doi:10.1038/nmeth.1322.
42. Ivanov, M. V; Bubis, J.A.; Gorshkov, V.; Tarasova, I.A.; Levitsky, L.I.; Lobas, A.A.; Solovyeva, E.M.; Pridatchenko, M.L.; Kjeldsen, F.; Gorshkov, M. V DirectMS1: MS/MS-Free Identification of 1000 Proteins of Cellular Proteomes in 5 Minutes. *Anal. Chem* **2022**, *18*, 49, doi:10.1021/acs.analchem.9b05095.
43. Meier, F.; Geyer, P.E.; Virreira Winter, S.; Cox, J.; Mann, M. BoxCar Acquisition Method Enables Single-Shot Proteomics at a Depth of 10,000 Proteins in 100 Minutes. *Nat. Methods* **2018**, *15*, 440–448, doi:10.1038/s41592-018-0003-5.
44. Richards, A.L.; Hebert, A.S.; Ulbrich, A.; Bailey, D.J.; Coughlin, E.E.; Westphall, M.S.; Coon, J.J. One-Hour Proteome Analysis in Yeast. *Nat. Protoc.* **2015**, *10*, 701–714, doi:10.1038/nprot.2015.040.
45. Adhikari, S. A High-Stringency Blueprint of the Human Proteome. *Nat. Commun.* **2020**, *11*, 1–16, doi:10.1038/s41467-020-19045-9.
46. Eng, J.K.; McCormack, A.L.; Yates, J.R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989, doi:10.1016/1044-0305(94)80016-2.
47. Perkins, D.N.; Pappin, D.J.C.; Creasy, D.M.; Cottrell, J.S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999**, *20*, 3551–3567, doi:10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.
48. Moore, R.E.; Young, M.K.; Lee, T.D. Qscore: An Algorithm for Evaluating SEQUEST Database Search Results. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 378–386.
49. Qian, W.-J.; Liu, T.; Monroe, M.E.; Strittmatter, E.F.; Jacobs, J.M.; Kangas, L.J.; Petritis, K.; Camp, D.G.; Smith, R.D. Probability-Based Evaluation of Peptide and Protein Identifications from Tandem Mass Spectrometry and SEQUEST Analysis: The Human Proteome The Results Demonstrate That the Present Criteria Provide Significantly Higher Levels of Confidence for Peptide Identifications from Mammalian Proteomes without Greatly Decreasing the Number of Identifications. *Nat. Biotechnol.* **2003**, *75*, 242–247, doi:10.1021/pr0498638.
50. Ludwig, C.; Gillet, L.; Rosenberger, G.; Amon, S.; Collins, B.C.; Aebersold, R. Data-independent Acquisition-based SWATH - MS for Quantitative Proteomics: A Tutorial . *Mol. Syst. Biol.* **2018**, *14*, doi:10.15252/MSB.20178126.

51. Vidova, V.; Spacil, Z. A Review on Mass Spectrometry-Based Quantitative Proteomics: Targeted and Data Independent Acquisition. *Anal. Chim. Acta* **2017**, *964*, 7–23, doi:10.1016/j.aca.2017.01.059.
52. Halder, A.; Verma, A.; Biswas, D.; Srivastava, S. Recent Advances in Mass-Spectrometry Based Proteomics Software, Tools and Databases. *Drug Discov. Today Technol.* **2021**, *39*, 69–79, doi:10.1016/j.ddtec.2021.06.007.
53. Tsou, C.C.; Avtonomov, D.; Larsen, B.; Tucholska, M.; Choi, H.; Gingras, A.C.; Nesvizhskii, A.I. DIA-Umpire: Comprehensive Computational Framework for Data-Independent Acquisition Proteomics. *Nat. Methods* **2015**, *12*, 258–264, doi:10.1038/nmeth.3255.
54. Li, Y.; Zhong, C.Q.; Xu, X.; Cai, S.; Wu, X.; Zhang, Y.; Chen, J.; Shi, J.; Lin, S.; Han, J. Group-DIA: Analyzing Multiple Data-Independent Acquisition Mass Spectrometry Data Files. *Nat. Methods* **2015**, *12*, 1105–1106, doi:10.1038/nmeth.3593.
55. Ting, Y.S.; Egertson, J.D.; Bollinger, J.G.; Searle, B.C.; Payne, S.H.; Noble, W.S.; MacCoss, M.J. PECAN: Library-Free Peptide Detection for Data-Independent Acquisition Tandem Mass Spectrometry Data. *Nat. Methods* **2017**, *14*, 903–908, doi:10.1038/nmeth.4390.
56. Demichev, V.; Messner, C.B.; Vernardis, S.I.; Lilley, K.S.; Ralser, M. DIA-NN: Neural Networks and Interference Correction Enable Deep Proteome Coverage in High Throughput. *Nat. Methods* **2019**, *17*, 41–44, doi:10.1038/s41592-019-0638-x.
57. Krasny, L.; Huang, P.H. Data-Independent Acquisition Mass Spectrometry (DIA-MS) for Proteomic Applications in Oncology. *Mol. Omi.* **2021**, *17*, 29–42, doi:10.1039/d0mo00072h.
58. Omenn, G.S.; Lane, L.; Lundberg, E.K.; Overall, C.M.; Deutsch, E.W. Progress on the HUPO Draft Human Proteome: 2017 Metrics of the Human Proteome Project. *J. Proteome Res.* **2017**, *16*, 4281–4287, doi:10.1021/acs.jproteome.7b00375.
59. Kulyyassov, A.; Fresnais, M.; Longuespée, R. Targeted Liquid Chromatography-Tandem Mass Spectrometry Analysis of Proteins: Basic Principles, Applications, and Perspectives. *Proteomics* **2021**, *21*, 1–20, doi:10.1002/pmic.202100153.
60. Mermelekas, G.; Vlahou, A.; Zoidakis, J. SRM/MRM Targeted Proteomics as a Tool for Biomarker Validation and Absolute Quantification in Human Urine. *Expert Rev. Mol. Diagn.* **2015**, *15*, 1441–1454, doi:10.1586/14737159.2015.1093937.
61. Mani, D.R.; Abbatiello, S.E.; Carr, S.A. Statistical Characterization of Multiple-Reaction Monitoring Mass Spectrometry (MRM-MS) Assays for Quantitative Proteomics. *BMC Bioinformatics* **2012**, *13*, 1–18, doi:10.1186/1471-2105-13-S16-S9.

62. Lemoine, J.; Fortin, T.; Salvador, A.; Jaffuel, A.; Charrier, J.P.; Choquet-Kastylevsky, G. The Current Status of Clinical Proteomics and the Use of MRM and MRM3 for Biomarker Validation. *Expert Rev. Mol. Diagn.* **2014**, *12*, 333–342, doi:10.1586/ERM.12.32.
63. 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. *Mol. Cell. Proteomics* **2002**, *1*, 376–386, doi:10.1074/mcp.M200025-MCP200.
64. Boersema, P.J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A.J.R. Multiplex Peptide Stable Isotope Dimethyl Labeling for Quantitative Proteomics. *Nat. Protoc.* **2009**, *4*, 484–494, doi:10.1038/nprot.2009.21.
65. Thompson, A.; Schaergen, K.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, T.; Neumann, T.; Hamon, C. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* **2003**, doi:10.1021/ac0262560.
66. Wiese, S.; Reidegeld, K.A.; Meyer, H.E.; Warscheid, B. Protein Labeling by ITRAQ: A New Tool for Quantitative Mass Spectrometry in Proteome Research. *Proteomics* **2007**, *7*, 340–350, doi:10.1002/PMIC.200600422.
67. Pratt, J.M.; Simpson, D.M.; Doherty, M.K.; Rivers, J.; Gaskell, S.J.; Beynon, R.J. Multiplexed Absolute Quantification for Proteomics Using Concatenated Signature Peptides Encoded by QconCAT Genes. *Nat. Protoc.* **2006**, *1*, 1029–1043, doi:10.1038/nprot.2006.129.
68. Gerber, S.A.; Rush, J.; Stemman, O.; Kirschner, M.W.; Gygi, S.P. Absolute Quantification of Proteins and Phosphoproteins from Cell Lysates by Tandem MS. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 6940–6945, doi:10.1073/PNAS.0832254100/SUPPL_FILE/2254SUPPORTINGMETHODS.HTML.
69. Zhang, H.; Bensaddek, D.; Chiapello, M.; Nikolovski, N.; Marondedze, C.; Labs, C. Narrow Precursor Mass Range for DIA-MS Enhances Protein Identification and Quantification in Arabidopsis. *Life* **2021**, *11*, 1–13, doi:10.3390/life11090982.
70. Guergues, J.; Wohlfahrt, J.; Stevens, S.M. Enhancement of Proteome Coverage by Ion Mobility Fractionation Coupled to PASEF on a TIMS–QTOF Instrument. *J. Proteome Res.* **2022**, *21*, 2036–2044, doi:10.1021/acs.jproteome.2c00336.
71. Brunner, A.M.; Lö, P.; Liu, F.; Huguet, R.; Mullen, C.; Yamashita, M.; Zabrouskov, V.; Makarov, A.; Maarten Altelaar, A.F.; Heck, A.J.R. Benchmarking Multiple Fragmentation Methods on an Orbitrap Fusion for Top-down Phospho-Proteome Characterization. *Anal. Chem.* **2015**, *87*, 4152–4158, doi:10.1021/acs.analchem.5b00162.
72. Schmitt, N.D.; Berger, J.M.; Conway, J.B.; Agar, J.N. Increasing Top-Down Mass Spectrometry Sequence Coverage by an Order of Magnitude through Optimized Internal Fragment Generation and Assignment. *Anal. Chem.* **2021**, *93*, 6355–6362, doi:10.1021/acs.analchem.0c04670.

73. Lu, L.; Scalf, M.; Shortreed, M.R.; Smith, L.M. Mesh Fragmentation Improves Dissociation Efficiency in Top-down Proteomics. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1319–1325, doi:10.1021/jasms.0c00462.
74. Leduc, R.D.; Taylor, G.K.; Kim, Y.-B.; Januszyk, T.E.; Bynum, L.H.; Sola, J. V; Garavelli, J.S.; Kelleher, N.L.; Keck, W.M. ProSight PTM: An Integrated Environment for Protein Identification and Characterization by Top-down Mass Spectrometry. *Nucleic Acids Res.* **2004**, *32*, W340–W345, doi:10.1093/nar/gkh447.
75. Zamdborg, L.; Leduc, R.D.; Glowacz, K.J.; Kim, Y.-B.; Viswanathan, V.; Spaulding, I.T.; Early, B.P.; Bluhm, E.J.; Babai, S.; Kelleher, N.L. ProSight PTM 2.0: Improved Protein Identification and Characterization for Top down Mass Spectrometry. *Nucleic Acids Res.* **2007**, *35*, 701–706, doi:10.1093/nar/gkm371.
76. Tsai, Y.S.; Scherl, A.; Shaw, J.L.; Mackay, C.L.; Shaffer, S.A.; Langridge-Smith, P.R.R.; Goodlett, D.R. Precursor Ion Independent Algorithm for Top-Down Shotgun Proteomics. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 2154–2166, doi:10.1016/j.jasms.2009.07.024.
77. Shen, Y.; Tolic, N.; Hixson, K.K.; Purvine, S.O.; Anderson, G.A.; Smith, R.D. De Novo Sequencing of Unique Sequence Tags for Discovery of Post-Translational Modifications of Proteins. *Anal. Chem.* **2008**, *80*, 7742–7754, doi:10.1021/ac801123p.
78. Liu, X.; Sirotkin, Y.; Shen, Y.; Anderson, G.; Tsai, Y.S.; Ting, Y.S.; Goodlett, D.R.; Smith, R.D.; Bafna, V.; Pevzner, P.A. Protein Identification Using Top-Down Spectra. *Mol. Cell. Proteomics* **2012**, *11*, 1–13, doi:10.1074/mcp.M111.008524.
79. Frank, A.M.; Pesavento, J.J.; Mizzen, C.A.; Kelleher, N.L.; Pevzner, P.A. Interpreting Top-Down Mass Spectra Using Spectral Alignment. *Anal. Chem.* **2008**, *80*, 2499–2505, doi:10.1021/ac702324u.
80. Smith, L.M.; Kelleher, N.L. Proteoform: A Single Term Describing Protein Complexity. *Nat. Methods* **2013**, *10*, doi:10.1038/nmeth.2369.
81. Smith, L.M.; Agar, J.N.; Chamot-rooke, J.; Danis, P.O.; Ge, Y.; Loo, J.A. The Human Proteoform Project: Defining the Human Proteome. *Sci. Adv.* **2021**, *7*, 1–8.
82. Human Proteoform Project & Atlas – Northwestern Proteomics Available online: <https://proteomics.northwestern.edu/services/human-proteoform-project/> (accessed on 16 September 2022).
83. Lee, J.E. Neuropeptidomics: Mass Spectrometry-Based Identification and Quantitation of Neuropeptides. *Genomics Inform.* **2016**, *14*, 12, doi:10.5808/gi.2016.14.1.12.
84. Leddy, O.K.; White, F.M.; Bryson, B.D. Leveraging Immunopeptidomics To Study and Combat Infectious Disease. *mSystems* **2021**, *6*, 1–9, doi:10.1128/msystems.00310-21.
85. Fricker, L.D. Limitations of Mass Spectrometry-Based Peptidomic Approaches. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1981–1991, doi:10.1007/s13361-015-1231-x.

86. Cunningham, R.; Ma, D.; Li, L. Mass Spectrometry-Based Proteomics and Peptidomics for Systems Biology and Biomarker Discovery. *Front. Biol. (Beijing)*. **2012**, *7*, 313–335, doi:10.1007/s11515-012-1218-y.
87. Fabre, B.; Combier, J.-P.; Plaza, S. Recent Advances in Mass Spectrometry-Based Peptidomics Workflows to Identify Short-Open-Reading-Frame-Encoded Peptides and Explore Their Functions. *Curr. Opin. Chem. Biol.* **2021**, *2021*, 122–130, doi:10.1016/j.cbpa.2020.12.002.
88. Menschaert, G.; Vandekerkhove, T.T.M.; Baggerman, G.; Schoofs, L.; Luyten, W.; Van Criekinge, W. Peptidomics Coming of Age: A Review of Contributions from a Bioinformatics Angle. *J. Proteome Res.* **2010**, *9*, 2051–2061, doi:10.1021/pr900929m.
89. Vizcaíno, J.A.; Kubiniok, P.; Kovalchik, K.A.; Ma, Q.; Rô, J.; Duquette, D.; Mongrain, I.; Deutsch, E.W.; Peters, B.; Sette, A.; et al. The Human Immunopeptidome Project: A Roadmap to Predict and Treat Immune Diseases In Brief. *Mol. Cell. Proteomics* **2020**, *19*, 31–49, doi:10.1074/mcp.R119.001743.
90. Admon, A.; Bassani-Sternberg, M. The Human Immunopeptidome Project, a Suggestion for yet Another Postgenome next Big Thing. *Mol. Cell. Proteomics* **2011**, *10*, doi:10.1074/MCP.O111.011833.
91. Mcgettigan, P.A. Transcriptomics in the RNA-Seq Era. *Chem. Biol.* **2013**, *17*, 4–11, doi:10.1016/j.cbpa.2012.12.008.
92. Cheung, P.K.; Ma, M.H.; Tse, H.F.; Yeung, K.F.; Tsang, H.F.; Chu, M.K.M.; Kan, C.M.; Cho, W.C.S.; Ng, L.B.W.; Chan, L.W.C.; et al. The Applications of Metabolomics in the Molecular Diagnostics of Cancer. *Expert Rev. Mol. Diagn.* **2019**, *19*, 785–793, doi:10.1080/14737159.2019.1656530.
93. Xiao, Y.; Ma, D.; Yang, Y.-S.; Yang, F.; Ding, J.-H.; Gong, Y.; Jiang, L.; Ge, L.-P.; Wu, S.-Y.; Yu, Q.; et al. Comprehensive Metabolomics Expands Precision Medicine for Triple-Negative Breast Cancer. *Cell Res.* **2022**, *32*, 477–490, doi:10.1038/s41422-022-00614-0.
94. Yeung, P.K. Metabolomics and Biomarkers for Drug Discovery. *Metabolites* **2018**, *8*, 1–3, doi:10.3390/metabo8010011.
95. Zhang, S.; Lee, H.; Chen, J.; Garmire, L.X.; Huang, S.; Chaudhary, K. More Is Better: Recent Progress in Multi-Omics Data Integration Methods. *Front. Genet.* **2017**, *1*, 84, doi:10.3389/fgene.2017.00084.
96. Quinn, R.A.; Navas-Molina, J.A.; Hyde, E.R.; Song, S.J.; Vázquez-Baeza, Y.; Humphrey, G.; Gaffney, J.; Minich, J.J.; Melnik, A. V.; Herschend, J.; et al. From Sample to Multi-Omics Conclusions in under 48 Hours. *mSystems* **2016**, *1*, doi:10.1128/MSYSTEMS.00038-16/FORMAT/EPUB.

97. Bechmann, N.; Watts, D.; Steenblock, C.; Wallace, P.W.; Schürmann, A.; Bornstein, S.R.; Wielockx, B.; Eisenhofer, G.; Peitzsch, M. Adrenal Hormone Interactions and Metabolism: A Single Sample Multi-Omics Approach. *Horm. Metab. Res.* **2021**, *53*, 326–334, doi:10.1055/a-1440-0278.
98. Bock, C.; Farlik, M.; Sheffield, N.C. Multi-Omics of Single Cells: Strategies and Applications. *Trends Biotechnol.* **2016**, *34*, 605–608, doi:10.1016/j.tibtech.2016.04.004.
99. Zielinski, J.M.; Luke, J.J.; Guglietta, S.; Krieg, C. High Throughput Multi-Omics Approaches for Clinical Trial Evaluation and Drug Discovery. *Front. Immunol.* **2021**, *12*, 783, doi:10.3389/fimmu.2021.590742.
100. Olivier, M.; Asmis, R.; Hawkins, G.A.; Howard, T.D.; Cox, L.A. The Need for Multi-Omics Biomarker Signatures in Precision Medicine. *Int. J. Mol. Sci.* **2019**, *20*, 1–13, doi:10.3390/ijms20194781.
101. Rose, G.D.; Fleming, P.J.; Banavar, J.R.; Maritan, A. A Backbone-Based Theory of Protein Folding. *Proc. Natl. Acad. Sci.* **2006**, *103*, 16623–16633, doi:10.1073/PNAS.0606843103.
102. Baldwin, R.L. Energetics of Protein Folding. *J. Mol. Biol.* **2007**, *371*, 283–301, doi:10.1016/J.JMB.2007.05.078.
103. Dahanayake, J.N.; Mitchell-Koch, K.R. How Does Solvation Layer Mobility Affect Protein Structural Dynamics? *Front. Mol. Biosci.* **2018**, *5*, 1–20, doi:10.3389/fmolb.2018.00065.
104. Davis, C.M.; Gruebele, M.; Sukenik, S. How Does Solvation in the Cell Affect Protein Folding and Binding? *Curr. Opin. Struct. Biol.* **2018**, *48*, 23–29, doi:10.1016/j.sbi.2017.09.003.
105. Debye, P.; Hückel, E.; by Michael Braus, T.J. Zur Theorie Der Elektrolyte. *Phys. Zeitschrift* **1923**, *9*, 185–206.
106. Otzen, D. Protein-Surfactant Interactions: A Tale of Many States. **2011**, doi:10.1016/j.bbapap.2011.03.003.
107. Otzen, D.E. Protein Unfolding in Detergents: Effect of Micelle Structure, Ionic Strength, PH, and Temperature. *Biophys. J.* **2002**, *83*, 2219–2230, doi:10.1016/S0006-3495(02)73982-9.
108. Cox, R.A.; Tanford, C. The Hydrophobic Effect: Formation of Micelles and Biological Membranes. *J. Chem. Educ.* **1980**, *232*, A246.
109. Ibel, K.; May, R.P.; Kirschner, K.; Szadkowski, H.; Mashcer, E.; Lundahl, P. Protein-decorated Micelle Structure of Sodium-dodecyl-sulfate–Protein Complexes as Determined by Neutron Scattering. *Eur. J. Biochem.* **1990**, *190*, 311–318, doi:10.1111/J.1432-1033.1990.TB15578.X.

110. Guo, X.H.; Zhao, N.M.; Chen, S.H.; Teixeira, J. Small-Angle Neutron Scattering Study of the Structure of Protein / Detergent Complexes. *Biopolymers* **1990**, *29*, 335–346, doi:10.1002/BIP.360290206.
111. Guo, X.H.; Chen, S.H. The Structure and Thermodynamics of Protein-SDS Complexes in Solution and the Mechanism of Their Transports in Gel Electrophoresis Process. *Chem. Phys.* **1990**, *149*, 129–139, doi:10.1016/0301-0104(90)80134-J.
112. Winogradoff, D.; John, S.; Aksimentiev, A. Protein Unfolding by SDS: The Microscopic Mechanisms and the Properties of the SDS-Protein Assembly. *Nanoscale* **2020**, *12*, 5422–5434, doi:10.1039/c9nr09135a.
113. Rundlett, K.L.; Armstrong, D.W. Mechanism of Signal Suppression by Anionic Surfactants in Capillary Electrophoresis-Electrospray Ionization Mass Spectrometry. *Anal. Chem* **1996**, *68*, 376.
114. Kebarle, P.; Tang, L. From Ions in Solution to Ions in the Gas Phase The Mechanism of Electrospray Mass Spectrometry. *Anal. Chem.* **1993**, *65*, 972–986, doi:10.1021/ac00070a001.
115. Botelho, D.; Wall, M.J.; Vieira, D.B.; Fitzsimmons, S.; Liu, F.; Doucette, A. Top-Down and Bottom-Up Proteomics of SDS-Containing Solutions Following Mass-Based Separation. *J. Proteome Res.* **2010**, *9*, 2863–2870, doi:10.1021/pr900949p.
116. Donnelly, D.P.; Rawlins, C.M.; DeHart, C.J.; Fornelli, L.; Schachner, L.F.; Lin, Z.; Lippens, J.L.; Aluri, K.C.; Sarin, R.; Chen, B.; et al. Best Practices and Benchmarks for Intact Protein Analysis for Top-down Mass Spectrometry. *Nat. Methods* **2019**, *16*, 587–594, doi:10.1038/s41592-019-0457-0.
117. Seddon, A.M.; Curnow, P.; Booth, P.J. Membrane Proteins, Lipids and Detergents: Not Just a Soap Opera. *Biochim. Biophys. Acta* **2004**, *1666*, 105–117, doi:10.1016/j.bbamem.2004.04.011.
118. Li, J.K.K.; Johnson, T.; Yang, Y.-Y.; Shore, V. Selective Separation of Virus Proteins and Double-Stranded RNAs by SDS-KC1 Precipitation. *J. Virol. Methods* **1989**, *26*, 3–16.
119. Jenö, P.; Scherer, P.E.; Manning-Krieg, U.; Horst, M. Desalting Electroeluted Proteins with Hydrophilic Interaction Chromatography. *Anal. Biochem.* **1993**, *215*, 292–298, doi:10.1006/abio.1993.1589.
120. Han, D.K.; Eng, J.; Zhou, H.; Aebersold, R. Quantitative Profiling of Differentiation-Induced Microsomal Proteins Using Isotope-Coded Affinity Tags and Mass Spectrometry. *Nat. Biotechnol.* **2001**, *19*, 946–951.
121. Kunz, W.; Henle, J.; Ninham, B.W. “Zur Lehre von Der Wirkung Der Salze” (about the Science of the Effect of Salts): Franz Hofmeister’s Historical Papers. *Curr. Opin. Colloid Interface Sci.* **2004**, *9*, 19–37, doi:10.1016/j.cocis.2004.05.005.

122. Hyde, A.M.; Zultanski, S.L.; Waldman, J.H.; Zhong, Y.-L.; Shevlin, M.; Peng, F. General Principles and Strategies for Salting-out Informed by the Hofmeister Series. *Org. Process Res. Dev.* **2017**, *21*, 1355–1370, doi:10.1021/acs.oprd.7b00197.
123. Cohn, E.J.; Strong, L.E.; Hughes, W.L.; Mulford, D.J.; Ashworth, J.N.; Melin, M.; Taylor, H.L. Preparation and Properties of Serum and Plasma Proteins. IV. A System for the Separation into Fractions of the Protein and Lipoprotein Components of Biological Tissues and Fluids ^{1a,b,c,D}. *J. Am. Chem. Soc.* **1946**, *68*, 459–475, doi:10.1021/ja01207a034.
124. Mellanby, J. The Precipitation of the Proteins of Horse Serum. *J. Physiol.* **1907**, *36*, 288–333, doi:10.1113/jphysiol.1907.sp001234.
125. Wang, P.; Anderko, A. Computation of Dielectric Constants of Solvent Mixtures and Electrolyte Solutions. *Fluid Phase Equilib.* **2001**, *186*, 103–122, doi:10.1016/S0378-3812(01)00507-6.
126. Fuoss, R.M.; Kraus, C.A. Properties of Electrolytic Solutions. III. The Dissociation Constant. *J. Am. Chem. Soc.* **1933**, *55*, 1019–1028.
127. Barritault, D.; Expert-Bezançon, A.; Guérin, M. -F; Hayes, D. The Use of Acetone Precipitation in the Isolation of Ribosomal Proteins. *Eur. J. Biochem.* **1976**, *63*, 131–135, doi:10.1111/j.1432-1033.1976.tb10215.x.
128. Thongboonkerd, V.; McLeish, K.R.; Arthur, J.M.; Klein, J.B. Proteomic Analysis of Normal Human Urinary Proteins Isolated by Acetone Precipitation or Ultracentrifugation. *Kidney Int.* **2002**, *62*, 1461–1469, doi:10.1046/j.1523-1755.2002.00565.x.
129. Jiang, L.; He, L.; Fountoulakis, M. Comparison of Protein Precipitation Methods for Sample Preparation Prior to Proteomic Analysis. *J. Chromatogr. A* **2004**, *1023*, 317–320, doi:10.1016/j.chroma.2003.10.029.
130. Fic, E.; Kedracka-Krok, S.; Jankowska, U.; Pirog, A.; Dziedzicka-Wasylewska, M. Comparison of Protein Precipitation Methods for Various Rat Brain Structures Prior to Proteomic Analysis. *Electrophoresis* **2010**, *31*, 3573–3579, doi:10.1002/elps.201000197.
131. Santa, C.; Anjo, S.I.; Manadas, B. Protein Precipitation of Diluted Samples in SDS-Containing Buffer with Acetone Leads to Higher Protein Recovery and Reproducibility in Comparison with TCA/Acetone Approach. *Proteomics* **2016**, *16*, 1847–1851, doi:10.1002/pmic.201600024.
132. Crowell, A.M.J.; Wall, M.J.; Doucette, A.A. Maximizing Recovery of Water-Soluble Proteins through Acetone Precipitation. *Anal. Chim. Acta* **2013**, *796*, 48–54, doi:10.1016/j.aca.2013.08.005.
133. Kachuk, C.; Stephen, K.; Doucette, A. Comparison of Sodium Dodecyl Sulfate Depletion Techniques for Proteome Analysis by Mass Spectrometry. *J. Chromatogr. A* **2015**, *1418*, 158–166, doi:10.1016/j.chroma.2015.09.042.

134. Pérez-Rodríguez, S.; Ramírez, O.T.; Trujillo-Roldán, M.A.; Valdez-Cruz, N.A. Comparison of Protein Precipitation Methods for Sample Preparation Prior to Proteomic Analysis of Chinese Hamster Ovary Cell Homogenates. *Electron. J. Biotechnol.* **2020**, *48*, 86–94, doi:10.1016/j.ejbt.2020.09.006.
135. Nickerson, J.L.; Doucette, A.A. Rapid and Quantitative Protein Precipitation for Proteome Analysis by Mass Spectrometry. *J. Proteome Res.* **2020**, *19*, 2035–2042, doi:10.1021/acs.jproteome.9b00867.
136. Baghalabadi, V.; Doucette, A.A. Mass Spectrometry Profiling of Low Molecular Weight Proteins and Peptides Isolated by Acetone Precipitation. *Anal. Chim. Acta* **2020**, *1138*, 38–48, doi:10.1016/j.aca.2020.08.057.
137. Zhang, Y.; Fonslow, B.R.; Shan, B.; Baek, M.-C.; Yates, J.R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013**, *113*, 2343–2394, doi:10.1021/cr3003533.
138. Vandermarliere, E.; Mueller, M.; Martens, L. Getting Intimate with Trypsin, the Leading Protease in Proteomics. *Mass Spectrom. Rev.* **2013**, *32*, 453–465, doi:10.1002/mas.21376.
139. Keil, B. *Specificity of Proteolysis*; Springer Berlin Heidelberg, 1992;
140. Huber, R.; Bode, W. Structural Basis of the Activation and Action of Trypsin. *Acc. Chem. Res.* **1978**, *11*, 114–122, doi:10.1021/ar50123a006.
141. Mares-Guiai, M.; Shaw, E. Studies on the Active Center of Trypsin The Binding of Amidine and Guanidines as Models of the Substrate Side Chain. *J. Biol. Chem.* **1965**, *240*, 1579–1585, doi:10.1016/S0021-9258(18)97474-0.
142. Inagami, T.; Mitsuda, H. The Mechanism of the Specificity of Trypsin Catalysis. II. Comparison of Trypsin and Alpha-Chymotrypsin in the Nonspecific Catalyses of the Hydrolysis of Acetylglycine Ethyl Ester. *J. Biol. Chem.* **1964**, *239*, 1388–1394, doi:10.1016/s0021-9258(18)9136-8.
143. Weiner, S.J.; Seibel, G.L.; Kollman, P.A. The Nature of Enzyme Catalysis in Trypsin. *Proc. Natl. Acad. Sci.* **1986**, *83*, 649–653, doi:10.1073/PNAS.83.3.649.
144. Hedstrom, L. Serine Protease Mechanism and Specificity. *Chem. Rev.* **2002**, *102*, 4501–4523, doi:10.1021/cr000033x.
145. Dufton, M.J. Could Domain Movements Be Involved in the Mechanism of Trypsin-like Serine Proteases? *FEBS Lett.* **1990**, *271*, 9–13, doi:10.1016/0014-5793(90)80360-U.
146. Radisky, E.S.; Lee, J.M.; Lu, C.J.K.; Koshland, D.E. Insights into the Serine Protease Mechanism from Atomic Resolution Structures of Trypsin Reaction Intermediates. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 6835–6840, doi:10.1073/PNAS.0601910103.

147. Kayode, O.; Wang, R.; Pendlebury, D.F.; Cohen, I.; Henin, R.D.; Hockla, A.; Soares, A.S.; Papo, N.; Caulfield, T.R.; Radisky, E.S. An Acrobatic Substrate Metamorphosis Reveals a Requirement for Substrate Conformational Dynamics in Trypsin Proteolysis. *J. Biol. Chem.* **2016**, *291*, 26304–26319, doi:10.1074/jbc.M116.758417.
148. Swaney, D.L.; Wenger, C.D.; Coon, J.J. Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics. *J. Proteome Res.* **2010**, *9*, 1323–1329, doi:10.1021/pr900863u.
149. Rovey, M. Chromatography of Trypsin to Remove Chymotrypsin, and of Chymotrypsin to Remove Trypsin. *Methods Enzymol.* **1967**, *11*, 231–236, doi:10.1016/S0076-6879(67)11027-6.
150. Titani, K.; Sasagawa, T.; Resing, K.; Walsh, K.A. A Simple and Rapid Purification of Commercial Trypsin and Chymotrypsin by Reverse-Phase High-Performance Liquid Chromatography. *Anal. Biochem.* **1982**, *123*, 408–412, doi:10.1016/0003-2697(82)90465-1.
151. Carpenter, F.H. Treatment of Trypsin with TPCK. *Methods Enzymol.* **1967**, *11*, 237, doi:10.1016/S0076-6879(67)11028-8.
152. Kostka, V.; Carpenter, F.H. Inhibition of Chymotrypsin Activity in Crystalline Trypsin Preparations. *J. Biol. Chem.* **1964**, *239*, 1799–1803.
153. Deng, Y.; van der Veer, F.; Sforza, S.; Gruppen, H.; Wierenga, P.A. Towards Predicting Protein Hydrolysis by Bovine Trypsin. *Process Biochem.* **2018**, *65*, 81–92, doi:10.1016/j.procbio.2017.11.006.
154. Brownridge, P.; Beynon, R.J. The Importance of the Digest: Proteolysis and Absolute Quantification in Proteomics. *Methods* **2011**, *54*, 351–360, doi:10.1016/j.ymeth.2011.05.005.
155. Thiede, B.; Lamer, S.; Mattow, J.; Siejak, F.; Dimmler, C.; Rudel, T.; Jungblut, P.R. Analysis of Missed Cleavage Sites, Tryptophan Oxidation and N-Terminal Pyroglutamylation after in-Gel Tryptic Digestion. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 496–502, doi:10.1002/(SICI)1097-0231(20000331)14:6.
156. Olsen, J. V.; Ong, S.E.; Mann, M. Trypsin Cleaves Exclusively C-Terminal to Arginine and Lysine Residues. *Mol. Cell. Proteomics* **2004**, *3*, 608–614, doi:10.1074/mcp.T400003-MCP200.
157. Rodriguez, J.; Gupta, N.; Smith, R.D.; Pevzner, P.A. Does Trypsin Cut Before Proline? *J. Proteome Res.* **2008**, 300–305, doi:10.1021/pr0705035.

158. Yen, C.Y.; Russell, S.; Mendoza, A.M.; Meyer-Arendt, K.; Sun, S.; Cios, K.J.; Ahn, N.G.; Resing, K.A. Improving Sensitivity in Shotgun Proteomics Using a Peptide-Centric Database with Reduced Complexity: Protease Cleavage and SCX Elution Rules from Data Mining of MS/MS Spectra. *Anal. Chem.* **2006**, *78*, 1071–1084, doi:10.1021/AC051127F/SUPPL_FILE/AC051127FSI20051204_100357.XLS.
159. Lawless, C.; Hubbard, S.J. Prediction of Missed Proteolytic Cleavages for the Selection of Surrogate Peptides for Quantitative Proteomics. *Omi. A J. Integr. Biol.* **2012**, *16*, 449–456, doi:10.1089/OMI.2011.0156/FORMAT/EPUB.
160. Walmsley, S.J.; Rudnick, P.A.; Liang, Y.; Dong, Q.; Stein, S.E.; Nesvizhskii, A.I. Comprehensive Analysis of Protein Digestion Using Six Trypsins Reveals the Origin of Trypsin As a Significant Source of Variability in Proteomics. *J. Proteome Res.* **2013**, *12*, 5666–5680, doi:10.1021/pr400611h.
161. Šlechtová, T.; Gilar, M.; Kalíková, K.; Tesařová, E. Insight into Trypsin Miscleavage: Comparison of Kinetic Constants of Problematic Peptide Sequences. *Anal. Chem.* **2015**, *87*, 7636–7643, doi:10.1021/acs.analchem.5b00866.
162. Havliš, J.; Thomas, H.; Šebela, M.; Shevchenko, A. Fast-Response Proteomics by Accelerated in-Gel Digestion of Proteins. *Anal. Chem.* **2003**, *75*, 1300–1306, doi:10.1021/ac026136s.
163. Finehout, E.J.; Cantor, J.R.; Lee, K.H. Kinetic Characterization of Sequencing Grade Modified Trypsin. *Proteomics* **2005**, *5*, 2319–2321, doi:10.1002/PMIC.200401268.
164. Heissel, S.; Frederiksen, S.J.; Bunkenborg, J.; Hojrup, P. Enhanced Trypsin on a Budget: Stabilization, Purification and High-Temperature Application of Inexpensive Commercial Trypsin for Proteomics Applications. *PLoS One* **2019**, 1–16.
165. Takemori, A.; Ishizaki, J.; Nakashima, K.; Shibata, T.; Kato, H.; Kodera, Y.; Suzuki, T.; Hasegawa, H.; Takemori, N. BAC-DROP: Rapid Digestion of Proteome Fractionated via Dissolvable Polyacrylamide Gel Electrophoresis and Its Application to Bottom-Up Proteomics Workflow. *J. Proteome Res.* **2021**, *20*, 1535–1543, doi:10.1021/acs.jproteome.0c00749.
166. Trampari, S.; Papagiannopoulos, A.; Pispas, S. Biochemical and Biophysical Research Communications Temperature-Induced Aggregation Behavior in Bovine Pancreas Trypsin Solutions. *Biochem. Biophys. Res. Commun.* **2019**, *515*, 282–288, doi:10.1016/j.bbrc.2019.05.124.
167. López-Ferrer, D.; Petritis, K.; Hixson, K.K.; Heibeck, T.H.; Moore, R.J.; Belov, M.E.; Camp, D.G.; Smith, R.D. Application of Pressurized Solvents for Ultrafast Trypsin Hydrolysis in Proteomics: Proteomics on the Fly. *J. Proteome Res.* **2008**, *7*, 3276–3281, doi:10.1021/pr7008077.

168. Yang, H.J.; Hong, J.; Lee, S.; Shin, S.; Kim, J.; Kim, J. Pressure-Assisted Tryptic Digestion Using a Syringe. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 901–908, doi:10.1002/RCM.4467.
169. Ruan, K.; Lange, R.; Bec, N.; Balny, C. A Stable Partly Denatured State of Trypsin Induced by High Hydrostatic Pressure. *Biochem. Biophys. Res. Commun.* **1997**, *239*, 150–154.
170. Lee, B.; Lopez-Ferrer, D.; Kim, B.C.; Na, H. Bin; Park, Y. Il; Weitz, K.K.; Warner, M.G.; Hyeon, T.; Lee, S.W.; Smith, R.D.; et al. Rapid and Efficient Protein Digestion Using Trypsin-Coated Magnetic Nanoparticles under Pressure Cycles. *Proteomics* **2011**, *11*, 309–318, doi:10.1002/pmic.201000378.
171. López-Ferrer, D.; Capelo, J.L.; Vázquez, J. Ultra Fast Trypsin Digestion of Proteins by High Intensity Focused Ultrasound. *J. Proteome Res.* **2005**, *4*, 1569–1574, doi:10.1021/pr050112v.
172. López-Ferrer, D.; Hixson, K.K.; Smallwood, H.; Squier, T.C.; Petritis, K.; Smith, R.D. Evaluation of a High-Intensity Focused Ultrasound-Immobilized Trypsin Digestion and O-Labeling Method for Quantitative Proteomics. *Anal. Chem.* **2009**, *81*, 6272–6277, doi:10.1021/ac802540s.
173. Carrera, M.; Benito, C.; Lopez-Ferrer, D.; Pineiro, C.; Vazquez, J.; Gallardo, J.M. Fast Monitoring of Species-Specific Peptide Biomarkers Using High-Intensity-Focused-Ultrasound-Assisted Tryptic Digestion and Selected MS/MS Ion Monitoring. *Anal. Chem.* **2011**, *83*, 5688–5695, doi:10.1021/ac200890w.
174. Dycka, F.; Bobal, P.; Mazanec, K.; Bobalova, J. Rapid and Efficient Protein Enzymatic Digestion: An Experimental Comparison. *Electrophoresis* **2012**, *33*, 288–295, doi:10.1002/ELPS.201100123.
175. Pramanik, B.N.; Mirza, U.A.; Ing, Y.H.; Liu, Y.-H.; Bartner, P.L.; Weber, P.C.; Bose, A.K. Microwave-Enhanced Enzyme Reaction for Protein Mapping by Mass Spectrometry: A New Approach to Protein Digestion in Minutes. *Protein Sci.* **2009**, *11*, 2676–2687, doi:10.1110/ps.0213702.
176. Sun, W.; Gao, S.; Wang, L.; Chen, Y.; Wu, S.; Wang, X.; Zheng, D.; Gao, Y. Microwave-Assisted Protein Preparation and Enzymatic Digestion in Proteomics. *Mol. Cell. Proteomics* **2006**, *5*, 769–776, doi:10.1074/mcp.T500022-MCP200.
177. Ha, N.Y.; Kim, S.H.; Lee, T.G.; Han, S.Y. Rapid Characterization of Protein Chips Using Microwave-Assisted Protein Tryptic Digestion and MALDI Mass Spectrometry. *Langmuir* **2011**, *27*, 10098–10105, doi:10.1021/la201812a.
178. Lin, S.; Yao, G.; Qi, D.; Li, Y.; Deng, C.; Yang, P.; Zhang, X. Fast and Efficient Proteolysis by Microwave-Assisted Protein Digestion Using Trypsin-Immobilized Magnetic Silica Microspheres. *Anal. Chem.* **2008**, *80*, 3655–3665, doi:10.1021/ac800023r.

179. Lin, S.; Yun, D.; Qi, D.; Deng, C.; Li, Y.; Zhang, X. Novel Microwave-Assisted Digestion by Trypsin-Immobilized Magnetic Nanoparticles for Proteomic Analysis. *J. Proteome Res.* **2008**, *7*, 1297–1307, doi:10.1021/pr700586j.
180. Atacan, K.; Kursunlu, A.N.; Ozmen, M. Preparation of Pillar[5]Arene Immobilized Trypsin and Its Application in Microwave-Assisted Digestion of Cytochrome C. *Mater. Sci. Eng. C* **2019**, *94*, 886–893, doi:10.1016/j.msec.2018.10.043.
181. Zhao, Q.; Fang, F.; Wu, C.; Wu, Q.; Liang, Y.; Liang, Z.; Zhang, L.; Zhang, Y. ImFASP: An Integrated Approach Combining in-Situ Filter-Aided Sample Pretreatment with Microwave-Assisted Protein Digestion for Fast and Efficient Proteome Sample Preparation. *Anal. Chim. Acta* **2016**, *912*, 58–64, doi:10.1016/j.aca.2016.01.049.
182. Hahn, H.W.; Rainer, M.; Ringer, T.; Huck, C.W.; Bonn, G.K. Ultrafast Microwave-Assisted in-Tip Digestion of Proteins. *J. Proteome Res.* **2009**, *8*, 4225–4230, doi:10.1021/pr900188x.
183. Juan, H.F.; Chang, S.C.; Huang, H.C.; Chen, S.T. A New Application of Microwave Technology to Proteomics. *Proteomics* **2005**, *5*, 840–842, doi:10.1002/pmic.200401056.
184. Reddy, P.M.; Hsu, W.Y.; Hu, J.F.; Ho, Y.P. Digestion Completeness of Microwave-Assisted and Conventional Trypsin-Catalyzed Reactions. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 421–424, doi:10.1016/j.jasms.2009.11.006.
185. Damm, M.; Nussold, C.; Cantillo, D.; Rechberger, G.N.; Gruber, K.; Sattler, W.; Kappe, C.O. Can Electromagnetic Fields Influence the Structure and Enzymatic Digest of Proteins? A Critical Evaluation of Microwave-Assisted Proteomics Protocols. *J. Proteomics* **2012**, *75*, 5533–5543, doi:10.1016/j.jprot.2012.07.043.
186. Min Tian, Z.; Xi Wan, M.; Pin Wang, S.; Qing Kang, J. Effects of Ultrasound and Additives on the Function and Structure of Trypsin. *Ultrason. Chem.* **2004**, *11*, 399–404, doi:10.1016/j.ultsonch.2003.09.004.
187. Zhang, N.; Li, L. Effects of Common Surfactants on Protein Digestion and Matrix-Assisted Laser Desorption/Ionization Mass Spectrometric Analysis of the Digested Peptides Using Two-Layer Sample Preparation. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 889–896, doi:10.1002/RCM.1423.
188. Zhang, N.; Chen, R.; Young, N.; Wishart, D.; Winter, P.; Weiner, J.H.; Li, L. Comparison of SDS- and Methanol-Assisted Protein Solubilization and Digestion Methods for Escherichia Coli Membrane Proteome Analysis by 2-D LC-MS/MS. *Proteomics* **2007**, *7*, 484–493, doi:10.1002/pmic.200600518.
189. Yu, Y.Q.; Gilar, M.; Lee, P.J.; Bouvier, E.S.P.; Gebler, J.C. Enzyme-Friendly, Mass Spectrometry-Compatible Surfactant for In-Solution Enzymatic Digestion of Proteins. *Anal. Chem.* **2003**, *75*, 6023–6028, doi:10.1021/ac0346196.

190. Takeda, K.; Moriyama, Y. Kinetic Aspects of Surfactant-Induced Structural Changes of Proteins-Unsolved Problems of Two-State Model for Protein Denaturation. *J. Oleo Sci.* **2015**, *64*, 1143–1158, doi:10.5650/jos.ess15157.
191. Ma, H.; Zou, T.; Li, H.; Cheng, H. The Interaction of Sodium Dodecyl Sulfate with Trypsin: Multi-Spectroscopic Analysis, Molecular Docking, and Molecular Dynamics Simulation. *Int. J. Biol. Macromol.* **2020**, *162*, 1546–1554, doi:10.1016/j.ijbiomac.2020.08.020.
192. Jian Zhou; Tieyang Zhou; Rui Cao; Zhen Liu; Jianying Shen; Ping Chen; Xianchun Wang, * and; Liang*, S. Evaluation of the Application of Sodium Deoxycholate to Proteomic Analysis of Rat Hippocampal Plasma Membrane. *J. Proteome Res.* **2006**, *5*, 2547–2553, doi:10.1021/PR060112A.
193. Masuda, T.; Tomita, M.; Ishihama, Y. Phase Transfer Surfactant-Aided Trypsin Digestion for Membrane Proteome Analysis. *J. Proteome Res.* **2008**, *7*, 731–740, doi:10.1021/pr700658q.
194. Lin, Y.; Zhou, J.; Bi, D.; Chen, P.; Wang, X.; Liang, S. Sodium-Deoxycholate-Assisted Tryptic Digestion and Identification of Proteolytically Resistant Proteins. *Anal. Biochem.* **2008**, *377*, 259–266, doi:10.1016/j.ab.2008.03.009.
195. Lin, Y.; Liu, H.; Liu, Z.; Liu, Y.; He, Q.; Chen, P.; Wang, X.; Liang, S. Development and Evaluation of an Entirely Solution-Based Combinative Sample Preparation Method for Membrane Proteomics. *Anal. Biochem.* **2013**, *432*, 41–48, doi:10.1016/j.ab.2012.09.023.
196. Lin, Y.; Wang, K.; Liu, Z.; Lin, H.; Yu, L. Enhanced SDC-Assisted Digestion Coupled with Lipid Chromatography-Tandem Mass Spectrometry for Shotgun Analysis of Membrane Proteome. *J. Chromatogr. B* **2015**, *1002*, 144–151, doi:10.1016/j.jchromb.2015.08.019.
197. Shahinuzzaman, A.D.A.; Chakrabarty, J.K.; Fang, Z.; Smith, D.; Kamal, A.H.M.; Chowdhury, S.M. Improved In-Solution Trypsin Digestion Method for Methanol–Chloroform Precipitated Cellular Proteomics Sample. *J. Sep. Sci.* **2020**, *43*, 2125–2132, doi:10.1002/jssc.201901273.
198. Moore, S.M.; Hess, S.M.; Jorgenson, J.W. Extraction, Enrichment, Solubilization, and Digestion Techniques for Membrane Proteomics. *J. Proteome Res.* **2016**, *15*, 1243–1252, doi:10.1021/ACS.JPROTEOME.5B01122.
199. Banazadeh, A.; Veillon, L.; Wooding, K.M.; Zabet-moghaddam, M.; Mechref, Y. Recent Advances in Mass Spectrometric Analysis of Glycoproteins. *Electrophoresis* **2017**, *38*, 162–189, doi:10.1002/ELPS.201600357.
200. Norrgran, J.; Williams, T.L.; Woolfitt, A.R.; Solano, M.I.; Pirkle, J.L.; Barr, J.R. Optimization of Digestion Parameters for Protein Quantification. *Anal. Biochem.* **2009**, *393*, 48–55, doi:10.1016/j.ab.2009.05.050.

201. Pop, C.; Mogosan, C.; Loghin, F. Evaluation of RapiGest Efficacy for the Digestion of Proteins from Cell Cultures and Heart Tissue. *Pharm. Clujul Med.* **2014**, *87*, 258–262, doi:10.15386/cjmed-367.
202. Chang, Y.H.; Gregorich, Z.R.; Chen, A.J.; Hwang, L.; Guner, H.; Yu, D.; Zhang, J.; Ge, Y. New Mass-Spectrometry-Compatible Degradable Surfactant for Tissue Proteomics. *J. Proteome Res.* **2015**, *14*, 1587–1599, doi:10.1021/pr5012679.
203. Brown, K.A.; Chen, B.; Guardado-Alvarez, T.M.; Lin, Z.; Hwang, L.; Ayaz-Guner, S.; Jin, S.; Ge, Y. A Photocleavable Surfactant for Top-down Proteomics. *Nat. Methods* **2019**, *16*, 417–420, doi:10.1038/s41592-019-0391-1.
204. Brown, K.A.; Tucholski, T.; Eken, C.; Knott, S.; Zhu, Y.; Jin, S.; Ge, Y. High-Throughput Proteomics Enabled by a Photocleavable Surfactant. *Angew. Chemie* **2020**, *132*, 8484–8488, doi:10.1002/ANGE.201915374.
205. Knott, S.J.; Brown, K.A.; Josyer, H.; Carr, A.; Inman, D.; Jin, S.; Friedl, A.; Ponik, S.M.; Ge, Y. Photocleavable Surfactant-Enabled Extracellular Matrix Proteomics. *Anal. Chem.* **2020**, *92*, 15693–15698, doi:10.1021/acs.analchem.0c03104.
206. Buck, K.M.; Roberts, D.S.; Aballo, T.J.; Inman, D.R.; Jin, S.; Ponik, S.; Brown, K.A.; Ge, Y. One-Pot Exosome Proteomics Enabled by a Photocleavable Surfactant. *Anal. Chem.* **2022**, *94*, 7164–7168, doi:10.1021/acs.analchem.2c01252.
207. Aballo, T.J.; Roberts, D.S.; Melby, J.A.; Buck, K.M.; Brown, K.A.; Ge, Y. Ultrafast and Reproducible Proteomics from Small Amounts of Heart Tissue Enabled by Azo and TimsTOF Pro. *J. Proteome Res.* **2021**, *20*, 4203–4211, doi:10.1021/acs.jproteome.1c00446.
208. Shin, H.W.; Nguyen, T. V.; Jung, J.Y.; Lee, G.H.; Jang, W.; Yoon, J.; Kim, S.T.; Min, C.W.; Gupta, R.; Won, H.; et al. Application of Mass-Spectrometry Compatible Photocleavable Surfactant for next-Generation Proteomics Using Rice Leaves. *J Plant Biotechnol* **2021**, *48*, 165–172, doi:10.5010/jpb.2021.48.3.165.
209. Proc, J.L.; Kuzyk, M.A.; Hardie, D.B.; Yang, J.; Smith, D.S.; Jackson, A.M.; Parker, C.E.; Borchers, C.H. A Quantitative Study of the Effects of Chaotropic Agents, Surfactants, and Solvents on the Digestion Efficiency of Human Plasma Proteins by Trypsin. *J. Proteome Res.* **2010**, *9*, 5422–5437, doi:10.1021/pr100656u.
210. Poulsen, J.W.; Madsen, C.T.; Young, C.; Poulsen, F.M.; Nielsen, M.L. Using Guanidine-Hydrochloride for Fast and Efficient Protein Digestion and Single-Step Affinity-Purification Mass Spectrometry. *J. Proteome Res.* **2012**, *12*, 1020–1030, doi:10.1021/PR300883Y.
211. Betancourt, L.H.; Sanchez, A.; Pla, I.; Kuras, M.; Zhou, Q.; Andersson, R.; Marko-Varga, G. Quantitative Assessment of Urea In-Solution Lys-C/Trypsin Digestions Reveals Superior Performance at Room Temperature over Traditional Proteolysis at 37 °C. *J. Proteome Res.* **2018**, *17*, 2556–2561, doi:10.1021/acs.jproteome.8b00228.

212. Dušeková, E.; Garajová, K.; Yavaşer, R.; Tomková, M.; Sedláková, D.; Dzurillová, V.; Kulik, N.; Fadaei, F.; Shaposhnikova, A.; Minofar, B.; et al. Modulation of Global Stability, Ligand Binding and Catalytic Properties of Trypsin by Anions. *Biophys. Chem.* **2022**, *288*, doi:10.1016/j.bpc.2022.106856.
213. Johnston, H.E.; Yadav, K.; Kirkpatrick, J.M.; Biggs, G.S.; Oxley, D.; Kramer, H.B.; Samant, R.S. Solvent Precipitation SP3 (SP4) Enhances Recovery for Proteomics Sample Preparation without Magnetic Beads. *Anal. Chem.* **2022**, *94*, 10320–10328, doi:10.1021/acs.analchem.1c04200.
214. Wessel, D.; Flügge, U.I. A Method for the Quantitative Recovery of Protein in Dilute Solution in the Presence of Detergents and Lipids. *Anal. Biochem.* **1984**, *138*, 141–143, doi:10.1016/0003-2697(84)90782-6.
215. Mattos, C.; Ringe, D. Proteins in Organic Solvents. *Curr. Opin. Struct. Biol.* **2001**, *11*, 761–764.
216. Russell, W.K.; Park, Z.Y.; Russell, D.H. Proteolysis in Mixed Organic-Aqueous Solvent Systems: Applications for Peptide Mass Mapping Using Mass Spectrometry. *Anal. Chem.* **2001**, *73*, 2682–2685, doi:10.1021/ac001332p.
217. Strader, M.B.; Tabb, D.L.; Hervey, W.J.; Pan, C.; Hurst, G.B. Efficient and Specific Trypsin Digestion of Microgram to Nanogram Quantities of Proteins in Organic-Aqueous Solvent Systems. *Anal. Chem.* **2006**, *78*, 125–134, doi:10.1021/ac051348l.
218. Hervey IV, W.J.; Strader, M.B.; Hurst, G.B. Comparison of Digestion Protocols for Microgram Quantities of Enriched Protein Samples. *J. Proteome Res.* **2007**, *6*, 3054–3061, doi:10.1021/pr070159b.
219. Liu, C.-C.; Liang, L.-H.; Yang, Y.; Yu, H.-L.; Yan, L.; Li, X.-S.; Chen, B.; Liu, S.-L.; Xi, H.-L. Direct Acetonitrile-Assisted Trypsin Digestion Method Combined with LC–MS/MS-Targeted Peptide Analysis for Unambiguous Identification of Intact Ricin. *J. Proteome Res.* **2020**, *20*, 369–380, doi:10.1021/acs.jproteome.0C00458.
220. Östin, A.; Tomas Bergström; Sten-Åke Fredriksson, A.; Nilsson, C. Solvent-Assisted Trypsin Digestion of Ricin for Forensic Identification by LC-ESI MS/MS. *Anal. Chem.* **2007**, *79*, 6271–6278, doi:10.1021/AC0701740.
221. Wall, M.J.; Crowell, A.M.J.; Simms, G.A.; Liu, F.; Doucette, A.A. Implications of Partial Tryptic Digestion in Organic-Aqueous Solvent Systems for Bottom-up Proteome Analysis. *Anal. Chim. Acta* **2011**, *703*, 194–203, doi:10.1016/j.aca.2011.07.025.
222. Sproß, J.; Yamashita, Y.; Gröger, H. Learning about Enzyme Stability against Organic Cosolvents from Structural Insights by Ion Mobility Mass Spectrometry. *ChemBioChem* **2020**, *21*, 1968–1971, doi:10.1002/cbic.201900648.

223. Espinosa, L.A.; Ramos, Y.; Andújar, I.; Torres, E.O.; Cabrera, G.; Martín, A.; Roche, D.; China, G.; Becquet, M.; González, I.; et al. In-Solution Buffer-Free Digestion Allows Full-Sequence Coverage and Complete Characterization of Post-Translational Modifications of the Receptor-Binding Domain of SARS-CoV-2 in a Single ESI-MS Spectrum. *Anal. Bioanal. Chem.* **2021**, *413*, 7559–7585, doi:10.1007/S00216-021-03721.
224. Castaneda-Agullo, M.; Del Castillo, L.M. The Influence of Medium Dielectric Strength Upon Trypsin Kinetics. *J. Gen. Physiol.* **1958**, *42*, 617–634, doi:10.1085/jgp.42.3.617.
225. Shen, X.; Sun, L. Systematic Evaluation of Immobilized Trypsin-Based Fast Protein Digestion for Deep and High-Throughput Bottom-Up Proteomics. *Proteomics* **2018**, *18*, 1–8, doi:10.1002/pmic.201700432.
226. Kecskemeti, A.; Gaspar, A. Particle-Based Immobilized Enzymatic Reactors in Microfluidic Chips. *Talanta* **2017**, *180*, 211–228, doi:10.1016/j.talanta.2017.12.043.
227. Massolini, G.; Calleri, E. Immobilized Trypsin Systems Coupled On-Line to Separation Methods: Recent Developments and Analytical Applications. *J. Sep. Sci.* **2005**, *28*, 7–21, doi:10.1002/JSSC.200401941.
228. Zhu, Y.; Chen, Q.; Shao, L.; Jia, Y.; Zhang, X. Microfluidic Immobilized Enzyme Reactors for Continuous Biocatalysis. *React. Chem. Eng.* **2020**, *5*, 9–32, doi:10.1039/c9re00217k.
229. Treetharnmathurot, B.; Ovartharnporn, C.; Wungsintaweekul, J.; Duncan, R.; Wiwattanapatapee, R. Effect of PEG Molecular Weight and Linking Chemistry on the Biological Activity and Thermal Stability of PEGylated Trypsin. *Int. J. Pharm.* **2008**, *357*, 252–259, doi:10.1016/j.ijpharm.2008.01.016.
230. Murphy, A.; Fágáin, C.Ó. Stability Characteristics of Chemically-Modified Soluble Trypsin. *J. Biotechnol.* **1996**, *49*, 163–171, doi:10.1016/0168-1656(96)01539-8.
231. Venkatesh, R.; Sundaram, P. V Modulation of Stability Properties of Bovine Trypsin after in Vitro Structural Changes with a Variety of Chemical Modifiers. *Protein Eng.* **1998**, *11*, 691–698, doi:10.1093/protein/11.8.691.
232. Vankova, H.; Pospisilova, M.; Ticha, M.; Turkova, J. Stabilization of Trypsin by Glycosylation. *Biotechnol. Tech.* **1994**, *8*, 375–380, doi:10.1007/BF00154306.
233. Sri Ram, J.; Terminiello, L.; Bier, M.; Nord, F.F. On the Mechanism of Enzyme Action. LVIII. Acetyltrypsin, a Stable Trypsin Derivative. *Arch. Biochem. Biophys.* **1954**, *52*, 464–477, doi:10.1016/0003-9861(54)90146-0.
234. Burkhart, J.M.; Schumbrutzki, C.; Wortelkamp, S.; Sickmann, A.; Zahedi, R.P. Systematic and Quantitative Comparison of Digest Efficiency and Specificity Reveals the Impact of Trypsin Quality on MS-Based Proteomics. *J. Proteomics* **2012**, *75*, 1454–1462, doi:10.1016/j.jpro.2011.11.016.

235. Lin, Y.-H.; Egeuz, R.V.; Torralba, M.G.; Singh, H.; Golusinski, P.; Golusinski, W.; Masternak, M.; Nelson, K.E.; Freire, M.; Yu, Y. Self-Assembled STRap for Global Proteomics and Salivary Biomarker Discovery. *J. Proteome Res.* **2019**, *18*, 1907–1915, doi:10.1021/acs.jproteome.9b00037.
236. Poachanukoon, O.; Roytrakul, S.; Koontongkaew, S. A Shotgun Proteomic Approach Reveals Novel Potential Salivary Protein Biomarkers for Asthma. *J. Asthma* **2022**, *59*, 243–254, doi:10.1080/02770903.2020.1850773/FORMAT/EPUB.
237. Núñ Ez Galindo, A.N.; Kussmann, M.; Dayon, L. Proteomics of Cerebrospinal Fluid: Throughput and Robustness Using a Scalable Automated Analysis Pipeline for Biomarker Discovery. *Anal. Chem.* **2015**, *87*, 10755–10761, doi:10.1021/acs.analchem.5b02748.
238. Eisenthal, R.; Peterson, M.E.; Daniel, R.M.; Danson, M.J. The Thermal Behaviour of Enzyme Activity: Implications for Biotechnology. *Trends Biotechnol.* **2006**, *24*, 289–292, doi:10.1016/j.tibtech.2006.05.004.
239. Georlette, D.; Blaise, V.; Collins, T.; D’Amico, S.; Gratia, E.; Hoyoux, A.; Marx, J.-C.; Sonan, G.; Feller, G.; Gerday, C. Some like It Cold: Biocatalysis at Low Temperatures. *FEMS Microbiol. Rev.* **2004**, *28*, 25–42, doi:10.1016/J.FEMSRE.2003.07.003.
240. Daniel, R.M.; Danson, M.J.; Eisenthal, R.; Lee, C.K.; Peterson, M.E. The Effect of Temperature on Enzyme Activity: New Insights and Their Implications. *Extremophiles* **2008**, *12*, 51–59, doi:10.1007/s00792-007-0089-7.
241. Ren, D.; Pipes, G.D.; Liu, D.; Shih, L.Y.; Nichols, A.C.; Treuheit, M.J.; Brems, D.N.; Bondarenko, P. V. An Improved Trypsin Digestion Method Minimizes Digestion-Induced Modifications on Proteins. *Anal. Biochem.* **2009**, *392*, 12–21, doi:10.1016/J.AB.2009.05.018.
242. Sun, S. Inhibition of Protein Carbamylation in Urea Solution Using Ammonium Containing Buffers Shisheng. *Anal Biochem.* **2008**, *6*, 2166–2171, doi:10.1016/j.ab.2013.10.024.Inhibition.
243. Müller, T.; Winter, D. Systematic Evaluation of Protein Reduction and Alkylation Reveals Massive Unspecific Side Effects by Iodine-Containing Reagents. *Mol. Cell. Proteomics* **2017**, *16*, 1173–1187, doi:10.1074/MCP.M116.064048.
244. Hains, P.G.; Robinson, P.J. The Impact of Commonly Used Alkylating Agents on Artifactual Peptide Modification. *J. Proteome Res.* **2017**, *16*, 3443–3447, doi:10.1021/acs.jproteome.7B00022.
245. Pan Fang; Mingqi Liu; Yu Xue; Jun Yao; Yang Zhang; Huali Shen; Pengyuan Yang Controlling Nonspecific Trypsin Cleavages in LC-MS/MS-Based Shotgun Proteomics Using Optimized Experimental Conditions. *Analyst* **2015**, *140*, 7613–7621, doi:10.1039/C5AN01505G.

246. Zheng, Y.Z.; DeMarco, M.L. Manipulating Trypsin Digestion Conditions to Accelerate Proteolysis and Simplify Digestion Workflows in Development of Protein Mass Spectrometric Assays for the Clinical Laboratory. *Clin. Mass Spectrom.* **2017**, *6*, 1–12, doi:10.1016/J.CLINMS.2017.10.001.
247. Kuznetsova, K.G.; Solovyeva, E.M.; Kuzikov, A. V.; Gorshkov, M. V.; Moshkovskii, S.A. Modification of Cysteine Residues for Mass Spectrometry-Based Proteomic Analysis: Facts and Artifacts. *Biochemistry* **2020**, *14*, 204–215, doi:10.1134/S1990750820030087.
248. Liu, F.J.; Ye, M.L.; Pan, Y.B.; Zou, H.F. High Concentration Trypsin Assisted Fast In-Gel Digestion for Phosphoproteome Analysis. *Chinese J. Anal. Chem.* **2015**, *43*, 1452–1458, doi:10.1016/S1872-2040(15)60864-7.
249. Gilliland, G.L.; Teplyakov, A. Structural Calcium (Trypsin, Subtilisin). *Encycl. Inorg. Bioinorg. Chem.* **2011**, doi:10.1002/9781119951438.EIBC0522.
250. Gorini, L. Rôle Du Calcium Dans Le Système Trypsine-Sérumalbumine. *Biochim. Biophys. Acta* **1951**, *7*, 318–334, doi:10.1016/0006-3002(51)90033-9.
251. Green, N.M.; Neurath, H. The Effects of Divalent Cations on Trypsin*. *J. Biol. Chem.* **1953**, *204*, 379–390, doi:10.1016/S0021-9258(18)66146-0.
252. Nord, F.F.; Bier, M.; Terminiello, L. On the Mechanism of Enzyme Action. LXI. The Self Digestion of Trypsin, Calcium-Trypsin and Acetyltrypsin. *Arch. Biochem. Biophys.* **1956**, *65*, 120–131, doi:10.1016/0003-9861(56)90182-5.
253. Sipos, T.; Merkel, J.R. An Effect of Calcium Ions on the Activity, Heat Stability, and Structure of Trypsin. *Biochemistry* **1970**, *9*, 2766–2775, doi:10.1021/bi00816a003.
254. Papaleo, E.; Fantucci, P.; Gioia, L. De Effects of Calcium Binding on Structure and Autolysis Regulation in Trypsins . A Molecular Dynamics Investigation. *J. Chem. Theory Comput.* **2005**, *1*, 1286–1297, doi:10.1021/ct050092o.
255. Erde, J.; Loo, R.R.O.; Loo, J.A. Enhanced FASP (EFASP) to Increase Proteome Coverage and Sample Recovery for Quantitative Proteomic Experiments. *J. Proteome Res.* **2014**, *13*, 1885–1895, doi:10.1021/pr4010019.
256. Yu, Y.; Suh, M.-J.; Sikorski, P.; Kwon, K.; Nelson, K.E.; Pieper, R. Urine Sample Preparation in 96-Well Filter Plates for Quantitative Clinical Proteomics. *Anal. Chem.* **2014**, *86*, 5470–5477, doi:10.1021/ac5008317.
257. Sandbaumhüter, F.A.; Nezhyva, M.; Eriksson, O.; Engberg, A.; Kreuger, J.; Andrén, P.E.; Jansson, E.T. Well-Plate MfASP for Proteomic Analysis of Single Pancreatic Islets. *J. Proteome Res.* **2022**, *21*, 1167–1174, doi:10.1021/acs.jproteome.2C00047.
258. Zougman, A.; Selby, P.J.; Banks, R.E. Suspension Trapping (STrap) Sample Preparation Method for Bottom-up Proteomics Analysis. *Proteomics* **2014**, *14*, 1006–1010, doi:10.1002/pmic.201300553.

259. Elinger, D.; Gabashvili, A.; Levin, Y. Suspension Trapping (S-Trap) Is Compatible with Typical Protein Extraction Buffers and Detergents for Bottom-Up Proteomics. *J. Proteome Res.* **2019**, *18*, 1441–1445, doi:10.1021/acs.jproteome.8b00891.
260. Wu, C.; Zhou, S.; Mitchell, M.I.; Hou, C.; Byers, S.; Loudig, · Olivier; Ma, J. Coupling Suspension Trapping-Based Sample Preparation and Data-Independent Acquisition Mass Spectrometry for Sensitive Exosomal Proteomic Analysis. *Anal. Bioanal. Chem.* **2022**, *1*, 3, doi:10.1007/s00216-022-03920-z.
261. Zougman, A.; Banks, R.E. C-STrap Sample Preparation Method—In-Situ Cysteinylyl Peptide Capture for Bottom-Up Proteomics Analysis in the STrap Format. *PLoS One* **2015**, *10*, e0138775, doi:10.1371/JOURNAL.PONE.0138775.
262. Duong, V.A.; Park, J.M.; Lee, H. Review of Three-Dimensional Liquid Chromatography Platforms for Bottom-up Proteomics. *Int. J. Mol. Sci.* **2020**, *21*, doi:10.3390/ijms21041524.
263. HaileMariam, M.; Vargas Eguez, R.; Singh, H.; Bekele, S.; Ameni, G.; Pieper, R.; Yu, Y. S-Trap, an Ultrafast Sample-Preparation Approach for Shotgun Proteomics. *J. Proteome Res.* **2018**, *17*, 2917–2924, doi:10.1021/acs.jproteome.8b00505.
264. Ludwig, K.R.; Schroll, M.M.; Hummon, A.B. Comparison of In-Solution, FASP, and S-Trap Based Digestion Methods for Bottom-Up Proteomic Studies. *J. Proteome Res.* **2018**, *17*, 2480–2490, doi:10.1021/acs.jproteome.8b00235.
265. Zougman, A.; Wilson, J.P.; Roberts, L.D.; Banks, R.E. Detergent-Free Simultaneous Sample Preparation Method for Proteomics and Metabolomics. *J. Proteome Res.* **2020**, *19*, 2838–2844, doi:10.1021/ACS.JPROTEOME.9B00662/.
266. Kulak, N.A.; Pichler, G.; Paron, I.; Nagaraj, N.; Mann, M. Minimal, Encapsulated Proteomic-Sample Processing Applied to Copy-Number Estimation in Eukaryotic Cells. *Nat. Methods* **2014**, *11*, 319–324, doi:10.1038/nmeth.2834.
267. Kassem, S.; Van Der Pan, K.; De Jager, A.L.; Naber, B.A.E.; De Laat, I.F.; Louis, A.; Van Dongen, J.J.M.; Teodosio, C.; Díez, P. Proteomics for Low Cell Numbers: How to Optimize the Sample Preparation Workflow for Mass Spectrometry Analysis. *J. Proteome* **2021**, *20*, 4217–4230, doi:10.1021/acs.jproteome.1c00321.
268. Hughes, C.S.; Moggridge, S.; Müller, T.; Sorensen, P.H.; Morin, G.B.; Krijgsveld, J. Single-Pot, Solid-Phase-Enhanced Sample Preparation for Proteomics Experiments. *Nat. Protoc.* **2019**, *14*, 68–85, doi:10.1038/s41596-018-0082-x.
269. Dagley, L.F.; Infusini, G.; Larsen, R.H.; Sandow, J.J.; Webb, A.I. Universal Solid-Phase Protein Preparation (USP3) for Bottom-up and Top-down Proteomics. *J. Proteome Res.* **2019**, *18*, 2915–2924, doi:10.1021/acs.jproteome.9b00217.

270. Sielaff, M.; Kuharev, J.; Bohn, T.; Hahlbrock, J.; Bopp, T.; Tenzer, S.; Distler, U. Evaluation of FASP, SP3, and IST Protocols for Proteomic Sample Preparation in the Low Microgram Range. *J. Proteome Res.* **2017**, *16*, 4060–4072, doi:10.1021/acs.jproteome.7b00433.
271. Supasri, K.M.; Kumar, M.; Mathew, M.J.; Signal, B.; Padula, M.P.; Suggett, D.J.; Ralph, P.J. Evaluation of Filter, Paramagnetic, and STAGETips Aided Workflows for Proteome Profiling of Symbiodiniaceae Dinoflagellate. *Processes* **2021**, *9*, 1–18.
272. Yang, Z.; Zhang, Z.; Chen, D.; Xu, T.; Wang, Y.; Sun, L. Nanoparticle-Aided Nanoreactor for Nanoproteomics. *Anal. Chem.* **2021**, *93*, 10568–10576, doi:10.1021/acs.analchem.1c01704.
273. Kelly, R.T. Single-Cell Proteomics: Progress and Prospects. *Mol. Cell. Proteomics* **2020**, *19*, 1739–1748, doi:10.1074/mcp.R120.002234.
274. Oedit, A.; Vulto, P.; Ramautar, R.; Lindenburg, P.W.; Hankemeier, T. Lab-on-a-Chip Hyphenation with Mass Spectrometry: Strategies for Bioanalytical Applications. *Curr. Opin. Biotechnol.* **2015**, *31*, 79–85, doi:10.1016/J.COPBIO.2014.08.009.
275. Woo, J.; Williams, S.M.; Markillie, L.M.; Feng, S.; Tsai, C.-F.; Aguilera-Vazquez, V.; Sontag, R.L.; Moore, R.J.; Hu, D.; Mehta, H.S.; et al. High-Throughput and High-Efficiency Sample Preparation for Single-Cell Proteomics Using a Nested Nanowell Chip. *Nat. Commun.* **2021**, *12*, doi:10.1038/s41467-021-26514-2.
276. Gebreyesus, S.T.; Siyal, A.A.; Kitata, R.B.; Chen, E.S.W.; Enkhbayar, B.; Angata, T.; Lin, K.I.; Chen, Y.J.; Tu, H.L. Streamlined Single-Cell Proteomics by an Integrated Microfluidic Chip and Data-Independent Acquisition Mass Spectrometry. *Nat. Commun.* **2022**, *13*, 1–13, doi:10.1038/s41467-021-27778-4.
277. Kecskemeti, A.; Gaspar, A. Preparation and Characterization of a Packed Bead Immobilized Trypsin Reactor Integrated into a PDMS Microfluidic Chip for Rapid Protein Digestion. *Talanta* **2017**, *166*, 275–283, doi:10.1016/j.talanta.2017.01.060.
278. Leipert, J.; Tholey, A. Miniaturized Sample Preparation on a Digital Microfluidics Device for Sensitive Bottom-up Microproteomics of Mammalian Cells Using Magnetic Beads and Mass Spectrometry-Compatible Surfactants. *Lab Chip* **2019**, *19*, 3490–3498, doi:10.1039/c9lc00715f.
279. Zhu, Y.; Clair, G.; Chrisler, W.B.; Shen, Y.; Zhao, R.; Shukla, A.K.; Moore, R.J.; Misra, R.S.; Pryhuber, G.S.; Smith, R.D.; et al. Proteomic Analysis of Single Mammalian Cells Enabled by Microfluidic Nanodroplet Sample Preparation and Ultrasensitive NanoLC-MS. *Angew. Chemie Int. Ed.* **2018**, *57*, 12370–12374, doi:10.1002/ANIE.201802843.
280. Zhou, M.; Uwugiaren, N.; Williams, S.M.; Moore, R.J.; Zhao, R.; Goodlett, D.; Dapic, I.; Paš, L.; Tolić, T.; Zhu, Y. Sensitive Top-Down Proteomics Analysis of a Low Number of Mammalian Cells Using a Nanodroplet Sample Processing Platform. *Anal. Chem.* **2020**, *92*, 7087–7095, doi:10.1021/acs.analchem.0c00467.

281. Ctortekca, C.; Hartlmayr, D.; Seth, A.; Mendjan, S.; Tourniaire, G.; Mechtler, K.; Biocenter, V. An Automated Workflow for Multiplexed Single-Cell Proteomics Sample Preparation at Unprecedented Sensitivity. *bioRxiv* **2022**, *4*, 1–17, doi:10.1101/2021.04.14.439828.
282. Crowell, A.M.J.; MacLellan, D.L.; Doucette, A.A. A Two-Stage Spin Cartridge for Integrated Protein Precipitation, Digestion and SDS Removal in a Comparative Bottom-up Proteomics Workflow. *J. Proteomics* **2015**, *118*, 140–150, doi:10.1016/j.jprot.2014.09.030.
283. Nickerson, J.L.; Baghalabadi, V.; Dang, Z.; Miller, V.A.; Little, S.L.; Doucette, A.A. Organic Solvent-Based Protein Precipitation for Robust Proteome Purification Ahead of Mass Spectrometry. *J. Vis. Exp.* **2022**, *180*, e63503, doi:10.3791/63503.
284. Lam Khuong, H.; Chen, C.-H.; Lin, J.-L.; Le, T.-N.; Hue Pham, T.; Bich Thao Le, T.; Canh Nguyen, X.; Chi Phan, V.; Ha Chu, H.; Wei-Wen Hsiao, W.; et al. Nanodiamond Solid-Phase Extraction and Triton X-114 Cloud Point Separation for Robust Fractionation and Shotgun Proteomics Analysis of the Human Serum Proteome. *J. Proteome Res.* **2022**, *21*, 67–76, doi:10.1021/acs.jproteome.1c00510.
285. Tran, J.C.; Doucette, A.A. Gel-Eluted Liquid Fraction Entrapment Electrophoresis: An Electrophoretic Method for Broad Molecular Weight Range Proteome Separation. *Anal. Chem.* **2008**, *80*, 1568–1573, doi:10.1021/ac702197w.
286. Unterlander, N.; Doucette, A.A. Membrane-Based SDS Depletion Ahead of Peptide and Protein Analysis by Mass Spectrometry. *Proteomics* **2018**, *18*, doi:10.1002/pmic.201700025.
287. Martinez, M.; Spitali, M.; Norrant, E.L.; Bracewell, D.G. Precipitation as an Enabling Technology for the Intensification of Biopharmaceutical Manufacture. *Trends Biotechnol.* **2019**, *37*, 237–241, doi:10.1016/j.tibtech.2018.09.001.
288. Sepperer, T.; Hernandez-Ramos, F.; Labidi, J.; Oostingh, G.J.; Bogner, B.; Petutschnigg, A.; Tondi, G. Purification of Industrial Tannin Extract through Simple Solid-Liquid Extractions. *Ind. Crops Prod.* **2019**, *139*, 111502, doi:10.1016/j.indcrop.2019.111502.
289. Xu, X.; Lan, J.; Korfmacher, W.A. Rapid LC/MS/MS Method Development for Drug Discovery. *Anal. Chem.* **2005**, *77*, 389A–394A, doi:10.1021/ac053476f.
290. Vuckovic, D. Current Trends and Challenges in Sample Preparation for Global Metabolomics Using Liquid Chromatography-Mass Spectrometry. *Anal. Bioanal. Chem.* **2012**, *403*, 1523–1548, doi:10.1007/s00216-012-6039-y.
291. Kanshin, E.; Thibault, P.; Hughes, C.S.; Foehr, S.; Garfield, D.A.; Furlong, E.E.; Steinmetz, L.M.; Krijgsveld, J. Ultrasensitive Proteome Analysis Using Paramagnetic Bead Technology. *Mol. Syst. Biol.* **2014**, doi:10.15252/msb.20145625.
292. Doucette, A.A.; Vieira, D.B.; Orton, D.J.; Wall, M.J. Resolubilization of Precipitated Intact Membrane Proteins with Cold Formic Acid for Analysis by Mass Spectrometry. *J. Proteome Res.* **2014**, *13*, 6001–6012, doi:10.1021/pr500864a.

293. Marcus, Y.; Hefter, G. Ion Pairing. *Chem. Rev.* **2006**, *106*, 4585–4621, doi:10.1021/cr040087x.
294. Askonas, B.A. The Use of Organic Solvents at Low Temperature for the Separation of Enzymes. Application to Aqueous Rabbit Muscle Extract. *Biochem. J.* **1951**, *48*, 42–48, doi:10.1042/bj0480042.
295. Bergman, L.W. Growth and Maintenance of Yeast. *Methods Mol. Biol.* **2001**, *177*, 9–14, doi:10.1385/1-59259-210-4:009.
296. Laemmli; U. Most Commonly Used Discontinuous Buffer System for SDS Electrophoresis. *Nature* **1970**, *227*, 680–686.
297. Neuhoff, V.; Arold, N.; Taube, D.; Ehrhardt, W. Improved Staining of Proteins in Polyacrylamide Gels Including Isoelectric Focusing Gels with Clear Background at Nanogram Sensitivity Using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **1988**, *9*, 255–262, doi:10.1002/elps.1150090603.
298. Wang, N.; Xie, C.; Young, J.B.; Li, L. Off-Line Two-Dimensional Liquid Chromatography with Maximized Sample Loading to Reversed-Phase Liquid Chromatography-Electrospray Ionization Tandem Mass Spectrometry for Shotgun Proteome Analysis. *Anal. Chem.* **2009**, *81*, 1049–1060, doi:10.1021/ac802106z.
299. Shevchenko, A.; Tomas, H.; Havlis, J.; Olsen, J. V; Mann, M. In-Gel Digestion for Mass Spectrometric Characterization of Proteins and Proteomes. *Nat. Protoc.* **2007**, *1*, 2856, doi:10.1038/nprot.2006.468.
300. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D.J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. The PRIDE Database and Related Tools and Resources in 2019: Improving Support for Quantification Data. *Nucleic Acids Res.* **2019**, *47*, doi:10.1093/nar/gky1106.
301. Zacharias, A.O.; Fang, Z.; Rahman, A.; Talukder, A.; Cornelius, S.; Chowdhury, S.M. Affinity and Chemical Enrichment Strategies for Mapping Low-Abundance Protein Modifications and Protein-Interaction Networks. *J. Sep. Sci.* **2021**, *44*, 310–322, doi:10.1002/JSSC.202000930.
302. Ctortocka, C.; Mechtler, K. The Rise of Single-cell Proteomics. *Anal. Sci. Adv.* **2021**, *2*, 84–94, doi:10.1002/ANSA.202000152.
303. Sobsey, C.A.; Ibrahim, S.; Richard, V.R.; Gaspar, V.; Mitsa, G.; Lacasse, V.; Zahedi, R.P.; Batist, G.; Borchers, C.H. Targeted and Untargeted Proteomics Approaches in Biomarker Development. *Proteomics* **2020**, *20*, 1900029, doi:10.1002/pmic.201900029.
304. Washburn, M.P.; Ulaszek, R.R.; Yates, J.R. Reproducibility of Quantitative Proteomic Analyses of Complex Biological Mixtures by Multidimensional Protein Identification Technology. *Anal. Chem.* **2003**, *75*, 5054–5061, doi:10.1021/ac034120b.

305. Collins, B.C.; Hunter, C.L.; Liu, Y.; Schilling, B.; Rosenberger, G.; Bader, S.L.; Chan, D.W.; Gibson, B.W.; Gingras, A.C.; Held, J.M.; et al. Multi-Laboratory Assessment of Reproducibility, Qualitative and Quantitative Performance of SWATH-Mass Spectrometry. *Nat. Commun.* **2017**, *8*, 1–12, doi:10.1038/s41467-017-00249-5.
306. Anderle, M.; Roy, S.; Lin, H.; Becker, C.; Joho, K. Quantifying Reproducibility for Differential Proteomics: Noise Analysis for Protein Liquid Chromatography-Mass Spectrometry of Human Serum. *Bioinformatics* **2004**, *20*, 3575–3582, doi:10.1093/BIOINFORMATICS/BTH446.
307. Tabb, D.L.; Vega-Montoto, L.; Rudnick, P.A.; Variyath, A.M.; Ham, A.-J.L.; Bunk, D.M.; Kilpatrick, L.E.; Billheimer, D.D.; Blackman, R.K.; Cardasis, H.L.; et al. Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography-Tandem Mass Spectrometry. *J. Proteome Res.* **2009**, *9*, 761–776, doi:10.1021/pr9006365.
308. Savitski, M.M.; Fischer, F.; Mathieson, T.; Sweetman, G.; Lang, M.; Bantscheff, M. Targeted Data Acquisition for Improved Reproducibility and Robustness of Proteomic Mass Spectrometry Assays. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1668–1679, doi:10.1016/j.jasms.2010.01.012.
309. Révész, Á.; Hevér, H.; Steckel, A.; Schlosser, G.; Szabó, D.; Vékey, K.; Drahos, L. Collision Energies: Optimization Strategies for Bottom-up Proteomics. *Mass Spectrom. Rev.* **2021**, doi:10.1002/MAS.21763.
310. Fernández-Costa, C.; Martínez-Bartolomé, S.; McClatchy, D.; Yates, J.R. Improving Proteomics Data Reproducibility with a Dual-Search Strategy. *Anal. Chem.* **2022**, *12*, 1697–1701, doi:10.1021/acs.analchem.9b04955.
311. Wang, W.; C-H Sue, A.; Goh, W.W. Feature Selection in Clinical Proteomics: With Great Power Comes Great Reproducibility. *Drug Discov. Today* **2017**, *22*, doi:10.1016/j.drudis.2016.12.006.
312. Ferna, C.; Martínez-Bartolome, S.; McClatchy, D.B.; Saviola, A.J.; Yu, N.-K.; Yates, J.R. Impact of the Identification Strategy on the Reproducibility of the DDA and DIA Results. *J. Proteome Res* **2020**, *19*, 58, doi:10.1021/acs.jproteome.0c00153.
313. Hu, A.; Noble, W.S.; Wolf-Yadlin, A. Technical Advances in Proteomics: New Developments in Data-Independent Acquisition. *F1000 Res.* **2016**, *5*, 1–12, doi:10.12688/f1000research.7042.1.
314. Cho, K.-C.; Oh, S.; Wang, Y.; Rosenthal, L.S.; Na, C.H.; Zhang, H. Evaluation of the Sensitivity and Reproducibility of Targeted Proteomic Analysis Using Data-Independent Acquisition for Serum and Cerebrospinal Fluid Proteins. *J. Proteome Res.* **2021**, *20*, 4291, doi:10.1021/acs.jproteome.1c00238.
315. Kachuk, C.; Doucette, A.A. The Benefits (and Misfortunes) of SDS in Top-down Proteomics. *J. Proteomics* **2018**, *175*, 75–86, doi:10.1016/j.jprot.2017.03.002.

316. Speicher, K.D.; Kolbas, O.; Harper, S.; Speicher, D.W. Systematic Analysis of Peptide Recoveries from In-Gel Digestions for Protein Identifications in Proteome Studies. *J. Biomol. Tech.* **2000**, *11*, 74.
317. Doellinger, J.; Schneider, A.; Hoeller, M.; Lasch, P. Sample Preparation by Easy Extraction and Digestion (SPEED) - A Universal, Rapid, and Detergent-Free Protocol for Proteomics Based on Acid Extraction. *Mol. Cell. Proteomics* **2020**, *19*, 209–222, doi:10.1074/mcp.TIR119.001616.
318. Tanca, A.; Abbondio, M.; Pisanu, S.; Pagnozzi, D.; Uzzau, S.; Addis, M.F. Critical Comparison of Sample Preparation Strategies for Shotgun Proteomic Analysis of Formalin-Fixed, Paraffin-Embedded Samples: Insights from Liver Tissue. *Clin. Proteomics* **2014**, *11*, 1–11, doi:10.1186/1559-0275-11-28.
319. Davalieva, K.; Kiprijanovska, S.; Dimovski, A.; Rosoklija, G.; Dwork, A.J. Comparative Evaluation of Two Methods for LC-MS/MS Proteomic Analysis of Formalin Fixed and Paraffin Embedded Tissues. *J. Proteomics* **2021**, *235*, 104117, doi:10.1016/j.jprot.2021.104117.
320. Loroch, S.; Kopczynski, D.; Schneider, A.C.; Schumbrutzki, C.; Feldmann, I.; Panagiotidis, E.; Reinders, Y.; Sakson, R.; Solari, F.A.; Vening, A.; et al. Toward Zero Variance in Proteomics Sample Preparation: Positive-Pressure FASP in 96-Well Format (PF96) Enables Highly Reproducible, Time-and Cost-Efficient Analysis of Sample Cohorts. *J. Proteome Res.* **2022**, *21*, 1181–1188, doi:10.1021/acs.jproteome.1c00706.
321. Guergues, J.; Zhang, P.; Liu, B.; Stevens, S.M. Improved Methodology for Sensitive and Rapid Quantitative Proteomic Analysis of Adult-Derived Mouse Microglia: Application to a Novel In Vitro Mouse Microglial Cell Model. *Proteomics* **2019**, *19*, 1800469, doi:10.1002/PMIC.201800469.
322. Fu, Q.; Kowalski, M.P.; Mastali, M.; Parker, S.J.; Sobhani, K.; Van Den Broek, I.; Hunter, C.L.; Van Eyk, J.E. Highly Reproducible Automated Proteomics Sample Preparation Workflow for Quantitative Mass Spectrometry. *J. Proteome Res.* **2017**, *17*, 420–428, doi:10.1021/acs.jproteome.7b00623.
323. Kruszewska, J.; Zajda, J.; Matczuk, M. How to Effectively Prepare a Sample for Bottom-up Proteomic Analysis of Nanoparticle Protein Corona? A Critical Review. *Talanta* **2021**, *226*, doi:10.1016/J.TALANTA.2021.122153.
324. Percy, A.J.; Chambers, A.G.; Smith, D.S.; Borchers, C.H. Standardized Protocols for Quality Control of MRM-Based Plasma Proteomic Workflows. *J. Proteome Res.* **2012**, *12*, 222–233, doi:10.1021/pr300893w.
325. Abbatiello, S.E.; Schilling, B.; Mani, D.R.; Zimmerman, L.J.; Hall, S.C.; MacLean, B.; Albertolle, M.; Allen, S.; Burgess, M.; Cusack, M.P.; et al. Large-Scale Interlaboratory Study to Develop, Analytically Validate and Apply Highly Multiplexed, Quantitative Peptide Assays to Measure Cancer-Relevant Proteins in Plasma. *Mol. Cell. Proteomics* **2015**, *14*, 2357–2374, doi:10.1074/mcp.M114.047050.

326. Addona, T.A.; Abbatiello, S.E.; Schilling, B.; Skates, S.J.; Mani, D.R.; Bunk, D.M.; Spiegelman, C.H.; Zimmerman, L.J.; Ham, A.-J.L.; Keshishian, H.; et al. Multi-Site Assessment of the Precision and Reproducibility of Multiple Reaction Monitoring-Based Measurements of Proteins in Plasma. *Nat. Biotechnol.* **2009**, *27*, 633–864, doi:10.1038/nbt.1546.
327. Kim, C.H.; Tworoger, S.S.; Stampfer, M.J.; Dillon, S.T.; Gu, X.; Sawyer, S.J.; Chan, A.T.; Libermann, T.A.; Eliassen, & A.H. Stability and Reproducibility of Proteomic Profiles Measured with an Aptamer-Based Platform. *Sci. Rep.* **2018**, *8*, doi:10.1038/s41598-018-26640-w.
328. Varnavides, G.; Madern, M.; Anrather, D.; Hartl, N.; Reiter, W.; Hartl, M. In Search of a Universal Method: A Comparative Survey of Bottom-Up Proteomics Sample Preparation Methods. *J. Proteome Res.* **2022**, doi:10.1021/acs.jproteome.2c00265.
329. Plotly Technologies Inc. *Collaborative data science* Montreal, QC <https://plot.ly>
330. Yang, Y.; Anderson, E.; Zhang, S. Evaluation of Six Sample Preparation Procedures for Qualitative and Quantitative Proteomics Analysis of Milk Fat Globule Membrane. *Electrophoresis* **2018**, *39*, 2332–2339, doi:10.1002/ELPS.201800042.
331. Webb-Robertson, B.-J.M.; Wiberg, H.K.; Matzke, M.M.; Brown, J.N.; Wang, J.; McDermott, J.E.; Smith, R.D.; Rodland, K.D.; Metz, T.O.; Pounds, J.G.; et al. Review, Evaluation, and Discussion of the Challenges of Missing Value Imputation for Mass Spectrometry-Based Label-Free Global Proteomics. *J. Proteome Res.* **2015**, *14*, 1993–2001, doi:10.1021/pr501138h.
332. Chiva, C.; Ortega, M.; Sabidó*, E.S. Influence of the Digestion Technique, Protease, and Missed Cleavage Peptides in Protein Quantitation. *J. Proteome Res.* **2014**, *13*, 3979–3986, doi:10.1021/pr500294d.
333. Li, M.; Smyth, G.K. Missing Values Are Informative in Label-Free Shotgun Proteomics Data: Estimating the Detection Probability Curve. *Biorxiv* **2022**, doi:10.1101/2022.07.02.498573.
334. Liu, M.; Dongre, A. Proper Imputation of Missing Values in Proteomics Datasets for Differential Expression Analysis. *Brief. Bioinform.* **2021**, *22*, doi:10.1093/BIB/BBAA112.
335. Jiang, X.; Shamshurin, D.; Spicer, V.; Krkhin, O. V. The Effect of Various S-Alkylating Agents on the Chromatographic Behaviour of Cysteine-Containing Peptides in Reversed-Phase Chromatography. *J. Chromatogr. B* **2013**, *915*, 57–63.
336. Nicora, G.; Vitali, F.; Dagliati, A.; Geifman, N.; Bellazzi, R. Integrated Multi-Omics Analyses in Oncology: A Review of Machine Learning Methods and Tools. *Front. Oncol.* **2020**, *10*, 1030, doi:10.3389/FONC.2020.01030/BIBTEX.

337. Hewelt-Belka, W.; Garwolińska, D.; Belka, M.; Bączek, T.; Namieśnik, J.; Kot-Wasik, A. Analytical Methods A New Dilution-Enrichment Sample Preparation Strategy for Expanded Metabolome Monitoring of Human Breast Milk That Overcomes the Simultaneous Presence of Low-and High-Abundance Lipid Species. *Food Chem.* **2019**, *288*, 154–161, doi:10.1016/j.foodchem.2019.03.001.
338. Sirén, K.; Siu Tze Mak, S.; Fischer, U.; Hansen, L.H.; Gilbert, M.T.P. Multi-Omics and Potential Applications in Wine Production. *Curr. Opin. Biotechnol.* **2019**, *56*, 172–178, doi:10.1016/j.copbio.2018.11.014.
339. Suvarna, K.; Salkar, A.; Palanivel, V.; Bankar, R.; Banerjee, N.; Gayathri, M.; Pai, J.; Srivastava, A.; Singh, A.; Khatri, H.; et al. A Multi-Omics Longitudinal Study Reveals Alteration of the Leukocyte Activation Pathway in COVID-19 Patients. *J. Proteome Res.* **2021**, *20*, 4667–4680, doi:10.1021/acs.jproteome.1c00215.
340. Nie, W.; Yan, L.; Lee, Y.H.; Guha, C.; Kurland, I.J.; Lu, H. Advanced Mass Spectrometry-Based Multi-Omics Technologies for Exploring the Pathogenesis of Hepatocellular Carcinoma. *Mass Spectrom. Rev.* **2016**, *35*, 331–349, doi:10.1002/MAS.21439.
341. Łuczykowski, K.; Warmuzińska, N.; Operacz, S.; Stryjak, I.; Bogusiewicz, J.; Jacyna, J.; Wawrzyniak, R.; Struck-Lewicka, W.; Markuszewski, M.J.; Bojko, B. Metabolic Evaluation of Urine from Patients Diagnosed with High Grade (HG) Bladder Cancer by SPME-LC-MS Method. *Molecules* **2021**, *26*, 2194, doi:10.3390/MOLECULES26082194.
342. Terai, Y.L.; Huang, C.; Wang, B.; Kang, X.; Han, J.; Douglass, J.; Hsiue, E.H.C.; Zhang, M.; Purohit, R.; Desilva, T.; et al. Valid-NEO: A Multi-Omics Platform for Neoantigen Detection and Quantification from Limited Clinical Samples. *Cancers (Basel)*. **2022**, *14*, 1–10, doi:10.3390/CANCERS14051243.
343. Reyes-Garces, N.; Gionfriddo, E. Recent Developments and Applications of Solid Phase Microextraction as a Sample Preparation Approach for Mass-Spectrometry-Based Metabolomics and Lipidomics. *Trends Anal. Chem.* **2019**, *113*, 172–181, doi:10.1016/j.trac.2019.01.009.
344. Roca, M.; Alcoriza, M.I.; Carlos Garcia-Ca, J.; Naveras; Lahoz, A. Reviewing the Metabolome Coverage Provided by LC-MS: Focus on Sample Preparation and Chromatography-A Tutorial. *Anal. Chim. Acta* **2020**, *1147*, 38–55, doi:10.1016/j.aca.2020.12.025.
345. Eisenbeiss, L.; Steuer, A.E.; Binz, T.M.; Baumgartner, M.R.; Kraemer, T. (Un)Targeted Hair Metabolomics: First Considerations and Systematic Evaluation on the Impact of Sample Preparation. *Anal. Bioanal. Chem.* **2019**, *411*, 3963–3977, doi:10.1007/s00216-019-01873-4.
346. Lu, H.; Chen, H.; Tang, X.; Yang, Q.; Zhang, H.; Chen, Y.Q.; Chen, W. Evaluation of Metabolome Sample Preparation and Extraction Methodologies for Oleaginous Filamentous Fungi *Mortierella Alpina*. *Metabolomics* **2019**, *15*, 1–10, doi:10.1007/s11306-019-1506-5.

347. Wawrzyniak, R.; Kosnowska, A.; Macioszek, S.; Bartoszewski, R.; Markuszewski, M.J. New Plasma Preparation Approach to Enrich Metabolome Coverage in Untargeted Metabolomics: Plasma Protein Bound Hydrophobic Metabolite Release with Proteinase K. *Sci. Rep.* **2018**, *8*, 1–10, doi:10.1038/s41598-018-27983-0.
348. Xiong, Y.; Shi, C.; Zhong, F.; Liu, X.; Yang, P. LC-MS/MS and SWATH Based Serum Metabolomics Enables Biomarker Discovery in Pancreatic Cancer. *Clin. Chim. Acta* **2020**, *506*, 214–221, doi:10.1016/j.cca.2020.03.043.
349. Dallas, D.C.; Guerrero, A.; Parker, E.A.; Robinson, R.C.; Gan, J.; German, J.B.; Barile, D.; Lebrilla, C.B. Current Peptidomics: Applications, Purification, Identification, Quantification, and Functional Analysis. *Proteomics* **2015**, *15*, 1026–1038, doi:10.1002/PMIC.201400310.
350. Fan, K.T.; Hsu, C.W.; Chen, Y.R. Mass Spectrometry in the Discovery of Peptides Involved in Intercellular Communication: From Targeted to Untargeted Peptidomics Approaches. *Mass Spectrom. Rev.* **2022**, e21789, doi:10.1002/MAS.21789.
351. Foreman, R.E.; George, A.L.; Reimann, F.; Gribble, F.M.; Kay, R.G. Peptidomics: A Review of Clinical Applications and Methodologies. *J. Proteome Res.* **2021**, *20*, 3782–3797, doi:10.1021/acs.jproteome.1c00295.
352. Manes, N.P.; Gustin, J.K.; Rue, J.; Mottaz, H.M.; Purvine, S.O.; Norbeck, A.D.; Monroe, M.E.; Zimmer, J.S.D.; Metz, T.O.; Adkins, J.N.; et al. Targeted Protein Degradation by Salmonella under Phagosome-Mimicking Culture Conditions Investigated Using Comparative Peptidomics. *Mol. Cell. Proteomics* **2007**, *6*, 717–727, doi:10.1074/mcp.M600282-MCP200.
353. MacLean, B.; Tomazela, D.M.; Shulman, N.; Chambers, M.; Finney, G.L.; Frewen, B.; Kern, R.; Tabb, D.L.; Liebler, D.C.; MacCoss, M.J. Skyline: An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments. *Bioinformatics* **2010**, *26*, 966, doi:10.1093/BIOINFORMATICS/BTQ054.
354. Adams, K.J.; Pratt, B.; Bose, N.; Dubois, L.G.; St. John-Williams, L.; Perrott, K.M.; Ky, K.; Kapahi, P.; Sharma, V.; Maccoss, M.J.; et al. Skyline for Small Molecules: A Unifying Software Package for Quantitative Metabolomics. *J. Proteome Res.* **2020**, *19*, 1447, doi:10.1021/ACS.JPROTEOME.9B00640.
355. Tulipani, S.; Llorach, R.; Urpi-Sarda, M.; Andres-Lacueva, C. Comparative Analysis of Sample Preparation Methods to Handle the Complexity of the Blood Fluid Metabolome: When Less Is More. *Anal. Chem.* **2013**, *85*, 341–348, doi:10.1021/AC302919T/SUPPL_FILE/AC302919T_SI_001.PDF.
356. Ooshima, H.; Sakata, M.; Harano, Y. Enhancement of Enzymatic Hydrolysis of Cellulose by Surfactant. *Biotechnol. Bioeng.* **1986**, *28*, 1727–1734, doi:10.1002/bit.260281117.

357. Lou, H.; Zeng, M.; Hu, Q.; Cai, C.; Lin, X.; Qiu, X.; Yang, D.; Pang, Y. Nonionic Surfactants Enhanced Enzymatic Hydrolysis of Cellulose by Reducing Cellulase Deactivation Caused by Shear Force and Air-Liquid Interface. *Bioresour. Technol.* **2018**, *249*, 1–8, doi:10.1016/J.BIORTECH.2017.07.066.
358. Rubingh, D.N. The Influence of Surfactants on Enzyme Activity Donn N Rubingh. *Curr. Opin. Colloid Interface Sci.* **1996**, *1*, 598–603, doi:10.1016/S1359-0294(96)80097-5.
359. Shome, A.; Roy, S.; Das, P.K. Articles Nonionic Surfactants: A Key to Enhance the Enzyme Activity at Cationic Reverse Micellar Interface. *Langmuir* **2007**, *23*, 4130–4136, doi:10.1021/la062804j.
360. An, B.; Zhang, M.; Johnson, R.W.; Qu, J. Surfactant-Aided Precipitation/on-Pellet-Digestion (SOD) Procedure Provides Robust and Rapid Sample Preparation for Reproducible, Accurate and Sensitive LC/MS Quantification of Therapeutic Protein in Plasma and Tissues. *Anal. Chem.* **2015**, *87*, 4023–4029, doi:10.1021/acs.analchem.5b00350.
361. Wu, F.; Sun, D.; Wang, N.; Gong, Y.; Li, L. Comparison of Surfactant-Assisted Shotgun Methods Using Acid-Labile Surfactants and Sodium Dodecyl Sulfate for Membrane Proteome Analysis. *Anal. Chim. Acta* **2011**, *698*, 36–43, doi:10.1016/j.aca.2011.04.039.
362. Savelli, G.; Spreti, N.; Profio, P. Di Enzyme Activity and Stability Control by Amphiphilic Self-Organizing Systems in Aqueous Solutions. *Curr. Opin. Colloid Interface Sci.* **2000**, *5*, 111–117.
363. Ghosh, S. Interaction of Trypsin with Sodium Dodecyl Sulfate in Aqueous Medium: A Conformational View. *Colloids Surfaces B Biointerfaces* **2008**, *66*, 178–186, doi:10.1016/J.COLSURFB.2008.06.011.
364. Wei, J.; Sun, J.; Yu, W.; Jones, A.; Oeller, P.; Keller, M.; Woodnutt, G.; Short, J.M. Global Proteome Discovery Using an Online Three-Dimensional LC-MS/MS. *J. Proteome Res.* **2005**, *4*, 801–808, doi:10.1021/pr0497632.
365. Loraine, J.; Alhumaidan, O.; Bottrill, A.R.; Mistry, S.C.; Andrew, P.; Mukamolova, G. V.; Turapov, O. Efficient Protein Digestion at Elevated Temperature in the Presence of Sodium Dodecyl Sulfate and Calcium Ions for Membrane Proteomics. *Anal. Chem.* **2019**, *91*, 9516–9521, doi:10.1021/ACS.ANALCHEM.9B00484.
366. Lin, Y.; Huo, L.; Liu, Z.; Li, J.; Liu, Y.; He, Q.; Wang, X.; Liang, S. Sodium Laurate, a Novel Potease- and Mass Spectrometry-Compatible Detergent for Mass Spectrometry-Based Membrane Proteomics. *Based Membr. Proteomics. PLoS ONE* **2009**, *8*, doi:10.1371/journal.pone.0059779.
367. Blinkhorn, C.; Jones, M.N. The Effect of Surfactant Structure on Ribonuclease A-Surfactant Interactions. *Biochem. J.* **1973**, *135*, 547–549.

368. Philippot, J.; Authier, M.H. Study of Human Red Blood Cell Membrane Using Sodium Deoxycholate II. Effects of Cold Storage, EDTA and Small Deoxycholate Concentrations on ATPase Activities. *Biochimica Biophys. Acta* **1973**, *298*, 887–900.
369. Porter, C.J.; Bereman, M.S. Comparison of Commercial LC-MS/MS Surfactants with Sodium Deoxycholate for Shotgun Proteomics. *J. Proteins Proteomics* **2014**, *5*, 151–161.
370. Serra, A.; Zhu, H.; Gallart-Palau, X.; Park, J.E.; Ho, H.H.; Tam, J.P.; Kwan Sze, S. Plasma Proteome Coverage Is Increased by Unique Peptide Recovery from Sodium Deoxycholate Precipitate. *Anal. Bioanal. Chem.* **2016**, *408*, 1963–1973, doi:10.1007/s00216-016-9312-7.
371. Scheerlinck, E.; Dhaenens, M.; Van Soom, A.; Peelman, L.; De Sutter, P.; Van Steendam, K.; Deforce, D. Minimizing Technical Variation during Sample Preparation Prior to Label-Free Quantitative Mass Spectrometry. *Anal. Biochem.* **2015**, *490*, 14–19, doi:10.1016/J.AB.2015.08.018.
372. Gracey, M.; Houghton, M.; Thomas, J. Deoxycholate Depresses Small-Intestinal Enzyme Activity. *Gut* **1975**, *16*, 53–56, doi:10.1136/gut.16.1.53.
373. Nakamura, S. Effect of Sodium Deoxycholate on 5'-Nucleotidase. *Biochim. Biophys. Acta* **1976**, *426*, 339–347.
374. Zaks, A.; Klibanov, A.M. The Effect of Water on Enzyme Action in Organic Media. *J. Biol. Chem.* **1988**, *263*, 8017–8021, doi:10.1016/S0021-9258(18)68435-2.
375. Zaks, A.; Klibanov, A.M. Enzymatic Catalysis in Nonaqueous Solvents. *J. Biol. Chem.* **1988**, *263*, 3194–3201, doi:10.1016/S0021-9258(18)69054-4.
376. Strader, M.B.; Tabb, D.L.; Hervey, W.J.; Pan, C.; Hurst, G.B. Efficient and Specific Trypsin Digestion of Microgram to Nanogram Quantities of Proteins in Organic–Aqueous Solvent Systems. *Anal. Chem.* **2005**, *78*, 125–134, doi:10.1021/AC051348L.
377. Blonder, J.; Chan, K.C.; Issaq, H.J.; Veenstra, T.D. Identification of Membrane Proteins from Mammalian Cell/Tissue Using Methanol-Facilitated Solubilization and Tryptic Digestion Coupled with 2D-LC-MS/MS. *Nat. Protoc.* **2007**, *1*, 2784–2790, doi:10.1038/nprot.2006.359.
378. Blonder, J.; Terunuma, A.; Conrads, T.P.; Chan, K.C.; Yee, C.; Lucas, D.A.; Schaefer, C.F.; Yu, L.R.; Issaq, H.J.; Veenstra, T.D.; et al. A Proteomic Characterization of the Plasma Membrane of Human Epidermis by High-Throughput Mass Spectrometry. *J. Invest. Dermatol.* **2004**, *123*, 691–699, doi:10.1111/j.0022-202X.2004.23421.x.
379. Blonder, J.; Rodriguez-Galan, M.C.; Chan, K.C.; Lucas, D.A.; Yu, L.R.; Conrads, T.P.; Issaq, H.J.; Young, H.A.; Veenstra, T.D. Analysis of Murine Natural Killer Cell Microsomal Proteins Using Two-Dimensional Liquid Chromatography Coupled to Tandem Electrospray Ionization Mass Spectrometry. *J. Proteome Res.* **2004**, *3*, 862–870, doi:10.1021/pr049927e.

380. Schwert, G.W.; Takenaka, Y. A Spectrophotometric Determination of Trypsin and Chymotrypsin. *BBA - Biochim. Biophys. Acta* **1955**, *16*, 570–575, doi:10.1016/0006-3002(55)90280-8.
381. Jones, M.N.; Skinner, H.A.; Tippingt, E.; Wilkinson, A. The Interaction between Ribonuclease A and Surfactants. *Biochem. J.* **1973**, *135*, 231–236, doi:10.1042/bj1350231.
382. Iwasaki, M.; Masuda, T.; Tomita, M.; Ishihama, Y. Chemical Cleavage-Assisted Tryptic Digestion for Membrane Proteome Analysis. *J. Proteome Res.* **2009**, *8*, 3169–3175, doi:10.1021/pr900074n.
383. Lin, Y.; Chen, Y.; Yang, X.; Xu, D.; Liang, S. Proteome Analysis of a Single Zebrafish Embryo Using Three Different Digestion Strategies Coupled with Liquid Chromatography–Tandem Mass Spectrometry. *Anal. Biochem.* **2009**, *394*, 177–185, doi:10.1016/J.AB.2009.07.034.
384. Li, D.; Farchone, A.; Zhu, Q.; Macchi, F.; Walker, D.E.; Michels, D.A.; Yang, F. Fast, Robust, and Sensitive Identification of Residual Host Cell Proteins in Recombinant Monoclonal Antibodies Using Sodium Deoxycholate Assisted Digestion. *Anal. Chem.* **2020**, *92*, 11888–11894, doi:10.1021/ACS.ANALCHEM.0C02258.
385. Bogdanova, L.R.; Gnezdilov, O.I.; Idiyatullin, B.Z.; Kurbanov, R.K.; Zuev, Y.F.; Us'yarov, O.G. Micellization in Sodium Deoxycholate Solutions. *Colloid J.* **2012**, *74*, 1–6, doi:10.1134/S1061933X12010036.
386. Rajagopalan, K. V; Fridovich, I. Competitive Inhibition of Enzyme Activity by Urea. *Artic. J. Biol. Chem.* **1961**, doi:10.1016/S0021-9258(18)64242-5.
387. Fan, Y.X.; Ju, M.; Zhou, J.M.; Tsou, C.L. Activation of Chicken Liver Dihydrofolate Reductase by Urea and Guanidine Hydrochloride Is Accompanied by Conformational Change at the Active Site. *Biochem. J.* **1996**, *315*, 97–102, doi:10.1042/BJ3150097.
388. Ghatge, M.S.; Deshpande, V. V. Evidence for Specific Interaction of Guanidine Hydrochloride with Carboxy Groups of Enzymes. *Biochem. Biophys. Res. Commun.* **1993**, 979–984.
389. Leonis, J. Investigation of the Behavior of Lysozyme in Urea Solutions. *Arch. Biochem. Biophys.* **1956**, *65*, 182–193.
390. Gomez, J.E.; Birnbaum, E.R.; Royer, G.P.; Darnall, D.W. The Effect of Calcium Ion on the Urea Denaturation of Immobilized Bovine Trypsin. *Biochim. Biophys. Acta* **1977**, *495*, 177–182.
391. Crewther, W.G. The Effect of PH and Cations on the Thermal Denaturation of Trypsin. *Aust. J. Biol. Sci.* **1953**, *6*, 597–616.


392. Nickerson, J.L.; Doucette, A.A. Maximizing Cumulative Trypsin Activity with Calcium at Elevated Temperature for Enhanced Bottom-Up Proteome Analysis. *Biology (Basel)*. **2022**, *11*, doi:10.3390/biology11101444.
393. Batra, R.; Gupta, M.N. Enhancement of Enzyme Activity in Aqueous-Organic Solvent Mixtures. *Biotechnol. Lett.* **1994**, *16*, 1059–1064.
394. Simon, L.M.; László, K.; Vértesi, A.; Bagi, K.; Szajáni, B. Stability of Hydrolytic Enzymes in Water-Organic Solvent Systems. *J. Mol. Catal. - B Enzym.* **1998**, *4*, 41–45, doi:10.1016/S1381-1177(97)00019-2.
395. Park, H.; Chi, Y.-M. Distinction between the Influence of Dielectric Constant and of Methanol Concentration on Trypsin-Catalyzed Hydrolysis and Methanolysis. *J. Microbiol. Biotechnol.* **1998**, *8*, 656–662.
396. DeSantis, G.; Jones, J.B. Chemical Modification of Enzymes for Enhanced Functionality. *Curr. Opin. Biotechnol.* **1999**, *10*, 324–330, doi:10.1016/S0958-1669(99)80059-7.
397. Gobom, J.; Nordhoff, E.; Ekman, R.; Roepstorff, P. Rapid Micro-Scale Proteolysis of Proteins for MALDI-MS Peptide Mapping Using Immobilized Trypsin. *Int. J. Mass Spectrom. Ion Process.* **1997**, *169–170*, 153–163, doi:10.1016/s0168-1176(97)00216-4.
398. Naldi, M.; Tramarin, A.; Bartolini, M. Immobilized Enzyme-Based Analytical Tools in the Omics Era: Recent Advances. *J. Pharm. Biomed. Anal.* **2018**, *160*, 222–237, doi:10.1016/j.jpba.2018.07.051.
399. Wouters, B.; Currivan, S.A.; Abdulhussain, N.; Hankemeier, T.; Schoenmakers, P.J. Immobilized-Enzyme Reactors Integrated into Analytical Platforms: Recent Advances and Challenges. *TrAC Trends Anal. Chem.* **2021**, *144*, 116419, doi:10.1016/J.TRAC.2021.116419.
400. Hua, L.; Low, T.Y.; Sze, S.K. Microwave-Assisted Specific Chemical Digestion for Rapid Protein Identification. *Proteomics* **2006**, *6*, 586–591, doi:10.1002/pmic.200500304.
401. Martins, G.; Fernández-lodeiro, J.; Djafari, J.; Lodeiro, C. Label-Free Protein Quantification after Ultrafast Digestion of Complex Proteomes Using Ultrasonic Energy and Immobilized-Trypsin Magnetic Nanoparticles. *Talanta* **2019**, *196*, 262–270, doi:10.1016/j.talanta.2018.12.066.
402. Poulsen, J.W.; Madsen, C.T.; Poulsen, F.M.; Nielsen, M.L. Using Guanidine-Hydrochloride for Fast and Efficient Protein Digestion and Single-Step Affinity-Purification Mass Spectrometry. *J. Proteome Res.* **2013**, *12*, 1020–1030, doi:10.1021/pr300883y.
403. Zhong, X.; Chen, H.; Zare, R.N. Ultrafast Enzymatic Digestion of Proteins by Microdroplet Mass Spectrometry. *Nat. Commun.* **2020**, *11*, doi:10.1038/s41467-020-14877-x.

404. Cannon, J.; Lohnes, K.; Wynne, C.; Wang, Y.; Edwards, N.; Fenselau, C. High-Throughput Middle-down Analysis Using an Orbitrap. *J. Proteome Res.* **2010**, *9*, 3886–3890, doi:10.1021/pr1000994.
405. Daniel, R.M.; Danson, M.J.; Eienthal, R. The Temperature Optima of Enzymes: A New Perspective on an Old Phenomenon. *Trends Biochem. Sci.* **2001**, *26*, 223–225, doi:10.1016/S0968-0004(01)01803-5.
406. Daniel, R.M.; Peterson, M.E.; Danson, M.J.; Price, N.C.; Kelly, S.M.; Monk, C.R.; Weinberg, C.S.; Oudshoorn, M.L.; Lee, C.K. The Molecular Basis of the Effect of Temperature on Enzyme Activity. *Biochem. J.* **2010**, *425*, 353–360, doi:10.1042/BJ20091254.
407. Arcus, V.L.; Prentice, E.J.; Hobbs, J.K.; Mulholland, A.J.; Kamp, M.W. Van der; Pudney, C.R.; Parker, E.J.; Schipper, L.A. On the Temperature Dependence of Enzyme-Catalyzed Rates. *Biochemistry* **2016**, *55*, 1681–1688, doi:10.1021/ACS.BIOCHEM.5B01094.
408. Crowell, A.M.J.; Stewart, E.J.; Take, Z.S.; Doucette, A.A. Critical Assessment of the Spectroscopic Activity Assay for Monitoring Trypsin Activity in Organic-Aqueous Solvent. *Anal. Biochem.* **2013**, *435*, 131–136, doi:10.1016/j.ab.2012.12.019.
409. Rundlett, K.L.; Armstrong, D.W. Mechanism of Signal Suppression by Anionic Surfactants in Capillary Electrophoresis-Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **1996**, *68*, 3493–3497, doi:10.1021/ac960472p.
410. Boja, E.S.; Fales, H.M. Overalkylation of a Protein Digest with Iodoacetamide. *Anal. Chem.* **2001**, *73*, 3576–3582, doi:10.1021/ac0103423.
411. Kuznetsova, K.G.; Levitsky, L.I.; Pyatnitskiy, M.A.; Ilina, I.Y.; Bubis, J.A.; Solovyeva, E.M.; Zgoda, V.G.; Gorshkov, M. V.; Moshkovskii, S.A. Cysteine Alkylation Methods in Shotgun Proteomics and Their Possible Effects on Methionine Residues. *J. Proteomics* **2021**, *231*, 104022, doi:10.1016/J.JPROT.2020.104022.
412. Lin, Z.; Ren, Y.; Shi, Z.; Zhang, K.; Yang, H.; Liu, S.; Hao, P. Evaluation and Minimization of Nonspecific Tryptic Cleavages in Proteomic Sample Preparation. *Rapid Commun. Mass Spectrom.* **2020**, *34*, doi:10.1002/rcm.8733.
413. Papaleo, E.; Fantucci, P.; De Gioia, L. Effects of Calcium Binding on Structure and Autolysis Regulation in Trypsins. A Molecular Dynamics Investigation. *J. Chem. Theory Comput.* **2005**, *1*, 1286–1297, doi:10.1021/ct050092o.
414. Brunius, G.; Sundbom, B.L. The Effect of Ca²⁺ on the Thermal Stability of Trypsin in Phosphate-Buffered Saline Solution Used for Harvesting of Human Embryonic Lung Fibroblast Cultures. *J. Biol. Stand.* **1987**, *15*, 265–270, doi:10.1016/0092-1157(87)90029-1.

415. Kotormán, M.; Laczkó, I.; Szabó, A.; Simon, L.M. Effects of Ca²⁺ on Catalytic Activity and Conformation of Trypsin and α -Chymotrypsin in Aqueous Ethanol. *Biochem. Biophys. Res. Commun.* **2003**, *304*, 18–21, doi:10.1016/S0006-291X(03)00534-5.
416. Melanson, J.E.; Avery, S.L.; Pinto, D.M. High-Coverage Quantitative Proteomics Using Amine-Specific Isotopic Labeling. *Proteomics* **2006**, *6*, 4466–4474, doi:10.1002/pmic.200600112.
417. Oliveros, J.C. -Venny-. Venn Diagrams for Comparing Lists. Available online: https://bioinfogp.cnb.csic.es/tools/venny_old/venny.php (accessed on 24 July 2022).
418. Crooks, G.E.; Hon, G.; Chandonia, J.-M.; Brenner, S.E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14*, 1188–1190, doi:10.1101/gr.849004.
419. Perez-Riverol, Y.; Bai, J.; Bandla, C.; García-Seisdedos, D.; Hewapathirana, S.; Kamatchinathan, S.; Kundu, D.J.; Prakash, A.; Frericks-Zipper, A.; Eisenacher, M.; et al. The PRIDE Database Resources in 2022: A Hub for Mass Spectrometry-Based Proteomics Evidences. *Nucleic Acids Res.* **2021**, *50*, doi:10.1093/nar/gkab1038.
420. Jeng, J.; Lin, M.F.; Cheng, F.Y.; Yeh, C.S.; Shiea, J. Using High-Concentration Trypsin-Immobilized Magnetic Nanoparticles for Rapid in Situ Protein Digestion at Elevated Temperature. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 3060–3068, doi:10.1002/rcm.3191.
421. Pan, Y.; Cheng, K.; Mao, J.; Liu, F.; Liu, J.; Ye, M.; Zou, H. Quantitative Proteomics Reveals the Kinetics of Trypsin-Catalyzed Protein Digestion. *Anal. Bioanal. Chem.* **2014**, *406*, 6247–6256, doi:10.1007/s00216-014-8071-6.
422. *Saccharomyces Cerevisiae* in PAXdb Available online: <https://pax-db.org/species/4932> (accessed on 29 July 2022).
423. Craik, C.S.; Largman, C.; Fletcher, T.; Rocznik, S.; Barr, P.J.; Fletterick, R.; Rutter, W.J. Redesigning Trypsin: Alteration of Substrate Specificity. *Science*. **1985**, *228*, 291–297.

Appendix A

Required Copyright Reproduction Agreements



Rapid and Quantitative Protein Precipitation for Proteome Analysis by Mass Spectrometry
Author: Jessica L. Nickerson, Alan A. Doucette
Publication: Journal of Proteome Research
Publisher: American Chemical Society
Date: May 1, 2020
Copyright © 2020, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#) [CLOSE WINDOW](#)

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Dec 16, 2022

This Agreement between Dalhousie University -- Jessica Nickerson ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	5450681471767
License date	Dec 16, 2022
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Mass Spectrometry Reviews
Licensed Content Title	Recent advances in top-down proteome sample processing ahead of MS analysis
Licensed Content Author	Alan A. Doucette, Ziheng Dang, Teresa S. McMillen, et al
Licensed Content Date	May 28, 2021
Licensed Content Volume	0
Licensed Content Issue	0
Licensed Content Pages	39

Copyright and Licensing

For all articles published in MDPI journals, copyright is retained by the authors. Articles are licensed under an open access Creative Commons CC BY 4.0 license, meaning that anyone may download and read the paper for free. In addition, the article may be reused and quoted provided that the original published version is cited. These conditions allow for maximum use and exposure of the work, while ensuring that the authors receive proper credit.

In exceptional circumstances articles may be licensed differently. If you have specific condition (such as one linked to funding) that does not allow this license, please mention this to the editorial office of the journal at submission. Exceptions will be granted at the discretion of the publisher.