

Development of miniaturized bioanalytical methods for the qualitative and quantitative analysis of proteins and proteomes by means of digital microfluidics and nanoLC-MS

Dissertation

zur Erlangung des Doktorgrades *Dr. rerum naturalium*

der Mathematisch-Naturwissenschaftlichen Fakultät

der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Max Konstantin Steinbach

Institut für Experimentelle Medizin

Kiel, 2024

Erstgutachter: Prof. Dr. Andreas Tholey

Zweitgutachter: Prof. Dr. Matthias Leippe

Datum der Einreichung: 08.10.2024

Datum der mündlichen Prüfung: 06.12.2024

Zum Druck genehmigt: 10.12.2024

“Keep your focus on your passion in life and the things you enjoy doing. You’ll get there. Trust me. I’m doing it.” – Tim Samaras

Acknowledgements

Writing an acknowledgements section for a PhD thesis is anything but easy when so many people come to mind who have all contributed to the success of this project in so many different ways.

First, I express my sincere thanks to my supervisor and mentor at the Institute for Experimental Medicine, Prof. Dr. Andreas Tholey. Andreas, not only were you willing to take a tremendous gamble by looking beyond grades on paper, believing those to belie the potential in me to see this mission through to the end, but also never ceased providing valuable ideas and guidance in crucial moments to keep everything going. I am overwhelmingly grateful for having had the opportunity to repay – and justify – your support, patience, and trust.

I likewise thank the Faculty of Mathematics and Natural Sciences at *Christian-Albrechts-Universität zu Kiel* for giving me the chance to prove myself as a PhD student. This has not only been the greatest endeavor of my life so far, but also a valuable experience on different levels to draw from in whatever challenging times that may lie ahead for me in the future.

Next, I sincerely thank Jan Leipert who, during our time together at the IEM, had to bear the dual burden of being my practical mentor – providing suggestions and hands-on assistance whenever necessary (and even letting me use gadgets constructed by himself) – all while keeping his own PhD project going. Looking back, he likely played the decisive role in making me feel “at home” in the field of digital microfluidics, and was a stalwart partner in scientific writing time and time again. Jan, there’s no telling how far I might have gotten on my own, but probably nowhere near as far as with your help, and what I do know for sure is that the value of your support cannot be overstated.

Of course, thanks also go to all other of my wonderful current and former colleagues (in no particular order) at the IEM – Patrick Kaleja, Konrad Winkels, Jerome Genth, Dr. Stephanie Bilke, Dr. Philipp Kaulich, Theo Matzanke, Dr. Christian Treitz, Dr. Liam Cassidy, Dr. Tomas Koudelka, Dr. Mohammad Abukhalaf, and Britta Steer – for more instances of invaluable advice, helping hands, joint work on publications, proofreading and discussions than I could possibly keep count of, as well as always being such a fun and friendly crew. Four years are a reasonable chunk of the average human lifetime, and definitely long enough for one to regret not investing that time well. I feel honored for having had the privilege to spend these four years in this group, with all of you.

Additional thanks go to Christine Blurton of the Zoological Institute at *Christian-Albrechts-Universität zu Kiel*, for the cultivation and preparation of *Caenorhabditis elegans*, as well as her constant willingness and patience in addressing any questions I had on them. The German saying “*Der Wurm ist in etwas drin*” (literally: The worm is in something), is an expression of frustration and despair, and one who uses it is typically on the verge of giving up. Ironically, the progression of this work rather depended on the presence of worms (or nematodes, to be precise), as well as guidance on their nomenclature and biology. Thank you, Christine, for providing both whenever it was necessary.

Next, I thank Prof. Dr. Matthias Leippe of the Zoological Institute at *Christian-Albrechts-Universität zu Kiel*, for agreeing to fill the role of second examiner, as well as be a member of the examination committee. I likewise thank him for providing the framework and resources required to even consider commencing experiments involving *C. elegans*, as well as for general guidance and many fruitful discussions.

Furthermore, both Prof. Leippe and Christine Blurton deserve additional thanks for the healthy collaboration between the IEM and Zoological Institute that would lay the foundation for two joint research articles I was glad to contribute to – in one case as first author (first time for me as well!). Thanks to both of you for making that possible (and everyone else involved).

Finally, my thanks go to my family for their continuous, unwavering support throughout this adventure, starting with my parents: You did not only encourage me to study chemistry from the start, but would never cease to support me in the years to come, no matter the circumstances. Even when on the verge of failure, you never questioned me, nor stopped believing that I would eventually pull through. For so many reasons, however I look at it, I would not have a thesis in front of me now to write my acknowledgements in, if it weren't for you – and for that, I shall be forever grateful.

Table of Contents

1.	Introduction	1
1.1.	Definition and relevance of proteomics	1
1.2.	Proteomics: Objectives and challenges	2
1.3.	Liquid chromatography	3
1.3.1.	General definitions and column designs.....	3
1.3.2.	Reverse-phase HPLC	4
1.4.	Mass spectrometry	5
1.4.1.	General introduction and terminology.....	5
1.4.2.	Ionization techniques	6
1.4.3.	Tandem mass spectrometry	7
1.4.4.	Ion species.....	8
1.4.5.	Mass filters & detectors	8
1.4.6.	High-field asymmetric-waveform ion mobility spectrometry (FAIMS).....	10
1.5.	MS-based proteomics: Basics, concepts and challenges.....	11
1.5.1.	MS-based Bottom-Up and Top-Down proteomics.....	11
1.5.2.	Data evaluation in quantitative MS-based proteomics	14
1.5.3.	MS and miniaturization: State of the art, challenges and perspectives.....	16
1.6.	Digital microfluidics.....	18
1.6.1.	Electrowetting on dielectric (EWOD)	19
1.6.2.	Biofouling: Definition, effects and prevention	20
1.6.3.	Air- vs. oil-based builds: Similarities and differences	22
2.	Aims of this thesis	24
3.	Method development for miniaturized proteomics on an oil-based digital microfluidics system	26
3.1.	Introduction	26
3.2.	Project aims & major experiments.....	26
3.3.	Materials & methods.....	27
3.3.1.	Chemicals & reagents.....	27
3.3.2.	Digi.Bio Desktop Biofoundry digital microfluidics device	28

3.3.3.	Jurkat T cell culture.....	29
3.3.4.	Preparation of Jurkat T cell pre-digest.....	30
3.3.5.	Custom drilling procedure.....	31
3.3.6.	Sample treatment to test custom drillings.....	31
3.3.7.	On-chip cell lysis protocol.....	32
3.3.8.	LC-MS/MS method.....	33
3.3.9.	Data evaluation.....	34
3.4.	Results & discussion.....	34
3.4.1.	General features of the device and performance tests.....	34
3.4.2.	Chip customization attempts.....	36
3.4.3.	Full on-chip lysis.....	38
3.5.	Conclusion.....	41
4.	Digital microfluidic sample preparation and low-input Bottom-Up proteomics of the model organism <i>Caenorhabditis elegans</i>	43
4.1.	Introduction.....	43
4.2.	Project aims & major experiments.....	45
4.2.1.	Investigation of various lysis and digestion conditions in oil-based DMF.....	45
4.2.2.	Lysis of single <i>C. elegans</i> specimens for the comparison of DMF- and tube-based methods.....	45
4.2.3.	Label-free quantification to compare two biological conditions.....	45
4.3.	Materials & methods.....	46
4.3.1.	Chemicals & reagents.....	46
4.3.2.	<i>C. elegans</i> preparation and handling.....	46
4.3.3.	Instrumentation.....	48
4.3.4.	DropBot DB3-120 digital microfluidics device.....	48
4.3.5.	Assessment of SDS removal following oil-based DMF.....	50
4.3.6.	Optimization of lysis conditions.....	51
4.3.7.	On-chip lysis protocol: Desktop Biofoundry.....	51
4.3.8.	On-chip lysis protocol: DropBot DB3-120.....	53
4.3.9.	Tube-adapted workflow.....	56
4.3.10.	LC-MS/MS: Orbitrap Q Exactive Plus.....	56

4.3.11. LC-MS/MS: Orbitrap Fusion Lumos	57
4.3.12. Data evaluation	58
4.4. Results & discussion.....	59
4.4.1. Optimization of lysis options	59
4.4.2. Testing of different lysis and digestion conditions (oil-based DMF)	62
4.4.3. Handling aspects of air- and oil-based DMF, and in-tube preparation	67
4.4.4. Air-based DMF vs in-tube: Qualitative results	68
4.4.5. Comparison of different biological conditions by label-free quantification.....	72
4.4.6. Biological interpretation	74
4.5. Conclusion	76
5. Development of microfluidics-supported methods for MS-based Top-Down nanoproteomics.....	77
5.1. Introduction	77
5.2. Project aims & major experiments.....	77
5.3. Materials & methods.....	78
5.3.1. Chemicals & reagents.....	78
5.3.2. <i>C. elegans</i> preparation and handling.....	78
5.3.3. TDP urea lysis buffer	79
5.3.4. Preparation of non-alkylated Jurkat T cell lysate as a low-input standard for TDP	79
5.3.5. Preparation of single <i>C. elegans</i> by DMF.....	80
5.3.6. LC-MS/MS method	83
5.3.7. Data evaluation	84
5.3.7.1. <i>Searching as individual files and subsequent reassembly.....</i>	<i>85</i>
5.3.7.2. <i>Searching CV-separated data as fractions</i>	<i>87</i>
5.3.7.3. <i>Independent database searches and evaluations per CV</i>	<i>88</i>
5.4. Results & discussion.....	89
5.4.1. Single <i>C. elegans</i> TDP LFQ	89
5.4.1.1. <i>Challenges & workarounds</i>	<i>89</i>
5.4.1.2. <i>Re-joining CV-sliced data.....</i>	<i>91</i>
5.4.1.3. <i>Individual evaluations using CV-separated subsets</i>	<i>93</i>
5.4.1.4. <i>Comparing different means of normalization</i>	<i>95</i>

5.4.1.5. <i>Proteoforms vs. inferred proteins</i>	96
5.5. Conclusion	101
6. Method development for microfluidics-supported MS-based Middle-Down nanoproteomics	102
6.1. Introduction	102
6.2. Project aims & major experiments	102
6.2.1. Fine tuning of critical workflow elements for low-input samples.....	102
6.2.2. Optimization of LC-MS/MS conditions	103
6.2.3. Label-free quantification	103
6.3. Materials & methods	103
6.3.1. Chemicals & reagents.....	103
6.3.2. Cultivation of CaCo-2 cells	104
6.3.3. Preparation of CaCo-2 cell samples	104
6.3.4. <i>C. elegans</i> preparation and handling.....	106
6.3.5. Preparation of single <i>C. elegans</i> using DMF	107
6.3.6. LC-MS/MS: Method development.....	108
6.3.7. LC-MS/MS: LFQ with single <i>C. elegans</i>	109
6.3.8. Data evaluation: General	110
6.3.9. Label-free quantification	111
6.3.10. Comparing protein sizes between MDP and TDP	112
6.4. Results & discussion	113
6.4.1. Proteoforms? Peptides? Considerations regarding nomenclature.....	113
6.4.2. Considerations for basic workflow aspects.....	115
6.4.3. Fine-tuning of critical workflow elements for low-input samples.....	116
6.4.4. Optimization of LC-MS/MS conditions.....	122
6.4.5. <i>C. elegans</i> : Middle-Down label-free quantification	124
6.4.5.1. <i>Challenges & workarounds</i>	124
6.4.5.2. <i>Qualitative data evaluation and LFQ</i>	126
6.5. Conclusion	129
7. Discussion	130

7.1.	Considerations for LC-MS/MS	130
7.2.	Evaluation of DMF approaches	130
7.3.	Comparison of proteomics approaches	134
8.	Conclusion	138
9.	Outlook.....	140
10.	References.....	143
	Declaration according to § 9 of the Doctoral Degree Regulations	xxiv

List of figures

Figure 1-1: A typical workflow in mass spectrometry-based proteomics, outlining the most common working steps and giving examples of what these may involve.	2
Figure 1-2: Characteristic structures of packed (a) and monolithic (b) columns.	4
Figure 1-3: Graphical demonstration of the significance of mass accuracy and error.	5
Figure 1-4: Schematic BUP workflow involving reduction, alkylation and digestion.....	12
Figure 1-5: Example of a typical two-plate DMF chip and electrode arrangement.....	18
Figure 1-6: Schematic demonstration of the EWOD effect.....	20
Figure 1-7: Comparison of the different DMF principles.....	23
Figure 3-1: Digi.Bio Desktop Biofoundry device.....	28
Figure 3-2: Schematic of Digi.Bio's DMF chip design at the time of the EAP	29
Figure 3-3: A DMF chip (Digi.Bio, Amsterdam, the Netherlands) filled with silicone oil saturated with the azo dye Sudan IV.	35
Figure 3-4: Schematic of the chip subjected to the custom drilling procedure.....	37
Figure 3-5: Chip schematic highlighting the preferred spatial distribution of samples	40
Figure 4-1: Lateral schematic views of the anatomy of adult <i>C. elegans</i>	44
Figure 4-2: Schematic of the <i>C. elegans</i> treatment prior to the LFQ experiment.....	48
Figure 4-3: Magnetic lens setup for the DropBot DB3-120.....	49
Figure 4-4: Graphical demonstration of the DMF sample preparation workflow.	55
Figure 4-5: Schematic of the in-tube BUP sample preparation workflow.....	56
Figure 4-6: Signal to noise ratios of SDS-d25 after multiple consecutive washing steps. 60	
Figure 4-7: Effects of five different lysis buffers on single live <i>C. elegans</i> over time.....	61
Figure 4-8: Average numbers of inferred protein identifications (a), protein groups (b), peptide groups (c) and PSMs (d) from the single- <i>C. elegans</i> lysis experiments on the Digi.Bio Desktop Biofoundry platform. Varying numbers of replicates were due to loss of complete samples.	64
Figure 4-9: Average numbers of proteins inferred with no, one or two missed cleavage events, compared across five experiments performed on the Digi.Bio platform.	66
Figure 4-10: Comparison of proteins inferred with high confidence (a) and peptide groups (b) in single <i>C. elegans</i> after BUP sample preparation on the DropBot DB3-120 DMF system or via an in-tube workflow.....	70
Figure 4-11: Overlap between replicates of inferred proteins from <i>C. elegans</i> after sample preparations via DMF- and in-tube protocols.....	71
Figure 4-12: Comparison of average filtered inferred proteins (a) and peptide groups (b) from “Heat” and “Control” samples after either DMF- or tube-based sample preparation. 72	

Figure 4-13: Volcano plots of inferred proteins from the DMF (a) and in-tube (b) datasets, highlighting biologically relevant proteins	74
Figure 5-1: DMF workflow applied to single <i>C. elegans</i> for TDP sample preparation.	82
Figure 5-2: DropBot DB3-120 setup in combination with the magnetic lens	83
Figure 5-3: Schematic of how CV slicing using multi-CV data files was achieved.	85
Figure 5-4: Processing of TDP data after searching as CV-separated subsets	86
Figure 5-5: TDP database search approach with CV slicing as well as pre-fractionation in Proteome Discoverer 2.5.	88
Figure 5-6: TDP data processing workflow with separate database searches per CV.	89
Figure 5-7: Volcano plot of quantified proteoforms with at least three valid values per group and no imputation.	92
Figure 5-8: Volcano plot of proteoforms quantified originally after a single fractionation...	93
Figure 5-9: Volcano plots of proteoforms per FAIMS CV, including only those proteoforms with at least three valid values per group.	95
Figure 5-10: Histograms of Log ₂ -transformed abundance values of the TDP dataset of single <i>C. elegans</i> after a search as fractions.....	96
Figure 5-11: Proteoforms of the 40S ribosomal protein S28 found to be significantly more abundant in the control (blue) or heat stress group (red) at a FAIMS CV of -50 V.....	98
Figure 5-12: Comparison of TDP with BUP results	99
Figure 6-1: Flow chart of the in-tube low-input MDP sample preparation workflow.....	106
Figure 6-2: Cleavage events observed in a Glu-C digest time series	118
Figure 6-3: Positions of the identified analytes in relation to the protein sequence.....	119
Figure 6-4: Overlap of Middle-Down peptides from CaCo-2 digests found in separate digest and eluate fractions, combined from four replicates.....	122
Figure 6-5: Volcano plot of Middle-Down peptides from single <i>C. elegans</i>	127

List of tables

Table 3-1: Numbers of inferred proteins, protein- and peptide groups, and PSMs, obtained on the Desktop Biofoundry device from preparations of Jurkat pre-digest (equivalent of ~1000 cells) and plain digest buffer, all removed via the same custom drill hole.	38
Table 3-2: Numbers of inferred proteins, protein- and peptide groups, and PSMs, obtained in a full cell lysis experiment (Jurkat T cells, ~100/sample) and blank run on the Desktop Biofoundry device.....	41
Table 4-1: Lysis agents and respective volumes tested on single live <i>C. elegans</i>	51
Table 4-2: Experimental parameters common and different across two DMF protocols...	67
Table 4-3: Numbers of inferred proteins, protein and peptide groups, and PSMs from the qualitative experiment series using single N2 adults.....	69
Table 4-4: Proteins quantified in the DMF- and in-tube datasets after applying valid value filters of varying strictness. Data originally published in (Steinbach et al., 2022).....	73
Table 5-1: Number of proteoforms per CV with at least three valid values per group.....	94
Table 6-2: Overview of proteins, Middle-Down peptides and PrSMs assigned for all single <i>C. elegans</i> replicates. Replicates chosen for LFQ evaluation are shaded green.....	111
Table 6-3: Proteins, isoforms, Middle-Down peptides and PrSMs for all replicates. Replicates chosen for comparison of isoform size with TDP data are shaded green.	113
Table 6-4: Numbers of proteins, peptides and PrSMs, as well as the average theoretical peptides masses and incidence of glutamate residues at the C termini of peptides, in a digestion time series.....	117
Table 6-5: Numbers of proteins, peptides, PrSMs and annotated (i.e. undigested) proteoforms, as well as the average theoretical peptides masses, in a series testing elution with different concentrations of FA	120
Table 6-6: Numbers of proteins, peptides, PrSMs, and annotated (i.e. undigested) proteoforms, as well as the incidence of glutamate residues at peptide C termini and the average theoretical peptides masses, in a series using ~1000 CaCo-2 cells as starting material with fractionation maintained.....	121
Table 6-7: Numbers of proteins, peptides and PrSMs, as well as the average theoretical peptides masses, in a series comparing different LC-MS/MS methods	124

The author was involved in the following publications:

Contributions to research articles

1. Leipert, J., **Steinbach, M. K.**, & Tholey, A. (2021). Isobaric Peptide Labeling on Digital Microfluidics for Quantitative Low Cell Number Proteomics. *Analytical Chemistry*, 93(15), 6278–6286. <https://doi.org/10.1021/acs.analchem.1c01205>

Author contributions – Max K. Steinbach: Method development and sample preparation.

2. **Steinbach, M. K.**, Leipert, J., Blurton, C., Leippe, M., & Tholey, A. (2022). Digital Microfluidics Supported Microproteomics for Quantitative Proteome Analysis of Single *Caenorhabditis elegans* Nematodes. *Journal of Proteome Research*, 21(8), 1986–1996. <https://doi.org/10.1021/acs.jproteome.2c00274>

Author contributions – Max K. Steinbach: Method development, sample preparation, data analysis, visualization, preparation of manuscript.

3. Leipert, J., Kaulich, P. T., **Steinbach, M. K.**, Steer, B., Winkels, K., Blurton, C., Leippe, M., & Tholey, A. (2023). Digital Microfluidics and Magnetic Bead-Based Intact Proteoform Elution for Quantitative Top-down Nanoproteomics of Single *C. elegans* Nematodes. *Angewandte Chemie International Edition*, 62(28), e202301969. <https://doi.org/https://doi.org/10.1002/anie.202301969>

Author contributions – Max K. Steinbach: Sample preparation, quantitative data analysis, interpretation, visualization, preparation of manuscript.

Review article

Steinbach, M. K., Leipert, J., Matzanke, T., & Tholey, A. (2024). Digital Microfluidics for Sample Preparation in Low-Input Proteomics. *Small Methods*, n/a(n/a), 2400495. <https://doi.org/https://doi.org/10.1002/smtd.202400495>

Author contributions – Max Steinbach: Literature search, preparation of manuscript, preparation of figures.

Conference poster

Steinbach, M.K., Leipert, J., Kaulich, P.T., Steer, B., Winkels, K., Blurton, C., Leippe, M., & Tholey, A. (2023). Bottom-up and top-down proteomics from single *C. elegans* nematodes using digital microfluidics and magnetic bead-based sample preparation.

Presented at the 54th annual conference of the *Deutsche Gesellschaft für Massenspektrometrie* (Dortmund, May 14th – 17th 2023).

List of Abbreviations

ACN: Acetonitrile

AmBiC: Ammonium bicarbonate

BSA: Bovine serum albumin

BUP: Bottom-Up proteomics

CaCo-2: Caucasian colon adenocarcinoma cell line

CID: Collision-induced dissociation

cRAP: common Repository of Adventitious Proteins

CV: Compensation voltage

DDM: n-Dodecyl- β -D-maltoside

DDOPM: 3-Dodecyloxypropyl-1- β -D-maltopyranoside

DISCO: Digital microfluidic Isolation of Single Cells for -Omics

DMF: Digital microfluidics

DMSO: Dimethylsulfoxide

DTT: Dithiothreitol

DV: Dispersion voltage

EIC: Emitter-integrated column

ESI: Electrospray ionization

EWOD: Electrowetting on dielectric

FA: Formic acid

FACS: Fluorescence activated cell sorting

FAIMS: High-field asymmetric-waveform ion mobility spectrometry

FDR: False discovery rate

HCCA: α -Cyanohydroxycinnamic acid

HCD: Higher-energy collisional dissociation

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

IAA: Iodoacetamide

ITO: Indium-tin oxide

iTRAQ: Isotope Tags for Relative and Absolute Quantitation

IUPAC: International Union of Pure and Applied Chemistry

LFQ: Label-free quantification

LIT: Linear ion trap

HPLC: High performance liquid chromatography

MALDI: Matrix-assisted laser desorption ionization

MBR: Match between runs

MDP: Middle-Down proteomics

MNA-C12: 1,2,5,6-tetra- β -D-glucopyranoside-3,4-O-Di-dodecyl-d-mannitol (Mannitol-Based Amphiphile C12)

MS: Mass spectrometry

MS/MS: Tandem mass spectrometry

m/z: Mass-to-charge ratio

nanoPOTS: Nanodroplet Processing in One pot for Trace Samples

NGM: Nematode growth medium

PBS: Phosphate buffered saline

PMMA: Polymethylmethacrylate

PrSM: Proteoform sequence match

PSM: Peptide-spectrum match(es)

PTM: Posttranslational modification

RPMI: Cell culture medium; named after Roswell Park Memorial Institute

SCP: Single-cell proteomics

SCoPE-MS: Single-Cell ProtEomics Mass Spectrometry

SDS: Sodium dodecyl sulfate

SILAC: Stable Isotope Labeling with Amino acids in Cell culture

SIM: Single ion monitoring

SP3: Single-pot solid phase-enhanced sample preparation

TDP: Top-Down proteomics

TFA: Trifluoroacetic acid

TMT: Tandem mass tag

ToF: Time-of-flight

Abstract

Constant technological advances in analytics enable proteomics research on an increasingly smaller scale, which inevitably requires suitable downscaling of sample preparation workflows as well. The overall goals of this thesis were to investigate and develop promising sample preparation methods for handling low-input biological samples, and adapt these methods to become suitable for different concepts of proteomics.

The first subproject primarily served the purpose of being an introduction to digital microfluidics (DMF) in general, while simultaneously providing the invaluable opportunity to test an oil-based DMF device in early access with regard to a multitude of different aspects relevant for potential use in nanoproteomics.

Next, the knowledge and experience acquired on the theory and practical use of DMF, as well as the preferred instrumental setup for LC-MS/MS, were applied in a series of Bottom-Up proteomics (BUP) experiments, using single specimens of the model organism *Caenorhabditis elegans* (*C. elegans*) as the preferred sample material. The aim was to pair a highly efficient DMF-based sample preparation with LC-MS conditions optimized for low-input proteomics to reproducibly achieve an unprecedented number of protein identifications from single nematodes. Once in place, the workflow was extended to also allow label-free quantification. Compared to an identical in-tube sample preparation, the DMF workflow led to a markedly higher number of inferred proteins considered differentially abundant between two groups.

Once a DMF-based workflow, viable both qualitatively and quantitatively, had been successfully established for BUP, the next goal was to transfer this mode of sample preparation to TDP and eventually MDP, and thus generate independent yet comparable quantitative datasets for single *C. elegans* according to three different approaches in proteomics (BUP, TDP and MDP). Furthermore, evaluation of these three datasets revealed high complementarity and clearly demonstrated the current respective strengths and limitations of each approach to proteomics.

Zusammenfassung

Durch ständigen technologischen Fortschritt in der Analytik wird Proteomikforschung in einem immer kleiner werdenden Maßstab möglich, was gleichzeitig die entsprechende Miniaturisierung der Probenvorbereitung erfordert. Die übergeordneten Ziele dieser Arbeit waren, vielversprechende Methoden zum Umgang mit biologischen „Low-Input“-Proben zu untersuchen und zu entwickeln, sowie diese Methoden zur Eignung für verschiedene Konzepte der Proteomik anzupassen.

Das erste Teilprojekt der Arbeit diente hauptsächlich der Einarbeitung in die digitale Mikrofluidik (DMF) im Allgemeinen. Gleichzeitig ergab sich die wertvolle Möglichkeit, im Rahmen eines Early-Access-Programms ein ölbasierendes DMF-Gerät in Bezug auf verschiedene für die Nanoproteomik relevante Aspekte zu testen.

Im weiteren Verlauf wurden die neu erworbenen Kenntnisse zur Theorie hinter DMF und im praktischen Umgang mit diesen Geräten in Form einer Messreihe für „Bottom-Up“-Proteomik (BUP) zur Anwendung gebracht, ebenso wie die dafür bevorzugte LC-MS/MS-Gerätekonfiguration. Hierbei stellten Einzelexemplare des Modellorganismus *Caenorhabditis elegans* (*C. elegans*) das bevorzugte Probenmaterial dar. Ziel war es, eine hocheffiziente DMF-basierte Probenvorbereitungsmethode mit für Low-Input-Proteomik optimierten LC-MS-Bedingungen zusammenzuführen, um so eine bisher unerreichte Anzahl an Proteinidentifikationen aus einzelnen Nematoden reproduzierbar zu erzielen. Nach Fertigstellung der neuen Methode wurde diese zudem erweitert, um auch Label-Free-Quantifizierung zu erlauben. Verglichen mit einer identischen Vorbereitung in herkömmlichen Reaktionsgefäßen führte der DMF-Workflow zu einer deutlich höheren Anzahl inferierter Proteine mit Abundanzunterschieden zwischen zwei Gruppen.

Nach der Etablierung eines sowohl für qualitative als auch quantitative Fragestellungen geeigneten DMF-basierten Probenvorbereitungsverfahrens für BUP war das nächste Ziel, diese Methode auf TDP und schließlich MDP zu übertragen. Schlussendlich sollten auf diese Art und Weise, für alle drei Grundprinzipien der Proteomik (BUP, TDP und MDP), unabhängige und trotzdem vergleichbare quantitative Datensätze aus einzelnen Exemplaren von *C. elegans* generiert werden. Darüber hinaus zeigte die Auswertung der drei Datensätze große Komplementarität auf und demonstrierte klar die derzeitigen Stärken und Limitationen der verschiedenen Prinzipien der Proteomik.

1. Introduction

1.1. Definition and relevance of proteomics

Modern-day bioanalysis has a strong focus on the “-omics” fields, such as genomics, transcriptomics and proteomics. These terms are used to describe the large-scale research of the associated analyte group, while the respective “-ome” represents the entirety of this group present in a defined biological entity (e.g. an organism or cell) at any given point in time. A particularly important example is the proteome, which is defined as the entire complement of proteins that can be expressed by a single genome (Wasinger et al., 1995). Proteins are built from the 22 so-called proteinogenic amino acids that contribute different biochemical properties to the resulting protein structure, characteristics and functions. The connected field of proteomics has developed into a fast-growing community with ever-diversifying research interests, and constitutes the focus of the work presented here.

The sequence of amino acid residues as encoded by DNA nucleotide chains is referred to as the canonical sequence of a protein, taking into account only the product to be expected based on this code and disregarding any possible deviations. However, mechanisms such as alternative splicing of RNA or different translation starting sites can lead to synthesis of different proteins (isoforms) from the same gene (Aebersold et al., 2018; Andreadis et al., 1987; Schlüter et al., 2009). Therefore, expression of an encoding gene may not necessarily result in the canonical protein. Additionally, a proteome is highly dynamic and responsive to outside factors, and all canonical proteins and isoforms are therefore prone to experiencing modifications after synthesis (posttranslational modifications [PTMs], e.g. phosphorylation) (Steinbach et al., 2024). As a result, multiple states of the same protein (referred to as proteoforms) are likely to be present at any given time (Smith et al., 2013; Smith & Kelleher, 2018). Furthermore, proteins and their isoforms also frequently experience artificial modifications introduced during sample preparation (Geiszler et al., 2021), such as formylation (Lenčo et al., 2020). In this thesis, the term “proteins” therefore refers to canonical proteins only.

Focusing on the final product of protein biosynthesis and thus on an analyte class directly reflecting the condition of a cell or organism, proteomics pursues the goal of understanding biological processes and accordingly maintains a high research interest. A key limitation to proteomics is that proteins cannot be amplified; therefore, analyses are

always limited to the material available at the outset. This can present a particular challenge whenever the sample material is highly limited by nature, e.g. in the case of rare cell types such as circulating tumor cells (CTCs) (Song et al., 2017). Conversely, while amplification options do exist in genomics or transcriptomics, the scope of both fields cannot extend beyond the process of protein biosynthesis, and they are therefore inherently unable to examine any changes proteins or associated isoforms may undergo once synthesized (PTMs), as well as the biological processes they are directly involved in.

1.2. Proteomics: Objectives and challenges

Proteomic analyses can pursue a number of goals, including straightforward identification of proteins, quantification, and characterization of modifications on the molecular level. Due to the complexity and dynamic nature of proteomes, and often challenging biological sample materials, achieving these goals requires numerous processing steps (Figure 1-1).

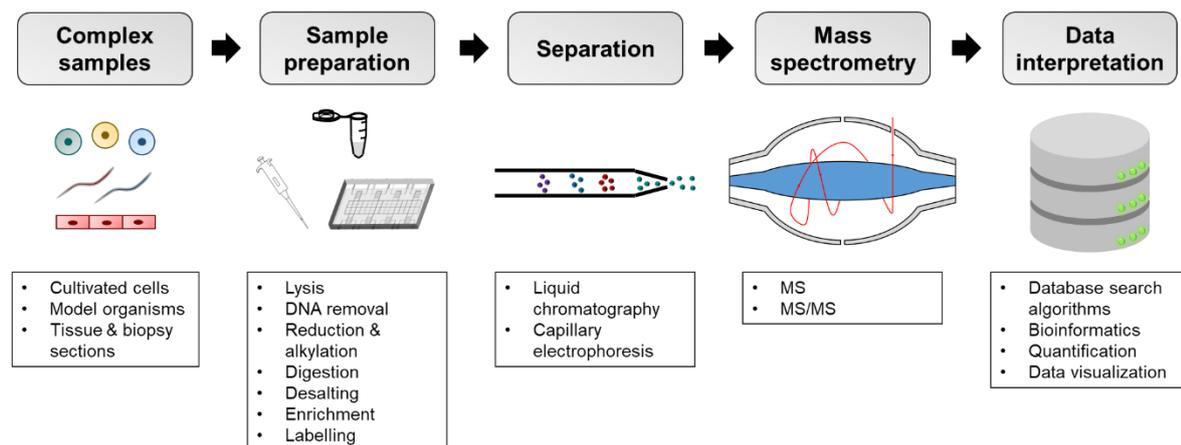


Figure 1-1: A typical workflow in mass spectrometry-based proteomics, outlining the most common working steps and giving examples of what these may involve. Adapted from (Steinbach et al., 2024). Use licensed under CC-BY-NC 4.0 (<https://creativecommons.org/licenses/by-nc/4.0/>). Copyright 2024 The Authors.

Regardless of the experiment objective, samples of biological materials such as cells, tissue sections, or even whole model organisms, have to undergo a thorough preparation process. Prior to being processed, the proteome as a whole first needs to be solubilized and isolated from other components of the sample. Once pretreated as required per the objective, proteins have to be separated in order to minimize interference with each other, for which purpose liquid chromatography and capillary electrophoresis are popular options. Separation is often followed by mass spectrometry, the data gained from which

needs to be interpreted by bioinformatical means. Due to the complex nature of these steps, each one of them is explained in greater detail in the following subchapters.

1.3. Liquid chromatography

1.3.1. General definitions and column designs

Liquid chromatography (LC) has evolved to be one of the most popular, diverse and widely applicable means of achieving chromatographic separation in the field of proteomics.

An LC system always consists of a solid stationary and liquid mobile phase. The former typically constitutes the essential means for analytes to be retained to different degrees (and thus separated) and is present within a column. The latter (also referred to as the eluent) serves to provide the coaxial flow required to both achieve elution of analytes from the stationary phase as well as ensure successful transport to downstream collection, processing or detection equipment. In many applications, the mobile phase consists of two components: an aqueous (polar) and a highly organic (non-polar) phase. Depending on experimental requirements, the mobile phase may be applied at a constant composition (isocratic method), or subject to a linear gradient, where the mobile phase composition is changed gradually over time (gradient method).

Frequent operation of LC systems at elevated eluent pressures gave rise to the term high-pressure liquid chromatography (HPLC). Although the abbreviation is still valid to date, the definition has since changed to read “high-performance liquid chromatography”.

The internal structure of an HPLC column varies depending on the separation principle and design. Chromatographic supports for LC may be based on inorganic or organic polymers, as well as hybrids thereof (Arshady, 1991), with the most widely used material being silica (Berthod, 1991). The support material has to be sufficiently stable mechanically in order to withstand common pressures in HPLC, and accommodate a high concentration of functional groups to retain analytes while itself being chemically inert to the latter. Additionally, the surface area and pore size have to be adequate for the application at hand (Nazario & Lanças, 2013).

The structure of the stationary phase and support within a column is typically divided into two categories, namely packed and monolithic columns. In packed columns, the stationary phase is supported by fine porous particles, which constitute the filling of the columns (Figure 1-2a). Conversely, monolithic columns contain a single solid and highly porous (sponge-like) structure (Figure 1-2b), advantages of which being reduced separation times by allowing higher flow rates without risking a damaging pressure increase (Nazario & Lanças, 2013).

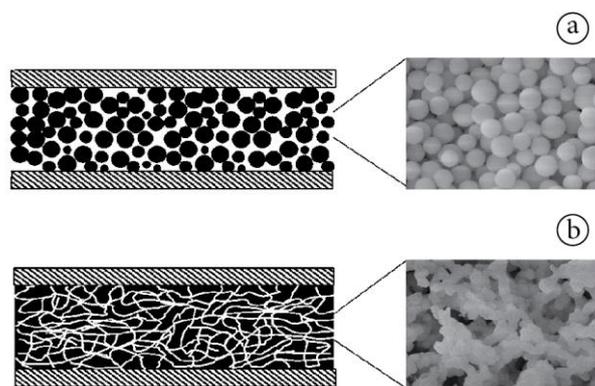


Figure 1-2: Characteristic structures of packed (a) and monolithic (b) columns. Reproduced with permission from (Nazario & Lanças, 2013). Copyright 2013 The Authors.

1.3.2. Reverse-phase HPLC

One of the most common separation systems for LC, reverse-phase HPLC (RP-HPLC) generally makes use of a relatively non-polar (i.e. hydrophobic) stationary phase. Contrarily, the mobile phase is typically applied at a relatively high polarity (i.e. low hydrophobicity) initially, before hydrophobicity is gradually increased through initiation of an organic solvent gradient. In a system like this, the retention time (RT, i.e. the time after which analytes elute from the column) increases with the hydrophobicity of the analytes: The more hydrophobic, the more strongly substances interact with the stationary phase, and the higher the proportion of organic solvent in the mobile phase needs to be to elute them.

This mode of HPLC is the preferred variant in proteomics, owing to the broad hydrophobicity range of proteins or peptides derived thereof. Accordingly, common stationary phases in proteomics consist of functionalized C_4 or C_{18} chains. Additionally, RP-HPLC analytical columns are often used in conjunction with trapping columns (or precolumns), containing a similar or identical stationary phase, by connecting these

columns to a multiport switching valve. This way, proteomes can be trapped on the precolumn while flushing common polar contaminants such as salts (Vissers et al., 2002).

1.4. Mass spectrometry

1.4.1. General introduction and terminology

Mass spectrometry (MS) is a measurement technique by which analytes present as ions are separated and characterized based on their respective mass-to-charge ratio (m/z). While build principles of mass spectrometers greatly vary, all have in common the presence of three essential regions – the ion source, mass filter and detector. The ion source ensures that analytes are present in a charged state upon leaving and thus able to be processed. The mass filter fulfils the purpose of sorting analyte ions based on their m/z value. In the detector, the presence of ions is converted to the readout later evaluated. A common type of detector are electron multipliers (Dietz, 1965).

Performance of MS devices is characterized by a variety of parameters. Mass resolution is a measure for the ability of a mass analyzer to distinguish different ions as baseline-separated signals, whereas mass error is defined as the difference between the reported and theoretical m/z value (Marty, 2022) (Figure 1-3). The lower the mass error, the closer a reported m/z value can be expected to be to the theoretical value.

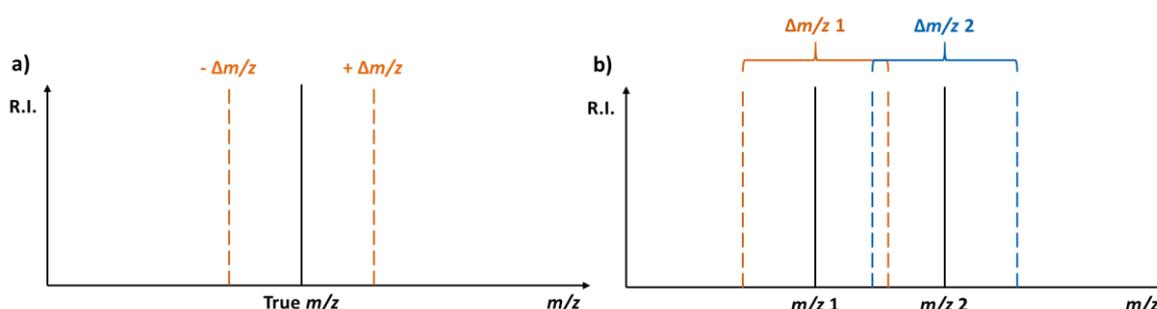


Figure 1-3: Graphical demonstration of the significance of mass accuracy and error. The mass accuracy of a device defines a margin of error around the true m/z of an analyte, within which the measured signal is expected to be (a). Unambiguous identification may be impossible in the case of analytes with similar m/z values, where the mass error can produce overlap zones (b). Mass spectra are highly simplified and idealized as line spectra; R.I. is relative intensity.

Due to the resolving power of MS devices, it is critical to consider definitions of mass values, the most important one for MS being the monoisotopic mass, where only the presence of the most abundant and stable naturally occurring isotopes is assumed. While generally accepted, this definition of the monoisotopic mass may be problematic in the context of isotopically modified substances, e.g. heavy water (D₂O) (Attygalle et al., 2022). For this reason, this mass value may be more suitably described as the default mass of a molecule with a specific isotopic composition, without any additional rare isotopes to consider.

1.4.2. Ionization techniques

All mass spectrometric analyses share the necessity of analytes being present with either a positive or negative charge; therefore, the first key component of each mass spectrometer is the ion source, which dictates by which means analyte ions are formed, and which coupling options can be applied. The most common ionization technique in LC-MS-based proteomics is electrospray ionization (ESI).

Known as a continuous ionization method, ESI always relies on a steady inflow of liquid containing the sample and is therefore most commonly coupled to an LC setup. Analyte ions are notably already largely present in the eluent stream, and ion formation at this stage is often aided by using volatile mobile phase additives (e.g. ammonium formate or formic acid for positive polarity). Upon reaching the ESI source, the sample passes through a heated capillary, to which a charge equal to the selected ion polarity is applied. Combined with the effect of an electric field beyond, the accumulation of identical charges at the end of the capillary leads to the solvent meniscus being drawn out into a stable cone shape (Taylor cone) (Banerjee & Mazumdar, 2012; Taylor, 1964).

Once detached from the capillary, the charged solvent droplets containing the analyte are drawn towards an orifice by means of a charge contrary to the selected polarity. During this process, the solvent gradually evaporates, decreasing the volume of the droplets and thus simultaneously increasing the charge density within the remaining liquid (Fenn et al., 1989). When the repulsion between the equal charges present within the droplets exceeds the cohesive force of the latter, droplets disintegrate (Coulomb explosion), releasing gas-phase analyte ions (Yamashita & Fenn, 1984). A key advantage of ESI is that ions are rarely fragmented prematurely; therefore, ESI is considered a “soft” ionization method.

In contrast, the highly popular ionization method known as matrix-assisted laser desorption ionization (MALDI), a variant of laser desorption (Karas et al., 1985), requires samples to be measured in combination with a high molar excess of a matrix substance. This sample/matrix mixture is spotted onto a steel plate, where all solvent evaporates, leaving the sample embedded within a crystalline matrix structure. The sample/matrix crystal is then targeted by a quick succession of laser pulses, causing the matrix structure to enter the gas phase and take the analyte with it (Karas et al., 2000). The matrix has several purposes: While it can aid in ionization by acting as a proton donor or acceptor, depending on the matrix and desired polarity, it also absorbs a substantial proportion of the energy from the laser, ensuring that the analyte is not fragmented completely. Although the latter still undergoes dissociation due to being impacted by the laser, those ions that do eventually end up being measured are likely to be intact, for which reason MALDI is also counted among the “soft” ionization methods.

1.4.3. Tandem mass spectrometry

Complex samples with a plethora of analytes and shared m/z values hinder the identification of a certain analyte based on its m/z alone. A popular solution to this problem is controlled fragmentation of ions followed by an additional mass filtering step, a process referred to as tandem mass spectrometry, MS/MS or MS^n , due to it essentially involving multiple consecutive applications of the same principle. Ions to be fragmented in this process are termed precursor or parent ions, whereas the resulting fragments are called product ions. MS/MS greatly increases specificity by making use of characteristic fragmentation patterns: Precursor ions non-distinguishable in MS due to identical or highly similar m/z values may well be distinguished in MS/MS due to the formation of different product ions.

A common way to achieve fragmentation is collision-induced dissociation (CID), where precursor ions travelling at high speeds in the gas phase impact inert gas molecules, thereby converting the kinetic energy of precursor ions to internal energy and thus promoting dissociation (Brodbelt, 2016; McLuckey, 1992). While popular in MS-based proteomics due to a high ionization efficiency, an important drawback of CID is bias towards cleavage of labile bonds, resulting in unspecific neutral losses such as water or ammonia. Additionally, CID used in ion trap mass analyzers (to be discussed in chapter 1.4.5) causes the lower m/z range to be truncated; a phenomenon also termed low mass cutoff (LMCO) (Brodbelt, 2016). Another fragmentation mode is higher-energy collisional

dissociation (HCD). Exclusive to ion traps, this method utilizes ion acceleration in a separate collision cell or multipole. A key advantage of HCD is that it alleviates the aforementioned LMCO problem (Brodbelt, 2016).

1.4.4. Ion species

One of the defining characteristics of the different ionization methods available are the ion species produced during this process. In atmospheric pressure ionization methods like ESI or MALDI, ions are mostly formed by protonation or deprotonation of molecules, with functional groups being particularly susceptible to such events. For instance, amines or ester moieties are most likely to be protonated, generating a cation, whereas carboxylic acids and sulfonates are usually deprotonated, producing an anion. Protonated or deprotonated molecules are given as $[M+H]^+$ or $[M-H]^-$, where M represents the original molecule, followed by an operator to indicate either an adduct or a loss, and finally the group involved in this event (here: a single proton). In positive polarity, the most common choice in MS-based proteomics, adduct ions can be observed whenever a charge-bearing group other than a proton is present (e.g. $[M+Na]^+$). Another common occurrence are isotopologues, which are analytes with a shared sum formula and structure, but a different isotopic composition.

1.4.5. Mass filters & detectors

One of the most frequently used mass filters is the quadrupole, which consists of four rod-shaped electrodes arranged coaxially in a diamond shape. In this configuration, two opposing rods share a potential polarity, forming an electrically connected pair. Furthermore, neighboring electrodes are subject to radio frequency (RF) alternating current (AC), as well as direct current (DC) potentials, based on a time-dependent function. Regulating the ratio of AC to DC potential ensures that only ions with an appropriate m/z ratio are able to maintain a stable flight path and thus clear the quadrupole (Hug, 2015).

A quadrupole is typically employed in one of two modes. In scan mode, AC frequency is adjusted in small steps, gradually allowing ions within a defined m/z window to pass through. Contrarily to this, single ion monitoring (SIM) sees the AC frequency set to a fixed value, thereby limiting access to only one m/z . Quadrupoles are often used consecutively

in groups of three, where the first and third (termed Q1 and Q3, respectively) serve as mass filters. The second quadrupole in such a setup (q2) serves as a collision cell, which is filled with an inert gas as a means to cause CID of precursor ions.

An inherently different principle of separating ions is by their time of flight (ToF). Used most frequently in combination with MALDI, ToF separation is based on ions being accelerated in an electric field. A key aspect to this is that all ions gain the identical amount of kinetic energy during this process. Therefore, ions of an identical charge state but different masses have different velocities and thus arrive at the detector after different flight times. Flight time calibration is achieved by introducing a mixture of standards of known mass, which allows an accurate translation of flight times to m/z values within a limited mass window.

MALDI-ToF MS devices may be operated in two modes, known as linear and reflector mode. The former is the most straightforward way of separation by ToF, where ions traverse the length of the flight tube in a linear path before reaching the detector at the opposite end. In reflector mode, ions initially follow the same flight path; however, instead of being detected at the distal end of the flight tube, they enter an electric field generated by the so-called reflectron and are propelled in the opposite direction, giving their path a near parabolic shape.

Another diverse category of mass filters are ion traps. Such devices are based on isolating a package of ions in a confined space, before eventually releasing the ions consecutively for detection. A popular design within this category is the linear ion trap (LIT). Storage of ions in an LIT requires both axial and radial confinement, with the former being achieved by means of stopping potentials at end electrodes, and the latter by applying two-dimensional radio frequency fields. This type of trap can therefore be constructed as a conversion of a linear multipole. LITs are standalone mass analyzers and highly versatile in their applications, able to be used in combination with other mass analyzers for MS/MS experiments, or to study ion-molecule chemistry (Douglas et al., 2005).

Whenever an inert gas is present in the trap, ions undergo a process known as collisional cooling. By repeatedly colliding with gas particles, ions gradually lose kinetic energy, slowing down in the process and moving towards the center of the quadrupole. Ions can then be released selectively via resonance excitation (Savaryn et al., 2016). The orientation of ion ejection depends on the build of ion trap, with the most straightforward mode being axial ion ejection. In contrast, radial ion ejection was first mentioned in 1995 as part of a proposal to use an LIT as a standalone mass spectrometer and thereby

increasing the volume of the trap (Bier et al., 1995). This idea would later be explored by Schwartz and co-workers (Schwartz et al., 2002).

One type of ion trap that has become especially popular in MS-based proteomics is the Orbitrap mass analyzer (Q. Hu et al., 2005). Introduced in 2005, this mass analyzer is based on the principle of the Kingdon (Kingdon, 1923) and Knight traps (Knight, 1981), while also bearing some resemblance to the Paul trap and Fourier transform ion cyclotron resonance (Q. Hu et al., 2005; Hug, 2015). The Orbitrap features axially symmetric inner and outer electrodes, and ions are injected perpendicular to the long axis of the inner electrode. Stable ion flight paths are notably made up not only of movement around the inner electrode, but also oscillation along the z-axis. The frequency of the latter oscillation is essential for deriving the m/z ratio in Orbitrap mass analyzers (Makarov, 2000).

Orbitraps are particularly renowned for their high mass resolution and –accuracy (Denisov et al., 2012; Q. Hu et al., 2005), making them invaluable for MS-based proteomics. Additionally, Orbitraps and LITs may be used in tandem during the same measurement, an example being the so-called Orbitrap-ion trap (OTIT) mode, where the Orbitrap is used to filter precursor ions and an LIT focuses on the analysis of product ions. Measurements performed in OTIT have been demonstrated to be advantageous in proteomics, in particular following fragmentation by HCD (Levy et al., 2018).

1.4.6. High-field asymmetric-waveform ion mobility spectrometry (FAIMS)

Prior to MS analysis, ions can be subjected to high-field asymmetric waveform ion mobility spectrometry (FAIMS), where ions are first directed between two electrodes by means of a carrier gas flow. The ions are then subjected to the influence of an oscillating electric field generated from a high-frequency waveform applied to the electrodes (referred to as dispersion voltage [DV]) (Creese & Cooper, 2012). Essential requirements are that the waveform is asymmetric and the net electric field between the combined electrodes is zero in every cycle of the waveform (i.e. the time per cycle spent within either the high- or low-field region has to be identical for both electrodes). Ions entering the gap between the electrodes are only able to pass through if their mobility in the high- and low-field regions is identical, whereas in any other case they gradually deviate from their original path, eventually resulting in a collision (Cooper, 2016). This collision course can be counteracted by applying a so-called compensation voltage (CV). The CV suitable for stabilizing the flight path of an ion depends directly on the latter's mobility in the high- and

low-field regions. For this reason, the FAIMS module may somewhat be used as a filter, with the associated benefits of an increased sensitivity and dynamic range. The selective consecutive application of multiple CVs allows ion fractionation, with each CV enabling ions of respective compatible mobility properties to pass on to the mass filter.

Ion fractionation by FAIMS after ESI is known to offer various advantages, one being that it is suitable for removing (mostly singly charged) contaminant ions that could otherwise interfere with measurements (Hebert et al., 2018; Venne et al., 2005), e.g. detergents introduced during sample preparation. FAIMS has become highly relevant in proteomics, especially as an element in low-input measurements, where it could recently be proven to substantially aid in the detection of proteins and peptide groups in as little as single mammalian cells (Cong et al., 2021; Hebert et al., 2018).

1.5. MS-based proteomics: Basics, concepts and challenges

1.5.1. MS-based Bottom-Up and Top-Down proteomics

MS-based proteomics is commonly divided into two concepts, known as Bottom-Up and Top-Down proteomics (BUP and TDP, respectively).

The general principle of BUP is that proteins are not measured intact, but instead hydrolyzed into characteristic peptides (Figure 1-4 and Figure 1-1). Most commonly, the first treatment step is the addition of a reducing agent (e.g. dithiothreitol [DTT]) to the sample. The purpose of this step is to disrupt disulfide bonds, reducing them to thiol groups. Following reduction, an alkylation agent (e.g. iodoacetamide [IAA]) is commonly applied to alkylate the newly formed thiol groups, ensuring they do not react to form new bonds (Hale et al., 2004). Notably, sample preparation techniques omitting reducing and alkylation have also been reported (Budnik et al., 2018). The defining working step in a BUP preparation is the addition of a protease and subsequent digestion. Since proteases favor different hydrolysis sites, the nature and number of peptides formed during this process depends on the protease introduced. One of the most popular proteases for BUP is trypsin, which hydrolyses at the C-terminus of lysine and arginine residues to create peptide molecules. These can then be separated and analyzed via MS/MS (Dongré et al., 1996; Huang et al., 2005; Swaney et al., 2010), after which the original proteins are inferred based solely on the peptides detected.

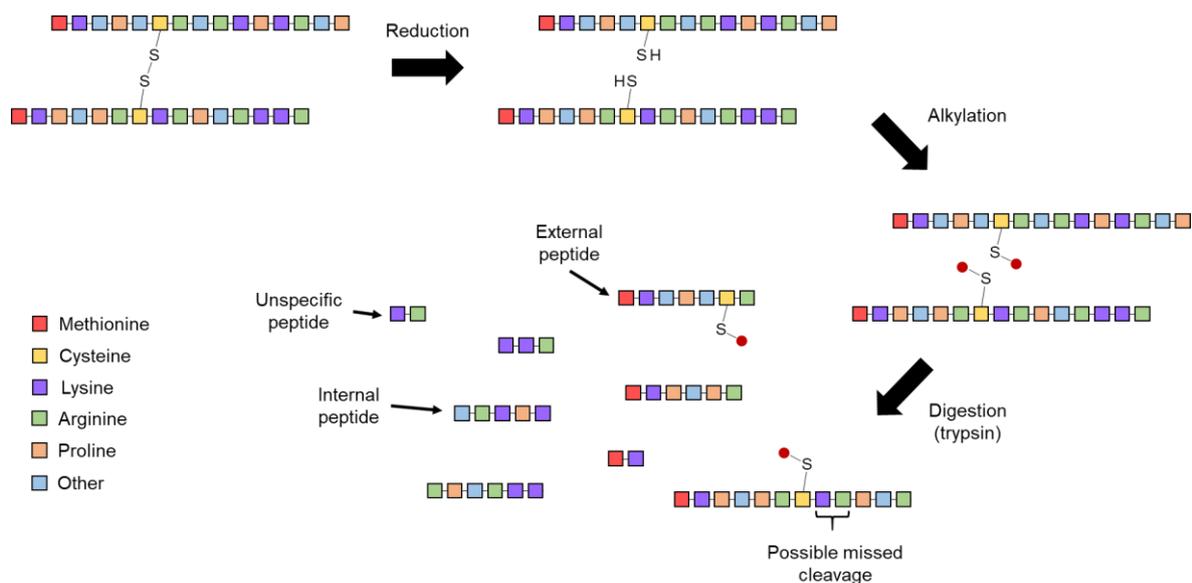


Figure 1-4: Schematic BUP workflow involving reduction, alkylation and digestion. Trypsin is used as an example enzyme. N-termini of amino acids or sequences thereof are to the left, C-termini to the right. Red orbs are meant to represent carbamidomethylation of cysteine residues.

Cleavage specificity of proteases may present a problem if proteins only possess a limited number of hydrolysis sites, leading to a number of peptides too low to confidently infer the associated protein. Conversely, the presence of a large number of possible hydrolysis sites per sequence portion may also present a problem, as resulting peptides may be too small to be specific for the protein. An approach to circumvent both issues is to apply multiple proteases to the same biological sample. By producing complementary peptides, this can lead to a higher overall peptide count and increased sequence coverage (Giansanti et al., 2016; Kaulich et al., 2021; Swaney et al., 2010).

The most important limitation of BUP is that proteins are inferred, but never identified directly (Nesvizhskii et al., 2003). Being by nature substantially shorter than the original proteins, the likelihood of different peptide sequences having similar molecular weights or high sequence homology may lead to the available MS/MS spectra being assigned to the wrong peptides, which in turn can cause proteins to be wrongly inferred. However, even with MS/MS spectra correctly assigned to peptides, protein inference may still be ambiguous if the peptides are shared across multiple protein sequence portions (e.g. due to sequence homology or multiple proteoforms of the same canonical protein) (Nesvizhskii et al., 2003; Nesvizhskii & Aebersold, 2005). Whether a protein is inferred unambiguously therefore depends on the quality of MS/MS spectra and number of assigned peptides, in particular that of unique peptides.

While results in BUP are most commonly reported as protein identifications, it therefore needs to be considered that these can only be achieved indirectly at best. Accordingly, BUP results published as “protein identifications” are treated as “inferred proteins” for the purpose of this thesis.

In contrast to BUP, TDP methods typically analyze intact proteins without the need for these to be digested by proteases (Winkels et al., 2021). This often not only substantially reduces sample preparation time compared to BUP, but also allows one to distinguish between multiple proteoforms corresponding to the same canonical protein, whereas this information is generally lost after cleavage of proteins into peptides (Aebersold et al., 2018; Fulcher et al., 2021; Smith et al., 2013; Smith & Kelleher, 2018).

Despite having multiple advantages, TDP is also accompanied by several challenges. The most important limitation of TDP compared to BUP is inherently lower sensitivity: In BUP, the total number of ions contributing to the inference of a single canonical protein is relatively high, due to the presence of associated peptides. Conversely, TDP is restricted to a comparatively low number of identifiable proteoforms, leading to limited signal intensity. Furthermore, ionized proteoforms are often present simultaneously in a much broader variety of charge states than peptides, limiting the intensity per charge state and increasing the complexity of the resulting spectra (Tholey & Becker, 2017). The factors described here narrow the proteoform size range for TDP measurements, with ~50 kDa currently considered an approximate upper limit (Takemori et al., 2024).

One alternative approach to BUP and TDP is known as Middle-Down proteomics (MDP). This approach mainly relies on enzymatic digestion of proteins, using proteases that target amino acid residues of lower abundance and/or are specific for only a single residue, such as Asp-N, Glu-C, Lys-C and Lys-N (Cristobal et al., 2017; Garcia, 2010), or chemical cleavage of peptide bonds. Additionally, enzyme activity can be further inhibited by adjusting the digestion temperature and time period. Middle-Down peptides therefore generally have molecular weights in excess of 3 kDa and are thus in a range above that typical for BUP peptides (Takemori et al., 2024).

1.5.2. Data evaluation in quantitative MS-based proteomics

The inherent complexity of samples in proteomics leads to a need for specialized data acquisition and processing methods. Different approaches to data acquisition are available for LC-MS-based proteomics, classified as data-dependent or data-independent acquisition (DDA and DIA, respectively) (Davies et al., 2021), the former currently being the most common one. In DDA, precursor ions are automatically selected in an MS scan (usually the 10 or 15 most intense ions), which is then followed by MS/MS to measure the corresponding product ions. While this comes with the risk of introducing bias towards the more abundant ions, and thus redundancy, this issue can be alleviated by a method termed dynamic exclusion, where precursor ions selected once are disregarded in consecutive scans (Chapman et al., 2014). Unlike DDA, preselection of precursors based on intensity does not take place in DIA. Accordingly, options in DIA are to either fragment all ions within a certain m/z window that are present in the MS device at a particular point in time (broadband DIA) (Chapman et al., 2014), or consecutively scan defined ranges within that m/z window (Chapman et al., 2014; Venable et al., 2004). Compared to DDA, important advantages of DIA analyses are increased numbers of proteins inferred and a higher sequence coverage per protein, as has been reported in several studies (Blackburn et al., 2010; Geromanos et al., 2009; Levin et al., 2011). Aside from these advantages of DIA, limitations exist as well. Database search times and false positive rates are likely to be higher in DIA, as precursor ions are not preselected and their mass accuracy therefore depends directly on the m/z window scanned (Venable et al., 2004). While analysis of datasets generated by DIA can be facilitated with spectral libraries, these typically require additional experimental effort and resources to create beforehand (Ludwig et al., 2018).

In quantitative proteomics, a common application is the comparison of biological specimens (e.g. cells or whole model organisms) after being subjected to different stimuli. These may involve being challenged with a particular pathogen (Treitz et al., 2015), exposed to chemical compounds (Dong et al., 2010; Leipert et al., 2021) or incubated at different temperatures (Bensaddek et al., 2016; Steinbach et al., 2022).

A plethora of quantification approaches for MS-based proteomics has been developed. Many of these rely on labelling, with prominent examples being “isobaric tags for relative and absolute quantitation” (iTRAQ) (Wiese et al., 2007), “stable isotope labelling by amino acids in cell culture” (SILAC) (Ong et al., 2002), and “tandem mass tags” (TMT) (Thompson et al., 2006). In contrast, quantitative analyses may also be achieved solely using abundance ratios of unmodified peptides or proteoforms, which are directly detected

based on their characteristic product ions. Known as label-free quantification (LFQ), this type of analysis has different advantages. Compared to labelling approaches, LFQ workflows are often less complex by design, while also usually requiring less specialized reagents, thus making them more cost-effective (Hamid et al., 2022).

Aside from these advantages, downsides to LFQ also exist. One of these is a relatively high incidence of missing values, the occurrence of which depending on factors such as sample input amount, preparation, measurement sensitivity, criteria applied during bioinformatics-supported data processing (e.g. database searches), and others (Hamid et al., 2022; Lazar et al., 2016). These may be of considerable detriment during subsequent statistical tests, especially when the dataset consists of a low total number of values (e.g. due to few technical replicates and/or strict filtering).

Missing values may be dealt with in different ways, examples being imputation or features such as Match Between Runs (MBR). Introduced as part of the MaxQuant software (Tyanova, Temu, & Cox, 2016), MBR has become a popular tool to aid in data evaluation whenever multiple technical replicates of the same sample are available. In a BUP analysis, a protein is typically considered identified in a sample when one or multiple unique peptides of this protein are detected in a measurement; however, whenever adequate peptide signals cannot be detected, proteins may be falsely considered absent. When applying MBR, signals that would not qualify as peptide signals on their own can still be linked to particular peptides if these have been confirmed in one or multiple other runs of the same series. Therefore, a protein otherwise considered absent from a replicate may still be assumed present if evidence from other replicates of the same series backs this assumption up.

In quantitative proteomics, multiple steps are required during raw data processing. Common steps are normalization and transformation (e.g. Log₂-transformation) of raw abundance values. Data of multiple replicates must also be grouped based on the conditions the samples were originally subject to. This is finally followed by a statistical test, e.g. a two-sample *t* test, which yields information regarding abundance differences of proteins between the two groups (i.e. fold change), as well as the statistical significance of these differences (Wang et al., 2008).

1.5.3. MS and miniaturization: State of the art, challenges and perspectives

In recent times, MS-based proteomics has been on a trend of miniaturization, leading to workflows allowing proteome analyses with ever decreasing amounts of starting material. Termed low-input, micro- or nanoproteomics, the aim of this approach is to perform reliable analyses with protein amounts in the mid-nanogram range or below, and even down to the single cell level (Couvillion et al., 2019; Gebreyesus et al., 2022; Zhu, Piehowski, Kelly, et al., 2018).

Currently, the widely accepted primary goal is to achieve reliable single-cell proteomics, thereby allowing the investigation of differences between proteomes of individual similar cells (e.g. from the same tissue or culture). Gene expression has been repeatedly shown to vary stochastically across single cells, resulting in the amount of RNA copies per gene and cell not correlating well with the abundance of associated proteins (Liu et al., 2020; Taniguchi et al., 2010). Investigation of such inter-cell differences is impossible when performing a traditional bulk preparation, during which the entire multitude of single-cell proteomes introduced is irreversibly merged, thereby masking the proteome characteristics of any single cells not conforming with the bulk average (Specht & Slavov, 2018). The same is true on a macroscopic scale when it comes to the analysis of multicellular organisms commonly prepared in bulk (e.g. *Caenorhabditis elegans* [*C. elegans*]) (Bensaddek et al., 2016).

Working with small amounts of protein-containing sample can pose several challenges, such as dilution of the available material by the mere addition of reagents. For instance, a common BUP workflow comprising a reduction, alkylation and digestion step would dilute the sample three times (once at each working step). Reagents also remain in the sample even beyond the duration of the working step they were meant for, leading to the risk of interference with downstream processes. For example, protease activity may be hampered by urea once part of the lysis reagent, or a detergent involved in the same step may cause ion suppression during MS/MS. A technique named single-pot solid-phase-enhanced sample preparation (SP3) has become renowned for addressing both of these problems (Hughes et al., 2014, 2019; Moggridge et al., 2018). Here, proteins are non-covalently bound to magnetic particles in the presence of an excess of organic solvent. These particles are pelleted or resuspended by selective application of a magnetic field, allowing contaminants to be washed away and isolated proteins to be processed afterwards in any way necessary.

In 2018, two particularly promising techniques in the field of low-input proteomics were introduced, one being “nanodroplet Processing in One pot for Trace Samples” (nanoPOTS) (Zhu, Piehowski, Zhao, et al., 2018), the other a sample preparation method based on an oil-air droplet chip (Li et al., 2018). While being used exclusively for BUP at the time of introduction, nanoPOTS would eventually also be used successfully for TDP involving input of as little as 100 mammalian cells (M. Zhou et al., 2020). Additionally, the nanoPOTS method would later be adapted to improve pipet accessibility (Weke et al., 2021) and automation potential (Liang et al., 2021).

Another example of a single-cell proteomics method is SingleCell ProtEomics by Mass Spectrometry (SCoPE-MS), which applied TMT labelling to peptides from the single cells of interest, as well as an excess of carrier peptides. Due to the use of a carrier channel, labelled peptides lost in the course of nanoLC separation were statistically likely not to be the sought-after single-cell peptides, ensuring that the majority of the latter remained available for measurement (Budnik et al., 2018).

Presently, an important challenge in single-cell proteomics remains the isolation of single cells, and thus the reproducibility of sample input, since cell counts during sample preparations with suspensions of loose cells are prone to error. A straightforward means to address this problem is the use of fluorescence-activated cell sorting (FACS) directly prior to the actual sample preparation (Hodne & Weltzien, 2015). In recent years, new approaches have been made to address the difficulties in acquiring single cells. An example for this is piezo acoustic liquid dispensing, which has seen widespread use in a variety of recent low-input proteomics applications (Ctortecka et al., 2023; Eshghi et al., 2023; Leduc et al., 2021, 2022; Petelski et al., 2021).

A similar interest in miniaturization exists for LC, with flow rates in the range of several hundred nL per minute at maximum (nanoLC), and inner column diameters at about 100 μm or lower, having become common in proteomics. While low flow rates are required to prevent excessive pressures that might damage nanoLC columns, they also benefit ESI-MS measurements by generating smaller droplets and thus increasing ESI efficiency (Y. Shen & Smith, 2005; Wilm & Mann, 1996). Downsides to this are slower column regeneration, as well as a higher gradient delay and effect of dead volumes.

Presently, numerous different approaches to increasing performance in nanoLC exist, and choosing the ideal option among these for the purpose of an experiment is a success-determining factor. Available options include common packed, monolithic, ultra-low-flow and emitter-integrated columns (EIC). EICs feature an ESI emitter equipped directly to the

front end of the column itself, effectively eliminating dead volumes between the column and ion source. The main disadvantage of EICs is that the emitter cannot be changed, and any damage to the latter, or simply natural ageing, renders the column useless.

1.6. Digital microfluidics

An especially promising and diverse solution to low-volume sample manipulation in bioanalytical research is digital microfluidics (DMF) (Kothamachu et al., 2020). The foundation of every DMF device is an array of electrodes, made from e.g. chromium (Fobel et al., 2014; Luk & Wheeler, 2009), indium tin oxide (ITO) (Samiei et al., 2016; Srinivasan et al., 2004), or gold (Barbulovic-Nad et al., 2008; Samiei et al., 2016), that can be manipulated individually (Figure 1-5). The most important feature of DMF devices is the ability to handle liquid samples and reagents without requiring any means of manual interaction (pipetting). Operators may therefore reduce sample loss on vessel walls to a high degree while maintaining full control over the position of the droplets and thus avoiding accidental mixing.

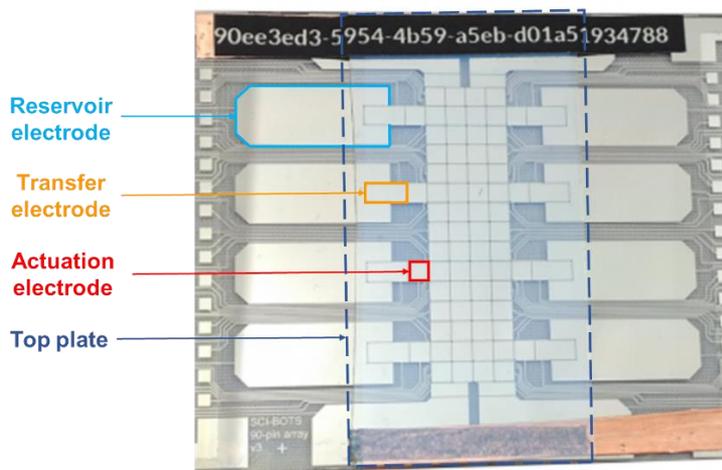


Figure 1-5: Example of a typical two-plate DMF chip and electrode arrangement (here: the 90-pin array v3 for the DropBot DB3-120 [SciBots Inc., ON, Canada]). The working area is defined by the top plate. Liquids are introduced via the reservoir electrodes by being placed against the hydrophobically coated edge of the top plate.

1.6.1. Electrowetting on dielectric (EWOD)

Droplet movement on DMF devices can be achieved in multiple ways, including electromagnetic movement of droplets containing magnetic particles (Kanitthamniyom et al., 2020; Zhang & Nguyen, 2017), or acoustic waves that spread across the surface bearing the liquid (Ding et al., 2013).

The most popular variant of DMF relies on the phenomenon of electrowetting on dielectric (EWOD) (Choi et al., 2012; Wheeler, 2008), utilizing an electrode array covered by an insulating layer (the dielectric), which is often equipped with an additional hydrophobic layer (e.g. Teflon AF, CYTOP or FluoroPel) to increase repulsion of polar solutions (Choi et al., 2012). Upon subjecting an electrode to a voltage, any liquid sample located on the coated dielectric directly above experiences an immediate change in contact angle. Simultaneously, the liquid is drawn towards the area actuated, spreading out evenly across the electrode and effectively being locked in place as long as the voltage is maintained. If even a small overlap with an adjacent electrode exists, activation of the latter can partially draw the liquid away from the originally activated electrode. The transition is completed when the original electrode is disengaged again (Figure 1-6 a-d).

EWOD DMF devices can further be categorized as one- or two-plate systems. A one-plate system essentially consists of a single surface layer on which liquids are placed, albeit without any cover (Figure 1-6). In contrast, a two-plate setup comprises two identical oppositely aligned surfaces that trap droplets in between (Zhang & Nguyen, 2017). In the latter configuration, the top plate typically consists of ITO, which is equipped with a hydrophobic coating and serves as a ground electrode (Choi et al., 2012; Shamsi et al., 2014).

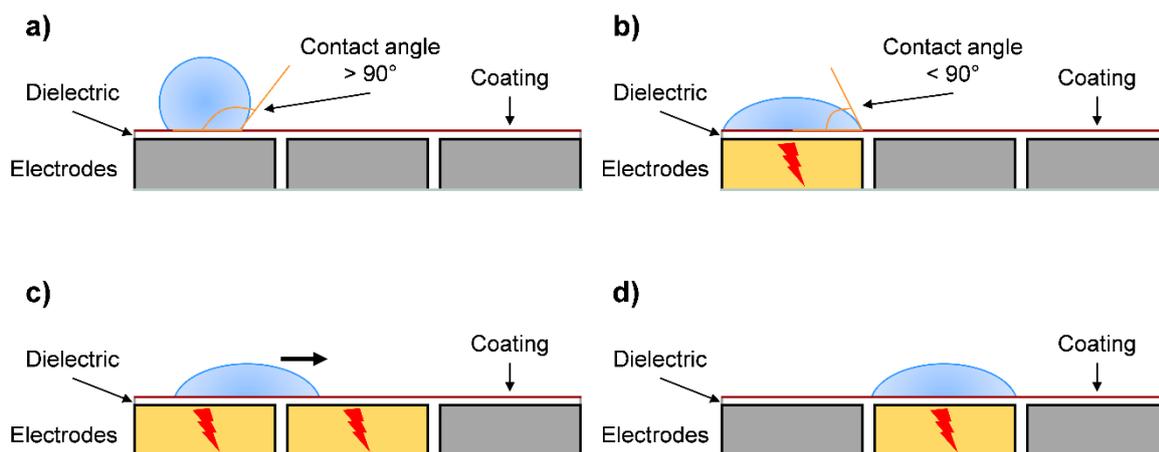


Figure 1-6: Schematic demonstration of the EWOD effect. a) A single droplet is present on the dielectric, while all electrodes are deactivated. **b)** The electrode beneath the droplet is actuated, causing an immediate change in the contact angle and spreading of the drop on the electrode. **c)** Activation of a neighboring electrode partially draws the droplet if sufficient electrode overlap exists. **d)** Deactivation of the original electrode allows the droplet to transit to the next spot. Actuated electrodes are shaded yellow and marked with a red lightning icon. Reproduced from (Steinbach et al., 2024). Use licensed under CC-BY-NC 4.0 (<https://creativecommons.org/licenses/by-nc/4.0/>). Copyright 2024 The Authors.

Since droplet manipulation requires the sample droplet to overlap with neighboring electrodes, samples too small cannot be moved via electrode actuation. For this reason, shape and size of electrodes are the crucial factors defining the range of working volumes that can be handled on a particular EWOD DMF device; additionally, gap width needs to be considered in the case of two-plate devices. The necessity of electrode overlap can also be a considerable complication in the context of evaporation: Drops originally of a sufficient volume may shrink during incubation steps and become immobile.

1.6.2. Biofouling: Definition, effects and prevention

Depending on the nature of the sample, added reagents, duration of the sample treatment and state of the dielectric, droplets and particularly biomolecules within may become increasingly prone to reversibly or irreversibly adsorbing to the surface. Termed biofouling, this process is detrimental in that it causes sample loss, as well as contamination of the chip, rendering the latter useless for later experiments. Simultaneously, moisture may infiltrate the dielectric and initiate electrolysis, causing electrode failure in the process (Eydelnant et al., 2012).

A common way to address the problems of droplet pinning and biofouling is by using detergents as additives to samples and reagents, which lower the surface tension and thus enhance droplet mobility. Having been shown to not only be effective in preventing biofouling, but also compatible even with live cells, poloxamer additives such as Pluronic F-68 or F-127 have seen widespread use in the development of DMF methods for bioanalysis (Au et al., 2011; Luk et al., 2008). A more recent example is ethylenediamine-tetrakis-(ethoxylate-*block*-propoxylate)-tetrol (Tetronic 90R4), a poloxamine detergent which has been used in the preparation of samples consisting of different bodily fluids (Dixon et al., 2016; Ng et al., 2018). The mechanism by which these two detergent classes inhibit biofouling has been hypothesized to be direct interaction with proteins (Jia et al., 2005; Lee et al., 2011); however, recent evidence suggests the detergent molecules to instead form a protective layer on the device surface, thereby making the latter inaccessible for proteins (Ho et al., 2023).

Detergents should be chosen carefully to ensure compatibility with the sample material itself, as well as any stages of sample preparation aside from DMF. For example, numerous detergents popular for DMF itself are incompatible with LC-MS. In the event of co-elution with the analytes, detergents may cause ion suppression and thus hamper the detection of lowly abundant ions. Such detergents should therefore be removed as efficiently as possible before commencing LC-MS. The aforementioned SP3 method has been shown to be a well-suited way to achieve this on DMF chips (Leipert & Tholey, 2019). Alternatively, LC-MS-compatible detergents such as 3-dodecyloxypropyl-1- β -D-maltopyranoside (DDOPM) or zwitterionic mixtures such as Invitrosol (IVS) may be used. In 2007, IVS was reported to elute after the vast majority of analytes such as peptides in a typical RP-HPLC gradient and thus not constitute a source of interference (E. I. Chen et al., 2007), with the same later found for DDOPM. Additionally, DDOPM has recently been shown to benefit BUP by reducing the number of missed cleavages during a tryptic digestion (Leipert et al., 2021). Another option are cleavable detergents such as the acid-cleavable RapiGest, which can be degraded during the often favorable acidification step preceding LC-MS (E. I. Chen et al., 2007), or the photodegradable 4-hexylphenylazosulfonate (otherwise known as "Azo"), which was only recently demonstrated to be compatible with DMF (Chan et al., 2023).

1.6.3. Air- vs. oil-based builds: Similarities and differences

DMF devices can be classified as air- and oil-based systems (see Figure 1-7). In air-based builds, chips are ready to use by default and do not require a liquid filler substance on the inside, only a stable electrical connection to the interface with the main device (Figure 1-7a). Liquids can then be loaded and experiments commenced without further preparation.

A downside to this design principle is a relatively high rate of evaporation, which is further accelerated if the device concerned is a one-plate system where drops are fully exposed (Abdelgawad et al., 2008). Evaporation becomes particularly relevant whenever sample preparation takes substantial time and/or requires the use of concentrated organic solvents. The issue has been limited in the past by enclosing DMF chips within a humidified chamber (thereby creating a saturated atmosphere) (Jebrail et al., 2010), or sealing them subject to a moderate vacuum (Steinbach et al., 2022). Another aspect to note is the inevitable direct contact between liquids and the chip surface, which means that droplet mobility depends directly on the degree of surface adhesion of the liquid in question, and may increase the risk of biofouling.

In contrast, oil-based DMF systems always rely on the chips being pre-filled with a non-polar liquid of low viscosity (Figure 1-7b), with silicone oils being among the most popular filler substances. Oil as a filler medium is known to possess different advantages. Sample droplets are permanently enveloped in a non-volatile substance, thereby effectively eliminating the risk of evaporation (S. Chen et al., 2021; Choi et al., 2012). Additionally, the oil serves as a protective layer between electrodes and sample drops, increasing droplet mobility and potentially reducing biofouling compared to air-based systems. An important requirement is that samples and reagents need to be completely immiscible with the oil (or any other filler medium), and interactions of biomolecules with the phase boundary must be minimal (Baret, 2012). While this is seldom a problem whenever aqueous samples are concerned, working steps involving common organic solvents (e.g. ACN, ethanol) often need to be planned around miscibility with oil (Pamula et al., 2005). A sample surrounded by any kind of filler medium also constitutes a high risk of contamination, since remnants of the medium may be carried along with the sample upon retrieval of the latter. Depending on the downstream methods and equipment to be used, extensive sample cleanup may be necessary to alleviate this problem.

The respective advantages of air- and oil-based DMF may also be united in the shape of a hybrid system (Brassard et al., 2008). Unlike the concept of a fully oil-based system, oil is

not used as a filler medium, but instead applied sparingly to individual droplets in the desired amount, creating a protective shell to both limit evaporation and increase droplet mobility, while leaving the bulk of the chip filled with air (Figure 1-7c).

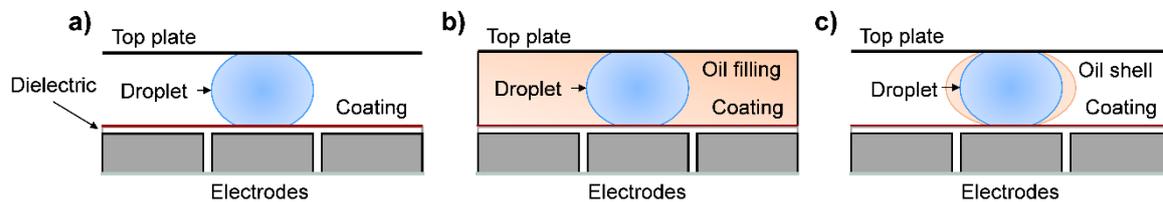


Figure 1-7: Comparison of the different DMF principles. a) Air-based system, with the sample directly added to the chip. b) Oil-based system, with the sample surrounded by an oil filling that occupies the entire chip. c) Hybrid build, where the sample is surrounded by a non-polar liquid shell, but the chip itself is filled with air. Adapted from (Steinbach et al., 2024). Use licensed under CC-BY-NC 4.0 (<https://creativecommons.org/licenses/by-nc/4.0/>). Copyright 2024 The Authors.

2. Aims of this thesis

The fact that proteins cannot be amplified like DNA or RNA remains the primary obstacle for low-input proteomics, and efficiency at two stages therefore is of utmost importance for such experiments. Firstly, sample preparation is a vital part of countless applications in biochemistry in general, where it represents not just the only stage during which operators directly interact with sample material, but also that with the highest risk of sample contamination or loss. This stage thus dictates the state and available amount of sample material. The second decisive step in this pipeline is downstream processing and eventually measurement of samples, which may comprise e.g. chromatographic separation followed by MS(/MS). Direct interaction with and modification of samples is no longer possible at this point, making it essential to instead optimize the analytical aspects of the process.

The overall aim of this thesis was to develop new methods for making low-input proteomics more reliable and accessible, emphasizing the two aforementioned experimental stages. Additionally, method development on both fronts – sample preparation and downstream processing – initially was to be done around the BUP concept.

Research in terms of downstream processing options would need to focus on optimization of LC separation and instrumental parameters. This would include comparing a number of inherently different column designs, while also taking into account the characteristics of the respective mass spectrometer and additional hardware options such as FAIMS ion fractionation.

The primary challenge of sample preparation evident from the start was the risk of sample loss to surfaces, a problem confounded further by the inherently low material amounts to be used throughout the project. Offering the possibility to forego a major amount of manual pipetting steps, DMF presented itself as a promising alternative to traditional tube-based sample preparation, with different design concepts (oil- and air-based) available at the time. A first vital task in optimizing DMF-based methods therefore was to investigate the advantages and limitations of each design to determine the potential of DMF as a substitute for tube-based preparation, and if so, whether one of the two designs would be especially advantageous for later proteomics experiments.

Another essential aspect of DMF method development was to investigate how the critical steps of a typical MS-based BUP workflow – reduction, alkylation, SP3 and digestion –

could be translated to the DMF format. This needed to take into account aspects such as the surface characteristics and spatial constraints of the chips, necessary additives to promote droplet mobility, and miscibility with the filler medium in the case of oil-based DMF. A final point of consideration was if and how samples, once prepared on-chip, would have to be processed for them to be safely measured. With regard to the latter aspect, designing workflows around the removal of potentially problematic reagents after their intended working step, as well as the use of downstream-compatible reagents post-removal, was essential.

Following the comparison of LC columns and DMF versions, the next cornerstone would be the combination of a novel DMF-based sample preparation method, optimized nanoLC separation, and MS/MS utilizing FAIMS, for a BUP experiment on a nanoproteomics scale. The next subproject would then have to focus on how the existing DMF workflow could be modified to become suitable for the two other proteomics approaches, TDP and MDP, while keeping complexity as low as possible. The ultimate goal was to use reliable, easy to execute and efficient DMF methods to generate complementary datasets for all three concepts of nanoproteomics.

3. Method development for miniaturized proteomics on an oil-based digital microfluidics system

3.1. Introduction

The technology behind DMF is constantly undergoing further development, be it to broaden the range of possible applications, downscale working volumes further (e.g. by implementing smaller electrodes), or refine instrument builds to better suit a particular purpose. The field of proteomics in particular can be expected to benefit from these advancements of DMF. While the focus has been increasingly shifting to low-input proteomics, the challenge of proteomes not being amplifiable remains a decisive obstacle in this field, highlighting the need for efficient preparation and limitation of sample losses. To this end, ensuring that samples not only stay mobile on chip surfaces, but also do not undergo a high degree of evaporation, is imperative. Oil-based DMF, where chips are filled with a water-insoluble and non-conductive filler substance such as silicone oil, addresses this by fully engulfing samples, while also somewhat improving droplet mobility. However, the introduction of a second liquid phase presents a new set of challenges to proteomics sample preparation that need to be taken into account, and circumventing these is a key element of the method development process.

3.2. Project aims & major experiments

For the following chapter, one particular oil-based DMF device (Desktop Biofoundry, Digi.Bio, the Netherlands) was acquired as part of an Early Access Program (EAP) and subjected to a series of experiments to benchmark the performance of the device in its current state of development, before putting the device to the test with newly developed sample preparation protocols for nanoproteomics.

The first objective was to test the oil-based DMF platform in terms of general handling, assess its potential for proteomics experiments, and identify limitations. Prior to attempting proteomics experiments, preliminary fluidics tests were necessary in order to test the mobility of different liquid solutions and thus establish the conditions required to prevent droplet pinning and biofouling on this oil-based DMF system. Another critical aspect of early testing was to investigate the solvent compatibility of the system, the key points of interest being how miscible particular solvents were with the silicone oil filling, how easily

they could absorb oil-soluble dye, and how they would interact with the polymer top plate. A second objective was to explore possible design improvements, with the goal of identifying low-cost alterations easy to implement on-site, and without damaging default chips beyond use. Following preliminary experiments, the potential of the Digi.Bio DMF device and unmodified chips for on-chip nanoproteomics was to be assessed by designing workflows for a complete cell lysis and proteolytic digest specifically for use on this system.

3.3. Materials & methods

3.3.1. Chemicals & reagents

Acetonitrile (ACN; Chromasolv quality), FA and TFA (both LC-MS-grade), were obtained from Honeywell Specialty Chemicals Seelze GmbH (Seelze, Germany), ethanol (absolute) from Th. Geyer GmbH & Co. KG (Renningen, Germany).

Gibco PBS buffer (10 ×) and Pierce borate buffer (20 ×) were from Thermo Fisher Scientific GmbH (Bremen, Germany).

The detergents n-dodecyl β -maltoside (DDM), 3-dodecyloxypropyl-1- β -D-maltopyranoside (DDOPM) were from Anatrace Inc. (Maumee, OH, USA), whereas NP-40 substitute and 3-(1-pyridinio)-1-propanesulfonate (NDSB-201), G-Biosciences (St. Louis, MO, USA) and Acros Organics (Schwerte, Germany), respectively.

Trypsin/Lys-C mix was from Promega GmbH (Walldorf, Germany).

SP3 procedures relied on carboxylate-modified magnetic particles (SeraMag SpeedBead magnetic carboxylate with an average particle size of 0.7-1.1 μm ; henceforth referred to as "SP3 beads"), of which two variants – hydrophilic and hydrophobic – were used. Hydrophilic and –phobic SP3 beads (both as 50 $\mu\text{g}/\mu\text{L}$ suspensions in 0.05% w/v azide) were always mixed to equal parts, washed twice with milli-Q-water, then resuspended at the desired concentration in water.

The beads of choice, silicone oil (5 cSt), as well as all other chemicals and reagents not explicitly mentioned, were purchased from Sigma Aldrich (Steinheim, Germany).

3.3.2. Digi.Bio Desktop Biofoundry digital microfluidics device

The DMF device by Digi.Bio presented here was notably still under development by the vendor at the time during which this subproject was undertaken (July 2020 – August 2021). For this reason, any performance- and feature assessments made here cannot reflect the state of development of the device beyond August 2021.

The device (Figure 3-1) was meant to enable the handling of small liquid volumes (in a range from 1-10 μL) by DMF. As additional features, the Biofoundry device comprised both an internal, retractable magnet and a temperature module that affected localized areas of the accompanying chips (purification and temperature modules, respectively). Other features included the ability to close the device around chips, thereby isolating them from both light and the environment, as well as optional internal lighting and camera support. Furthermore, chips were supposed to be filled completely with silicone oil (5 cSt).

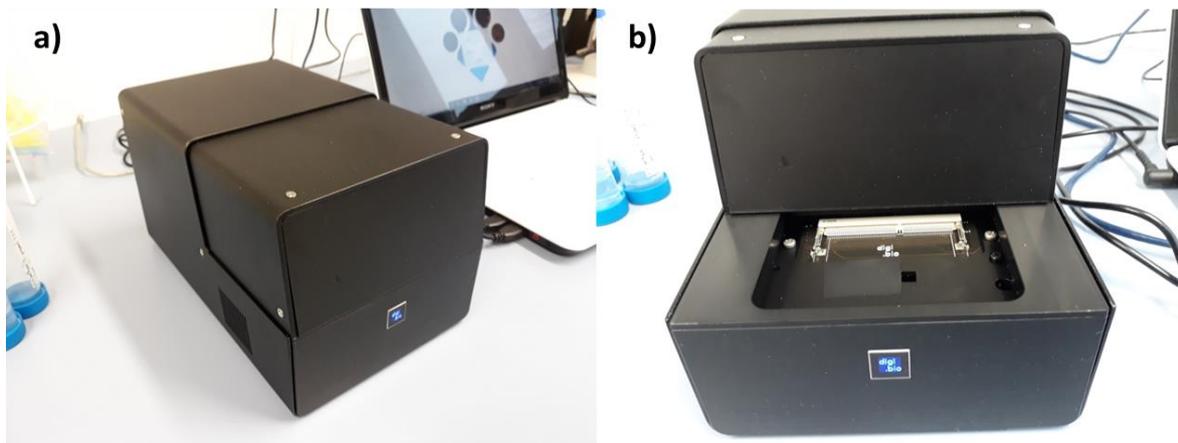


Figure 3-1: Digi.Bio Desktop Biofoundry device with closed (a) and opened cover (b), the latter image also showing the chip connection interface.

The DMF chips (Figure 3-2) used throughout this subproject comprised an array of 127 electrodes (107 regular, 20 reservoir electrodes). Regular electrodes were square-shaped, with a diameter of ~ 2.0 to 2.2 mm. Reservoir electrodes were arranged in pairs with a total length of ~ 6 mm and width of ~ 6.5 mm. The electrodes were powered via independent connections, enabling the user to selectively actuate single electrodes. To limit droplet pinning and biofouling, a hydrophobic film covering the dielectric constituted the surface of the chip interior where sample handling would occur.

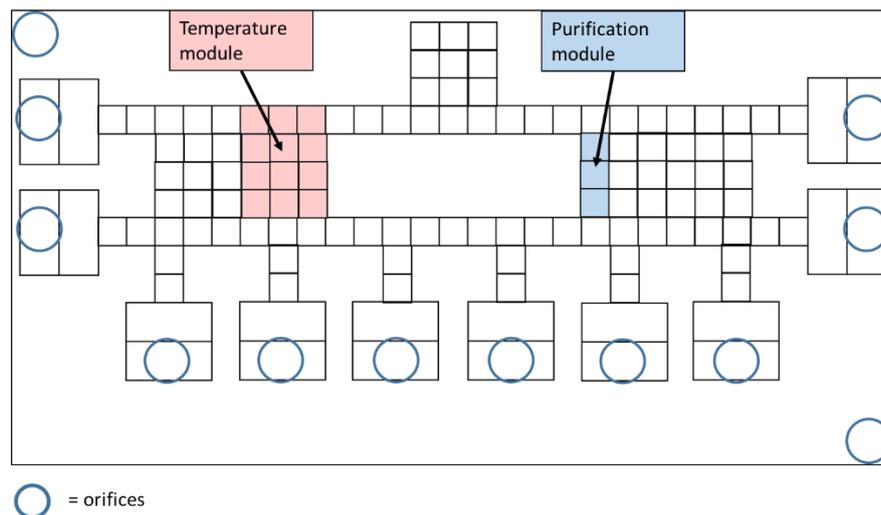


Figure 3-2: Schematic of Digi.Bio's DMF chip design at the time of the EAP, highlighting the position of the entry orifices, the temperature- and the magnet zones.

Above the surface, a plastic top plate was positioned, creating a gap between the two layers (exact plate height not disclosed). Single regular electrodes could accommodate about 1 μL , single reservoir electrodes up to 5 μL of liquid (adding up to 10 μL per loading zone). The gap between top plate and dielectric was only partially sealed, with the sides left open by default to facilitate interior cleaning by acting as additional liquid exits. Once connected, the chips could be powered with voltages in a range from 150 to 300 V.

3.3.3. Jurkat T cell culture

Jurkat T cells (ATCC, Clone E6-1, ATCC TIB-152) were obtained from LGC Standards GmbH (Wesel, Germany). Cells were grown in liquid culture at 37 °C and 5% atmospheric CO₂ using sterile non-coated T25 bottles (Th. Geyer GmbH & Co. KG, Renningen, Germany). Nutrition was provided by Roswell Park Memorial Institute (RPMI) 1640 culture medium supplemented with 2 mmol/L glutamine, 25 mmol/L 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer, 0.013 mmol/L phenol red, 10% v/v fetal calf serum (FCS) and 1% v/v penicillin/streptomycin mix (Pen/Strep). Depending on the state of the culture, passaging of cells was performed in regular intervals of 48-72 h. The culture bottle was first inspected visually for signs of contamination (e.g. fungal infestation). Upon confirming absence of these, the culture was mixed and transferred to a sterile 15 mL Falcon centrifuge tube (Greiner Bio-One GmbH, Frickenhausen, Germany). A sample of the culture was then diluted 1:5 in sterile Trypan Blue to stain any dead cells present in the medium, before counting live cells with the aid of a Neubauer

haemocytometer (Paul Marienfeld GmbH, Lauda-Königshofen, Germany), thus establishing the concentration. At a concentration of 7×10^5 cells/mL or above, the cells would then be passaged to create a new subculture with a concentration of 2.5×10^5 cells/mL in a total volume of 10 mL. For this purpose, the volume of cell culture containing the required number of cells was always filled up with fresh medium, before returning the new subculture to the incubator until the next passage. Cell culture maintenance up to this point was done by Britta Steer (AG Tholey).

After passaging, the remainder of the previous subculture was used as starting material for proteomics experiments immediately, or aliquotted and stored at $-80\text{ }^\circ\text{C}$ until further use. In any case, the cells were first centrifuged for 5 min at $4\text{ }^\circ\text{C}$ and $200 \times g$. After removing the supernatant, the cells were washed $3 \times$ with $1 \times$ PBS cooled to $4\text{ }^\circ\text{C}$, and centrifuged in the same manner every time. The PBS volume for washing was always matched to the original volume of the culture. Following the addition of PBS for the third washing step, the cells were counted again to estimate losses during washing.

After the final wash, the supernatant was again removed and the cells resuspended in an adjusted volume of $1 \times$ PBS to achieve the desired concentration. For long-term storage, the cells were resuspended to a concentration of $\sim 1 \times 10^6$ cells/mL, transferred to 2 mL protein LoBind tubes (Eppendorf SE, Hamburg, Germany) in 1 mL aliquots and centrifuged under the same conditions as during washing. After removing the supernatant, the cells were frozen as pellets of about 1×10^6 cells each.

Note: Dead cells stained by Trypan Blue were not counted on principle since they were not viable in a new subculture. While this counting method bore the potential of introducing error into the preparation of proteomics experiments (dead cells would still contribute protein content), the risk of this scenario was considered negligible, as any cultures containing a substantial proportion of dead cells were generally discontinued and replaced.

3.3.4. Preparation of Jurkat T cell pre-digest

A pellet of $\sim 1 \times 10^6$ cells was resuspended in 400 μL cell sample buffer ($1 \times$ PBS, 0.1% w/v Pluronic F-68) for a concentration of ~ 2500 cells/ μL . 150 μL of this suspension were mixed in a new vial with 150 μL lysis buffer (~ 5.35 mol/L urea, 5 mmol/L DTT, 1% w/v NP-40, 0.1% w/v Tetronic 90R4). The cells were then lysed for 15 min in a Bioruptor Pico sonication device (Diagenode SA, Seraing, Belgium) at $4\text{ }^\circ\text{C}$ (15 cycles,

30 s on, 30 s off). Afterwards, the lysis was continued at 37 °C for another 30 min without the aid of the BioRuptor. 150 µL of alkylation buffer (25 mmol/L IAA, 50 mmol/L borate buffer, 0.1% w/v Tetronic 90R4) were then added for a 45-minute alkylation at room temperature (approximately 20 °C) in the dark.

144 µL of cell lysate were transferred to a 2 mL LoBind tube (Eppendorf SE, Hamburg, Germany) containing 48 µL of SP3 bead suspension (20 µg/µL in 0.08% w/v Pluronic F-127). To the lysate on beads, 576 µL ACN were added, followed by gentle mixing and protein binding at room temperature for 10 min. The supernatant was discarded and the beads washed first with 1200 µL 70% ethanol, then 600 µL ACN. Eventually, 240 µL digest buffer (0.22 mmol/L SB3-14, 25 mmol/L NDSB-201, 25 mmol/L 3-(benzyltrimethylammonio)propanesulfonate [NDSB-256], 25 mmol/L borate buffer, 2 ng/µL trypsin/Lys. C mix) were added to the protein-laden bead pellets. The sample was incubated for 4 h at 37 °C, then split into 12 µL aliquots (approximately 500 cells/µL), which were stored at -80 °C until further processing.

3.3.5. Custom drilling procedure

To outfit the top plate of a DMF chip with new drillings, the plate was first carefully removed from the main body of the chip, refraining from the use of sharp-edged tools. The plate was then placed on a 96-well plate (coated side up) and underneath the drill. The designated drill site was aligned with one of the wells, to provide an empty volume to drill into and thus ensure that every part of the new drill channel had the same diameter. Upon completion of the process, plastic shavings were carefully scraped off to prevent them from coming loose on the inside of the chip. Finally, the top plate was reattached to the chip using the original dried glue as a mold.

3.3.6. Sample treatment to test custom drillings

Frozen Jurkat T cell pre-digest (see chapter 3.3.4) was thawed and incubated overnight (~16 h) at 37 °C to promote a complete digestion. The magnetic beads still present within the sample vials were immobilized using a rod magnet, the supernatant then aspirated and loaded into a customized DMF chip (filled with stained silicone oil) via an unmodified reservoir. The sample was split into equal portions (~3 µL) by targeted electrode actuation. Individual drops were then consecutively moved towards and aspirated through

the drilling, taking care to avoid oil transfer and keep the liquid coherent. The samples were transferred to new 0.5 mL LoBind tubes (Eppendorf SE, Hamburg, Germany), each containing 100 µg SP3 beads (pre-placed as 5 µL of a 20 µg/µL suspension in water, which was removed before the sample transfer). To each sample, 57 µL of 100% v/v ethanol were added, before mixing by brief, slow vortexing. The vials were then left for 10 min at room temperature and shaken at 800 rpm to induce peptide binding. Afterwards, the beads were immobilized and the supernatant removed, before washing three additional times with 57 µL of ethanol. To elute peptides, 5 µL of 2% DMSO were added to each sample before shaking at 1400 rpm for 1 min at room temperature, a step based on a procedure reported earlier (Sielaff et al., 2017). The beads were then immobilized and the supernatants transferred to 100 µL glass LC inserts (Th. Geyer GmbH & Co. KG, Renningen, Germany). After fully drying the samples down using a SpeedVac vacuum concentrator (Eppendorf SE, Hamburg, Germany), they were resuspended in 5 µL of loading buffer (3% ACN, 0.1% TFA).

3.3.7. On-chip cell lysis protocol

A chip was connected to the power source interface, filled with stained silicone oil and subjected to a voltage of 280-300 V. Lysis buffer (0.2% w/v DDOPM, 5 mmol/L DTT) was loaded onto an actuated reservoir electrode and split into volumes of approximately 1 µL, which were then dispensed to other reservoir electrodes meant to serve as starting points for the lysis. One µL of cell suspension (cells in 1 × PBS, 1 ppm DDOPM) was loaded onto an activated reservoir electrode, taking care to complete loading in one motion to avoid splitting the droplet. Once merged, mixing was performed by moving the drops over manually actuated electrodes. The chip was then disconnected from the device interface and transferred to a CTO-20AC column oven (Shimadzu Europa GmbH, Duisburg, Germany) for a 1 h-incubation at 50 °C. After reconnecting the chip, 1 µL alkylation buffer (50 mmol/L borate buffer, 1 ppm DDOPM, 25 mmol/L IAA) was added to each sample. The chip was left connected to the interface during alkylation at room temperature for 1 h, with the front cover closed to protect the samples from light. Following alkylation, 1 µL of digest buffer (8 ng/µL trypsin/Lys. C mix, 25 mmol/L borate buffer, 1 ppm DDOPM) was added. The samples were mixed and the chip transferred to a column oven for overnight (~16 h) incubation at 37 °C.

The samples were retrieved using 10 μL micropipettes and transferred to 0.5 mL LoBind tubes (Eppendorf SE, Hamburg, Germany) containing 125 μg of SP3 beads (pre-placed as 5 μL of a 25 $\mu\text{g}/\mu\text{L}$ suspension in water, which was removed before the sample transfer). Peptide binding was induced by adding 96 μL of ethanol, then mixing by slow vortexing and incubating for 8 min at room temperature while shaking at 800 rpm. The beads were immobilized and the supernatant discarded, before washing the beads three times with the same volume of ethanol. Five μL of 2% DMSO were then added to the beads, which were vortexed and shaken for 1 min at room temperature and 1400 rpm. The beads were immobilized and the eluates transferred to LC glass inserts (Th. Geyer GmbH & Co. KG, Renningen, Germany) containing 1 μL 5% FA.

3.3.8. LC-MS/MS method

Samples were injected onto an Ultimate 3000 RSLCnano UHPLC System (Thermo Fisher Scientific GmbH, Bremen, Germany). For desalting, a PepMap C₁₈ trapping column (5 mm \times 0.3 mm, 5 μm , 100 \AA , Thermo Fisher Scientific GmbH, Bremen, Germany) was installed, followed by an Acclaim PepMap C₁₈ analytical column (50 cm \times 75 μm , 2 μm , 100 \AA , Thermo Fisher Scientific GmbH, Bremen, Germany), the latter being kept at a constant 45 $^{\circ}\text{C}$. Subsequent MS/MS measurements were carried out on a Q Exactive Plus Orbitrap mass spectrometer. To enable ESI, columns were equipped with PicoTip™ emitters, SilicaTip™ with P200P coating (New Objective, Littleton, MA, USA).

Depending on the nature of the starting material and the sample handling, different injection volumes were applied. In those cases where only pre-made cell lysate was processed on-chip, the volume was set to 2 μL . In contrast, samples prepared exclusively on-chip were injected entirely, and the injection volume was therefore set to 7.5 μL .

Eluents used throughout the method were 0.05% FA (eluent A) and 80% ACN with 0.04% FA (eluent B). The flow rate was set to a constant 300 nL/min, and the starting composition at the time of injection was 5% eluent B. Desalting was performed for 5 min at this eluent composition, followed by a 60-minute linear gradient from 5 to 50% eluent B. For washing purposes, another gradient to 95% B was run over 5 min, and this composition was maintained for a further 10 min. After washing, the column was equilibrated for 15 min. MS was performed in positive polarity at a resolution of 70 000, an AGC target of 3×10^6 , a maximum injection time of 80 ms and within a scan range from m/z 300 to 2000. Data dependent MS₂ of the top 15 most abundant precursors was done

at a resolution of 17 500, a normalized collision energy of 27, an AGC target of 1×10^5 , a maximum injection time of 120 ms, a loop count of 15, an isolation window of 3 m/z , and within a scan range from m/z 200 to 2000.

3.3.9. Data evaluation

Raw data was searched using the Proteome Discoverer software (v. 2.2.0.388) and the Sequest HT algorithm (Enzyme: Trypsin [full], Precursor Mass Tolerance: 10 ppm, Fragment Mass Tolerance: 0.6 Da, Fixed modification: Carbamidomethyl, Dynamic modification: Oxidation [O]).

Searches were run against FASTA databases of human and common laboratory contaminant proteins (common Repository of Adventitious Proteins [cRAP]), downloaded from Uniprot on April 7th 2017. Qualitative values compared were the numbers of protein identifications (i.e. inferred proteins), protein and peptide groups, and peptide-spectrum matches (PSMs).

3.4. Results & discussion

3.4.1. General features of the device and performance tests

Fluidics tests were performed with aqueous dye solutions (in 0.5 or 2 mol/L of NaCl) supplied by Digi.Bio, as well as colorless surfactant solutions previously used on an air-based DMF system (DropBot DB3-120, Sci-Bots Inc., ON, Canada) and known to exhibit sufficient mobility there (Leipert et al., 2021; Leipert & Tholey, 2019). Chips were initially tested both with and without a silicone oil filling. All solutions thus tested were found to be highly mobile on oil-filled DMF chips and could be easily manipulated in whichever way necessary, and surfactant solutions remarkably even more so than on the DropBot under comparable conditions.

The manipulation of colorless drops on the chips proved to be a challenge due to the silicone oil itself being colorless, making visual recognition of drops difficult. This was considered problematic especially in the context of removing samples without accidentally carrying over substantial quantities of oil. To alleviate this issue, the attempt was made to stain the oil with the azo dye 1- $[[2\text{-methyl-4-}[(2\text{-methylphenyl})\text{diazonyl}]\text{phenyl}]\text{diazonyl}]$

naphthalen-2-ol (also known as Sudan IV) to the point of oversaturation (~100 µg/mL). Preliminary photometric tests with the dye had suggested negligible solubility in water, for which reason it could be assumed not to diffuse into aqueous sample-containing droplets. Staining the oil with Sudan IV provided an adequate contrast to both colored and colorless aqueous drops (Figure 3-3), substantially facilitating withdrawal of colorless samples while avoiding a transfer of excessive amounts of oil.

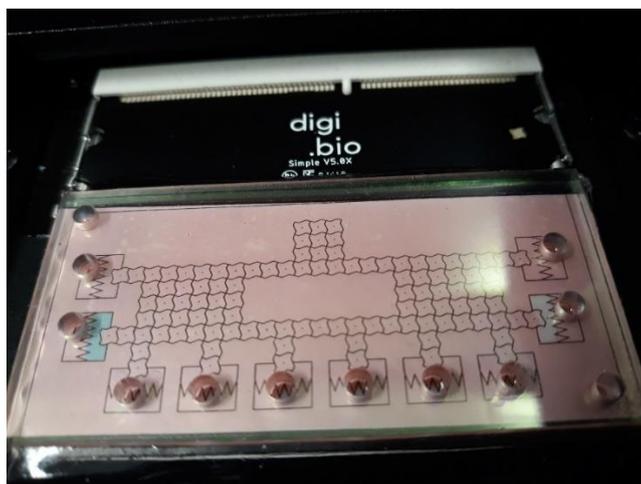


Figure 3-3: A DMF chip (Digi.Bio, Amsterdam, the Netherlands) filled with silicone oil saturated with the azo dye Sudan IV. For demonstration purposes, two reservoir electrodes are loaded with 0.1% w/v Tetronic 90R4, the latter in one case stained with 20 µg/mL erioglaucine (left).

Furthermore, aqueous solutions of common detergents already known to be suitable for on-chip proteomics (e.g. Tetronic 90R4) (Leipert & Tholey, 2019), as well as organic solvents at low concentrations (ACN, methanol) turned out to not readily take up dye from the surrounding oil. This especially benefitted future proteomics experiments by eliminating any need to stain the samples themselves, thus avoiding the risk of interactions between water-soluble dyes and proteins.

In order to qualify for use in sample solutions on oil-filled chips, organic solvents were also required to be completely immiscible with the silicone oil, as even poor miscibility would not only have promoted sample loss due to diffusion into the oil, but also contaminated the entire filling at the same time. This requirement was met by water, ACN, methanol and dimethylsulfoxide (DMSO), all of which play a role in biochemical applications. Furthermore, the same was true for solutions of the detergents Tetronic 90R4, Pluronic F-68, DDM (all at 0.1% w/v) and SB3-14 (0.22 mmol/L), the use of which in DMF-based proteomics on air-filled devices had already been reported earlier (Leipert et al., 2021; Leipert & Tholey, 2019). Conversely, off-chip testing of the alcohols ethanol and 2-propanol, as well as acetone and ethylethanoate (EA) showed these substances to be

highly miscible with the oil, making them likely choices to aid in cleaning the chips after use.

Aside from immiscibility with silicone oil, another requirement to any solvents meant for use on the chips was chemical compatibility with the dielectric and top plate. The latter in particular was found to be a limiting factor in terms of solvent compatibility. While the identity of the top plate material had not been disclosed, solvent damage observed during preliminary tests pointed to polymethylmethacrylate (PMMA) being the principal, if not exclusive, component. In an attempt to use ACN to clean the chip interior, notably immediately after an identical step with 100% ethanol, the underside of the top plate visibly reacted with the organic solvent. Combinations of lesser alcohols and ACN have been shown to cause dissolution of PMMA (González-Benito & Koenig, 2006; Groele et al., 1991), and remnants of ethanol would therefore have formed a mixture with ACN that PMMA could be assumed sensitive to.

As a consequence, the chip top plates were henceforth treated as being made from PMMA, and solvents for use on the chips were chosen with respect to their compatibility with this particular polymer, meaning that ACN, acetone, EA and DMSO were excluded from any working steps to take place within the chips. For interior cleaning of chips, the lesser alcohols ethanol and 2-propanol were selected due to both being miscible with silicone oil and compatible with PMMA when used individually (Groele et al., 1991).

3.4.2. Chip customization attempts

One of the recurring limitations encountered during early testing was the size and positioning of the orifices above the reservoir electrodes. These proved sufficient for the retrieval of sample volumes sufficient to cover the majority of the reservoirs (~5 μL); however, smaller volumes would not extend fully to the orifices, making a complete retrieval difficult. The thickness of the top plate also hindered sample removal by means of flexible GELoader tips (Eppendorf SE, Hamburg, Germany), which had previously been used successfully for this purpose on air-filled chips (Leipert & Tholey, 2019).

To address this problem, the top plate was outfitted with additional drillings further away from the reservoir locations, where they would be above the regular electrodes and volumes down to 1 μL could be accessed more easily. A diameter of 1.2 mm appeared to be ideal to allow access by 10 μL pipette tips (Eppendorf SE, Hamburg, Germany). Additionally, by removing the top plate for this purpose only to be reattached later, it was

possible to drill holes without the risk of directly damaging the electrodes, as well as rid the top plate of plastic shavings prior to reassembling the chip.

Despite the chip presumably remaining undamaged throughout this procedure, the drilling and reassembly process remained flawed. The most important concern was the risk of reassembling the chip with the wrong gap width, thereby potentially altering the liquid volume per electrode, and thus leading to errors whenever liquids were dispensed from a larger volume.

Interestingly, tests with 0.1% w/v Tetronic 90R4 showed an increased incidence of droplet pinning near the first custom drill site, a pattern to which could not be observed. While the first drill site was surrounded to such an extent by pinning zones that it became completely inaccessible, a second drill site on the same chip turned out to be completely unaffected (Figure 3-4).

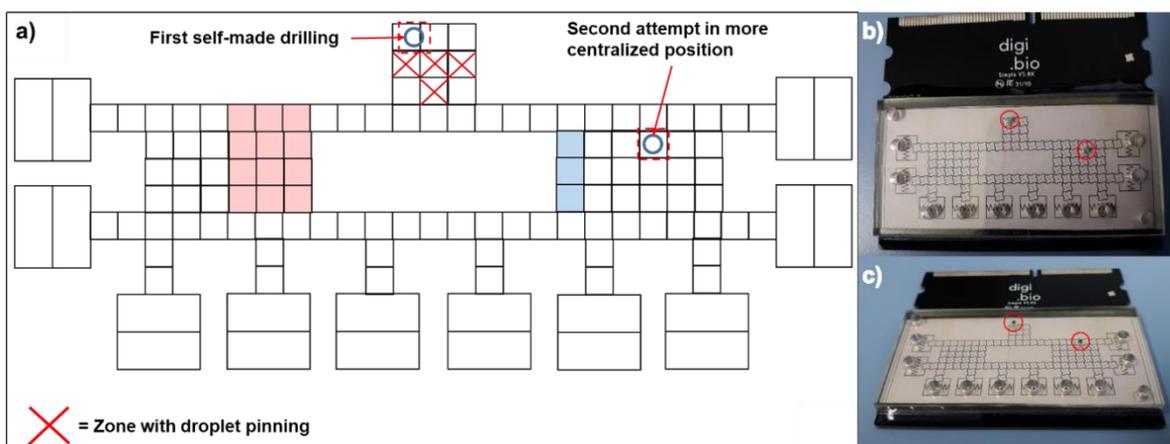


Figure 3-4: Schematic of the chip subjected to the custom drilling procedure (a), highlighting the positions of the new holes and sites where droplet pinning was observed. Note the complete absence of problematic zones around the second hole. For a comparison, the actual chip is shown from above (b) and at an angle (c).

While tests suggested custom drillings to be better suited for the withdrawal of sample volumes of approximately 1-3 μL than the regular reservoirs, one concern was the inevitable lack of a hydrophobic coating on the inside of the drill channels, and localized damage to the existing coating on the underside of the top plate. Both factors were expected to promote biofouling, and an aspect of particular interest therefore was the extent of cross-contamination in a scenario where multiple consecutive samples would be withdrawn from the same exit orifice.

In order to test the suitability of the custom drillings for nanoproteomics, samples of pre-digest from Jurkat T cells, containing protein/peptide equivalents of ~1000 cells, were therefore loaded onto the customized chip for processing (chapters 3.3.4 and 3.3.6), then removed consecutively via the same route and through the new drilling. Qualitative values obtained from four replicates suggested a high repeatability across samples; however, a drop of plain digest buffer, having been retrieved from the chip in the same manner, revealed a concerning degree of contamination (Table 3-1). Since the Jurkat pre-digest loaded onto the chip had contained various sulfobetaine detergents as part of the digest buffer, biofouling along the retrieval route was considered an unlikely cause of the contamination observed. Instead, the latter was hypothesized to be a result of cross-contamination at the custom drill site, which therefore also indicated incomplete sample withdrawal.

Table 3-1: Numbers of inferred proteins, protein- and peptide groups, and PSMs, obtained on the Desktop Biofoundry device from preparations of Jurkat pre-digest (equivalent of ~1000 cells) and plain digest buffer, all removed via the same custom drill hole.

Sample	Replicate	Inferred proteins	Protein groups	Peptide groups	PSMs
Jurkat digest	1	1562	1700	7376	8915
	2	1479	1617	6827	8220
	3	1433	1558	6440	7739
	4	1422	1548	7077	8573
Digest buffer	1	979	1104	3785	4438

Due to the lack of consistency and other expected risks associated with the process, custom drillings were subsequently abandoned and methods instead henceforth developed around the default chip design.

3.4.3. Full on-chip lysis

During earlier testing, the oil filling added to the chips proved to largely increase droplet mobility. Making use of this, presence of detergents in buffer solutions for full lysis protocols on the given platform could be limited to the LC-MS-compatible maltoside DDOPM. Furthermore, working concentrations of the latter were largely reduced, with solutions containing as little as 1 ppm DDOPM still displaying facile droplet movement.

Five experiments were performed, during which the sample preparation workflow was refined in an iterative process, with the final version described in chapter 3.3.7. Compared to the final version, the first lysis protocol was substantially different as far as the compositions of the lysis- (5 mmol/L DTT, 1 ppm DDOPM, ~5 mol/L urea), alkylation- (25 mmol/L IAA, 50 mmol/L borate buffer, 1 ppm DDOPM) and digest buffers (2 ng/μL trypsin/Lys C mix, 0.2% DDOPM, 25 mmol/L borate buffer) were concerned. In this version, urea was notably used to aid in the purely chemical lysis of cells early in the workflow. Since the protocol did not include an on-chip SP3 procedure prior to the addition of proteases, the latter were subsequently exposed to urea, albeit diluted down to ~1.25 mol/L. This was considered acceptable in the light of successful tryptic digestion in the presence of urea having been described previously, albeit at a lower concentration (Daly et al., 2018). Results from a first lysis experiment under these conditions, however, showed unexpectedly low numbers of inferred proteins, and especially peptide groups, suggesting changes to the protocol to be necessary. In response, later iterations of the protocol used a higher protease concentration in the digest buffer (4 to 8 ng/μL) while omitting urea in the lysis buffer.

A major problem observed during development of the final lysis workflow was leakage of silicone oil from the chips. While this could generally occur at any time during the procedure, they were notably observed most frequently during either the lysis or digestion step. It was hypothesized that the elevated temperatures used for either of these steps (70 °C, later 50 °C for the lysis, 37 °C for the digestion) made oil more prone to leaking from the chips by lowering the viscosity. Additional presumed risk factors were the act of transferring the chip in itself, but also the positioning within the column oven, where any unnoticed unevenness or incline would have increased the risk of leaks.

The final lysis protocol was designed for a maximum of four replicates per chip with respect to the number of electrodes and their spatial distribution on the chips. In order to prevent cross-contamination, samples were only loaded at the horizontally distal reservoir electrodes. Additionally, mixing zones were designated in a way to ensure samples would not pass over the same electrodes as others, and additional space between samples was allocated whenever incubation steps without electrode actuation were performed to decrease the risk of samples merging due to random drifting (Figure 3-5).

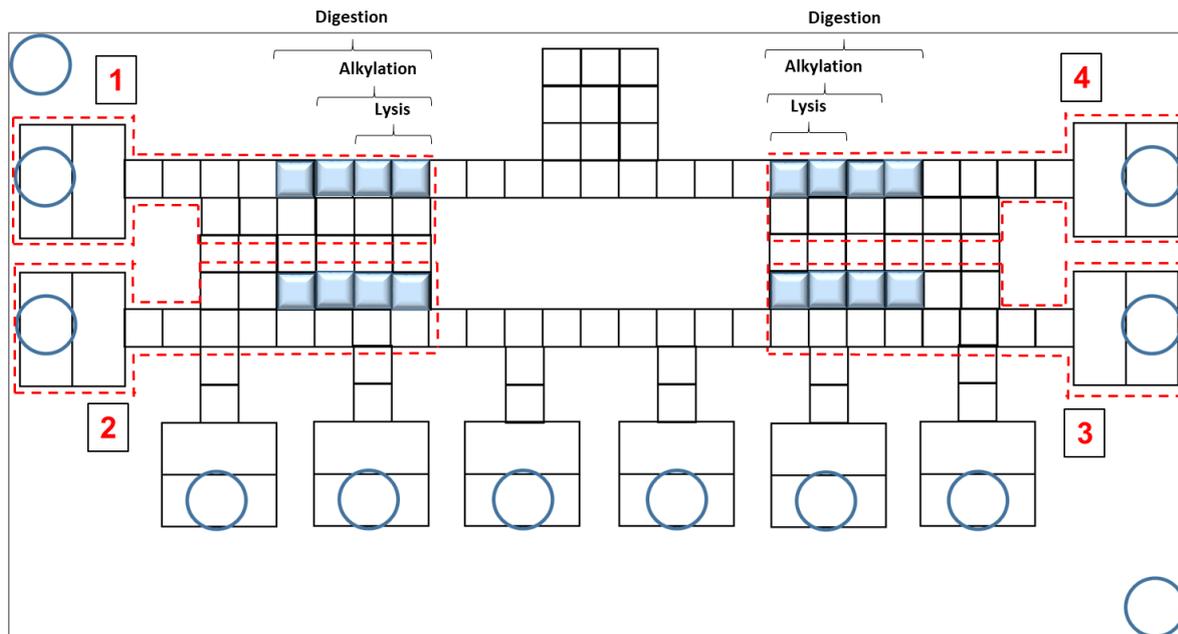


Figure 3-5: Chip schematic highlighting the preferred spatial distribution of samples during the on-chip lysis experiment series, in this case with four replicates. Marked in blue are the zones where samples were placed for incubation steps. Due to the consecutive addition of reagents, these would contain a minimum of 2 and a maximum of 4 μL , depending on the stage of the workflow. Zone numbers 1 through 4 correspond to replicate designations later used.

Using this electrode occupation scheme, the final lysis protocol (as described in chapter 3.3.7) was run once using ~ 100 Jurkat T cells as starting material. To assess the extent of sample carryover with this protocol, the sample preparation was repeated immediately after the previous digests had been retrieved, using the same chip, electrodes and oil filling, but substituting the cell suspension for plain sample buffer. The blank preparation was thus performed assuming the worst-case scenario of all possible contamination routes adding to the risk of carryover, as well as maximum contact time with previously used electrodes.

Evidence from the experiment strongly suggested cross-contamination to indeed be a major influence on results under the given conditions (Table 3-2), as had previously also been observed during the testing of custom drillings. After sharing movement routes and storage locations previously used for samples, sample buffer blanks fully injected at once would yield up to 208 proteins from 232 protein groups, inferred from 796 peptide groups.

Table 3-2: Numbers of inferred proteins, protein- and peptide groups, and PSMs, obtained in a full cell lysis experiment (Jurkat T cells, ~100/sample) and blank run (sample buffer only) on the Desktop Biofoundry device.

Sample	Replicate	Inferred proteins	Protein groups	Peptide groups	PSMs
Jurkat digest	1	539	606	2523	4579
	2	561	638	2195	3979
	3	790	880	3146	5798
	4	900	974	3713	5610
Buffer only	1	208	232	796	1489
	2	82	89	381	998
	3	122	148	744	2471

These findings hinted at the oil filling itself not being sufficient in preventing contamination of electrodes by biofouling entirely, making it necessary instead to replace the oil filling and/or clean used electrodes between experiments.

3.5. Conclusion

Overall, the Desktop Biofoundry and accessories were found to offer several features considered advantageous for nanoproteomics. Two aspects that stood out in particular were the high droplet mobility and negligible evaporation during extensive on-chip portions of protocols. Compared to air-based DMF, not only did this allow a much more sparing use of detergents, but also grant a greater flexibility in terms of reaction temperatures. In practice, this meant that proteins were less likely to be affected by components of reagent mixtures, as well as proteases being usable at their respective optimum temperatures with impunity.

Despite these advantages, a certain number of (at the time) insurmountable obstacles led to the decision to continue method development for DMF-based low-input proteomics on another device, the DropBot DB3-120 (Sci-Bots Inc., ON, Canada), thereby also abandoning oil-based DMF entirely. The decisive limiting factor was considered to be the necessity to execute off-chip peptide SP3 following sample retrieval, with a total of four solvent additions (one for SP3, followed by three washes) necessary to ensure sufficient removal of silicone oil. This added a substantial amount of manual pipetting and associated sample loss by surface adsorption, thereby somewhat defeating the purpose of adopting DMF for low-input proteomics in the first place. Conversely, the presence of

silicone oil, as well as limited solvent compatibility of the top plate, strongly narrowed down the amount of options for on-chip SP3, meaning that the choice of reagents in early working steps would have a high influence on downstream processes, especially digestions (see also chapter 4.4.2). Another obstacle was the positioning of the loading orifices above the reservoir electrodes in combination with the thickness of the top plate, which hampered the removal of small sample volumes and increased the likelihood of losses due to droplet splitting.

Although these combined aspects led to the decision of abandoning method development on this device in its current state, the final verdict was that the device could be considered highly promising, especially when allowing for the fact it was still under development at the time of testing. Future modifications to the design and/or materials based on the experience gained during the early access program may well see the device become versatile enough to allow the execution of low-input proteomics experiments as envisioned at the outset.

4. Digital microfluidic sample preparation and low-input Bottom-Up proteomics of the model organism *Caenorhabditis elegans*

4.1. Introduction

The nematode *Caenorhabditis elegans* (*C. elegans*; Figure 4-1) is an organism most commonly associated with rotting fruit and vegetables, where microorganisms constitute its primary food source, and from which they have shown to be isolated the most easily (Barrière & Félix, 2014; Félix & Braendle, 2010). *C. elegans* are ~0.25 mm long as newly hatched larvae, and grow up to ~1 mm long as adults. The nematodes have a fast life cycle: At an ambient temperature of 25 °C, they will reach adulthood three days after hatching. Additionally, the majority of *C. elegans* are self-fertilizing hermaphrodites, with males constituting only ~0.2% of the population (Corsi et al., 2015). First established by Sydney Brenner as a model organism in 1974, *C. elegans* has since become an essential research subject worldwide, being popular for its ease of handling, short generation cycle and widely understood genome (Brenner, 1974; Kirienko et al., 2014). Adult *C. elegans* are notably eutelic (Kimble & Hirsh, 1979; Sulston et al., 1983; Sulston & Horvitz, 1977), with adult hermaphrodites possessing 959 and males 1031 somatic cells. Additionally, adult nematodes possess ~1500 germ cells, thereby placing the estimated total adult cell count at ~2500 (Bensaddek et al., 2016). The protein content of adult *C. elegans* has been examined previously for a bulk amount via a Bradford assay (Bradford, 1976), based on which an amount of ~550 to 650 µg protein per 1000 nematodes was reported (Kohl et al., 2018). While the exact protein content per single nematode can therefore only be stated within a margin of error, the organism still constitutes a convenient source of material in low-input proteomics, since the sampling even of intact individuals yields a highly reproducible starting amount of protein without requiring an exact cell counting method (e.g. FACS).

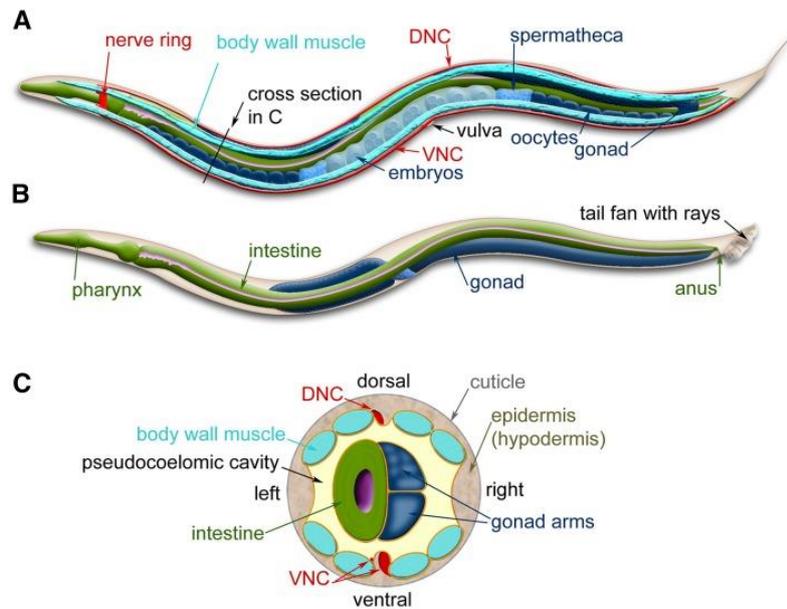


Figure 4-1: Lateral schematic views of the anatomy of adult *C. elegans* hermaphrodites (a) and males (b), and cross-section of the hermaphrodite body (c), highlighting the dorsal (DNC) and ventral nerve cords (VNC). Reproduced with permission from (Corsi et al., 2015). Copyright 2015 Corsi, Wightman, and Chalfie. Image originally modified from those found at WormAtlas.org (Altun et al., 2002).

In proteomics studies involving *C. elegans*, the nematode has traditionally been prepared in bulk amounts to generate sufficient starting material for separation by e.g. 2D gel electrophoresis prior to processing (Mádi et al., 2003). Conversely, the analysis of single nematodes is an interesting research subject, since various differences between individuals can be observed that would be masked in a bulk experiment. In a 2005 study, fluorescence measurements in single *C. elegans* originating from an isogenic population indicated a normal distribution in the abundance of the heat shock protein Hsp-16.2 based on the expression of the associated gene, thus suggesting an individually different response to heat stress (Rea et al., 2005). With proteins being directly involved in biological processes and interactions, comparable discoveries at the protein level can thus be expected to provide deeper understanding of how this model organism responds to environmental stimuli, and making miniaturized proteomics more accessible will most likely be a valuable aid to support this line of research.

4.2. Project aims & major experiments

Research efforts were devoted to the investigation of different sample preparation and measurement methods for BUP of single *C. elegans*. Although strongly emphasizing DMF applications, the subproject also aimed to provide solutions applicable with more common in-tube workflows.

4.2.1. Investigation of various lysis and digestion conditions in oil-based DMF

In the first part of this subproject, an oil-based DMF system (Digi.Bio Desktop Biofoundry device; see chapter 3.3.2) was tested for its potential to support a reproducible DMF-based preparation of single *C. elegans* specimens. For this purpose, objectives were to first establish conditions suitable for the lysis of live *C. elegans* in oil-based DMF, then optimize conditions for the on-chip digestion step.

4.2.2. Lysis of single *C. elegans* specimens for the comparison of DMF- and tube-based methods

Following method development for oil-based DMF, the air-based version was to be used in a new experimental series relying on single *C. elegans*, where two objectives were pursued. The first of these was to determine how different aspects of handling single *C. elegans* would change on the air- compared to the oil-based DMF system. The second objective was to directly compare sample preparations of single nematodes by air-based DMF and in-tube regarding qualitative results of subsequent LC-MS/MS measurements. To achieve this, the DMF lysis protocol originally developed for use on the oil-based Desktop Biofoundry platform was to serve as a roadmap to be modified for compatibility with the DropBot, before devising a comparable version of the new lysis protocol to resemble a typical in-tube preparation.

4.2.3. Label-free quantification to compare two biological conditions

In the next stage, the newly developed DMF- and in-tube preparation methods developed previously were to be compared in terms of suitability for label-free quantification. This experiment required *C. elegans* to be present in two different biological conditions at the outset, while keeping subsequent processing steps identical for both of these. Following

measurement of the samples, data evaluation would require addressing different points of consideration, such as normalization, filtering and appropriate statistical tests.

Parts of this study are the basis of a later publication (Steinbach et al., 2022).

4.3. Materials & methods

4.3.1. Chemicals & reagents

Acetonitrile (ACN; Chromasolv quality), FA and TFA (both LC-MS-grade), were obtained from Honeywell Specialty Chemicals Seelze GmbH (Seelze, Germany), ethanol (absolute) from Th. Geyer GmbH & Co. KG (Renningen, Germany).

Gibco PBS buffer (10 x) and Pierce borate buffer (20 x) were from Thermo Fisher Scientific GmbH (Bremen, Germany).

The detergents 3-dodecyloxypropyl-1- β -D-maltopyranoside (DDOPM) and NP-40 substitute were from Anatrace Inc. (Maumee, OH, USA) and G-Biosciences (St. Louis, MO, USA), respectively.

Trypsin/Lys-C mix was from Promega GmbH (Walldorf, Germany).

All other chemicals and reagents, including SeraMag Speed Beads (hydrophilic and –phobic, carboxylate-modified), were from Sigma Aldrich (Steinheim, Germany).

Hydrophilic and –phobic SP3 beads (both as 50 μ g/ μ L suspensions in 0.05% w/v azide) were always mixed to equal parts, washed twice with milli-Q-water, then resuspended at the desired concentration in water (for oil-based DMF on the Desktop Biofoundry) or 0.08% w/v Pluronic F-127 (for air-based DMF on the DropBot DB3-120).

4.3.2. *C. elegans* preparation and handling

All experiments described here depended on the availability of single organisms isolated beforehand from their original population. Throughout the subproject, multiple batches of *C. elegans* (all of the Bristol N2 strain) were prepared by Christine Blurton (AG Leippe) at the Zoological Institute of the *Christian-Albrechts-Universität zu Kiel*. These were cultured on nematode growth medium (NGM) plates populated with a lawn of *Escherichia coli*

(*E. coli*) OP50, a B-type *E. coli* mutant constituting a common nutrient source for *C. elegans* (Neve et al., 2020).

All populations of *C. elegans* used throughout the subproject were synchronized by first dissolving gravid hermaphrodites using 12% w/v NaClO and 5 mol/L NaOH, leaving only eggs behind. These were then allowed to hatch overnight in M9 buffer at 20 °C. L1 larvae were transferred to OP50-seeded NGM plates and incubated at 20 °C until reaching the L4 or adult stage, depending on the experiment to follow.

For the investigation of lysis conditions on the Digi.Bio Desktop Biofoundry platform, single specimens of *C. elegans* (N2 adults) were pre-placed in 0.2 mL PCR tubes (Eppendorf SE, Hamburg, Germany) surrounded by 1 µL of M9 buffer, then flash-frozen by briefly dipping the bottoms of the closed vials into liquid nitrogen. The specimens were transported on ice for about 45 min to the proteomics laboratory, then stored at –80 °C until further use.

For the comparison of an adapted DMF protocol on the DropBot DB3-120 device and a tube-based equivalent, individual live specimens of *C. elegans* (N2 adults) were placed in 0.2 mL PCR tubes (Eppendorf SE, Hamburg, Germany) surrounded by 1 µL of M9 buffer, then transported at approximately 15 °C. In this state, they were either processed immediately or flash-frozen in ethanol cooled to –80 °C for long-term storage at the same temperature.

For label-free quantification, a synchronized population of *C. elegans* was grown under default conditions until the L4 stage of development; then, two subpopulations were created. One of these was incubated for about 1 h at 33 °C, based on a recently published method (Golden et al., 2020), whereas the other was kept at default conditions. Single nematodes were then isolated in 0.2 mL PCR tubes (Eppendorf SE, Hamburg, Germany), immersed in 1 µL of M9 buffer, and flash-frozen in liquid nitrogen (Figure 4-2). To improve stability of the sample material, the single specimens were transported on dry ice and stored at –80 °C.

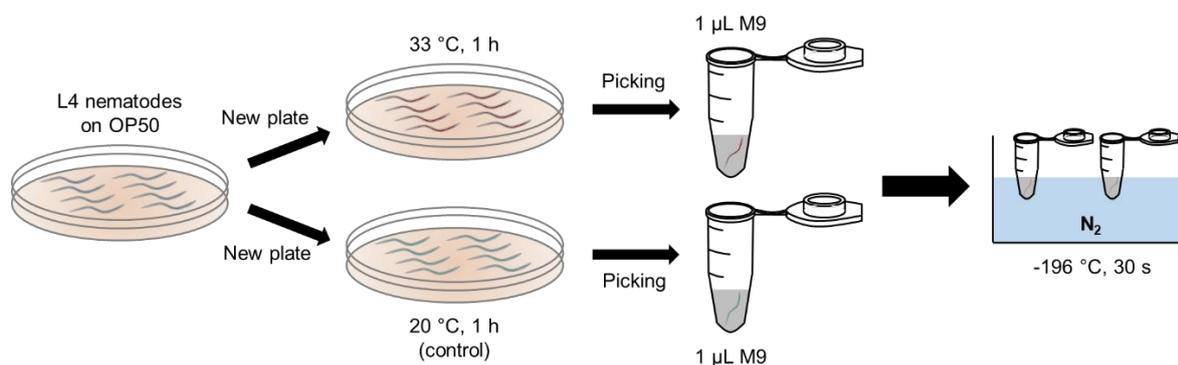


Figure 4-2: Schematic of the *C. elegans* treatment prior to the LFQ experiment. Reprinted with permission from (Steinbach et al., 2022). Copyright 2022 American Chemical Society.

4.3.3. Instrumentation

Experiments contributing to the subproject described in chapter 4.2.1 heavily relied on sample preparation using the Digi.Bio Desktop Biofoundry device that had already comprised the main body of the research described in chapter 3.

LC-MS/MS measurements were performed on a Q Exactive Plus and an Orbitrap Fusion Lumos (Thermo Scientific, Bremen, Germany), coupled to an Ultimate 3000 RSLCnano UHPLC System (Thermo Fisher Scientific GmbH, Bremen, Germany) to achieve chromatographic separation. An Acclaim PepMap column (50 cm × 75 µm, 2 µm, 100 Å; Thermo Fisher Scientific GmbH, Bremen, Germany) was used for all measurements on the Q Exactive Plus, and an Aurora column (25 cm × 75 µm, 1.6 µm, 100 Å; IonOpticks, Fitzroy, Victoria, Australia) for those on the Fusion Lumos.

MALDI-ToF MS measurements were performed on a 5800 TOF/TOF mass spectrometer (AB Sciex) using the 4000 Series Explorer software (AB Sciex).

4.3.4. DropBot DB3-120 digital microfluidics device

The majority of work presented in this chapter was carried out on the DropBot DB3-120 DMF device (Kitchener, ON, Canada) equipped with 2" × 3" 90-pin (v4) chips. The system was controlled via the MicroDrop v2.19 software.

Additionally, a number of working steps performed on the DropBot relied on the aid of a magnetic lens, an approach described earlier (Choi et al., 2013; Leipert et al., 2021). To

implement this concept, 90-pin chips were secured on a frame built from 10 mm diameter MakerBeam profiles (chartup.com System Engineering GmbH, Bergkirchen, Germany) and connected to the DropBot itself by a two-way interface (Figure 4-3). The magnetic lens was positioned below the chip on a platform secured on opposite ends by a pair of rails, which was equipped with magnetic discs on each corner, whereas the chip platform above was outfitted with oppositely aligned magnetic discs. This setup enabled the magnetic lens to snap into place whenever it was moved upwards for maximum effect on the chip; subsequent disengagement of the lens was possible by manually pulling the chip platform downwards with enough strength to overcome the magnetic force. This magnetic lens setup had been devised, constructed and used earlier by Jan Leipert (AG Tholey).

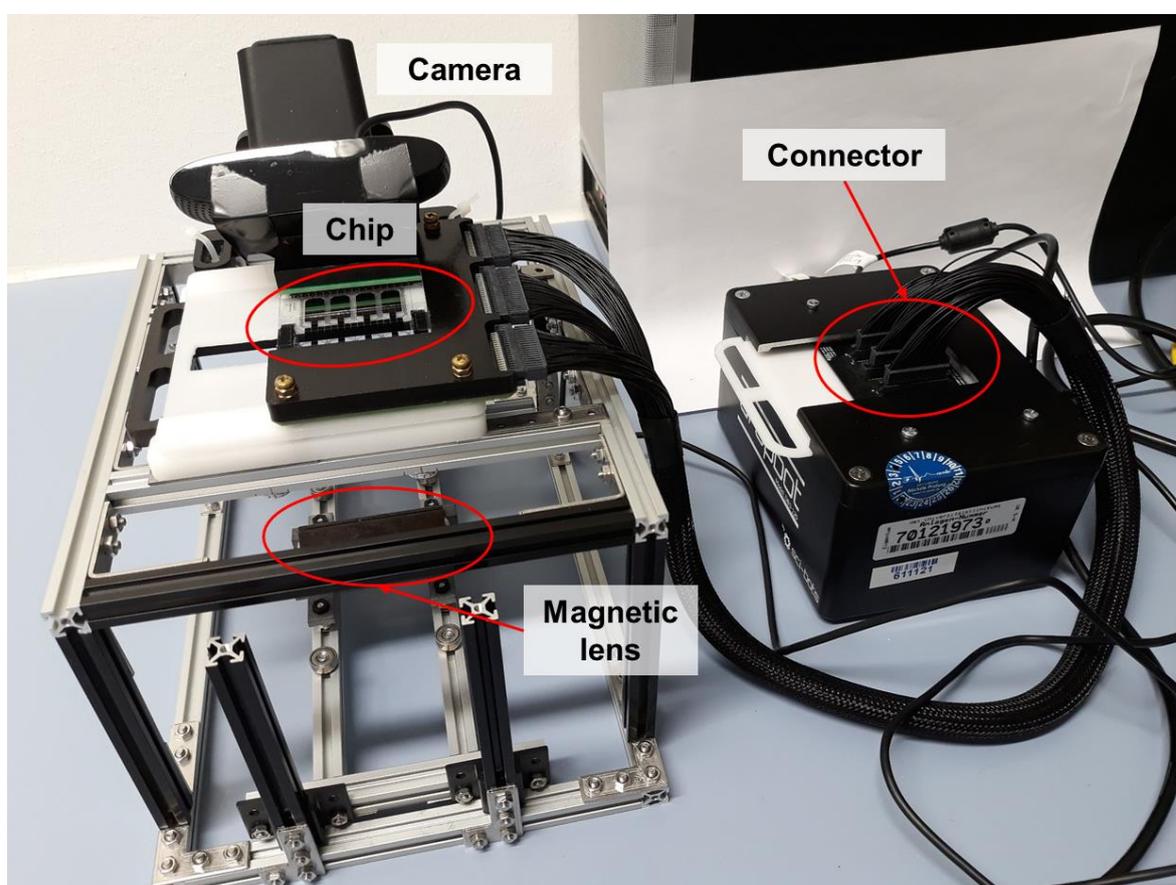


Figure 4-3: Magnetic lens setup for the DropBot DB-3-120 (designed, constructed and used previously by Jan Leipert). Instead of being placed within the DropBot itself, chips would be connected via a two-way interface, allowing more customization options for the lens platform. Adapted with permission from (Leipert et al., 2023). Copyright 2023 The Authors.

New chips were used for each experiment and tested for functionality before the experiment was performed. For testing, different solutions (1% DDOPM, then 30% ACN and finally 100% ACN) were loaded onto a chip to occupy all reservoirs on one side of the

chip. The liquid was then gradually moved through the chip, making sure to actuate all electrodes at least once. The liquid was then aspirated from the opposing row of reservoirs, where in turn the next solution was loaded to repeat the procedure. Through this procedure, functionality of electrodes for solutions of different polarity could be confirmed, while simultaneously cleaning the chip on the inside.

4.3.5. Assessment of SDS removal following oil-based DMF

2.5 μL of BSA solution (2 mg/mL) were mixed with 5 μL DTT (20 mmol/L in 100 mmol/L AmBiC) and 10 μL SDS-d25 solution. Four spiking concentrations of SDS-d25 (1, 0.5, 0.1, 0.01% w/v) and a blank (0% w/v) were tested, resulting in actual starting concentrations of 0.4, 0.2, 0.04, 0.004 and 0% w/v. Samples were then incubated at 56 $^{\circ}\text{C}$ for 30 min. 2.5 μL IAA (100 mmol/L) were then added, followed by alkylation in the dark at room temperature (approximately 20 $^{\circ}\text{C}$) for 30 min. Five μL digest buffer (20 ng/ μL trypsin/Lys. C mix, 25 mmol/L borate buffer) were then added and the samples incubated overnight (~16 h) at 37 $^{\circ}\text{C}$.

The digest was transferred completely to 0.5 mL vials containing 125 μg of SP3 beads (pre-placed as 5 μL of a 25 $\mu\text{g}/\mu\text{L}$ suspension in water, with the latter removed afterwards). 60 μL ethanol were added, the samples slowly vortexed, and then shaken at room temperature and 800 rpm for 8 min. Three more washes with 60 μL of ethanol were performed, with all supernatants saved in separate new 0.5 mL vials. For peptide elution, 5 μL of 2% DMSO were added to the beads before shaking at room temperature at 1400 rpm for 1 min. Eluates were transferred to new 0.5 mL vials. All wash supernatants and eluates were dried in a SpeedVac vacuum concentrator (Eppendorf SE, Hamburg, Germany) (settings: V-AQ, 45 $^{\circ}\text{C}$). Afterwards, the dried samples were resuspended in 10 μL HCCA matrix solution (3 mg/mL in 70% v/v ACN with 0.1% v/v TFA) and spotted as 1 μL drops on a 384-well Opti-TOF MALDI insert (AB Sciex), then allowed to dry for approximately 15 min at room temperature.

MALDI-ToF MS measurements were performed in positive polarity, with each sample/matrix co-crystal exposed to 400 laser pulses with an intensity of 65%. Spectra were acquired in a mass range from m/z 100 to 1200.

4.3.6. Optimization of lysis conditions

Single live *C. elegans* specimens (N2 adults) were transferred to a 60-well HLA Terasaki plate, where they were placed in individual wells and immersed in 1 μ L of M9 buffer. The following lysis agents were then immediately added according to Table 4-1:

Table 4-1: Lysis agents and respective volumes tested on single live *C. elegans*. Concentrations stated here are the original concentrations (before adding reagents to the nematodes in M9 buffer).

Condition	Composition	Volume in μ L
1	6 mg/mL DTT, 0.1% w/v SDS	2
2	6 mg/mL DTT, 0.2% w/v DDOPM	2
3	6 mg/mL DTT, ~8 mol/L urea, 1% w/v NP-40	2
4	ACN, 100%	4
5	DMSO, 100%	4
6	1 \times PBS (control)	4

The process was documented by photographs, which were taken immediately after exposure, then in 5-minute intervals for a total of 25 min.

4.3.7. On-chip lysis protocol: Desktop Biofoundry

To single *C. elegans* in 1 μ L of M9 buffer, 60 μ L of 0.1% w/v DDOPM were added. Nematodes were aspirated in a 5 μ L volume using a 10 μ L Eppendorf micropipette. The pipette was then held vertically, allowing the nematodes to sink downwards. Injection into the chip reservoir followed when the organisms were just above the tip mouth. As a control measure, liquid in the original vessels was repeatedly pumped using the same tip to confirm absence of the specimens (and presence on the chip). The sample drops were immediately moved inwards by electrode actuation and further moved around until the single organisms could be spotted within the liquid, at which point these were isolated in portions of approximately 1 μ L, whereas the remainder of the liquid was removed.

Eight μ L of lysis buffer (by default 6 mg/mL DTT in 0.1% w/v SDS) were then loaded in an isolated location. Portions of approximately 1 μ L were split off and moved within close proximity of the sample drops, before removing the remainder of the lysis buffer through the original opening. The samples and lysis buffer drops were then merged and mixed in quick succession; incubation was for 45 min at room temperature while maintaining a voltage of 200 V to keep the drops in place.

Eight μL of alkylation buffer (by default 25 mmol/L IAA in 50 mmol/L borate buffer and 1 ppm DDOPM) were then added, dispensed and merged in the same manner. The alkylation was carried out under near identical conditions as the lysis, but for 1 h and in the dark.

The digest buffer (by default 8 ng/ μL trypsin/Lys. C mix, 25 mmol/L borate buffer and 1 ppm DDOPM) was added to the samples in the same way, followed by an overnight (~16 h) incubation at room temperature and with the drops held in place by a voltage of 200 V.

The samples were retrieved by means of a micropipette and transferred to 0.5 mL vials containing 125 μg of SP3 beads (placed as an aqueous 25 $\mu\text{g}/\mu\text{L}$ suspension with the water removed afterwards). 96 μL of ethanol were added, the vials slowly vortexed and shaken at room temperature for 10 min at 800 rpm to induce binding of peptides. The beads were then immobilized and the supernatant discarded. This was followed by three more washes with the identical volume of ethanol. Afterwards, 5 μL of 2% DMSO were added to the beads, which were then vortexed at high speed and shaken for 1 min at room temperature and 1400 rpm. The beads were immobilized and the supernatants transferred to glass inserts (Th. Geyer GmbH & Co. KG, Renningen, Germany) containing 1 μL of 5% FA. For measurement via nanoLC-MS/MS, the samples were injected entirely as approximately 7.5 μL of liquid in total.

This workflow was applied to a series of five individual experiments, where alterations to individual conditions were made after each iteration:

- Experiment 1 was performed exactly according to the default workflow described in this chapter (referred to as “0.1% SDS” in the discussion).
- Experiment 2: The lysis buffer was replaced with a variant with 6 mg/mL DTT in ~8 mol/L urea and 1% w/v NP-40 (referred to as “NP-40, RT” in the discussion). All other conditions were kept identical.
- Experiment 3: Building upon experiment 2, changes to the digest parameters were made. Instead of a digest at room temperature (approximately 20 °C) with the samples subjected to a constant voltage, a digest at 37 °C was performed in the column oven (referred to as “NP-40, 37 °C” in the discussion).
- Experiment 4: Compared to experiment 3, the protease concentration was doubled, but no other alterations were made (referred to as “NP-40, 2x trypsin” in the discussion).

- Experiment 5: In an alteration to experiment 4, the protease concentration was reverted to the original value. Instead, the urea concentration in the lysis buffer was halved to approximately 4 mol/L (referred to as “NP-40, 4 M urea” in the discussion).

4.3.8. On-chip lysis protocol: DropBot DB3-120

At the start of the experiment, the reservoir and transfer electrodes to be used were turned on subject to a voltage of 95 V at a frequency of 1 kHz. 1 μ L of lysis buffer (~5 mol/L urea, 6 mg/mL DTT, 0.1% w/v Tetric 90R4, 1% w/v NP-40) per sample was then pre-placed and moved onto the respective transfer electrode, which was kept actuated while also reactivating the reservoir electrode (see also Figure 4-4).

50 μ L of Pluronic F-127 solution (0.08% w/v or 0.8 mg/mL) were added to the sample vessel. A 2.5 μ L pipette was then used to aspirate the nematode in a 1 μ L volume. This volume containing the nematode was then added to a droplet of lysis buffer already on the chip from the previous step. The merged sample and lysis buffer were then mixed by moving the drop over consecutively actuated electrodes beyond the transfer zone. Following this, lysis and protein reduction was performed for 50 min at room temperature.

One μ L of alkylation buffer (25 mmol/L IAA, 50 mmol/L borate buffer, 0.1% w/v Tetric 90R4) was added to each sample and mixed via the same method. The chip was removed from the device and kept at room temperature in the dark for a 50-minute alkylation step.

Following alkylation, the chip was reconnected to the DMF device. For SP3, the samples were first moved further inwards (away from the transfer electrodes). Two μ L of SP3 bead suspension per sample were then loaded onto the respective transfer electrodes. A magnetic lens was then used to pellet and immobilize the beads before removing the supernatant via the reservoir electrodes.

The samples were moved onto the respective bead pellets before adding 4.5 μ L of ACN to induce protein binding. To create a homogenous suspension, the magnetic lens was disengaged and the bead-containing drops moved via an automated route (Repeats: 10, Voltage: 95 V, Frequency: 1 kHz, Duration: 3000 ms, Trail length: 2). The chip was then placed on an elevated platform within a humidified chamber containing 10 mL of ACN to limit evaporation during the following 30-minute binding step at room temperature.

Following protein binding, the chip was replaced into the DMF device and the beads were immobilized. The supernatants were removed and discarded before washing the beads with 4.5 μ L of 70% ethanol and ACN (in this order), where the same automated mixing route was employed. After washing the bead pellets, 2 μ L of digest buffer (4 ng/ μ L trypsin/Lys-C mix, 25 mmol/L borate buffer, 0.01% w/v DDOPM) were added to the bead pellets and briefly mixed. The chip was removed from the device again and wrapped in two layers of Parafilm[®] (Bemis Company, Inc., Neenah, WI, USA) to help prevent evaporation, with one layer applied only to the central region of the chip comprising the top plate, and the other around the entire chip. The latter was then placed in a vacuum sealing bag (Allpax GmbH & Co. KG, Papenburg, Germany), which was initially sealed at the top using a heat sealer. The bag was then punctured below the seal and remaining air withdrawn using a syringe, taking care not to accidentally merge samples by exerting excessive pressure on the chip. Following sufficient withdrawal of air, a second seal was applied below the puncture site. In this state, the chip was left overnight (~16 h) for the digestion step at room temperature.

The chip was unpacked carefully and reconnected to the DropBot. Prior to sample retrieval, the extent of evaporation was assessed and compared across the samples visually, based on the surface area of the drops in relation to the electrode diameter. In order to gain drops of approximately equal volumes, varying quantities of 30% ACN (or none at all) were added to each sample, effectively cancelling out differences in the extent of evaporation overnight. Samples were then retrieved via 20 μ L GELoader tips (Eppendorf SE, Hamburg, Germany) and transferred to 100 μ L glass LC inserts (Th. Geyer GmbH & Co. KG, Renningen, Germany) with 1 μ L of 5% FA. Using another tip, 1.5 μ L of 0.05% FA were then used to flush the original tip and thus ensure maximum retrieval efficiency. With a total volume of approximately 4.5 μ L, the samples were then measured directly. A graphical representation of the on-chip workflow is provided in Figure 4-4.

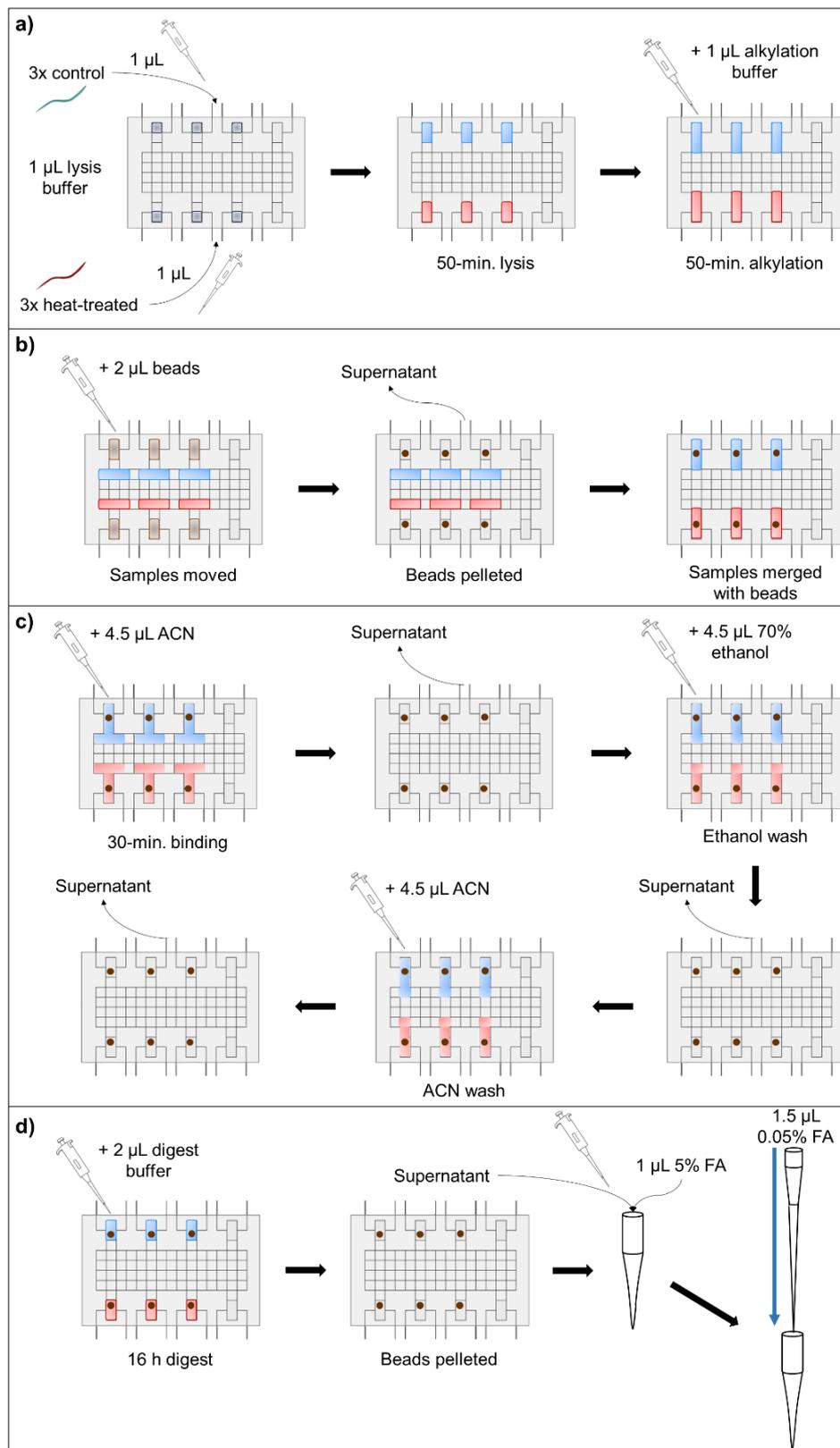


Figure 4-4: Graphical demonstration of the DMF sample preparation workflow. a) Loading of nematodes, lysis and alkylation. b) Addition of SP3 beads and merging with samples. c) SP3 sample treatment and washes. d) Digest and sample retrieval. Adapted with permission from (Steinbach et al., 2022). Copyright 2022 American Chemical Society.

4.3.9. Tube-adapted workflow

The DMF protocol established for the DropBot DB3-120 (chapter 4.3.8) was adapted to be used in a nearly identical fashion for sample preparation in standard 0.2 mL PCR tubes (Eppendorf SE, Hamburg, Germany; Figure 4-5). To make execution of the working steps reasonably consistent, all volumes, save those of the solvents for binding and washing during SP3, were trebled. Mixing of new reagents into samples was always achieved by repeated pipetting.

44 μL of ACN were added to the sample/bead suspension to initiate protein binding, creating a solvent/sample ratio used in the past for tube-based SP3 with low numbers of mammalian cells (Leipert et al., 2021). The samples were then left for 30 min at room temperature. Subsequent washing steps were performed with 100 μL 70% ethanol and 50 μL ACN, respectively.

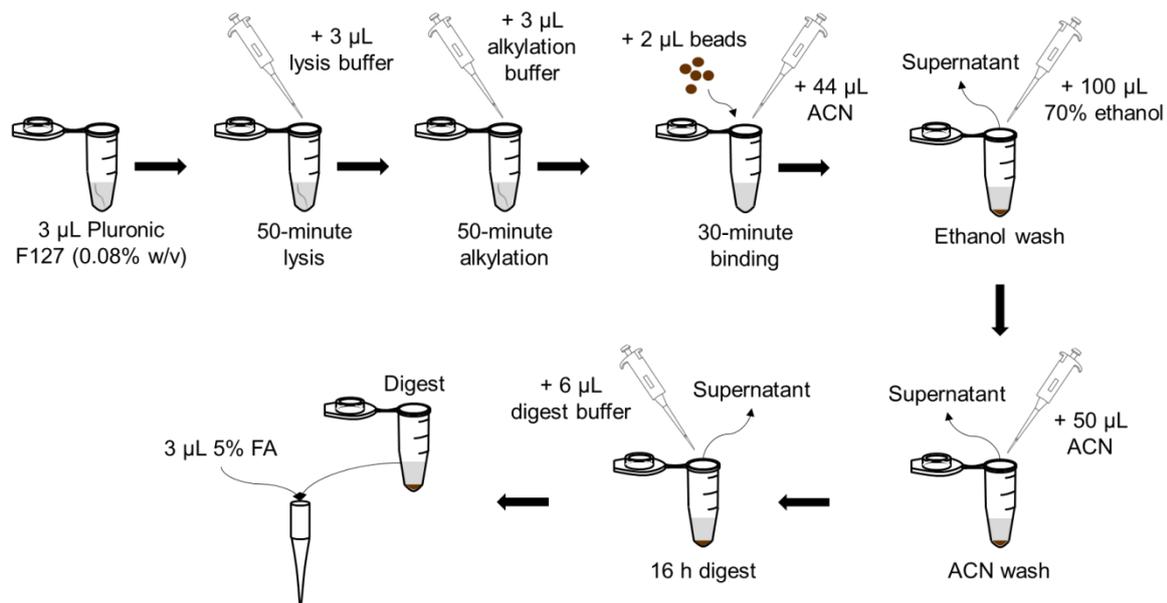


Figure 4-5: Schematic of the in-tube BUP sample preparation workflow for single *C. elegans*. Reprinted with permission from (Steinbach et al., 2022). Copyright 2022 American Chemical Society.

4.3.10. LC-MS/MS: Orbitrap Q Exactive Plus

Since this method was meant for the injection of samples processed on the Digi.Bio Desktop Biofoundry, the equipment setup was identical to that described in chapter 3.3.8.

Measurements were performed in a single-shot fashion, with injection volumes of 7.5-8 μL and all sample volume injected at once. The initial eluent composition was 97% eluent A

(0.05% FA) and 3% eluent B (80% ACN with 0.04% FA), and the flow rate was kept at a constant 300 $\mu\text{L}/\text{min}$. Desalting was performed for 5 min at this eluent composition, followed by a 60-minute linear gradient from 5 to 50% eluent B. For washing purposes, another gradient to 95% B was run over 5 min, and this composition was maintained isocratically for another 10 min. After washing, the column was equilibrated for 15 min.

MS was performed in positive polarity at a resolution of 70 000, an AGC target of 3×10^6 , a maximum injection time of 80 ms and within a scan range from m/z 300 to 2000. Data dependent MS2 of the top 15 most abundant precursors was done at a resolution of 17 500, a normalized collision energy of 27%, an AGC target of 1×10^5 , a maximum injection time of 120 ms, a loop count of 15, an isolation window of 3 m/z , and within a scan range from m/z 200 to 2000.

4.3.11. LC-MS/MS: Orbitrap Fusion Lumos

Samples prepared on the DropBot device or in-tube for direct comparison were exclusively measured on an Orbitrap Fusion Lumos tribrid mass spectrometer, to which a FAIMS Pro interface (Thermo Fisher Scientific GmbH, Bremen, Germany) was additionally equipped.

Identically to measurements on the Q Exactive Plus, samples were injected onto an Ultimate 3000 RSLCnano UHPLC System (Thermo Fisher Scientific GmbH, Bremen, Germany). A PepMap C₁₈ trapping column (5 mm \times 0.3 mm, 5 μm , 100 \AA , Thermo Fisher Scientific GmbH, Bremen, Germany) was used for desalting, followed by an Aurora-series analytical EIC (25 cm \times 75 μm , 1.6 μm , 100 \AA ; IonOpticks, Fitzroy, Victoria, Australia). To achieve a constant temperature of 45 $^\circ\text{C}$, the column was placed in a Butterfly Heater (Phoenix S&T, Chadds Ford, PA, USA).

Due to the different final volumes after DMF- and in-tube preparation, injection volumes were 5 and 9 μL , respectively. The initial mobile phase composition was 97% eluent A (0.05% FA) and 3% eluent B (80% ACN with 0.04% FA), and the flow rate was kept at a constant 200 $\mu\text{L}/\text{min}$. A 2-minute desalting step at this eluent composition was followed by a first 90-minute linear gradient from 3-35% eluent B and another 30-minute gradient from 35-50% B. After the separation, the eluent B content was increased to 90% over the course of 1 min, which was maintained for another 12 min. The system was then equilibrated for 15 min.

Measurements were in positive polarity within a scan range from m/z 375-1500, with the maximum injection time set to 50 ms, the resolution to 60 000 and the normalized AGC target to 100%. FAIMS was used with two CVs (-50 and -70 V). HCD fragmentation used a maximum injection time of 35 ms, an AGC target of 100%, and a normalized collision energy of 30%. After detection in the Orbitrap, precursor ions were fragmented over a cycle time of 0.75 s per FAIMS CV cycle. Product ions were then detected in the linear ion trap (OTIT mode) with a dynamic exclusion time of 60 s.

4.3.12. Data evaluation

Raw data was always loaded into Proteome Discoverer v. 2.2.0.388 and searched against the FASTA databases of *C. elegans*, *E. coli* (strain B/BL21-DE3), and the cRAP list of common laboratory contaminants, all of which were downloaded on February 17th 2017 (Cassidy et al., 2018). Any follow-up processing and evaluation of the data after the search was performed using Proteome Discoverer v. 2.4.0.305.

In the case of experiments involving oil-based DMF (see also chapter 4.2.1), raw data files were searched individually and compared based on inferred proteins, protein- and peptide groups, PSMs, and digestion efficiency (evident from the extent of missed cleavage events). Individual values for each category were averaged across the replicates of the respective experiment, then compared.

For LFQ evaluation, only inferred proteins were considered that originated from *C. elegans* (i.e. with descriptions containing “*elegans*”), and were represented by at least two peptides, of which at least one was unique. Furthermore, inferred proteins were required to have valid abundance ratios between the two conditions “Control” and “Heat”. Normalized abundance values were then exported to Microsoft Excel, where an additional median normalization was applied. Following this, the data was then imported to the Perseus analysis software v. 1.6.15.0 (Tyanova, Temu, Sinitcyn, et al., 2016). Here, the normalized abundances were Log_2 -transformed, and grouped as “Control” and “Heat”. This was followed by a two-sample t test (Student’s t test, FDR: 0.05, number of randomizations: 250, s_0 : 0.1). For visualization, the data was exported to Microsoft Excel after the t test, before plotting $-\text{Log}_{10}(\text{p-value})$ against the Log_2 -transformed fold change.

4.4. Results & discussion

4.4.1. Optimization of lysis options

Due to both the difficulty of performing on-chip protein SP3 on the Digi.Bio chips and the necessity of removing silicone oil efficiently following sample retrieval, the less used peptide SP3 involving multiple washing steps was essential as an alternative in any sample preparation on this platform. Peptide SP3 thus played a major role in designing suitable lysis protocols, since it represented the only way to remove not only oil, but also any reagents not compatible with LC-MS. Furthermore, the choice of solvent was limited to pure ethanol as miscibility with silicone oil was essential.

A first step in method development therefore was to determine how efficiently typical components of lysis reagents could be removed by washing steps in peptide SP3 if only using ethanol. SDS was chosen as an example for this preliminary experiment due to it being considered a likely candidate for the final lysis reagent mix at the time. In a mock BUP preparation of BSA as an analyte, varying mass concentrations of SDS-d25 were introduced as part of the lysis buffer, followed by alkylation, digestion and peptide SP3. Wash supernatants and eluates from the SP3 step were then analyzed via MALDI-ToF MS to detect any SDS-d25 remnants. Three technical replicates were prepared per SDS concentration, and four MALDI spots prepared for each replicate, resulting in 12 data points per concentration and stage. The S/N ratio of the SDS-d25 signal in each sample was then plotted against the respective spot number (Figure 4-6).

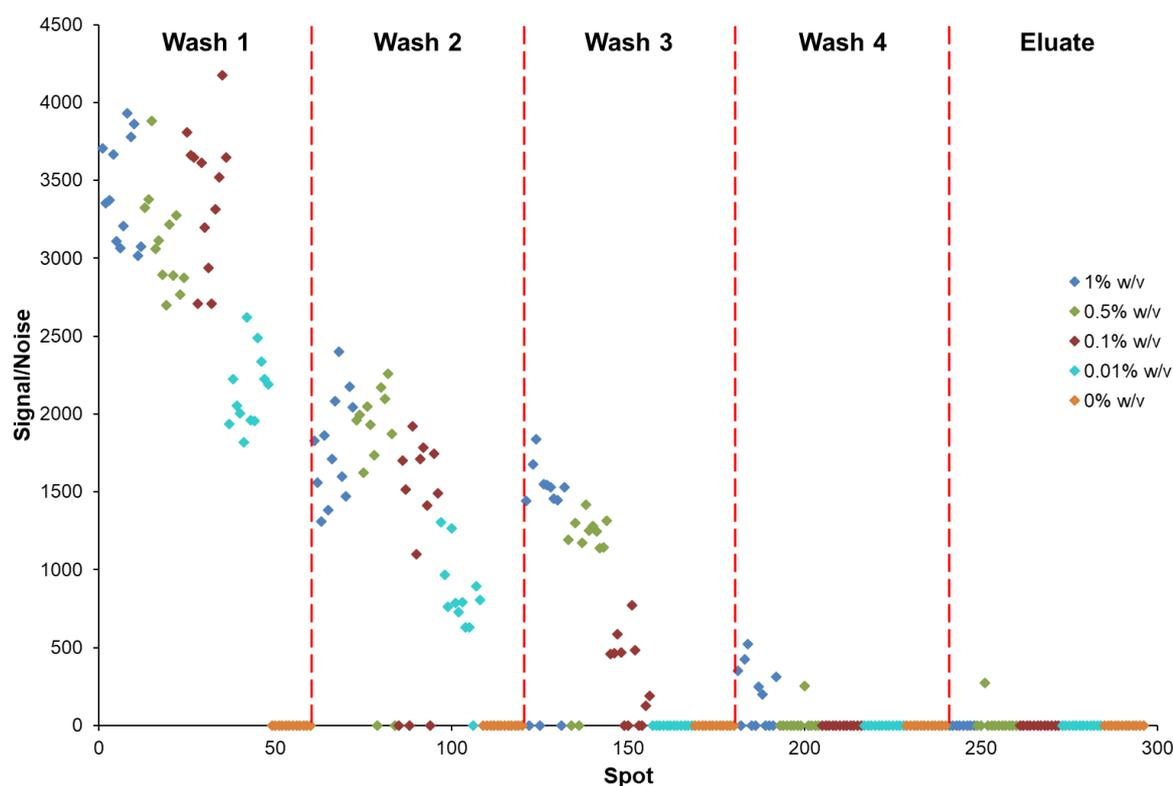


Figure 4-6: Signal to noise ratios of SDS-d25 after multiple consecutive washing steps. Concentrations stated in the legend are the spiking concentrations of SDS-d25.

Based solely on S/N ratios, SDS-d25 was efficiently removed by washing steps using ethanol only. At an initial spiking concentration of 0.1% w/v, three wash runs appeared to ensure sufficient removal, with the identifying signal entirely absent from the fourth wash supernatant.

After establishing the concentration range within which SDS could be removed by three washing steps, the next stage of method development not only comprised testing an SDS-containing lysis reagent mix on live *C. elegans* specimens, but also directly comparing performance of the former with that of other reagents or chemicals. For this purpose, six conditions were tested in total, of which one represented the control with 1 × PBS. The lysis progress was monitored by taking photographs in intervals of ~5 min (see Figure 4-7).

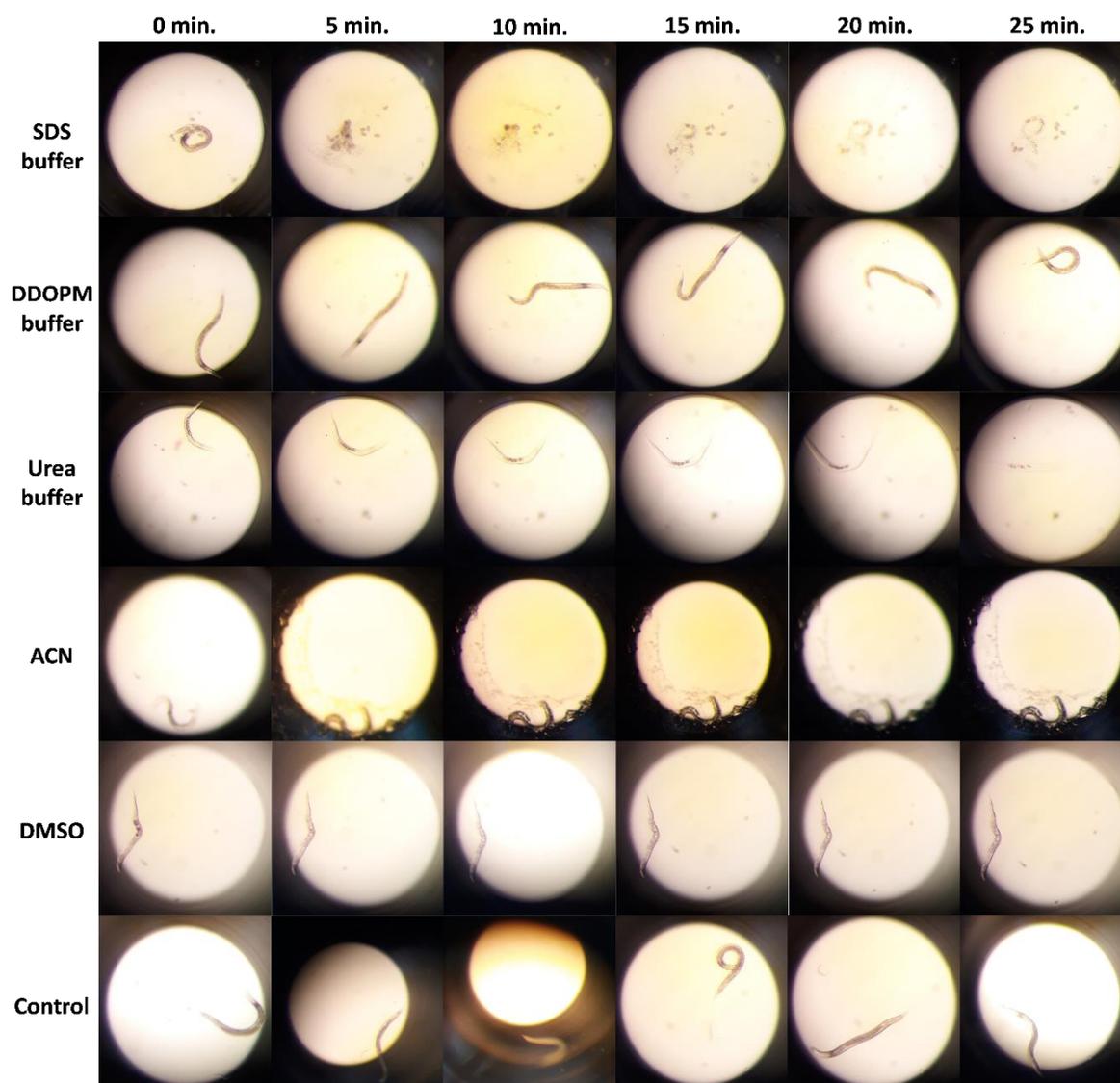


Figure 4-7: Effects of five different lysis buffers on single live *C. elegans* over time, including the SDS- and urea-based buffers later used. Conditions were tested in triplicate and one replicate was subsequently included in the montage. Adapted with permission from (Steinbach et al., 2022). Copyright 2022 American Chemical Society.

As expected, all reagents (excluding the PBS control) tested in this manner proved lethal to the nematodes, in most cases within less than a minute of being exposed, as was observable from a complete loss of motility. An exception was the DDOPM-containing lysis reagent, exposure to which saw nematodes able to survive for ~20 min without apparent damage to the cuticle or internal body structures, albeit accompanied by a gradual decrease in movement speed that eventually led to complete stasis. Similarly, the organic solvents ACN and DMSO, although apparently fast to take effect on the organisms, were not observed to cause external or internal lysis, and also had to be excluded categorically from experiments on the Desktop Biofoundry device due to

incompatibility with the top plate material. Moreover, ACN notably evaporated within the first 10 minutes of the experiment. This left two options – 6 mg/mL DTT, dissolved in either 0.1% w/v SDS or ~8 mol/L urea and 1% w/v NP-40 – as likely lysis reagent mixtures. Interestingly, while both of these mixtures were lethal almost immediately, their observed effects on the structural integrity of *C. elegans* turned out to be highly different. The SDS-based reagent presumably led to the complete dissolution of the cuticle within 5 min, releasing any tissue portions still unaffected into the surrounding medium. In contrast, the urea/NP-40-based mixture did not appear to cause rapid disintegration of the cuticle within any point in time during the experiment. Instead, nematodes viewed under the microscope would gradually lose all opacity, suggesting a slower dissolution of the cuticle. This was suspected to promote internal lysis of the specimens by making the cuticles sufficiently permeable for the lysis reagent to take effect within the nematode bodies, potentially causing an even more thorough lysis than was apparent for the SDS-based mixture.

4.4.2. Testing of different lysis and digestion conditions (oil-based DMF)

Based on the results of the lysis trial, reagent mixtures based on SDS and NP-40 were considered for further use in DMF-based proteomics. Five full lyses of single *C. elegans* specimens was performed on the Desktop Biofoundry device. For all of these experiments, the lysis workflow described in chapter 4.3.7 served as the default template, based on which alterations to single conditions were made after each experiment. By default, each sample was exposed to 8 ng trypsin; based on previously reported values for the protein content of adult *C. elegans* (Kohl et al., 2018), this amounted to a protein/protease mass ratio of approximately 75:1.

A noteworthy factor and recurring theme throughout this series was the frequency of complete replicate losses. All of the five experiments discussed below were originally intended to comprise four replicates; however, in three of these experiments, one replicate was lost for reasons such as incomplete retrieval or oil leaks. Results presented here therefore need to be viewed in the context of inconsistent replicate numbers.

While both of the pre-selected lysis reagents had suggested favorable performance during the trials already described, LC-MS/MS measurements of the reduced, alkylated and digested lysates showed markedly different yields in terms of inferred proteins, protein and peptide groups, and PSMs (Figure 4-8). In the first experiment under default conditions (see chapter 4.3.7), numbers of inferred protein identifications and protein groups reached

~200 on average ($n = 3$), with peptide groups and PSMs accordingly averaging at around 500. This surprisingly low yield contradicted expectations set by the SDS lysis mix, the latter having been the only one tested capable of causing a visually confirmed disintegration of the cuticle with subsequent release of interior cells. Based on these results, the lysis mix was switched out for the urea/NP-40 version, while otherwise maintaining the exact conditions stated in the default protocol (chapter 4.3.7). Surprisingly, this replacement alone caused yields to increase significantly across the board, with the numbers of inferred protein identifications and protein groups being more than quadrupled. Peptide groups experienced the most pronounced gain, with their number displaying an eight-fold increase. These observations were attributed largely to the direct impact of the urea/NP-40-based mixture on the lysis itself; however, the possibility of this mixture being less inhibitive than the SDS variant towards the proteases later introduced could not be excluded either.

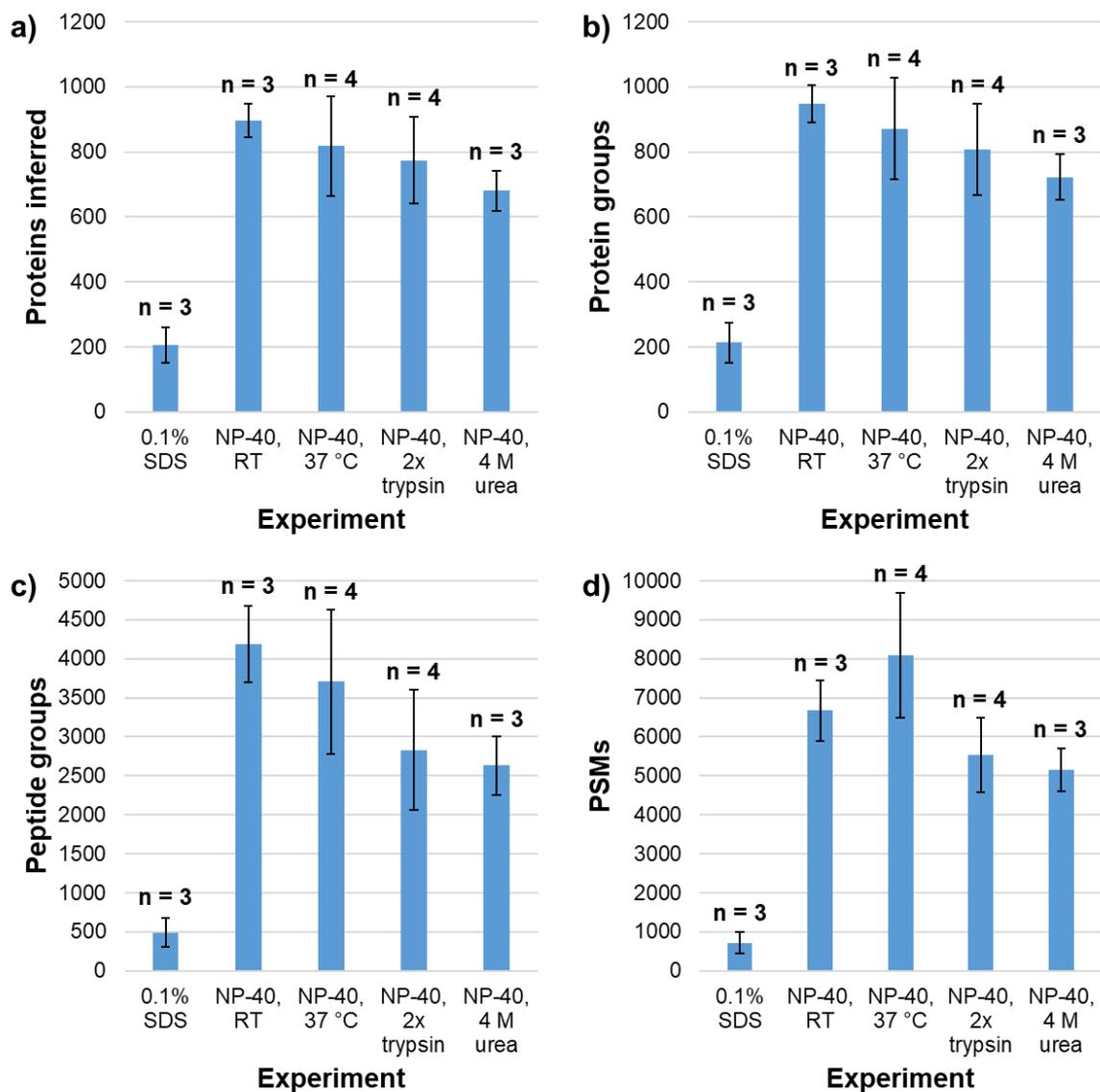


Figure 4-8: Average numbers of inferred protein identifications (a), protein groups (b), peptide groups (c) and PSMs (d) from the single-*C. elegans* lysis experiments on the Digi.Bio Desktop Biofoundry platform. Varying numbers of replicates were due to loss of complete samples.

With an optimized lysis reagent mix established after the first two experiments, it was then investigated how further adjustments to the workflow would affect the digestion efficiency in terms of missed cleavage events. The first approach was to increase the temperature during the digestion step to 37 °C, matching the temperature optimum of trypsin. In a second setup, the trypsin concentration was doubled compared to the default concentration, while maintaining a digestion temperature of 37 °C. Here, the aim was to further increase digestion efficiency by increasing the likelihood of proteins being exposed to active trypsin. For the third approach, the original trypsin concentration was used at

37 °C; however, the urea concentration in the lysis buffer was halved, with the intention of lowering the overall chaotrope concentration to which the protease was exposed.

Surprisingly, the highest average numbers for all parameters considered were in fact almost exclusively obtained in the second experiment of the series (urea/NP-40 lysis and overnight digestion at room temperature). An exception to this were the numbers of PSMs, which peaked following the switch to digestion at 37 °C.

Next, it was investigated how the different reaction conditions tested affected the incidence of missed cleavage events. Based thereupon, the composition of the lysis reagent mix could be assumed to have a considerable impact on the protease activity during the digestion step. Switching out the SDS- for the urea/NP-40 lysis buffer caused the incidence of no missed cleavages to drop by ~10% (compare Figure 4-9a and b), suggesting the latter mixture to inhibit the proteases used even more than the SDS-containing variant. This simultaneously supported the assumption of a urea/NP-40-based buffer being beneficial to the lysis of *C. elegans*. Furthermore, as expected, increasing the digestion temperature to 37 °C from room temperature resulted in a marked decrease of missed cleavage events (compare Figure 4-9b and c). An even higher digestion efficiency was achieved when the trypsin concentration was doubled (Figure 4-9d), while again maintaining 37 °C. Conversely, reducing the starting concentration of urea in the lysis buffer by half only had a limited effect on the percentage of missed cleavage events (Figure 4-9e).

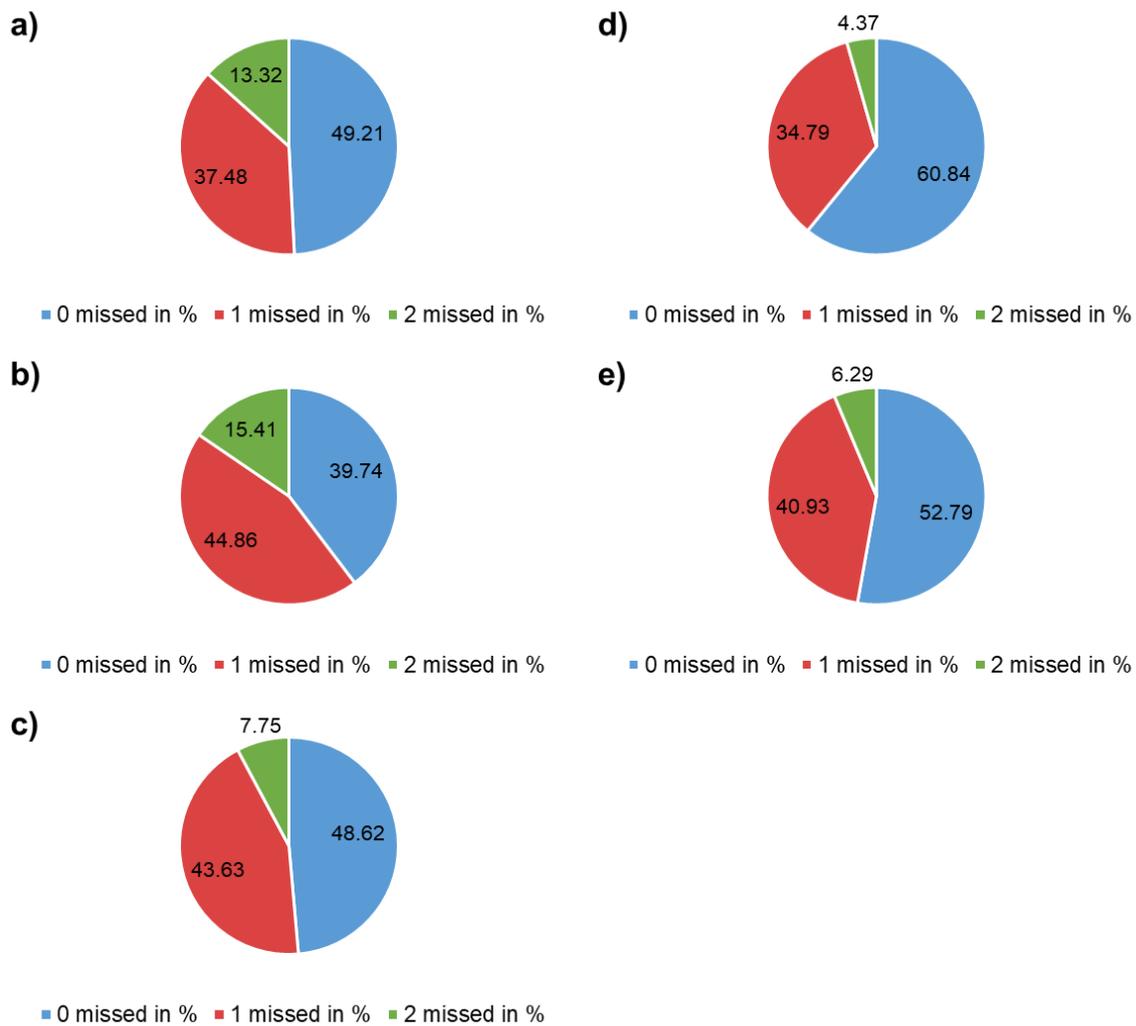


Figure 4-9: Average numbers of proteins inferred with no, one or two missed cleavage events, compared across five experiments performed on the Digi.Bio platform. a) Experiment conducted according to the default protocol. b) Lysis buffer replaced with NP-40 variant, digest at room temperature. c) Similar to B, digest at 37 °C. d) Similar to C, protease concentration doubled. e) Similar to C, urea concentration reduced to ~4 mol/L.

In the light of evidence provided by the qualitative evaluation of the experiments described here, as well as by the incidence of missed cleavages, the lysis mixture based on urea/NP-40 was selected as the default for future LFQ experiments with *C. elegans*.

4.4.3. Handling aspects of air- and oil-based DMF, and in-tube preparation

Regarding the preparation of single nematodes, the workflow devised for the DropBot DB3-120 could be compared against two others, those being a tube-based approach to represent a traditional off-chip sample preparation, as well as the method developed for the Desktop Biofoundry device (chapter 4.3.7) to represent oil-based DMF. Regarding working volumes on the chips, both DMF systems proved remarkably similar, with regular electrodes accommodating ~1 μL , and reservoir electrodes 5-10 μL . Workflows could thus be transferred from one device to the other with relative ease and no adjustments to volumes. It should be noted, however, that the sample preparation methods on these different DMF devices were succeeded by measurements on different instrumental setups, which generally limits comparability between the two methods beyond aspects concerning sample handling itself.

Adsorption of sample material, especially biomolecules, to plastic container walls is a known disadvantage of handling minute sample amounts in common reaction vessels such as tubes (Kasuga et al., 2017), in particular as a result of repeated manual pipetting required to add and mix reagents into samples, and retrieve the latter for analysis. The major aim of DMF method development for this subproject therefore was to limit such working steps as much as possible. In this regard, air-based DMF was found to outperform its oil-based peer, where mandatory off-chip peptide SP3 to remove oil alone accounted for five additional pipetting steps that directly interacted with samples (Table 4-2).

Table 4-2: Experimental parameters common and different across two DMF protocols devised for use on the air- and oil-based platforms (DropBot DB3-120 and Desktop Biofoundry, respectively).

Experimental aspect	Air-based DMF	Oil-based DMF
Reagent volumes	1 μL by default, except digest buffer (2 μL)	1 μL by default throughout procedure
Evaporation	Possible (samples directly exposed to the air)	Negligible due to silicone oil filling
Manner of SP3	On-chip protein SP3 (to remove lysis/alk. reagents)	Off-chip peptide SP3 (to remove oil)
Fate of lysis/alkylation reagents	Removed by SP3 prior to digest	Still present upon start of digest, can affect proteases
Volume during digest	2 μL from digest buffer alone	4 μL due to accumulation
Chip handling for digest	Sealing overnight, or short digest (evaporation!)	Overnight digest possible, risk of oil leaks
Off-chip pipetting events	2 (sample transfers to and from chip)	7 (transfers + washing steps for peptide SP3)

Another aspect in favor of air-based DMF was the ability to perform SP3 sample cleanup directly on-chip, which made it possible to remove any harsh reagents before introducing proteases. Compatibility issues between preferred SP3 solvents with polymer-based top plates and/or silicone oil led to the decision to forego on-chip SP3 during oil-based DMF. This brought protease-containing drops into direct contact with potentially detrimental reagents, which in turn had a detectable impact on results (Figure 4-9).

As expected, evaporation of liquids was an important limiting factor in experiments on the DropBot DB3-120, resulting in the necessity of direct countermeasures such as enclosing chips and even replenishing lost volume (see chapter 4.3.8), all of which proved entirely omissible in oil-based experiments. Conversely, the use of oil came with the possibility of leaks, potentially leading not only to sample loss, but also to a severely increased risk of oil contamination of any equipment used downstream. Additionally, the lack of both evaporation and removal steps on the oil-based platform led to higher spatial demands since total sample volumes only ever increased with reagent additions, which limited the number of replicates that could be prepared per experiment without the risk of mixing or cross-contamination.

Being non-reliant on space on an electrode grid, tube-based sample preparation of *C. elegans* notably had the advantage of being the least restrictive in terms of replicates per experiment. However, one important risk factor with this approach, aside from the notorious potential for sample loss due to surface adsorption, was seen in the possibility of aspirating not or only partially lysed nematode remains in the early stages of the experiment during any instance of pipetting (e.g. to cause appropriate mixing). Reasons for this were poor visibility of the specimen caused both by the sample itself due to turbidity and/or random air bubbles, or by the surrounding container due to material opacity and light reflection. While this risk was alleviated in both DMF procedures described here, specimen loading and processing on the air-based platform proved to be especially convenient due to the excellent visibility offered by the thin glass top plate.

4.4.4. Air-based DMF vs in-tube: Qualitative results

The air-based DMF- and in-tube workflows presented in chapters 4.3.8 and 4.3.9 were initially tested on single N2 adults of *C. elegans*.

Absolute yields in terms of inferred proteins, protein-/peptide groups and PSMs substantially exceeded those observed in the previous oil-based DMF series. This was

attributed largely to the changes in instrumentation. When compared directly, the air-based DMF and in-tube methods were found to perform similarly in terms of inferred protein identifications and -groups. Additionally, DMF was observed to notably outperform the in-tube protocol regarding peptide groups and, to a lesser extent, PSMs.

Another aspect where DMF proved beneficial during these trials was variance between replicates. While the tube-based workflow generally resulted in numbers comparable to the DMF version for all parameters, accumulated errors during sample handling, presumably mainly due to the predominance of manual pipetting, led to an increased risk of outliers and were thus detrimental to repeatability. This was directly reflected by the qualitative parameters of the tube-based series, an example for which is replicate 2, where values in all compared categories were significantly lower than in the other four replicates (Table 4-3).

Important: Replicate 2 of the in-tube dataset was treated as an outlier and disregarded for the further discussion in this chapter, with the intention of equalizing the number of replicates per dataset discussed. This directly affects Figure 4-10 and Figure 4-11.

Table 4-3: Numbers of inferred proteins, protein and peptide groups, and PSMs from the qualitative experiment series using single N2 adults, following “per file” searches. The proteins and peptide groups are of *C. elegans* only, the former being represented by at least two peptides (one unique, one other) with a high FDR confidence. The outlier in the in-tube dataset (later disregarded) is highlighted.

Method	Replicate	Inferred proteins	Protein groups	Peptide groups	PSMs
DMF	1	4388	5811	45 242	57 299
	2	4279	5697	44 430	55 954
	3	4099	5464	41 383	51 097
	4	4207	5773	43 298	54 568
In-tube	1	4394	5935	42 687	54 122
	2	3446	5222	30 169	38 074
	3	4211	5874	39 766	50 039
	4	4425	6451	42 142	54 782
	5	4050	5698	38 136	48 948

In previously published BUP studies, numbers of inferred protein identifications were shown to experience a notable increase when matching results obtained from consecutive replicates of the same series (e.g. the MBR function of the MaxQuant software) (Zhu, Piehowski, Zhao, et al., 2018). Aiming to achieve a similar effect in Proteome Discoverer, the dataset described here was searched again using the Minora Feature Detector node, with the results obtained then compared directly to the average numbers from the batch

search. By applying feature mapping in this manner, a slight increase in terms of inferred protein identifications and a significant increase in terms of peptide groups was observed for both the DMF- and in-tube dataset (Figure 4-10). Interestingly, when disregarding replicate 2 of the in-tube dataset, the standard deviation of inferred proteins with and without feature mapping, as well as peptide groups with feature mapping only, turned out to be lower than for the DMF data. In qualitative terms alone, this suggested a comparable performance of the in-tube method in relation to DMF, provided that errors introduced during sample preparation do not create outliers among the technical replicates.

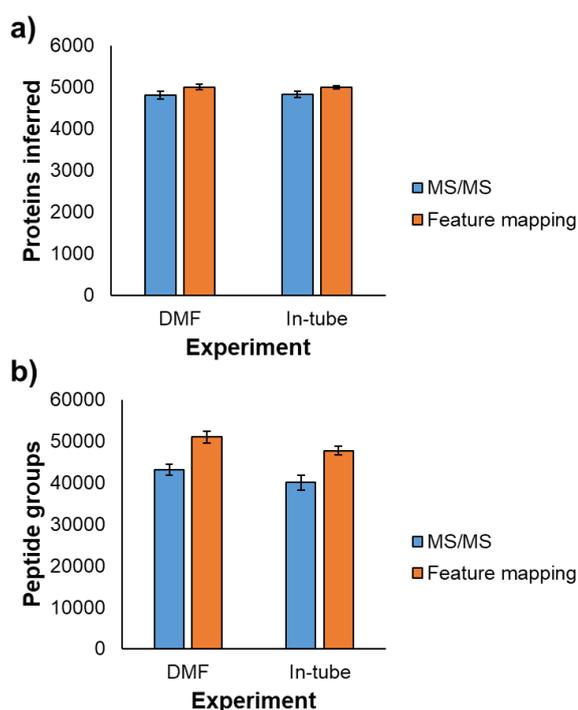


Figure 4-10: Comparison of proteins inferred with high confidence (a) and peptide groups (b) in single *C. elegans* after BUP sample preparation on the DropBot DB3-120 DMF system or via an in-tube workflow (numbers per replicate shown in Table 4-3). Datasets are shown with and without MBR simulated by means of the Minora Feature Detector. Error bars are defined by the standard deviation. Replicate 2 of the in-tube dataset was disregarded.

Simultaneously, the protein identification overlap between replicates increased likewise when compared to the absence of feature mapping. As expected, this was observed for both sample preparation methods. In the case of DMF preparation, the percentage of proteins inferred in all four replicates was 80.5% without and 90.2% with feature mapping (Figure 4-11). In-tube sample preparation notably did in fact lead to marginally higher values of 85.2 and 93.2%, respectively, despite the overall similar performance in terms of total proteins and peptide groups, although this again needs to be viewed with respect to

the outlier removed from the in-tube dataset (see also Table 4-3). The latter reason in particular was considered an indication of the DMF workflow being more robust than the in-tube variety, which is also supported by a similar earlier comparison (Leipert & Tholey, 2019).

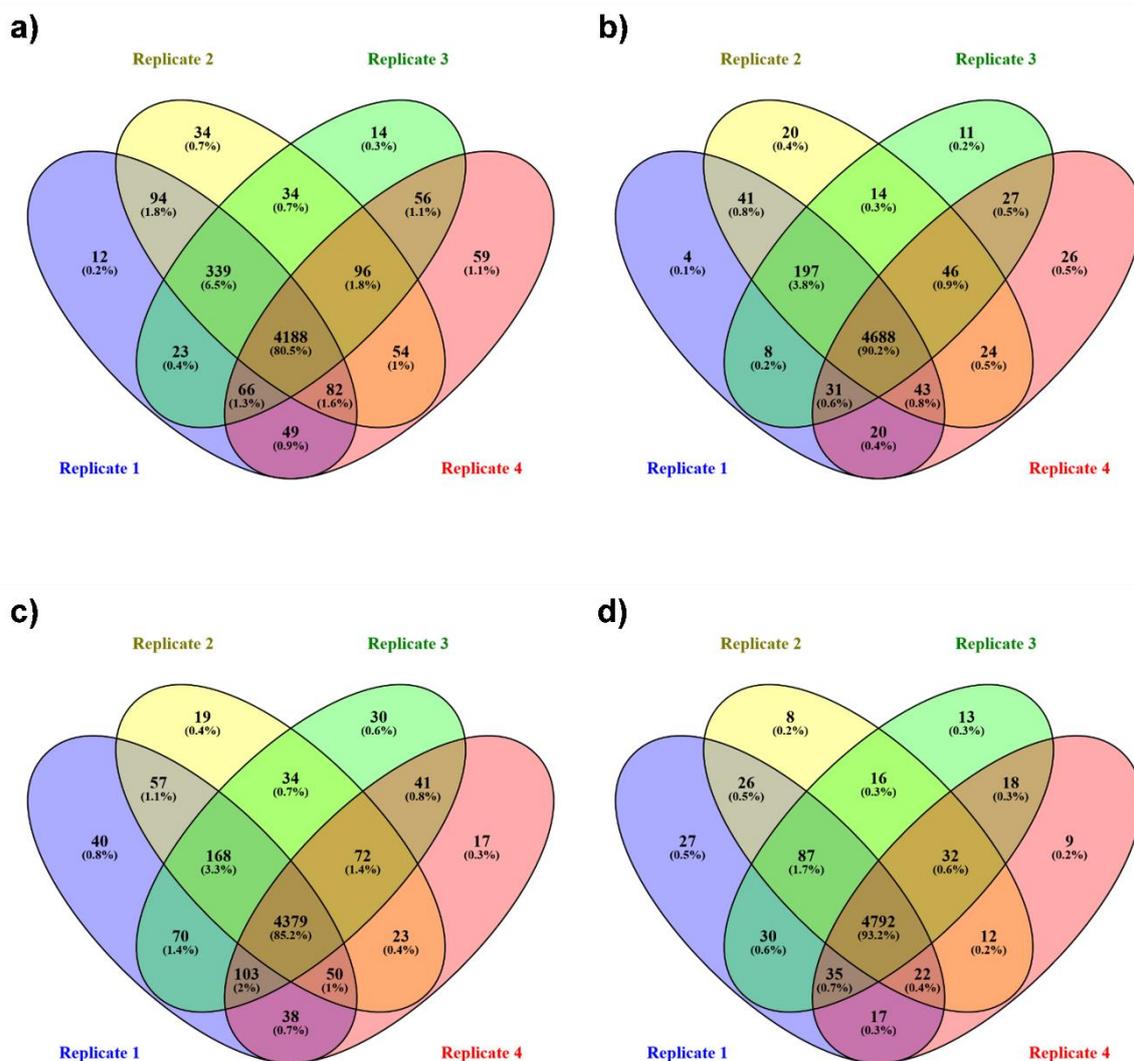


Figure 4-11: Overlap between replicates of inferred proteins from *C. elegans* after sample preparations via DMF- and in-tube protocols, both with and without feature mapping. Replicate 2 of the in-tube dataset was disregarded (see also Table 4-3). a) DMF, MS/MS only. b) DMF, feature mapping. c) In-tube, MS/MS only. d) In-tube, feature mapping. Venn diagrams were created using the Venny 2.1 online tool (Oliveros, 2007).

4.4.5. Comparison of different biological conditions by label-free quantification

Using the DMF workflow for the DropBot DB-120 (chapter 4.3.8) and the in-tube counterpart (chapter 4.3.9), single L4-stage *C. elegans* specimens were prepared following a 1-hour exposure to heat or constant growth under control conditions (20 °C). Six replicates were prepared per condition and method. Due to the spatial constraints of the DMF chips, two of the latter were used, with the replicates for both conditions divided evenly between the two (three “Control”, three “Heat” samples per chip) to avoid batch effects.

As observed during qualitative trial runs (chapter 4.4.4), the application of feature mapping notably led to a significant increase both in the numbers of proteins identified and peptide groups. Feature mapping was also found to reduce variance between replicates regardless of the sample preparation method (Figure 4-12), thereby confirming the qualitative results once more.

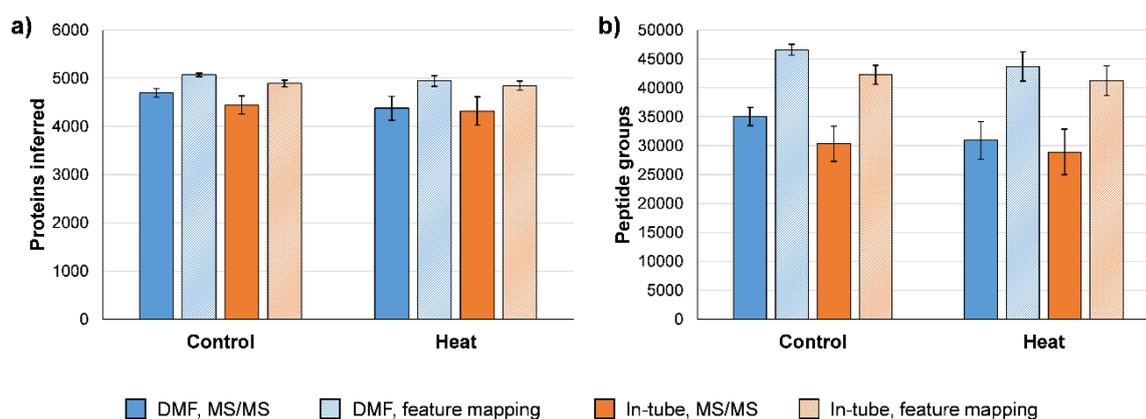


Figure 4-12: Comparison of average filtered inferred proteins (a) and peptide groups (b) from “Heat” and “Control” samples after either DMF- or tube-based sample preparation, both with and without feature mapping. Adapted with permission from (Steinbach et al., 2022). Copyright 2022 American Chemical Society.

The complete datasets of quantifiable proteins consisted of ~5000 entries for both methods, with the count being marginally lower in the case of the in-tube preparation. Inspection of the data with regard to the ratio of missing and valid values showed that each set comprised only ~5% missing values. Accordingly, almost all quantified proteins in either dataset were represented by at least two valid values per group, and ~80% did not lack a single value (see also Table 4-4). Due to this data quality, reducing the datasets based on valid values was omitted entirely in favor of retaining as many quantifiable protein entries as possible.

Table 4-4: Proteins quantified in the DMF- and in-tube datasets after applying valid value filters of varying strictness. Data originally published in (Steinbach et al., 2022).

Dataset	Minimum values per group	Proteins quantified	% of total
DMF	0	5067	100.00
	1	5065	99.96
	2	5065	99.96
	3	4937	97.43
	4	4778	94.30
	5	4513	89.07
	6	4051	79.95
In-Tube	0	4916	100.00
	1	4914	99.96
	2	4914	99.96
	3	4778	97.19
	4	4622	94.02
	5	4383	89.16
	6	3963	80.61

In both datasets combined, 45 inferred proteins were found to be differentially abundant following heat treatment, compared to control conditions, with 16 being more and 29 less abundant. The evidence from this experiment strongly suggested the DMF protocol to considerably benefit the detection of abundance changes in inferred proteins: Of the 16 proteins more abundant after heat stress, nine were identified as such exclusively after the DMF- and one after the in-tube protocol, while six showed this kind of abundance difference after both workflows. A similar trend was observed for the 29 proteins with lower abundance after heat stress, where this abundance difference tendency was observed for 21 (DMF), four (in-tube) and four (both) proteins, respectively.

The markedly higher yield of the DMF sample preparation in terms of differentially abundant proteins is particularly interesting when considering the otherwise similar performance of the two methods. Both methods yielded comparable results in terms of inferred proteins from a qualitative point of view, with a greater advantage of the DMF method observed at peptide group level, thereby matching the results of earlier trials with single adult *C. elegans* specimens. Additionally, inferred proteins with a significant detected abundance difference between groups following both preparations were often found to display comparable difference values across both methods, with some difference values in fact being higher in the case of the in-tube dataset (Figure 4-13).

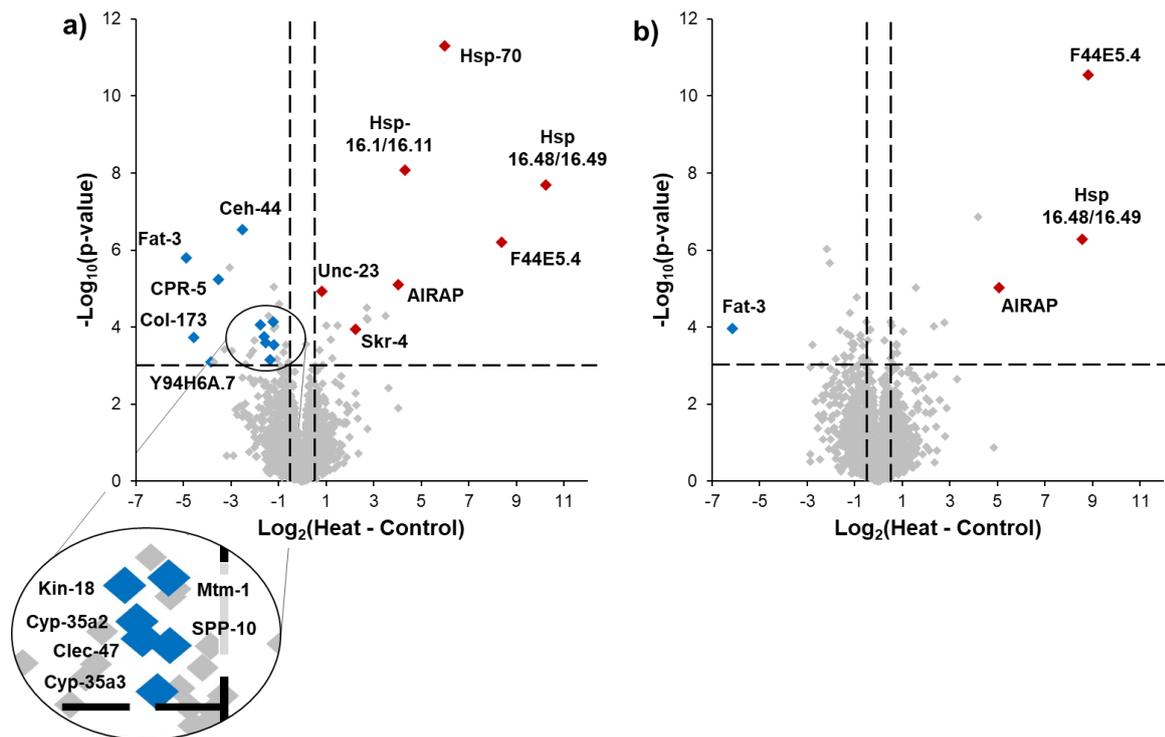


Figure 4-13: Volcano plots of inferred proteins from the DMF (a) and in-tube (b) datasets, highlighting biologically relevant proteins discussed in the associated research article, as well as in chapter 4.4.6 of this thesis. Adapted with permission from (Steinbach et al., 2022). Copyright 2022 American Chemical Society.

4.4.6. Biological interpretation

In general, the proteins more and less abundant after heat treatment appeared to match both expectations and previous reports with regard to the biological response of nematodes to heat stress. As expected, several heat shock proteins (Hsp 16.48/16.49, Hsp-70 and Hsp-16.1/16.11) were significantly more abundant after heat stress, which became especially apparent in the DMF dataset. This confirmed an earlier study where the genes associated with these heat shock proteins were reported to show increased expression levels after heat stress for 30 min at 35 °C (Jovic et al., 2017).

Among the other proteins more abundant after heat stress were Unc-23 (a BAG domain-containing protein) and F44E5.4. The former is known to alter the ATPase domain of Hsp-70, whereas the latter also belongs to the *hsp-70* family, for which upregulation at the gene expression level during heat stress had already been shown (GuhaThakurta et al., 2002). A similar abundance difference was found for Skr-4, a promoter of thermotolerance (Panek et al., 2020), as well as the arsenite-inducible RNA-associated protein (AIRAP).

The corresponding gene of the latter has been shown in the past to be upregulated by different stress factors, in particular presence of arsenite (Yun et al., 2008); however, earlier results did not suggest a correlation between heat stress and increased abundance of arsenite mRNA (Sok et al., 2001).

The most pronounced negative abundance difference was observed for delta(6)-fatty-acid desaturase (Fat-3), an enzyme responsible for the synthesis of eicosatetraenoic acids (Napier et al., 1998; Watts & Browse, 2002). Abundance of the Col-cuticle-N-domain-containing protein Col-173, assumed to be a structural constituent of the cuticle (Shaye & Greenwald, 2011), was affected in the same way, suggesting an alteration of the cuticle composition as a response to heat stress.

Also less abundant after the heat treatment was the C-type lectin domain-containing protein Clec-47, a protein playing a role in the synthesis of antimicrobial and detoxifying proteins by increasing the transcription of associated genes (Pan et al., 2021). Accordingly, the saposin-like protein SPP-10, being in a family of antimicrobial peptides (Roeder et al., 2010), presented a similar abundance difference.

The protein Y94H6A.7, associated with a low molecular weight phosphatase (LMWPC) family domain, may correlate negatively with the heat shock resistance of *C. elegans*. Based on an earlier report, the resistance to different stress factors such as heat or oxidative stress, but also to DNA damage, increases if activity of the underlying gene is downregulated (L. Zhou et al., 2018). This also appears to be reflected by the results of this study, where Y94H6A.7 was in fact present at a lower abundance following heat stress, compared to the control group.

The myotubularin-related protein Mtm-1, as well as the serine/threonine protein kinase Kin-18, inhibit apoptosis, and the decreased abundance of these two proteins after heat stress may therefore suggest beginning apoptosis as a result (Neukomm et al., 2011; Takayama et al., 2003; Yin et al., 2016; Zou et al., 2009).

Aside from the inferred proteins discussed here in detail, a general observation was that those proteins less abundant after heat stress were made up to a notable proportion of enzymes, with proteases being particularly strongly represented, which lines up with previous findings (Bensaddek et al., 2016).

4.5. Conclusion

In this subproject, numerous methods to prepare single *C. elegans* for BUP with the aid of DMF were tested and reaction conditions optimized. Two different builds of DMF devices (air- and oil-based) were used, demonstrating their distinct advantages and limitations. Especially with regard to the latter, method development for oil-based DMF was eventually abandoned in favor of the air-based alternative.

Making use of FAIMS to precede MS/MS measurements proved highly beneficial regarding sensitivity, no matter whether samples had been prepared in-tube or on-chip. Additional benefits were achieved through feature mapping in a manner similar to the MBR option of the MaxQuant software. However, while the performance of traditional in-tube preparation in this setup was found encouraging, sample preparation relying on air-based DMF not only matched, but slightly exceeded the number of proteins inferred by comparison, with a more pronounced difference observed on peptide group level.

The potential of DMF-based sample preparation for BUP was underlined even further in the quantitative evaluation of the same datasets. In terms of inferred proteins with significant abundance differences between two groups, the DMF method outperformed the tube-based equivalent, yielding approximately three times the number of total data points.

Overall, both the air-DMF- and in-tube method proved their strengths in different areas, with the main advantage of the latter method being sample throughput. Despite present shortcomings in that regard, the air-based DMF-BUP method presented in this chapter produced encouraging results, especially concerning quantitative evaluation, and thus paved the way for further development, in particular adaptation of the method to allow DMF-TDP and eventually -MDP.

5. Development of microfluidics-supported methods for MS-based Top-Down nanoproteomics

5.1. Introduction

To date, the most popular approach to MS-based proteome analyses follows the concept of BUP, at the heart of which lies protein inference based on characteristic peptides generated by proteolytic digestion. The latter therefore constitutes an indispensable stage of sample treatment, as do common preparatory steps such as disulfide bond reduction and sulfhydryl alkylation.

Compared to the BUP principle, TDP offers several advantages, one being that the absence of a digestion step allows workflows to be shortened significantly compared to BUP. More importantly, multiple proteoforms of the same canonical protein can be distinguished in TDP. Combined with the decisive advantage of low-input or even single-cell proteomics (i.e. avoidance of masking information on individual proteomes by a bulk average), this opens up new avenues when it comes to investigating biological processes. Tailoring sample preparation in the low-input realm to the requirements of TDP is therefore an emerging and promising research subject. On the other hand, TDP method development needs to address a different set of challenges. Due to their size, proteoforms are present in a greater variety of charge states and isotopologues, and are generally more heterogenous in their chemical properties, than peptides in BUP. These aspects therefore make maximizing the yield from limited input even more essential in TDP.

5.2. Project aims & major experiments

Following the single-nematode BUP experiments involving DMF-based sample preparation, as well as both a qualitative and quantitative data evaluation (chapter 4), the next step was to transfer the existing DMF-BUP concept to TDP, followed by the preparation of single *C. elegans* in two groups, and MS-based LFQ. The ultimate goal was the comparison of two DMF-based methods suitable for preparing single *C. elegans* for either BUP or TDP in terms of information gained, while also showing the advantages and inherent limitations of each proteomics approach.

The DMF-TDP-LFQ results described in this chapter were a contribution to a subsequent publication (Leipert et al., 2023).

5.3. Materials & methods

5.3.1. Chemicals & reagents

Acetonitrile (ACN; Chromasolv quality), FA and TFA (both LC-MS-grade), were obtained from Honeywell Specialty Chemicals Seelze GmbH (Seelze, Germany), ethanol (absolute) from Th. Geyer GmbH & Co. KG (Renningen, Germany).

Gibco PBS buffer (10 ×) and Pierce borate buffer (20 ×) were from Thermo Fisher Scientific GmbH (Bremen, Germany).

Mannitol-based amphiphile C12 (MNA-C12) was from Anatrace Inc. (Maumee, OH, USA).

All other chemicals and reagents, including SeraMag Speed Beads (hydrophilic and –phobic, carboxylate-modified), were from Sigma Aldrich (Steinheim, Germany).

Hydrophilic and –phobic SP3 beads (both as 50 µg/µL suspensions in 0.05% w/v azide) were always mixed to equal parts, washed twice with milli-Q-water, then resuspended at the desired concentration in 0.08% w/v Pluronic F-127.

5.3.2. *C. elegans* preparation and handling

C. elegans were prepared by Christine Blurton (AG Leippe) at the Zoological Institute of the *Christian-Albrechts-Universität zu Kiel* as described in chapter 4.3.2.

For label-free quantification, a synchronized population of *C. elegans* was grown under default conditions until the L4 stage of development; then, two subpopulations were created. One of these was incubated for about 1 h at 33 °C, based on a recently published method (Golden et al., 2020), whereas the other was kept at default conditions. Single nematodes were then isolated in 0.2 mL PCR tubes (Eppendorf SE, Hamburg, Germany) immersed in 1 µL of M9 buffer, and flash-frozen in liquid nitrogen. To improve the stability of the sample material, the single specimens were transported on dry ice and stored at –80 °C.

5.3.3. TDP urea lysis buffer

A major part of the sample preparation during this subproject relied on the use of a lysis buffer (proposed and tested by Jan Leipert, AG Tholey). This buffer consisted of ~5 mol/L urea, 1% w/v NP-40, 0.1% w/v Tetriconic 90R4, 5 mmol/L DTT, benzonase 1:5000, which were dissolved or diluted in different manners prior to being mixed. Urea was dissolved to a concentration of ~8.7 mol/L in a solution of 1 × cOmplete (EDTA-free), which was diluted to 5 mol/L in the final buffer. Benzonase was diluted first by 100 in a mixture of 100 mmol/L MgCl₂ and 250 mmol/L borate buffer, then by 50 in the final buffer. All other substances were initially dissolved in water (DTT, 25 mmol/L and Tetriconic 90R4, 1% w/v) or added to the buffer as a readily available solution (NP-40, 10% w/v).

5.3.4. Preparation of non-alkylated Jurkat T cell lysate as a low-input standard for TDP

Jurkat T cells in nutrient medium were transferred to 2 mL protein LoBind tubes (Eppendorf SE, Hamburg, Germany), with amounts of ~1 × 10⁶ cells loaded per tube (three tubes in total). The cells were centrifuged for 5 min at 4 °C and 200 × g, and the supernatant was discarded. Cell pellets were washed three times with 1 × PBS pre-cooled on ice, using 1 mL PBS per wash and centrifuging under the same conditions as applied before. Afterwards, each cell pellet was resuspended in 400 µL PBS, for a concentration of approximately 2.5 × 10⁶ cells/mL. All cell suspensions were then pooled.

200 µL lysis buffer (see chapter 5.3.3) were pre-placed in six new 2 mL protein LoBind tubes, before adding 200 µL of pooled cell suspension to each (for a concentration of ~1.25 × 10⁶ cells/mL). All vials were then placed in a Bioruptor Pico sonication device (Diagenode SA, Seraing, Belgium) for 15 min at 4 °C (15 cycles, 30 s on, 30 s off). Afterwards, the samples were incubated for 15 min at 37 °C. All lysates were then pooled and transferred to 0.5 mL SafeLock tubes in 160 µL aliquots.

To prepare Jurkat low-input standard, 3 µL of SP3 bead suspension (20 µg/µL in water, equal parts type A and B) were pre-placed in 0.5 mL LoBind tubes. Eight µL of lysate were then added to each tube, the latter in turn vortexed slowly. 33 µL of ACN were added to the tubes and mixed by manual pipetting, before leaving the vials at room temperature (approximately 20 °C) for 10 min and subsequently removing the supernatant. The beads were then washed with 100 µL of 70% ethanol and 50 µL ACN. 20 µL of 40% FA (-20 °C) were added to the tubes, which were then immediately stored at -20 °C for 15 min. The

tubes were then placed in an aluminum rack chilled to $-20\text{ }^{\circ}\text{C}$, and $80\text{ }\mu\text{L}$ of water ($4\text{ }^{\circ}\text{C}$) were added to each tube. The eluates were transferred to $100\text{ }\mu\text{L}$ LC glass inserts (Th. Geyer GmbH & Co. KG, Renningen, Germany). In this state, they comprised a protein equivalent of $\sim 100\text{ cells}/\mu\text{L}$ and could either be used directly or stored at $-80\text{ }^{\circ}\text{C}$ until further use.

Cell lysate and low-input standard were prepared as described here by Britta Steer (AG Tholey).

5.3.5. Preparation of single *C. elegans* by DMF

The method described here is an adaptation of one originally proposed and tested qualitatively by Jan Leipert (AG Tholey).

Single specimens of *C. elegans* were prepared using a DropBot DB3-120 DMF device (Kitchener, ON, Canada) equipped with $2'' \times 3''$ 90-pin (v4) chips, in combination with a custom-made magnetic lens setup as described in chapter 4.3.4.

The reservoir- and adjacent transfer electrodes at the designated sample loading sites were actuated with a voltage of 80 V . Single nematodes stored at $-80\text{ }^{\circ}\text{C}$ in $1\text{ }\mu\text{L}$ of M9 buffer were thawed before adding $50\text{ }\mu\text{L}$ of 0.08% w/v Pluronic F-127. Individuals were then aspirated in volumes of $1\text{ }\mu\text{L}$ and loaded onto the activated reservoir electrodes. After loading these specimens for the desired number of replicates, $1\text{ }\mu\text{L}$ of lysis buffer was added to each sample and mixed with the latter, before leaving the chip at room temperature for 50 min without the electrodes being actuated.

Two μL of SP3 bead suspension ($10\text{ }\mu\text{g}/\mu\text{L}$ in 0.08% Pluronic F-127) were then added via the original reservoir electrodes, taking care not to mix prematurely with the samples. The beads were pelleted using a magnetic lens and the supernatant removed via the entry route. Each sample was moved onto a bead pellet and the magnetic lens disengaged to allow resuspension of the beads. The latter was achieved via an automated mixing route (Voltage: 95 V , Frequency: 1 kHz , Repeats: 20 , Delay: 2500 ms). $4.5\text{ }\mu\text{L}$ of ACN were added to each sample before transferring the chip to a chamber with an ACN-saturated atmosphere where it was left for 30 min at room temperature to induce protein binding.

The chip was reconnected to the DropBot and the occupied electrodes immediately actuated. After pelleting the beads, the supernatant was removed and the beads washed consecutively with $4.5\text{ }\mu\text{L}$ 70% ethanol, then $4.5\text{ }\mu\text{L}$ ACN. Prior to removing the final ACN

supernatant, 4 μL portions of MNA-C12 (0.01% w/v) were loaded from the opposite side of the chip still unused at that point and moved into close proximity of the samples. Important: Due to this particular step, experiments performed subject to this protocol always required the use of brand new DMF devices. Failure to do so bore the risk of contaminating MNA-C12 drops with sample or reagent residue from previous experiments and thus falsifying results.

After removing the final ACN supernatant, the magnetic lens was disengaged and the bead pellets moved onto the outward facing end of the transfer electrodes. 1 μL of 40% FA pre-chilled to $-20\text{ }^{\circ}\text{C}$ was added to each pellet to initiate elution, before immediately transferring the loaded chip to a freezer drawer for incubation at $-20\text{ }^{\circ}\text{C}$ for 15 min.

Following elution, the chip was reconnected to the DropBot and all electrodes with MNA-C12 immediately actuated. To achieve dilution, movement of the MNA-C12 drops towards the reservoir electrodes was initiated while simultaneously using the magnetic lens to pull the beads in the opposite direction, taking the eluates with them, ideally colliding with the MNA-C12 drops almost immediately. Until all samples were diluted and in position for retrieval, constant cooling was applied by placing an aluminum rod pre-chilled to $-20\text{ }^{\circ}\text{C}$ on the top plate above the sample drops. A graphical demonstration of this workflow is provided in Figure 5-1, and the cooling setup demonstrated in Figure 5-2.

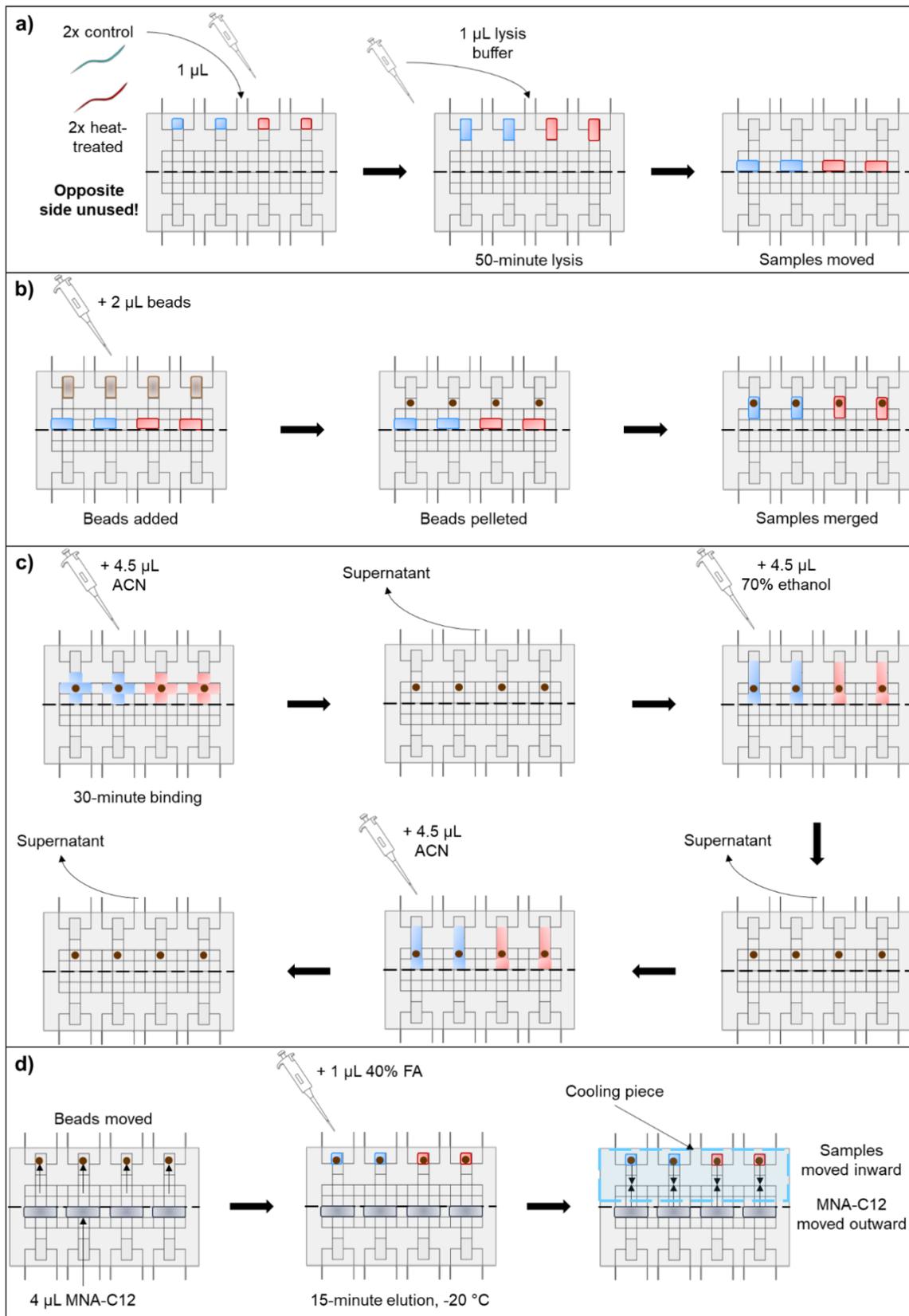


Figure 5-1: DMF workflow applied to single *C. elegans* for TDP sample preparation. a) Loading of nematodes and lysis. b) Addition of SP3 beads. c) SP3 treatment and washes. d) Protein elution and dilution with MNA-C12.

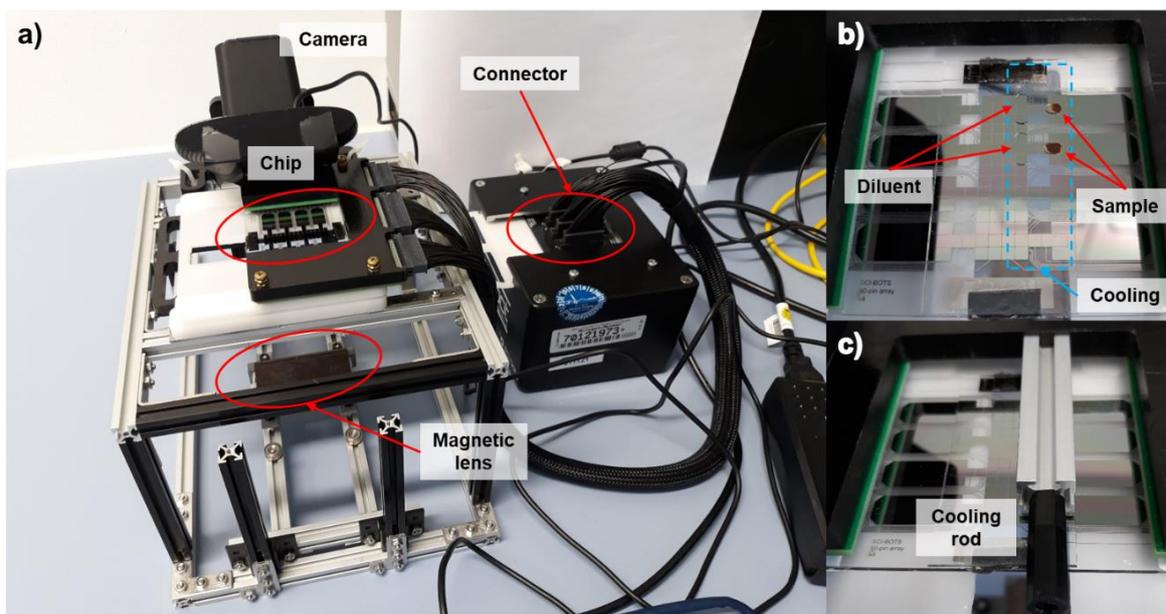


Figure 5-2: DropBot DB3-120 setup in combination with the magnetic lens (a). Simulation of the positioning of sample and diluent with two droplet pairs, highlighting the cooling zone (b). Demonstration of the actual cooling setup with the pre-chilled aluminum rod in place (c). Reproduced from (Leipert et al., 2023). Use licensed under CC-BY-NC 4.0 (<https://creativecommons.org/licenses/by-nc/4.0/>). Copyright 2023 The Authors.

Samples were recovered by aspiration from the edge of the top plate via 20 μ L GELoader tips (Eppendorf SE, Hamburg, Germany), then transferred to 100 μ L glass LC inserts (Th. Geyer GmbH & Co. KG, Renningen, Germany).

5.3.6. LC-MS/MS method

Processing of *C. elegans* lysates for TDP was done on an Ultimate 3000 RSLCnano UHPLC system, with a PepMap C₄ trap cartridge (5 mm \times 0.3 mm, 5 μ m, 300 Å, Thermo Fisher Scientific GmbH, Bremen, Germany) and an Accucore C₄ analytical column (50 cm \times 75 μ m, 2.6 μ m, 150 Å, Thermo Fisher Scientific) used for chromatographic separation. The LC setup was coupled to an Orbitrap Fusion Lumos tribrid mass spectrometer equipped with a FAIMS Pro interface (both Thermo Fisher Scientific GmbH, Bremen, Germany). The injection volume was set to 6.2 μ L, thereby ensuring each sample was injected entirely in one run. The column temperature was kept at a constant 45 °C.

LC eluents were 0.05% FA (eluent A) and 80% ACN with 0.04% FA (eluent B). The flow rate was 300 nL/min. Each run started with a 5-minute desalting step in forward flush mode, where a starting eluent composition of 4% eluent B was maintained. Next, the

eluent B flow was increased to 15% over 2 min, followed by a 180-minute linear gradient from 15 to 55% eluent B. Afterwards, the composition was changed to 98% eluent B over 6 s and maintained that way for 16 min, before equilibrating the column with 4% eluent B for 15 min.

To enable ESI, columns were equipped with fused silica emitters (outer diameter: 360 μm , inner diameter: 20 μm , tip length: 10 μm , total length: 10 cm; CoAnn Technologies, Richland, WA, USA).

FAIMS was used with internal CV stepping, a procedure described previously (Kaulich et al., 2022). This involved four scan cycles of 3 s each, with CVs of -20, -30, -40 and -50 V. MS acquisition was performed in the Orbitrap over a scan range from m/z 400 to 1800 with a resolution of 120 000, an AGC target of 200% and a maximum injection time of 246 ms. For the CV of -50 V, resolution and maximum injection time were set to 60 000 and 118 ms, respectively.

Data-dependent MS/MS (ddMS²) scans using Orbitrap detection took into account the most intense ions with charge states from +4 to +50, using a CID collision energy of 25%, a resolution of 60 000, an AGC target of 400% and a maximum injection time of 250 ms. Adjustments were again made in the case of CV -50 V, where the resolution was 50 000 and the maximum injection time 125 ms.

5.3.7. Data evaluation

Raw data from four replicates per condition was first loaded into the Freestyle 1.5 software, where MS- and MS/MS spectra were filtered per FAIMS CV (Figure 5-3) and saved as .raw files. The original eight files were thus split into 32 by creating four subsets per original replicate (referred to as CV slicing).

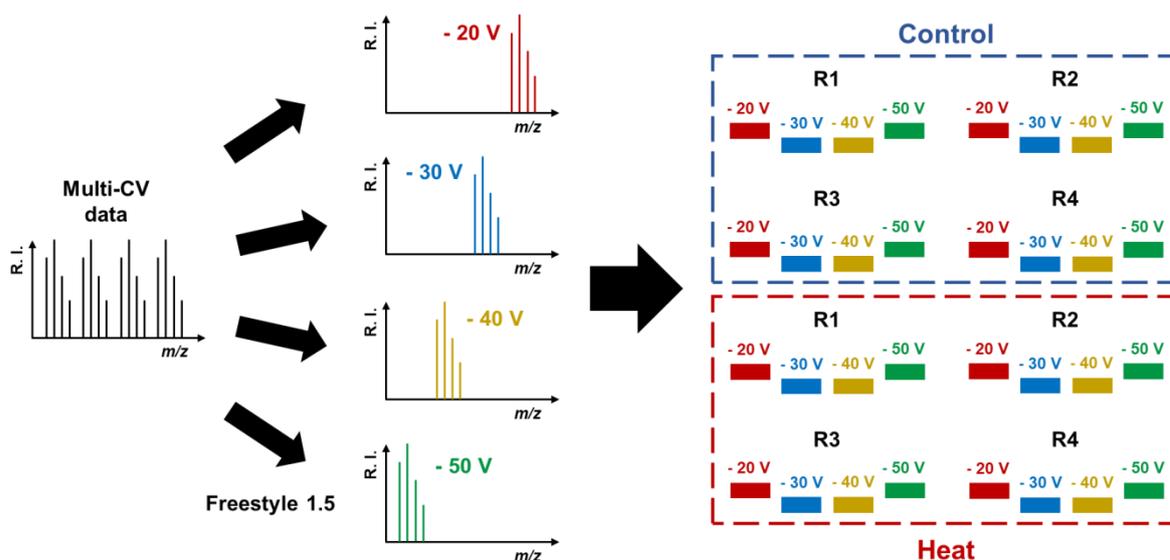


Figure 5-3: Schematic of how CV slicing using multi-CV data files was achieved. Chromatograms and mass spectra were filtered by CV, then saved as individual .raw files, leading to each original multi-CV replicate (R1-R4) being sliced into four new filtered files per replicate. Colors indicate the different FAIMS CVs, R = replicate.

After CV filtering of files originally united as multi-CV data (Figure 5-3), the new files were subjected to different data evaluation approaches, which shall be elaborated on in the following subchapters.

5.3.7.1. Searching as individual files and subsequent reassembly

The database search was performed in Proteome Discoverer v. 2.5.0.400 with ProSight v. 4.0 (Proteinaceous, Evanston, IL, USA) against a FASTA database of *C. elegans* (downloaded from Uniprot on August 26th 2021), using the ProSightPD High/High cRAWler and ProSightPD High Resolution Feature Detector node, with the 32 files grouped as “Control” and “Heat”. For the purpose of the database search, these files were treated as independent entities, effectively creating a search with 2 × 16 replicates. Both a ProSightPD Annotated Proteoform Search and a ProSightPD Subsequence Search were executed in parallel, with precursor- and fragment mass tolerances of 10 ppm. Additionally, the thresholds for “Minimum Number of Charge States Detected” and “Minimum Number of Sliding Window Detections” were set to 2. In the consensus workflow, the Feature Presence in Files threshold was set to 10%. Apart from these alterations, default settings were maintained.

Following the database search, proteoforms were filtered to only include entries with valid abundance ratios (i.e. proteoforms with sufficient abundance values per group to generate a ratio), then exported. Raw (i.e. not pre-normalized) abundance values of quantified proteoforms were then subjected to a median normalization. Next, the previously applied CV slicing was reverted by summing up the abundance values per replicate, thereby restoring the original composition of 2 × 4 samples (Figure 5-4a).

In an alternative evaluation, proteoform abundance values were processed in the opposite order, where CV-separated values were first summed up per replicate, and median normalization then performed on the sums (Figure 5-4b).

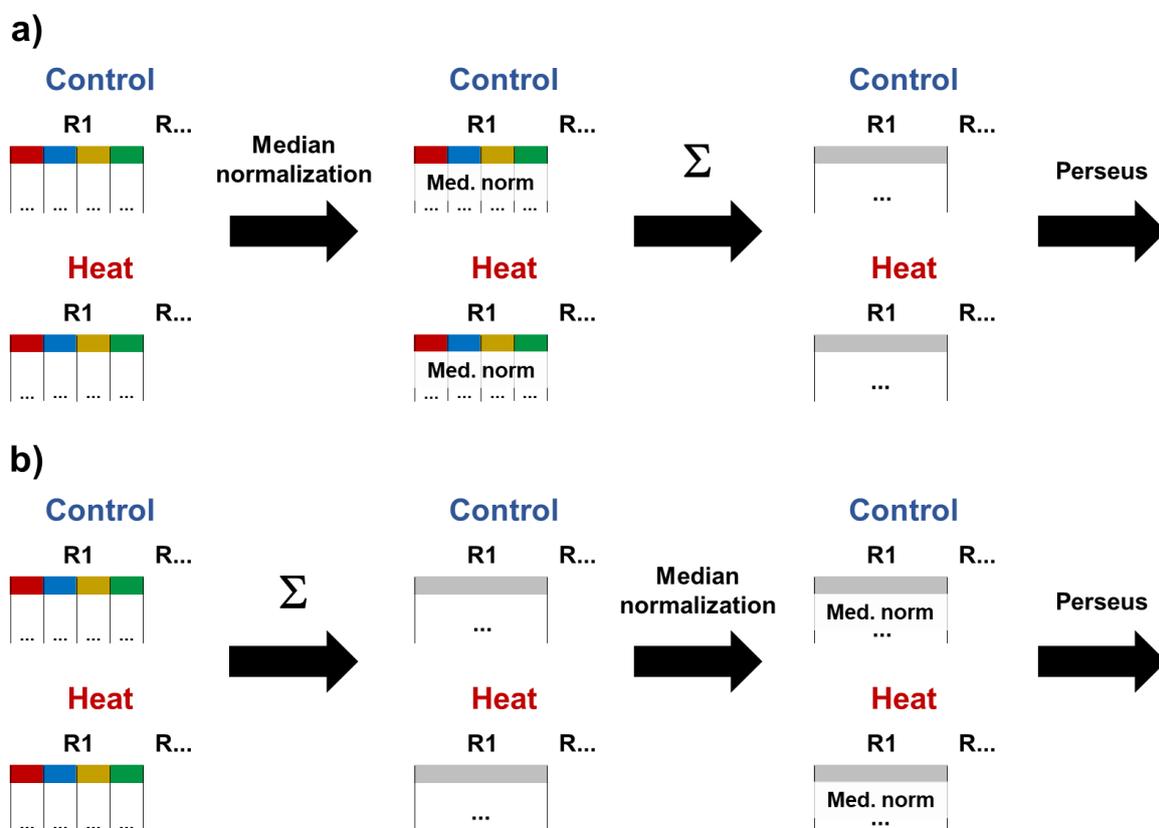


Figure 5-4: Processing of TDP data after searching as CV-separated subsets, with reverting of CV-slicing at different stages. a) CV-separated abundance values were first subjected to a median normalization, then summed up per replicate to revert the CV slicing. b) CV-separated abundance values were summed up per replicate prior to the median normalization, which was accordingly performed on the sums. Colors represent different CVs, grey indicates summed data.

In both cases, proteoform data was then exported to the Perseus software (v. 1.6.15.0), where the median-normalized and summed abundance values were Log_2 -transformed, and the replicates grouped based on the treatment of the live *C. elegans* (“Heat” vs “Control”). A filter for missing values was then applied, leaving only proteoforms with valid

abundance values in at least three replicates per group. The reduced dataset was subjected to a two-sample Student's *t* test (s0: 0, Number of randomizations: 250, Correction: Permutation-based FDR), the results of which were exported to Microsoft Excel. The $-\text{Log}_{10}$ -transformed p-value was then plotted against the Log_2 -transformed fold change. Difference thresholds were placed arbitrarily at -1 and 1 (corresponding to a 2-fold higher or lower abundance), whereas the significance threshold was placed based on the lowest $-\text{Log}_{10}$ -transformed p-value at which a proteoform was still considered significant (Figure 5-7). Accordingly, the accuracy of these significance thresholds varied based on the density of data points in the boundary region between significant and non-significant proteoforms.

5.3.7.2. Searching CV-separated data as fractions

In an alternative database search, the newly generated CV-sliced .raw files (formerly united as multi-CV datasets) were not treated as individual files once loaded into PD v. 2.5.0.400, but instead as fractions assigned to the original replicates (Figure 5-5). Following the database search, proteoforms were again filtered to only include entries with valid abundance ratios, then exported to Microsoft Excel. Since the CV-sliced abundance values had been assigned to replicates as fractions prior to the database search, no addition step was required. The abundance values were median-normalized, exported to Perseus (v. 1.6.15.0), Log_2 -transformed and filtered for missing values. For statistical evaluation, a two-sample Student's *t* test (s0: 0, Number of randomizations: 250, Correction: Permutation-based FDR) was applied. For visualization, the data was exported to Microsoft Excel after the *t* test, before plotting $-\text{Log}_{10}(\text{p-value})$ against the Log_2 -transformed fold change.

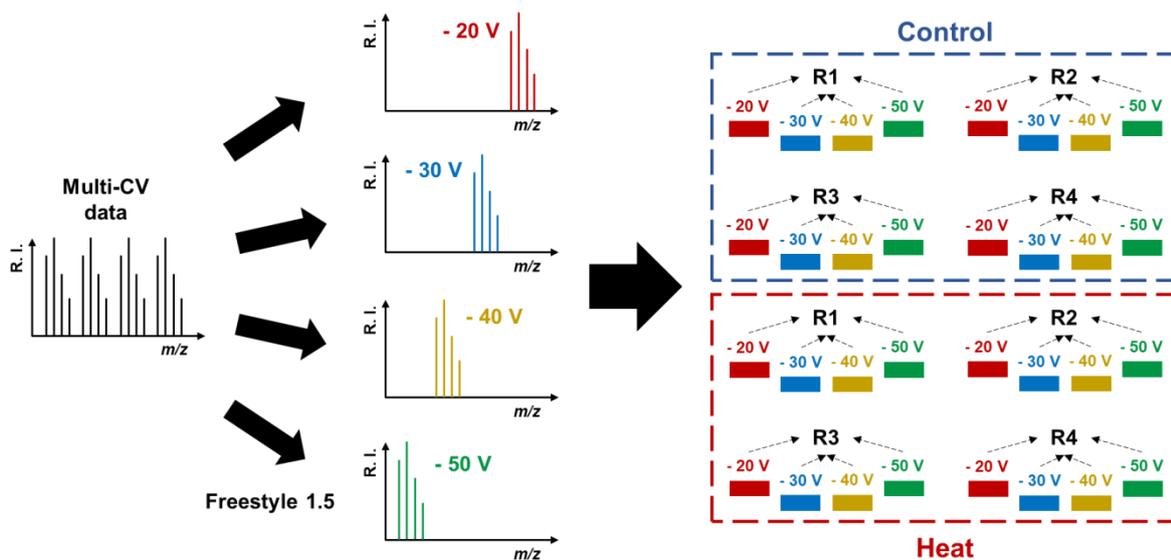


Figure 5-5: TDP database search approach with CV slicing as well as pre-fractionation in Proteome Discoverer 2.5. CV-separated files were linked accordingly to their original replicates prior to the database search itself. Colors indicate the different FAIMS CVs, R = replicate.

5.3.7.3. Independent database searches and evaluations per CV

Raw files for each CV were also subjected to database searches separately. This approach effectively created four independent subsets: One for each CV, with 2 × 4 replicates (Figure 5-6). Since the CV slicing was maintained, abundance values from each subset were only median normalized. Processing in Perseus v. 1.6.15.0 was performed as described in chapters 5.3.7.1 and 5.3.7.2.

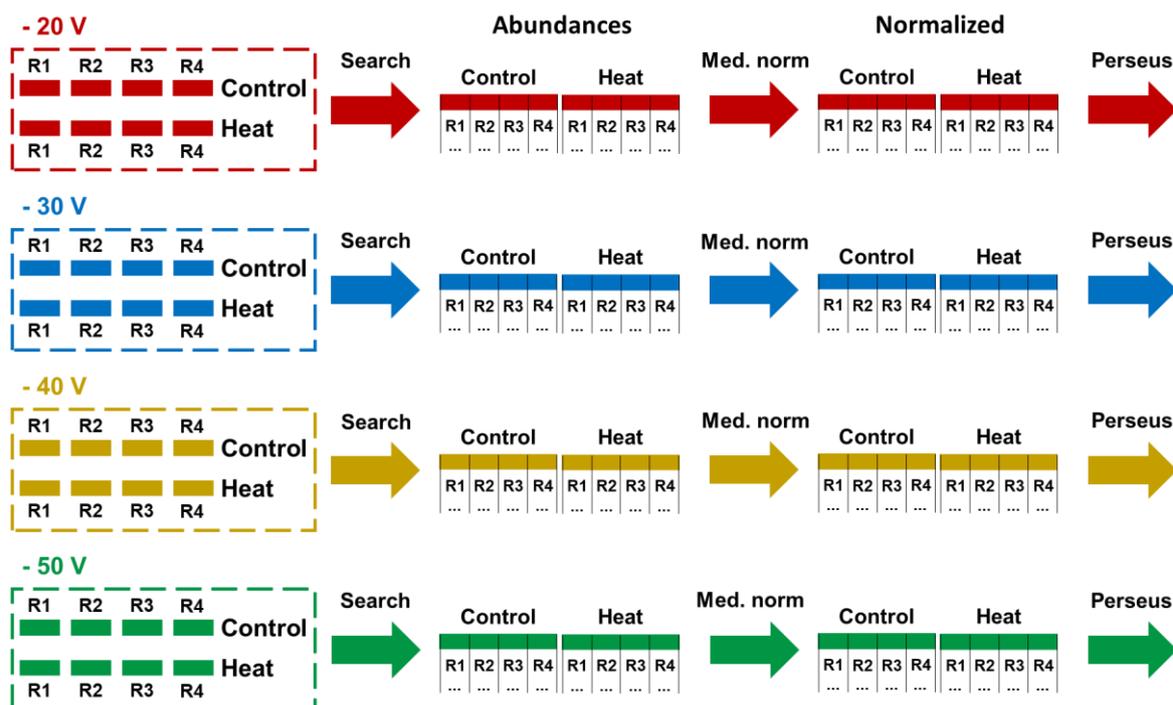


Figure 5-6: TDP data processing workflow with separate database searches per CV. R = replicate, Med. norm = median normalization.

5.4. Results & discussion

5.4.1. Single *C. elegans* TDP LFQ

TDP experiments with single *C. elegans* and the associated data evaluation were performed considering various aspects, which shall be elaborated on in detail in the following subchapters.

5.4.1.1. Challenges & workarounds

In general, sample preparation and data evaluation for DMF-TDP experiments using single *C. elegans* as starting material required dealing with a multitude of challenges.

Due to the absence of an enzymatic reaction step and thus short duration of the protocol, as well as performing all reactions at room temperature, evaporation notably did not prove detrimental to sample handling. It was therefore possible to almost always incubate chips without a closed chamber providing a humidified atmosphere. The only exception was the

SP3 binding step, where an ACN-saturated atmosphere was required to prevent evaporation of the volatile SP3 solvent. This sparing use of humidified chambers lowered the risk of the dielectric being infiltrated by moisture, which can potentially lead to electrode failure (Eydelnant et al., 2012; Leipert & Tholey, 2019). Despite this, storage of the chip at $-20\text{ }^{\circ}\text{C}$ for the elution of proteoforms, in combination with the sudden transition to room temperature, constituted another risk factor in that condensation would form during the subsequent dilution step, generating moisture in the direct vicinity of the electrode array.

In contrast to similar SP3-based BUP workflows, where peptides are resolubilized as proteins are digested, the TDP variety required a designated elution step to elute proteoforms from SP3 beads. This has been achieved in the past with an aqueous solution of 80% FA (Dagley et al., 2019; Doucette et al., 2014). While the use of FA for protein/proteoform resolubilization at room temperature, especially at such high concentrations, is associated with a high risk of formylation events (Beavis & Chait, 1990; Klunk & Pettegrew, 1990), it has been shown that the latter is effectively prevented entirely when maintaining a temperature of $-20\text{ }^{\circ}\text{C}$ during elution and subsequently diluting the eluate by a factor of 10, bringing down the FA concentration to 8% (Doucette et al., 2014).

To adapt this approach for DMF-based TDP, several modifications were inevitable. Due to the spatial constraints of the DMF chips, dilution of resolubilized proteoforms by a factor of 10 was considered impractical. Results from earlier method development (Jan Leipert, Philipp Kaulich, Konrad Winkels, Britta Steer; all AG Tholey) suggested a concentration of only 40% FA to also be sufficient for resolubilizing proteoforms on beads (Leipert et al., 2023). The previously reported concentration of 80% FA during elution was therefore halved, equally reducing the required dilution volume. Additionally, the diluent was 0.01% w/v MNA-C12 instead of water to ensure droplet mobility while keeping the resulting sample compatible with LC-MS/MS. During the elution, cooling was achieved by incubating in a freezer cabinet, providing a constant temperature of $-20\text{ }^{\circ}\text{C}$. For the dilution step, aluminum rods pre-chilled to $-20\text{ }^{\circ}\text{C}$ were placed on the glass top plate of the chip above the samples (Figure 5-2). Effectiveness of cooling during this stage therefore depended on the positioning of the rods, as well as the duration of sample handling, thereby increasing the risk of cooling being inconsistent.

To prevent substantially delaying dilution of FA-containing eluates, MNA-C12 drops were loaded prior to initiating the elution, then moved within close proximity of the samples to

allow easy access. One complication was that MNA-C12 solution, at the concentration employed, was prone to freezing on-chip during the elution step, causing minor delays upon resuming sample treatment while the MNA-C12 solution thawed, and possibly damage to the hydrophobic coating of the top plate and dielectric due to the previous formation of ice.

The multi-CV TDP dataset could not be analyzed quantitatively by Proteome Discoverer since the Feature Detector node available at the time was unsuitable for handling multi-CV datasets. As a workaround, the original data was split into individual files for each CV, while maintaining grouping by treatment condition, with the intention of joining the separated data back together after the search. Different approaches to processing this data outside of Proteome Discoverer were then investigated for their influence on quantification.

5.4.1.2. Re-joining CV-sliced data

In the original approach to evaluating this dataset, the CV-sliced data was subjected to a database search grouped only based on treatment of the original nematodes (“Control” and “Heat”), while completely disregarding replicates within each group, thereby effectively creating a search with 2×16 replicates (see also Figure 5-3). This method led to 925 proteoforms being identified, of which 577 were quantifiable (i.e. featured valid abundance ratios between the two groups).

Prior to the actual quantitative evaluation, the data was subjected to a pre-treatment using Microsoft Excel. First, the CV-slicing was reversed by adding the raw, single-CV abundance values back together per original replicate. Then, the raw abundance values were also subjected to a median normalization as described in chapter 4.3.12. A critical aspect to this was in which order these two steps – addition of abundances and median normalization – were to be performed, and two variants of the process (median normalization first, then addition, or vice versa) were compared for this reason (see also Figure 5-4).

From a quantitative point of view, it indeed played a role whether the CV-sliced abundance values were median-normalized or added together per replicate first, as shown in Figure 5-7a and b. Following the aforementioned pre-treatment of the data in Excel, both evaluations continued in the same fashion, with the data being Log_2 -transformed, grouped, filtered for missing values and subjected to a two-sample Student’s *t* test.

Missing value filters and *t* test parameters were notably identical for both evaluations of the dataset presented in Figure 5-7, ensuring identical numbers of data points as a statistical foundation. Despite this, only five proteoforms, all of which more abundant in the control group, showed up as significant when CV-separated data was median-normalized *before* adding up abundances (Figure 5-7a). Furthermore, the statistical test in this scenario appeared to be biased towards the control group in general, observable as a shift to the left in the resulting plot. Conversely, 24 proteoforms in total were significant when the order of data pre-treatment steps was reversed, with 15 proteoforms being more, nine less abundant after heat treatment (Figure 5-7b), and the plot was observed to retain the expected centered distribution. Additionally, comparing values of $-\text{Log}_{10}(\text{p-value})$ for significant proteoforms between both approaches suggested a markedly lower significance threshold when abundance values were median-normalized after being summed up.

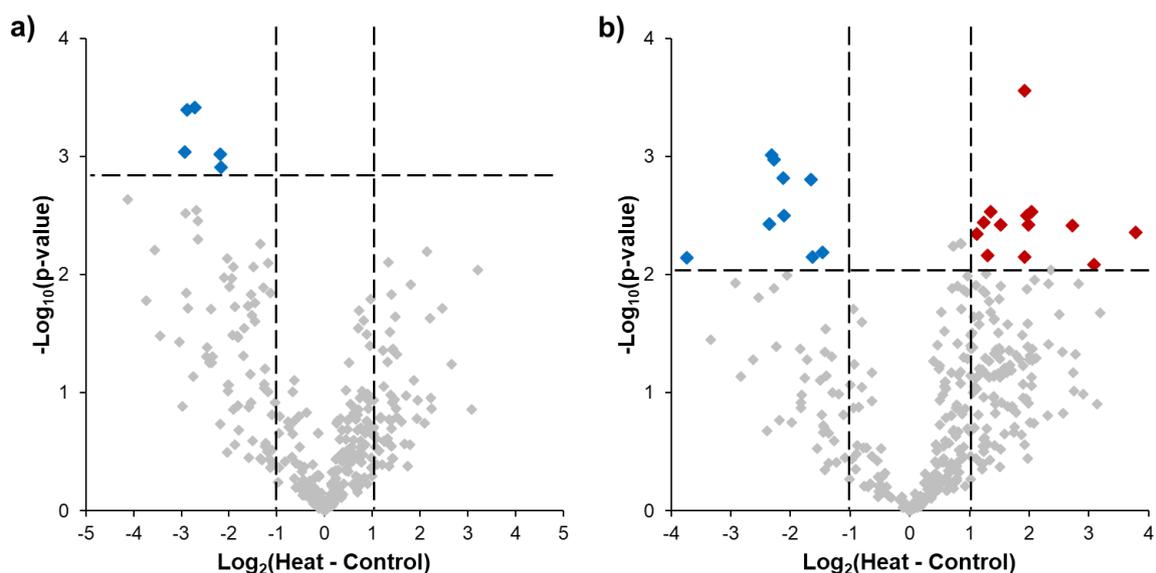


Figure 5-7: Volcano plot of quantified proteoforms with at least three valid values per group and no imputation. The original raw data was searched as separate files for each CV. Abundance values were then added together based on the original replicates, with a median normalization performed either before (a) or after the addition (b).

Since the splitting of the original files by FAIMS CV had been meant to be reversible from the start, the possibility was also considered to treat the separated data as fractions of the original replicates, to which they were assigned prior to the database search (see also Figure 5-5). This method made it possible to reduce the resulting amounts of proteoform abundance values back to 8 (from 32), while still circumventing compatibility issues with the ProSight Feature Detector node by keeping the input files separate. Searching with

these settings again led to 925 total proteoform identifications; however, the number of quantifiable ones increased to 761.

While data processed in this manner yielded a similar picture following a statistical test as the purely unfractionated and re-added data had done, a noteworthy difference was that no proteoforms were considered significant in the case of the data treated as fractions (Figure 5-8). For that reason, a significance threshold could not be estimated following this data evaluation approach.

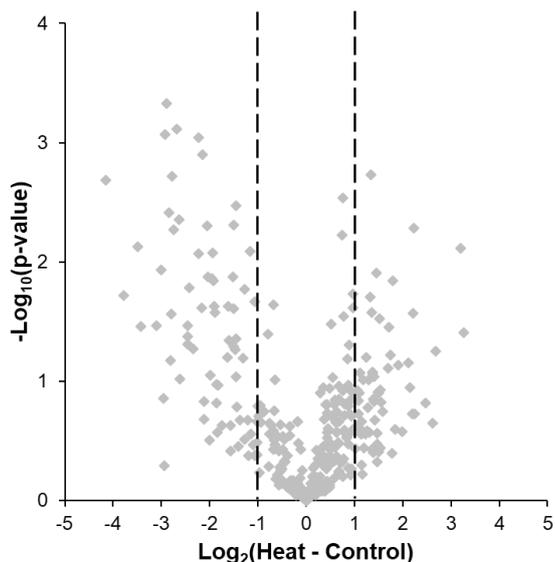


Figure 5-8: Volcano plot of proteoforms quantified originally after a single fractionation, where data for individual CVs had been assigned to replicates as fractions within Proteome Discoverer 2.5. Proteoforms had at least three valid values per group, and no imputation was performed.

5.4.1.3. Individual evaluations using CV-separated subsets

Since questions regarding the correct approach to the evaluation of the TDP dataset remained, especially in the light of the conflictive results discussed in chapter 5.4.1.2, data for each CV was also searched separately to provide another comparison as a basis for the final quantitative evaluation (see also Figure 5-6). The resulting subsets, despite originating from the same multi-CV measurements, were thus completely unrelated to each other for the purpose of the database search. An expected risk of directly comparing single-CV subsets in this manner was the possibility of introducing bias, since internal CV-stepping could have indirectly led to grouping of proteoforms by their molecular weight (Fulcher et al., 2021; Kaulich et al., 2022). Additionally, depending on the proteoform yield per CV, individual subsets would possibly end up with numbers of data points low enough

to be detrimental during statistical tests. Despite the aforementioned risk, this approach also made it possible to narrow down the CV range in this experiment within which proteoforms were the most likely to be isolated. Information could also be gained on the mass range representing the bulk of proteoforms detected per CV.

Independent database searches of subsets separated by CV resulted in highly variable yields in terms of proteoforms with a significant abundance difference between the two groups, as well as proteoforms quantifiable in general (Table 5-1 and Figure 5-9). Taking into account only proteoforms with 3 or more valid abundance values per group, the highest number of entries was observed at a CV of -40 V (253 proteoforms). At -50 and -30 V, lower numbers of proteoforms remained (125 and 114, respectively), whereas a CV of -20 V only yielded 54 proteoforms when applying the same filter settings. The data displayed a notable trend towards proteoforms in a relatively low mass region ranging from approximately 3 to 10 kDa, the majority of which were observed at CVs of -40 and -50 V and thus contributed heavily to the relatively high numbers of proteoforms obtained under these conditions. Conversely, proteoforms quantified at CVs of -20 and -30 V predominantly had masses of approximately 10 to 20 kDa, with only one considered significant (Transcription factor BTF3 homolog; accession Q18885).

Table 5-1: Number of proteoforms per CV with at least three valid values per group, as well as the numbers of associated proteoforms with significant abundance differences.

FAIMS CV in V	Proteoforms after filter	More abundant in "Heat"	More abundant in "Control"
-20	54	0	1
-30	114	0	0
-40	253	0	7
-50	125	9	17

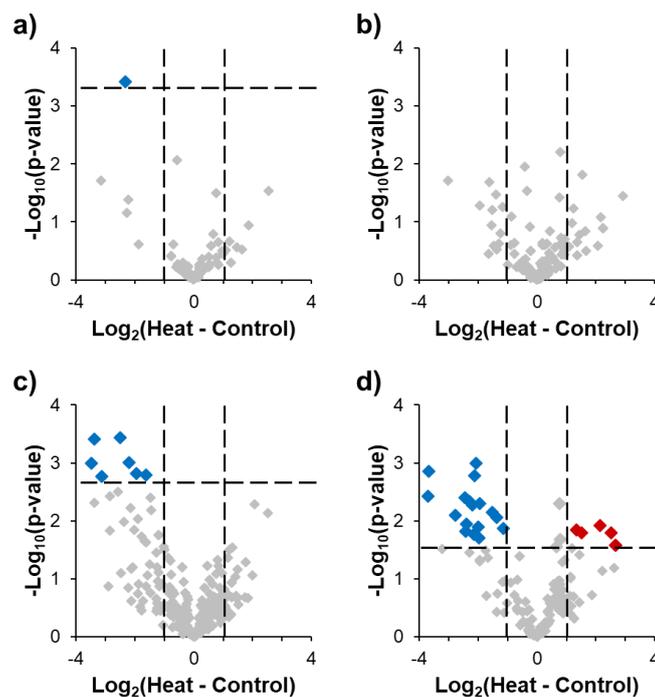


Figure 5-9: Volcano plots of proteoforms per FAIMS CV, including only those proteoforms with at least three valid values per group. No imputation was performed. The plots depict the results for -20 V (a), -30 V (b), -40 V (c) and -50 V (d). Adapted from (Leipert et al., 2023). Use licensed under CC-BY-NC 4.0 (<https://creativecommons.org/licenses/by-nc/4.0/>). Copyright 2023 The Authors.

Of the proteoforms considered showing a significant abundance difference, the majority was observed to be more abundant in the control group following heat stress (see Figure 5-9 c and d), and the single largest group of proteoforms among the latter could be attributed to ribosomal proteins.

5.4.1.4. Comparing different means of normalization

With respect to the limited number of data points obtainable from the TDP experiment at hand, particular attention was also paid to the effect of various methods of normalization. For the evaluation presented here, raw abundance values were exported from Proteome Discoverer v. 2.5.0.400 and subjected to a median normalization externally using Microsoft Excel. Since Proteome Discoverer v. 2.5.0.400 also readily delivers normalized abundance values of its own, this scenario was compared to the external median normalization. The final case investigated was that of no normalization being performed at all. For visualization, abundance values were Log_2 -transformed after each normalization and plotted as a histogram (Figure 5-10).

Overall, the different normalization approaches appeared to have markedly different effects. Compared to the non-normalized data, the abundance histogram after median normalization appeared nearly identical, presumably with a minor correction towards lower abundance values. Contrarily, the normalization applied by Proteome Discoverer appeared to induce a marked shift towards higher abundance values (Figure 5-10).

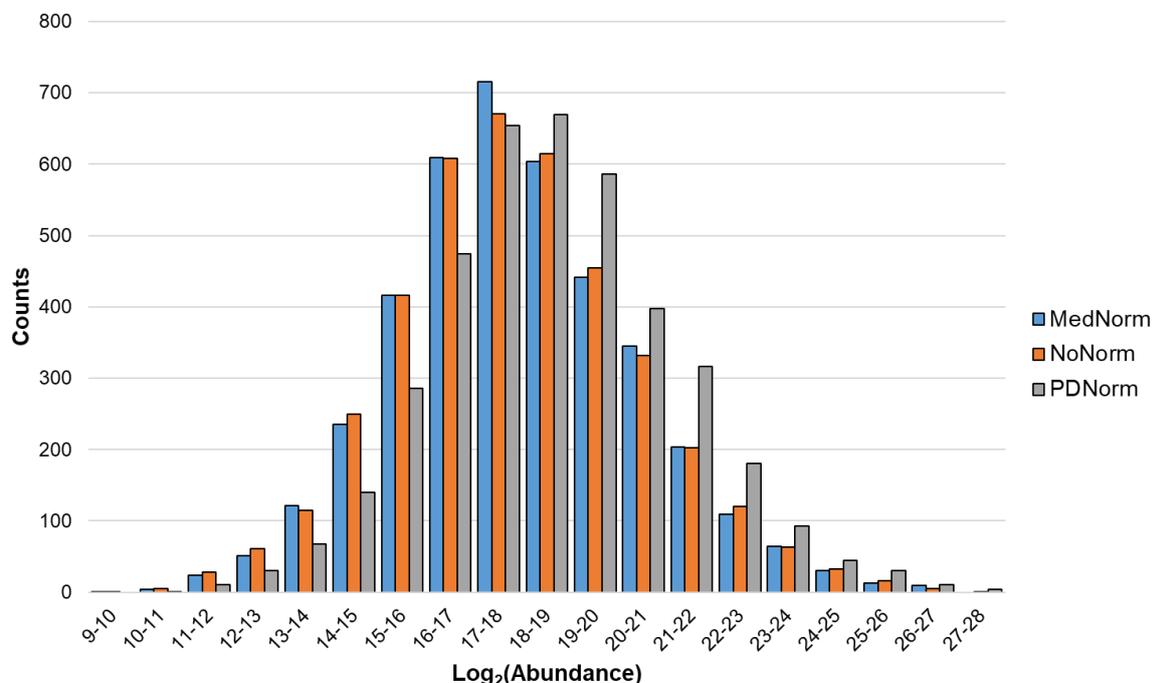


Figure 5-10: Histograms of Log₂-transformed abundance values of the TDP dataset of single *C. elegans* after a search as fractions. Histograms are shown for a manual median normalization of raw abundances (MedNorm), values as normalized by Proteome Discoverer (PDNorm), or not normalized at all (NoNorm).

This comparison highlighted the potential for various normalization approaches to have different effects, depending on the available data. In any case, the results shown here were interpreted as confirmation of median normalization as a standard method, due to the only minor correction tendency observed.

5.4.1.5. Proteoforms vs. inferred proteins

An inherent limitation of BUP is that proteins can only be inferred from peptides at the cost of losing proteoform-related information. Substantial amino acid sequence portions are usually shared among multiple co-existing proteoforms, or even all of them. Generation of

peptides from such shared regions therefore leads to the loss of any information on the original proteoforms, since these peptides cannot be assigned unambiguously.

An example for this scenario from this particular TDP experiment was the 40S ribosomal protein S28. In the data subset of a FAIMS CV of -50 V, not only were five proteoforms of this canonical protein found to be differentially abundant following exposure to heat stress, but the abundance differences observed also turned out to be conflictive. Most notably, only one of these proteoforms, effectively representing the canonical sequence with a Met1 acetylation, was more abundant after this treatment, whereas the four other proteoforms were exclusively less abundant following the heat stress event. These four latter proteoforms were notably almost completely unmodified except for different degrees of N- and possibly C-terminal truncation, with all of these proteoforms lacking the canonically C-terminal residue Arg65.

Next, these TDP results for the 40S ribosomal protein S28 at proteoform level were compared to BUP results for the same protein. Precisely, tryptic peptides generated during the study described in chapter 4 of this work (Steinbach et al., 2022) were compared to the proteoforms discussed in this chapter in terms of their positioning within the canonical amino acid sequence.

In total, seven peptides were included in this comparison. Being fully tryptic, the sequence portions they were derived from made them entirely unspecific concerning the various proteoforms (Figure 5-11). This became especially apparent with regard to the N termini (where the proteoforms discussed bore the highest amount of variation). In ascending order, N termini of proteoforms shown in Figure 5-11 were at Met1, Leu6, Ala7, Val9 and Lys11; conversely, the least advanced peptide N terminus was at Thr16. For this reason, tracing the origin of any peptide to a particular proteoform based on the degree of N-terminal truncation was impossible in this case. Hypothetically, specific peptides in this scenario would have had to be semitryptic and originating from either canonical terminus of the protein, in which case assignment to the acetylated proteoform (highlighted red in Figure 5-11) would have been possible.

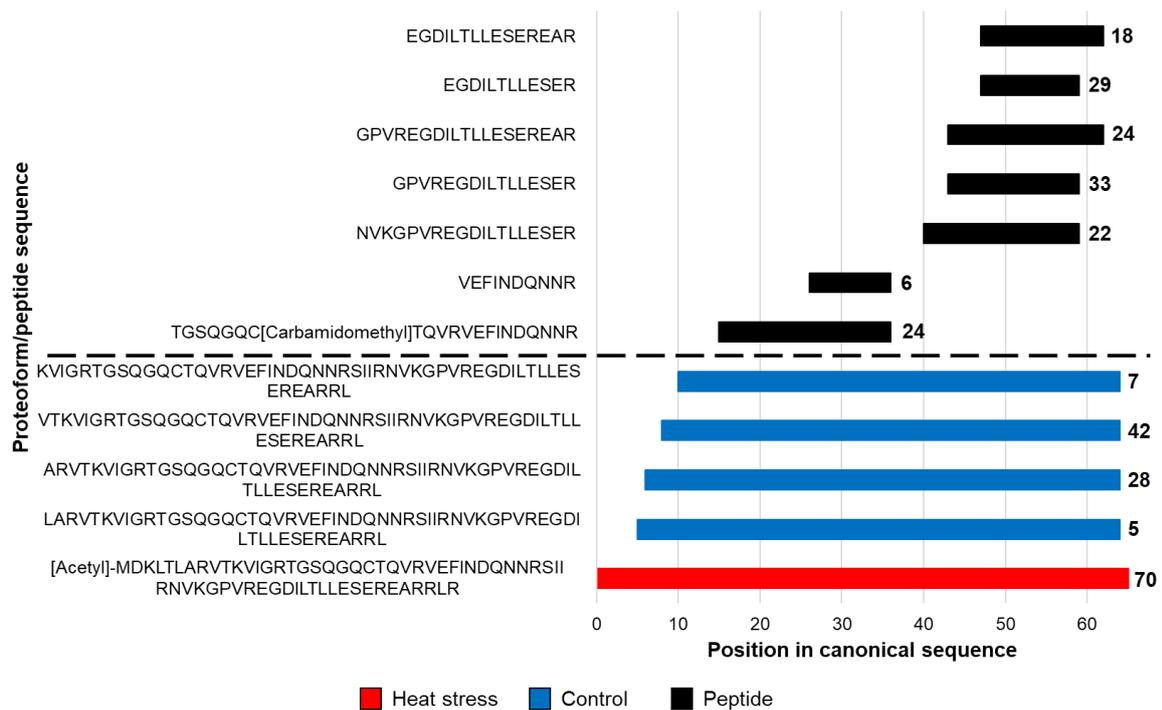


Figure 5-11: Proteoforms of the 40S ribosomal protein S28 found to be significantly more abundant in the control (blue) or heat stress group (red) at a FAIMS CV of -50 V. Peptides detected in an earlier DMF-BUP experiment (black) were notably unspecific and could not be assigned unambiguously to any proteoform. Shown to the right are the respective numbers of PSMs or PrSMs. Reproduced from (Leipert et al., 2023). Use licensed under CC-BY-NC 4.0 (<https://creativecommons.org/licenses/by-nc/4.0/>). Copyright 2023 The Authors.

Given that conflictive abundance differences observed at the proteoform level could not be resolved at peptide level for the reasons stated, it was hypothesized that the simultaneous presence of peptides derived from a single canonical protein, but a multitude of proteoforms, would have obscured abundance differences of individual proteoforms, placing the inferred protein well below the difference and significance thresholds. The comparison of the TDP data presented in this chapter with the BUP data published earlier showed this to indeed be the case for the 40S ribosomal protein S28 discussed here (Figure 5-12).

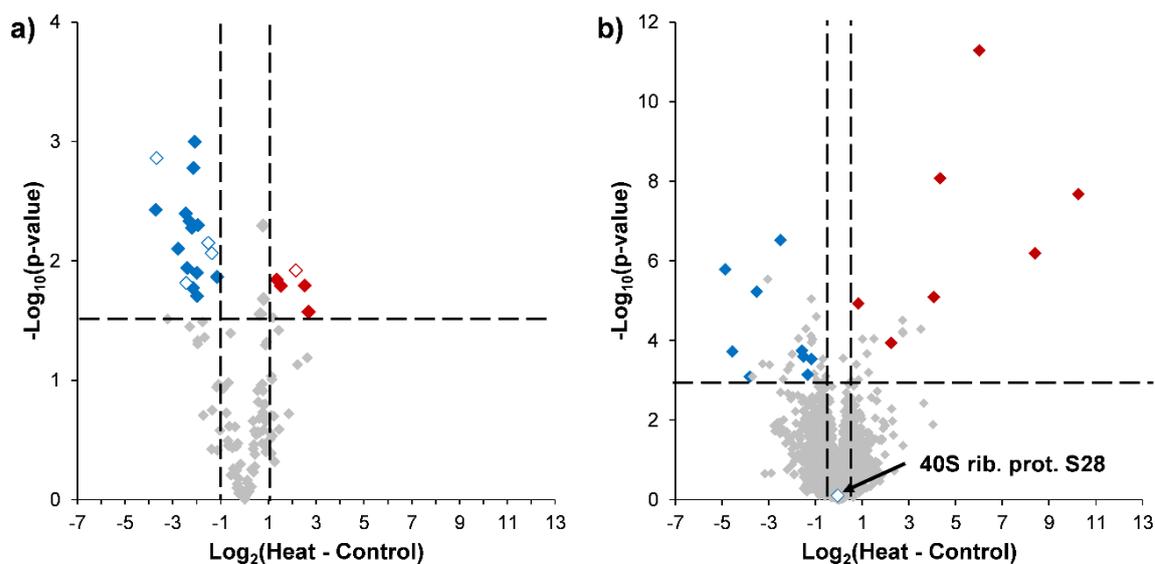


Figure 5-12: Comparison of TDP with BUP results: a) Significant proteoforms in the -50 V CV TDP subset, with the proteoforms of 40S ribosomal protein S28 highlighted as open diamonds. Reproduced from (Leipert et al., 2023). Use licensed under CC-BY-NC 4.0 (<https://creativecommons.org/licenses/by-nc/4.0/>). Copyright 2023 The Authors. b) Quantified inferred proteins in the DMF dataset generated for the previously described BUP study (Steinbach et al., 2022), highlighting proteins discussed there. Note the position of the inferred 40S ribosomal protein S28 almost at position (0|0). Adapted with permission from (Steinbach et al., 2022). Copyright 2022 American Chemical Society.

In general, a striking observation when comparing BUP and TDP results was the lack of overlap between proteoforms and inferred proteins considered significant, which was a recurring theme with both TDP data rejoiner after the database search, as well as CV-separated subsets. Interestingly, while a scenario as demonstrated here for the 40S ribosomal protein S28 provides an explanation as to why multiple proteoforms with significant abundance differences would not correspond to such a difference on the inferred protein level, other examples of this kind were not observed throughout the entire TDP dataset. Notably, this must also be viewed in the context of the limited amount of starting material per sample and long cycle times during the multi-CV FAIMS method, as well as the currently still notoriously limited sensitivity of TDP. Therefore, one may argue that further examples of the observation described would likely be encountered in a bulk TDP experiment, especially after further optimization of the number of CVs and/or cycle time.

Related to the same confounding factors, another notable limitation of the TDP analysis presented here was the impact of missing value filtering. Due to the lower overall sensitivity in detecting proteoforms compared to peptides, as well as possibly suboptimal

deconvolution, proteoforms quantified were more likely to miss valid abundance values than proteins inferred from peptides. As a result, prior to any filtering, the CV-separated subsets of the TDP analysis were comprised of missing values in a range from ~22 to ~42%, depending on the subset. This did not only severely hamper any statistical tests, but also disqualified imputation of missing values as an acceptable solution (Bennett, 2001).

As a compromise, only those proteoforms were considered for statistical tests that were represented by at least three valid abundance values in both groups. While this filter reduced the missing value rate to ~5% in every subset, making the execution of statistical tests possible even without imputation of the remaining values, it came at the cost of disregarding up to ~50% of the initially quantified proteoforms. This notably led to the exclusion of any proteoforms with abundance values present only in one group. In contrast, the DMF- and in-tube BUP datasets consisted of missing values to the same ~5% even at the outset, enabling statistical tests without disregarding any inferred proteins.

Following the evaluation of the significant proteoforms per CV, the investigation was therefore extended to also take into account proteoforms with a one-sided distribution of valid values (i.e. valid values in only one group). As an additional threshold, the respective group had to contain valid values from at least three (out of four) replicates, to confirm the presence of the proteoform in question within the respective group. The aim here was to isolate any proteoforms with abundance differences between groups sufficiently pronounced for valid values to be missing from one group entirely, thereby attempting to compensate for the elimination of the affected proteoforms from statistical tests.

Most notably, the majority of proteoforms only represented by valid values in samples from heat-treated *C. elegans* turned out to be of heat shock proteins, which was observed for all CVs except -50 V. This confirmed the organisms' response heat stress already observed with BUP data, while also explaining the absence of heat shock proteins in the quantitative evaluation of the TDP dataset. In the case of proteoforms quantified exclusively in the control group, a similar trend was observed for ribosomal proteins, which represented the majority of control-exclusive proteoforms in the CV -40 and -50 V subsets.

5.5. Conclusion

In this subproject, the compatibility of DMF with state-of-the-art TDP sample preparation methods could be demonstrated, highlighting once again the versatility of the DMF technology, as well as its potential in terms of enabling low-input TDP.

Quantitative evaluation of TDP data showed a host of proteoforms to be of significantly different abundance in both groups. Furthermore, significant proteoforms could frequently be assigned to the same accession. In one case, this included proteoforms with conflictive abundance differences, thereby highlighting the key strength of TDP, as well as a crucial area where BUP lacked.

Aside from the obvious advantages, the subproject also shone a light on the inherent limitations of quantitative TDP, especially with regard to low-input samples, with the relatively low number of data points and associated high impact of missing values perhaps being the most noteworthy factors. Statistical tests and graphical representations thereof (Volcano plots) relied on the majority of values being valid, a condition achievable only through strict filtering for missing values at the cost of losing a host of proteoforms with a particularly one-sided distribution between groups. Here, BUP provided an invaluable fallback due to the sheer size of the dataset mitigating the effect of missing values.

To conclude, the data presented not only showed examples for the different advantages of TDP compared to BUP, but also underlined the potential of both approaches when used complementarily.

6. Method development for microfluidics-supported MS-based Middle-Down nanoproteomics

6.1. Introduction

BUP and TDP are well-known yet inherently different concepts in MS-based proteomics with their own individual strengths and weaknesses. The approach of MDP aims to bridge the gap between the former two principles by generating peptides substantially larger (> 3 kDa) than is common in BUP (typically 0.7-3 kDa). Means to do so include enzymatic reactions with proteases hydrolyzing at relatively rare amino acid residues, as well as chemical cleavage of peptide bonds such as by microwave-accelerated acid hydrolysis (Cristobal et al., 2017; Garcia, 2010). With MDP thus offering a compromise between analyte size and proteoform information, this approach is an attractive target for miniaturization using DMF.

6.2. Project aims & major experiments

6.2.1. Fine tuning of critical workflow elements for low-input samples

The first stage of this subproject pursued the goal of developing a DMF-based sample preparation method for low-input MDP, which initially required preparations of CaCo-2 cells as samples consisting of approximately 1000 cells. Using these samples, the objective was to tune individual aspects of the workflow and obtain LC-MS/MS data as feedback comparable to later analyses. The ultimate aim was to finalize the sample preparation itself based on complete MS/MS datasets to allow the reliable generation of Middle-Down peptides in a favorable size range (3-10 kDa) from small sample amounts such as single *C. elegans* nematodes.

6.2.2. Optimization of LC-MS/MS conditions

After establishing a viable sample preparation readily suitable for analyses, the next objective was to optimize the downstream separation and detection process. Here, different adjustments to LC, FAIMS and MS/MS conditions were made with the intention of optimizing sensitivity during measurements of low-input samples, and the goal of finalizing the LC-MS/MS method to be used for subsequent quantification experiments.

6.2.3. Label-free quantification

The final objective of this subproject was to employ the newly developed and optimized DMF-TDP method for a quantitative analysis of single *C. elegans* in two differently treated groups, with the objective of generating a new MDP dataset for LFQ, similar to the BUP and TDP experiments described in chapters 4 and 5 of this thesis. Eventually, this was supposed to allow a rounded comparison between the three different approaches to proteomics.

6.3. Materials & methods

6.3.1. Chemicals & reagents

Acetonitrile (ACN; Chromasolv quality), FA and TFA (both LC-MS-grade), were obtained from Honeywell Specialty Chemicals Seelze GmbH (Seelze, Germany), ethanol (absolute) from Th. Geyer GmbH & Co. KG (Renningen, Germany).

Mannitol-based amphiphile C12 (MNA-C12) was from Anatrace Inc. (Maumee, OH, USA).

Sequencing-grade Glu-C was from Promega GmbH (Walldorf, Germany).

All other chemicals and reagents, including SeraMag Speed Beads (hydrophilic and –phobic, carboxylate-modified), were from Sigma Aldrich (Steinheim, Germany).

Hydrophilic and –phobic SP3 beads (both as 50 µg/µL suspensions in 0.05% w/v azide) were always mixed to equal parts, washed twice with milli-Q-water, then resuspended at the desired concentration in 0.08% w/v Pluronic F-127.

6.3.2. Cultivation of CaCo-2 cells

CaCo-2 cells (ATCC number HTB-37™) were obtained from CLS Cell Lines Service GmbH (Eppelheim, Germany). Cells were grown at a temperature of 37 °C in a humidified atmosphere containing 5% CO₂. The cultures were contained in Nunclon-coated T75 flasks to promote the formation of an adherent culture. The nutrient medium was RPMI-1640 medium supplemented with 25 mmol/L HEPES, 2 mmol/L L-glutamine, 0.013 mmol/L phenol red, 10% FCS and 1% Pen/Strep. Passaging was performed when the cells had formed a confluent layer covering approximately 90-100% of the flask surface. Culture bottles were first inspected visually for signs of contamination. This was followed by decanting of the medium and washing of the adherent cells with 25 mL 1 × PBS (37 °C). After decanting the PBS, 2 mL of TrypLE dissociation reagent (37 °C) were added to promote detachment of the cells from the coated surface, for which step the culture was incubated for 8 min at 37 °C and 5% CO₂.

10 mL of medium (37 °C) were added to the resuspended cells, which were transferred to a sterile 15 mL Falcon tube (Greiner Bio-One GmbH, Frickenhausen, Germany). The cells were then centrifuged for 5 min at 200 × g and ~25 °C. After removing the supernatant, the cells were resuspended in 10 mL of fresh medium. An aliquot of cell suspension was then diluted 1:5 in Trypan Blue for counting by means of a Neubauer haemocytometer. After establishing the cell concentration, a new Nunclon-coated T75 flask was seeded with 2 × 10⁶ cells and filled up to 25 mL, with the remaining volume consisting of fresh medium. The new subculture was then incubated under the same conditions as before the passage.

The remaining cell suspension was centrifuged under the same conditions as before and the supernatant removed. Afterwards, the cells were washed three times with 1 × PBS (37 °C), the volume of which was matched to the original value every time. After the third resuspension in PBS, the cells were split into aliquots containing 3-4 × 10⁶ cells, then centrifuged. The supernatant was removed and the cells stored as pellets at -70 °C.

6.3.3. Preparation of CaCo-2 cell samples

CaCo-2 cells stored as pellets at -70 °C were resuspended in 0.08% w/v Pluronic F-127 to a concentration of ~333 cells/μL. Three μL of cell suspension were placed in 0.2 mL PCR tubes (Eppendorf SE, Hamburg, Germany). Three μL lysis buffer (composition as described in chapter 5.3.3, replacing borate with ammonium bicarbonate [AmBiC] buffer)

were then added and mixed with the samples by pipetting. Lysis and protein reduction were performed for 50 min at room temperature (approximately 20 °C).

Three μL alkylation buffer (30 mmol/L IAA, 0.1% w/v Tetracarb 90R4, 75 mmol/L AmBiC) were mixed into the samples by pipetting, before incubating for 50 min at room temperature and in the dark.

Two μL SP3 beads (10 $\mu\text{g}/\mu\text{L}$ in 0.08% w/v Pluronic F-127) were added to the samples, followed by 44 μL ACN. Mixing was achieved by slow vortexing, with the samples then left for 30 min at room temperature. After removing the supernatant, samples were washed with 100 μL 70% ethanol and 50 μL ACN, respectively.

Two μL digest buffer (5 ng/ μL Glu-C, 0.01% w/v MNA-C12, 25 mmol/L AmBiC) were added to the bead pellets before mixing by pipetting, with the subsequent digestion step performed at room temperature (default duration: 1 h). Afterwards, the beads were immobilized and the supernatant transferred to 100 μL LC glass inserts (Th. Geyer GmbH & Co. KG, Renningen, Germany) containing 1 μL 3% v/v FA.

Two μL 40% v/v FA (pre-chilled to $-20\text{ }^{\circ}\text{C}$) were added to the bead pellets. Samples were then incubated for 15 min at $-20\text{ }^{\circ}\text{C}$. Eight μL MNA-C12 (0.01% w/v) were added immediately afterwards, the beads immobilized and the eluate reunited with the digest fraction.

This protocol represents a default workflow, which was used for LC-MS/MS method development (chapter 6.3.6), as well as for a series of experiments where adjustments to individual conditions were made as required (see also Figure 6-1):

- Experiment 1: Glu-C digestion time was varied to test different durations (0, 5, 15, 30, 45, 60 and 120 min). Additionally, Glu-C was left out entirely for the 0 min time. All other workflow parameters were kept as described.
- Experiment 2: Different FA concentrations for the elution step were tested (0, 10, 20, 30 and 40% v/v), while otherwise maintaining default conditions (including dilution volumes after the elution step). The eluent for the 0% v/v FA condition was plain milli-Q-water.
- Experiment 3: Default conditions (Glu-C digestion for 60 min at approximately $20\text{ }^{\circ}\text{C}$ and elution with 40% v/v FA) were applied, but samples divided into digest and eluate fractions, rather than merging these.

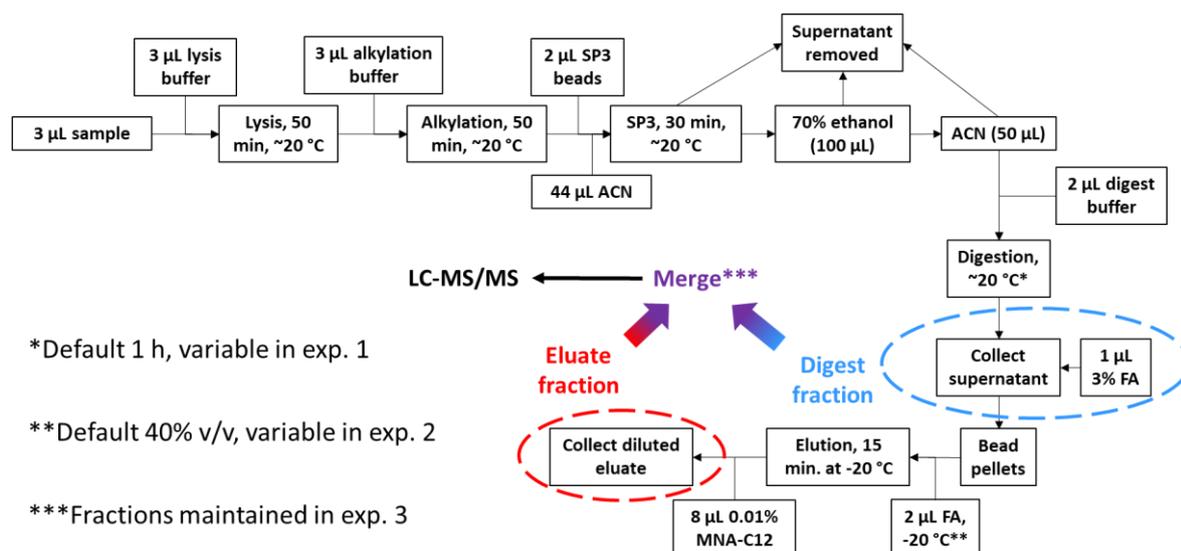


Figure 6-1: Flow chart of the in-tube low-input MDP sample preparation workflow used during method development, highlighting alterations made for individual experiments (abbreviated as exp.). The numbering scheme of the experiments is identical to that used in chapter 6.3.3.

6.3.4. *C. elegans* preparation and handling

C. elegans were raised and prepared by Christine Blurton (AG Leippe) at the Zoological Institute of the *Christian-Albrechts-Universität zu Kiel*.

Throughout the subproject, two different batches of nematodes were used as samples. Preliminary experiments for method development and optimization of reaction conditions relied on adult specimens of the Bristol N2 strain, which were grown on NGM agar populated with an *E. coli* OP50 lawn to provide the nematodes with nutrition.

C. elegans used for LFQ experiments were grown in a synchronized population until the L4 stage of development under default conditions; then, one subpopulation was incubated for about 1 h at 33 °C based on a recently published method (Golden et al., 2020). Single nematodes were then isolated in 0.2 mL PCR tubes (Eppendorf SE, Hamburg, Germany) immersed in 1 µL of M9 buffer, and flash-frozen in liquid nitrogen. To improve stability of the sample material, the single specimens were transported on dry ice and stored at -80 °C.

6.3.5. Preparation of single *C. elegans* using DMF

DMF sample preparation was performed using the DropBot DB3-120 device and respective chips. Prior to experiments, individual *C. elegans* frozen in 1 μL of M9 buffer were thawed, resuspended in 50 μL 0.08% w/v Pluronic F-127, and stored on ice until commencing sample preparation. Single nematodes were aspirated in 1 μL of liquid, loaded onto actuated reservoir electrodes, and then moved onto the adjacent transfer electrodes. One μL of lysis buffer (see chapter 6.3.3) was added to the sample drop, before briefly mixing by movement across several actuation electrodes. The electrodes were turned off and the chip incubated for 50 min at room temperature for the lysis and protein reduction step. One μL of alkylation buffer (30 mmol/L IAA, 0.1% w/v Tetrionic 90R4, 75 mmol/L AmBiC) was added to the samples and the chip protected from light for the 50-min alkylation step at room temperature.

Two μL SP3 bead suspension (10 $\mu\text{g}/\mu\text{L}$ in 0.08% w/v Pluronic F-127) per sample were added away from the sample drops, the beads immobilized and the supernatant removed. The samples were merged with the bead drops and the beads resuspended by mixing via an automated route (Repeats: 20, Voltage: 90 V, Frequency: 8 kHz, Duration: 2500 ms, Trail length: 2). 4.5 μL ACN were then added to each sample and mixed with it by manual electrode actuation. The chip was then placed in a container with an ACN-saturated atmosphere and incubated there for 30 min at room temperature. Afterwards, the beads were immobilized and the supernatant removed. Bead pellets were then washed with 4.5 μL 70% ethanol and 100% ACN, respectively.

Two μL digest buffer (5 ng/ μL Glu-C, 0.01% w/v MNA-C12, 25 mmol/L AmBiC) were added to each sample, followed by mixing using manual electrode actuation. Samples were incubated on-chip for 1 h at room temperature while leaving the chip in the open. Afterwards, 1 μL of 60% v/v FA ($-20\text{ }^\circ\text{C}$) was added to the samples, briefly mixed, and the chip immediately transferred to a freezer cabinet, where it was left for the 15-minute elution step at $-20\text{ }^\circ\text{C}$.

Following elution, the chip was reconnected to the DropBot and 4 μL 0.01% w/v MNA-C12, pre-cooled on ice, were added to each sample via the original reservoir electrodes. Samples were then retrieved using 20 μL GELoader tips (Eppendorf SE, Hamburg, Germany) and transferred to 100 μL LC glass inserts already containing 3 μL of water, for a final volume of approximately 10 μL . Constant cooling was applied throughout this procedure by placing an aluminum rod chilled to $-20\text{ }^\circ\text{C}$ on the top plate (see also Figure 5-2).

The sample preparation described here was performed multiple times on separate DMF chips until all desired replicates had been obtained. In total, eight replicates per condition were measured and evaluated, with replicates for both conditions prepared on five chips, respectively.

6.3.6. LC-MS/MS: Method development

MDP measurements of CaCo-2 digests were done by injecting complete samples on an Ultimate 3000 RSLCnano UHPLC system, with a PepMap C₄ trap cartridge (5 mm × 0.3 mm, 5 μm, 300 Å, Thermo Fisher Scientific GmbH, Bremen, Germany) and an Accucore C₄ analytical column (50 cm × 75 μm, 2.6 μm, 150 Å, Thermo Fisher Scientific GmbH, Bremen, Germany) used for chromatographic separation. The LC setup was coupled to an Orbitrap Fusion Lumos tribrid mass spectrometer equipped with a FAIMS Pro interface (Thermo Fisher Scientific GmbH, Bremen, Germany). Eluents were 0.05% v/v FA (eluent A) and 80% ACN with 0.04% FA (eluent B). The nanoLC flow rate was set to a constant 300 nL/min.

The starting flow composition was 4% eluent B, which was maintained for 5 min to achieve desalting. This was followed by an increase to 15% B over 2 min. A gradient then commenced to increase the ratio of eluent B from 15 to 50% over a time of 120 min. Within 2 min, the proportion of B was raised to 90%, which was then maintained for another 11 min, before eventually equilibrating for 10 min to restore the starting conditions.

To enable ESI, columns were equipped with fused silica emitters (outer diameter: 360 μm, inner diameter: 20 μm, tip length: 10 μm, total length: 10 cm; CoAnn Technologies, Richland, WA, USA).

FAIMS was used with internal CV stepping, a procedure described previously (Kaulich et al., 2022). This involved four scan cycles with CVs of -35, -45, -55 and -65 V, using a cycle time of 3 s.

MS acquisition was performed in the Orbitrap over a scan range from *m/z* 400 to 1800 with a resolution of 60 000, an AGC target of 200%, a maximum injection time of 118 ms, and a microscan number of 1. For the CV of -35 V, a resolution of 120 000, a microscan number of 2, and a maximum injection time of 246 ms were chosen.

Data-dependent MS/MS (ddMS²) was also performed in the Orbitrap, using CID as the activation mode with a normalized collision energy of 25% and scanning a range from *m/z* 150 to 2000. Charge states included were +4 to +50 for CVs of -35 and -45 V, and +3 to +50 for CVs of -55 and -65 V. Other parameters used were a resolution of 50 000, a maximum injection time of 200 ms, a normalized AGC target of 800% and a microscan number of 2. For the CV of -35 V, the maximum injection time was changed to 250 ms, the Orbitrap resolution to 60 000, and the microscan number to 4.

This method constituted the default version used throughout the MDP workflow development process. The following adjustments were made during later optimization:

- Variant 1: The FAIMS CV of -65 V was replaced with -60 V, while retaining all parameters of the former. Additionally, the ddMS² microscan number was set to 2 for CV -35 V.
- Variant 2: Building upon variant 1, the number of MS microscans was set to 1 for CV -35 V. The maximum injection time for ddMS² were set to 250 ms for all CVs, and the ddMS² microscan number to 1 (2 for CV -60 V).
- Variant 3 and 4: Based on the default method, LC gradient lengths were adjusted to 180 (variant 3) and 60 min (variant 4), respectively.
- Variant 5: Building upon variant 2, HCD replaced CID as the activation mode for ddMS², using a normalized collision energy of 27% for all CVs.

The numbering scheme introduced here to distinguish different LC-FAIMS-MS/MS methods will also be used for the purpose of discussion (Table 6-6).

6.3.7. LC-MS/MS: LFQ with single *C. elegans*

Measurements of MDP digests of single *C. elegans* for the purpose of LFQ were performed using largely the same instrumental setup and RP-HPLC gradient as described in chapter 6.3.6. One important exception was that the analytical LC column was replaced with a prototype C₄ analytical EIC (25 cm × 75 μm, 3 μm, 100 Å; IonOpticks, Fitzroy, Victoria, Australia). Accordingly, to achieve a constant temperature of 45 °C, the column was placed in a Butterfly Heater (Phoenix S&T, Chadds Ford, PA, USA).

Out of the LC-FAIMS-MS/MS methods described in chapter 6.3.6, variant 2 was chosen as the new default for the purpose of the LFQ sequence.

6.3.8. Data evaluation: General

For evaluations as a part of method development (i.e. without LFQ), raw data was searched in Proteome Discoverer v. 3.0.0.757 with ProSight v. 4.2.0.42 (Proteinaceous, Evanston, IL, USA) either against a *C. elegans*- or human protein database (downloaded on March 1st 2023 and February 6th 2023, respectively) using the ProSight High/High cRAWler node. Both a ProSightPD Annotated Proteoform Search and a ProSightPD Subsequence Search were executed in parallel, setting precursor- and fragment mass tolerances to 10 ppm. Carbamidomethylation of cysteine residues was selected as a static modification.

Despite the raw data consisting of signals of Middle-Down peptides, these searches would always yield numbers of proteins, isoforms, proteoforms and PrSMs, due to the ProSight node originally being meant for TDP. Therefore, any entries within the “proteoform” category would in fact be Glu-C peptides (treated as “Middle-Down peptides”).

Qualitative parameters always compared were numbers of proteins, isoforms, Middle-Down peptides and PrSMs. Furthermore, the extent of Glu-C digestion during method development was evaluated based on average and median theoretical peptide masses, numbers of peptides with C-terminal glutamate residues, as well as numbers of annotated (i.e. undigested) proteoforms remaining.

The Python script “Middle Down Analyzer”, devised and developed by Philipp Kaulich (AG Tholey), and available online at <https://github.com/PhilippKaulich/MiddleDownAnalyzer>, was used as a quality control tool during method development, where various factors were considered particularly important to assess method performance. “Potential Cleavage” offered an overview of cleavage events detected within proteoform sequences by plotting the incidence of specific amino acid residues at the C-termini of presumed cleavage products against the incidence of N-terminal residues. This plot could therefore be used to estimate activity and hydrolysis specificity of a particular protease. Another tool to assess the extent of digestion was “Proteoform Information”, which provided the ratio of cases where Middle-Down peptides were internal, located at the N- or C-terminus of the original proteoform, or in fact made up the proteoform itself, the latter representing the scenario of a completely undigested proteoform.

6.3.9. Label-free quantification

Of the eight replicates originally prepared per condition based on chapter 6.3.5, six were eventually used for LFQ. These replicates were chosen based on the number of Middle-Down peptides assigned in the preliminary database search (Table 6-1).

Table 6-1: Overview of proteins, Middle-Down peptides and PrSMs assigned for all single *C. elegans* replicates. Replicates chosen for LFQ evaluation are shaded green.

Sample	Proteins	Middle-Down peptides	PrSMs
C1	385	1082	2771
C2	420	1162	2816
C3	352	928	2400
C4	232	511	1151
C5	421	1134	2955
C6	433	1137	2842
C7	506	1518	3475
C8	19	22	31
H1	426	1168	2915
H2	468	1332	3497
H3	507	1458	3601
H4	194	477	960
H5	346	804	1850
H6	470	1260	3223
H7	407	1093	2625
H8	536	1502	3624

First, raw data files were subjected to CV slicing as described for the quantitative evaluation of TDP data (chapter 5.3.7), leading to 48 total files obtained from an original 12 (2 × 6). The database search was performed in Proteome Discoverer v. 3.0.0.757 with ProSight v. 4.2 (Proteinaceous, Evanston, IL, USA) against an XML database of *C. elegans* (downloaded from Uniprot on March 1st 2023), using the ProSightPD High/High cRAWler and ProSightPD High Resolution Feature Detector node, with the 48 files grouped as “Control” and “Heat”. Both a ProSightPD Annotated Proteoform Search and a ProSightPD Subsequence Search were executed in parallel, with precursor- and fragment mass tolerances of 10 ppm. As opposed to the TDP approach described in chapter 5, carbamidomethylation of cysteine residues was selected as a static modification. Additionally, the thresholds for Minimum Number of Charge States Detected and Minimum Number of Sliding Window Detections were set to 2. In the consensus workflow, the

Feature Presence in Files threshold was set to 10%. Apart from these alterations, default settings were maintained.

Following the database search, Middle-Down peptides (treated as “proteoforms” by ProSight) were filtered to retain only those with a valid abundance ratio between the groups “Control” and “Heat”. Non-normalized abundance values for these peptides were exported to Microsoft Excel and summed up per replicate, with a median normalization then performed on the sums (see also Figure 5-4b). Normalized abundance values were then exported to Perseus (v. 1.6.15.0), where Log_2 -transformation, grouping and valid value filtering were applied in this order. Peptides with at least 4 valid values per group were subjected to a two-sample Student’s *t* test (s0: 0, Number of randomizations: 250, Correction: Permutation-based FDR). For visualization, the results were exported to Microsoft Excel, where $-\text{Log}_{10}(\text{p-value})$ was plotted against the Log_2 -transformed fold change. Difference thresholds were placed arbitrarily at -1 and 1 (corresponding to a 2-fold higher or lower abundance), whereas the significance threshold was placed based on the lowest $-\text{Log}_{10}$ -transformed p-value at which a proteoform was still considered significant.

6.3.10. Comparing protein sizes between MDP and TDP

Important: Having been part of the original output of the ProSight database search, the column “isoforms” was used here for the purpose data evaluation, as well as result presentation and discussion; however, the use of this term should be viewed with caution for reasons later discussed (see chapter 6.4.1).

Isoforms from the MDP data were compared with those directly identified in TDP, regarding molecular weight and number of amino acid residues. For this purpose, the MDP data was first limited to four replicates per group (i.e. “Heat” and “Control”) to match the size of the TDP dataset, choosing those with the highest numbers of Middle-Down peptides, and accordingly, inferred proteins (Table 6-2).

Table 6-2: Proteins, isoforms, Middle-Down peptides and PrSMs for all replicates. Replicates chosen for comparison of isoform size with TDP data are shaded green.

Sample	Proteins	Isoforms	Middle-Down peptides	PrSMs
C1	385	387	1082	2771
C2	420	423	1162	2816
C3	352	354	928	2400
C4	232	232	511	1151
C5	421	423	1134	2955
C6	433	435	1137	2842
C7	506	509	1518	3475
C8	19	19	22	31
H1	426	428	1168	2915
H2	468	470	1332	3497
H3	507	509	1458	3601
H4	194	194	477	960
H5	346	348	804	1850
H6	470	472	1260	3223
H7	407	409	1093	2625
H8	536	538	1502	3624

Isoforms from the TDP and MDP data were combined per group (“Control” and “Heat”) and all duplicates removed, then compared in terms of overlap between methods. Those isoforms only present in one of the two datasets were then evaluated further, averaging their molecular weight and number of amino acid residues.

6.4. Results & discussion

6.4.1. Proteoforms? Peptides? Considerations regarding nomenclature

The unique role of MDP as a bridge between BUP and TDP gave rise to different general points of consideration even before the experimental stage, the first of which concerned the nature of the analytes themselves, as well as their appropriate treatment during data evaluation. Since MDP shares certain aspects with both other approaches to proteomics, arguments exist to support the comparison of the former with either of the latter. Relying on time-limited or incomplete proteolysis, often with the aid of proteases, MDP bears substantial resemblance to BUP, making the size range of generated peptides the only criterion by which to distinguish these two proteomics principles. On the other hand, the

inhibitive digestion conditions applied in MDP mean that undigested proteoforms may remain. Therefore, MDP samples may partially qualify for a TDP evaluation as far as the size range and sequence coverage are concerned.

This overlap with both BUP and TDP led to the related question of how to evaluate MDP data adequately. Since no dedicated MDP data evaluation workflow was available at the time, the only options were to treat the MDP dataset as either BUP (e.g. using the SequestHT algorithm for the database search) or TDP data (e.g. using ProSight). The obvious reliance of MDP on peptides, the latter being tantamount to BUP peptides with a substantial number of missed cleavage events, pointed to the use of SequestHT (as described for the earlier BUP study; see chapter 4). Conversely, the use of ProSight offered the opportunity to gain additional insight into modifications on the large MDP peptides, and was the most straightforward option for any undigested proteoforms that remained. Furthermore, preliminary searches suggested a superior performance of the ProSight algorithm concerning the number of Middle-Down peptides identified. For these three reasons, ProSight became the method of choice for MDP data evaluation.

Using ProSight in this manner notably came with the downside of erroneous nomenclature. Originally meant for TDP, the algorithm would provide a readout of proteins, isoforms, proteoforms and PrSMs. The term “proteoforms” in particular was considered a problem since all Middle-Down peptides detected were treated as such for the purpose of data evaluation. While it may be argued that the small population of annotated (i.e. undigested) proteoforms in the dataset could indeed be considered positively identified, thereby justifying the term, the same was not true for the vastly larger proportion of Middle-Down peptides that, like in BUP, could only serve the purpose of protein inference. In the context of this chapter only, the term “proteoforms” therefore was essentially a replacement for “peptide groups” in a SequestHT search.

The term “isoforms” was likewise considered problematic, especially in the context of MDP. According to IUPAC (International Union of Pure and Applied Chemistry) nomenclature, isoforms are defined as variations on the genetic, but not the protein level (Jungblut et al., 2008; Schlüter et al., 2009). While this definition automatically disregards any changes to the amino acid sequence once assembled (PTMs), genome-based variation alone leaves a high potential for protein isoforms to deviate from the canonical genomic sequence. Since protein inference only links peptides to canonical proteins, isoforms, like proteoforms, therefore cannot be distinguished by the inference-based MDP approach. This is directly relevant for the size comparison presented in chapters 6.3.10

and 6.4.5.2, where the “isoforms” readout was used to gain a larger number of available data points.

6.4.2. Considerations for basic workflow aspects

With the focus of MDP being the retention of a high degree of proteoform information, the generation of peptides well above the mass range of those typical in BUP was a crucial aspect. Method development was therefore required to maintain the balance between hydrolyzing intact proteoforms and not causing a near complete digestion tantamount to a Bottom-Up scenario.

The protease Glu-C is known to hydrolyze at the C termini of aspartic or glutamic acid residues (Drapeau et al., 1972). Here, it was used in combination with an AmBiC buffer due to the resulting increased hydrolysis specificity for only one site, namely the C terminus of glutamate residues (Jakoby et al., 2012). In a recent report, incomplete proteolysis using Glu-C for MDP was performed over 4 h at a temperature of 37 °C, with a mass ratio of sample protein/protease of ~12:1 (Takemori et al., 2024). In order to further limit the extent of proteolysis, these conditions were altered to incubate samples mixed with Glu-C at room temperature, which was also considered beneficial with respect to later air-based DMF applications, where evaporation would have to be controlled. As another measure to facilitate accommodation of the digestion step on DMF chips, reducing the digestion time to 1 h at most, while maintaining a similar protein/protease ratio, was considered.

Since the desired incomplete proteolysis was expected to yield peptides within a broad mass range, the workflow was designed to allow fractionation after the digestion step. A first fraction would consist of the acidified Glu-C digest, containing smaller peptides readily detached from the SP3 beads during the proteolysis (see also Figure 6-1). The second fraction would comprise the eluate from the subsequent FA-based elution, containing larger peptides still aggregated on the beads after the digest. Both fractions could therefore be expected to contain different size distributions, thereby enabling a crude separation while also making it possible to investigate the interaction of peptides with SP3 beads and elution behavior.

While the BUP workflow for DMF-based LFQ of single *C. elegans* discussed earlier (chapter 4.3.8) relied on a tryptic digest in the presence of DDOPM, the latter detergent was replaced by 0.01% w/v MNA-C12 during MDP method development. MNA-C12 would

be used later as a diluent following elution by FA; therefore, this adjustment was made to avoid the introduction of an additional detergent and thus limit sample complexity. Preliminary tests using bulk amounts of protein and SDS-PAGE for visualization suggested MNA-C12 in the given concentration to not be detrimental towards Glu-C in the intended buffer environment.

6.4.3. Fine-tuning of critical workflow elements for low-input samples

With a preliminary version of a protocol established, the attention was focused on those individual stages deemed decisive especially with regard to low-input preparations. Aiming to optimize a method specifically for the processing of samples from single *C. elegans*, the material amount per sample was downscaled to ~1000 CaCo-2 cells and samples prepared specifically for LC-MS/MS processing.

The first parameter to be tested at this stage was the digestion time. Having already narrowed down a promising timeframe from 30 to 120 min for Glu-C digestion at room temperature in trials using bulk amounts, this window was investigated further, while also taking into account shorter timeframes. Qualitative data evaluation primarily focused on the overall number of peptides assigned per sample. Other aspects of interest were the average theoretical mass (m_{Theo}) and the incidence of glutamate residues at the C termini of assigned peptides, both of which were used as a simple measure to estimate the digestion efficiency of Glu-C over a given time (Table 6-3).

Table 6-3: Numbers of proteins, peptides and PrSMs, as well as the average theoretical peptides masses and incidence of glutamate residues at the C termini of peptides, in a digestion time series from 0 to 120 min using approximately 1000 CaCo-2 cells as starting material.

Sample	Rep.	Proteins	Peptides	PrSMs	m^{Theo} (\emptyset) in kDa	# Glu at C-term. (%)
0 min (no Glu-C)	1	113	233	1540	10.0	3.0
	2	130	277	1699	9.8	2.9
5 min	1	236	513	2083	7.7	35.9
	2	242	508	2082	7.3	37.8
15 min	1	264	598	2198	7.2	44.0
	2	282	618	2275	7.0	47.4
30 min	1	305	628	1971	6.6	55.3
	2	327	711	2214	6.3	56.7
45 min	1	346	722	2304	6.4	56.5
	2	342	768	2427	6.4	56.8
60 min	1	366	785	2334	6.2	60.4
	2	346	706	2030	6.3	58.1
120 min	1	399	824	2095	5.8	66.9
	2	456	1039	2812	5.7	67.8

The data generated in this experiment clearly demonstrated an antiproportional correlation between the average theoretical mass of peptides and the duration of the Glu-C digestion step.

An investigation of cleavage events within proteoforms suggested a favorable performance of Glu-C both with regard to specificity of hydrolysis sites, and overall activity under the chosen reaction conditions. For the negative control (i.e. incubation for mere seconds and in absence of Glu-C; essentially a TDP experiment), cleavage events were expected to be negligible in number and not show any pattern (D. Chen et al., 2019; Kaulich et al., 2022). Both of these expectations were in fact confirmed by the analysis (Figure 6-2). At a temperature of approximately 20 °C – well below the Glu-C incubation temperature of 37 °C previously used in MDP (Takemori et al., 2024) – this protease was expected to display substantially reduced activity. Despite this, incubation at this temperature for only 5 min already saw the overall number of cleavage events rise significantly, with a clear majority of them now occurring at the C-termini of glutamate residues, thereby confirming activity of the enzyme at the given conditions. As expected, both trends would continue at higher digestion times.

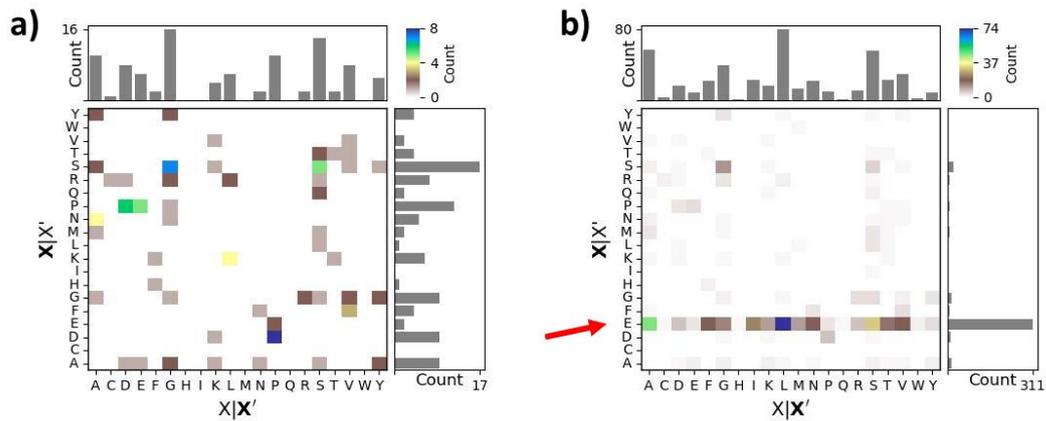


Figure 6-2: Cleavage events observed in a Glu-C digest time series involving samples of ~1000 CaCo-2 cells, following no digestion at all (a), or digestion for 5 min at approximately 20 °C (b).

While the cleavage specificity shifted towards C-terminal glutamate as digestion time increased, the prevalence of undigested, intact proteoforms decreased accordingly. In the control condition, where Glu-C had been left out entirely, the majority of identified analytes was made up of full-length proteoforms (at ~59%), whereas the proportion of proteoforms that were both N- and C-terminally truncated (“Internal”) appeared negligible as expected (at ~4%). After 30 min of Glu-C digestion at approximately 20 °C, the incidence of undigested proteoforms had dropped substantially to 19%, with internal peptides in turn seeing the greatest increase to ~39% (Figure 6-3).

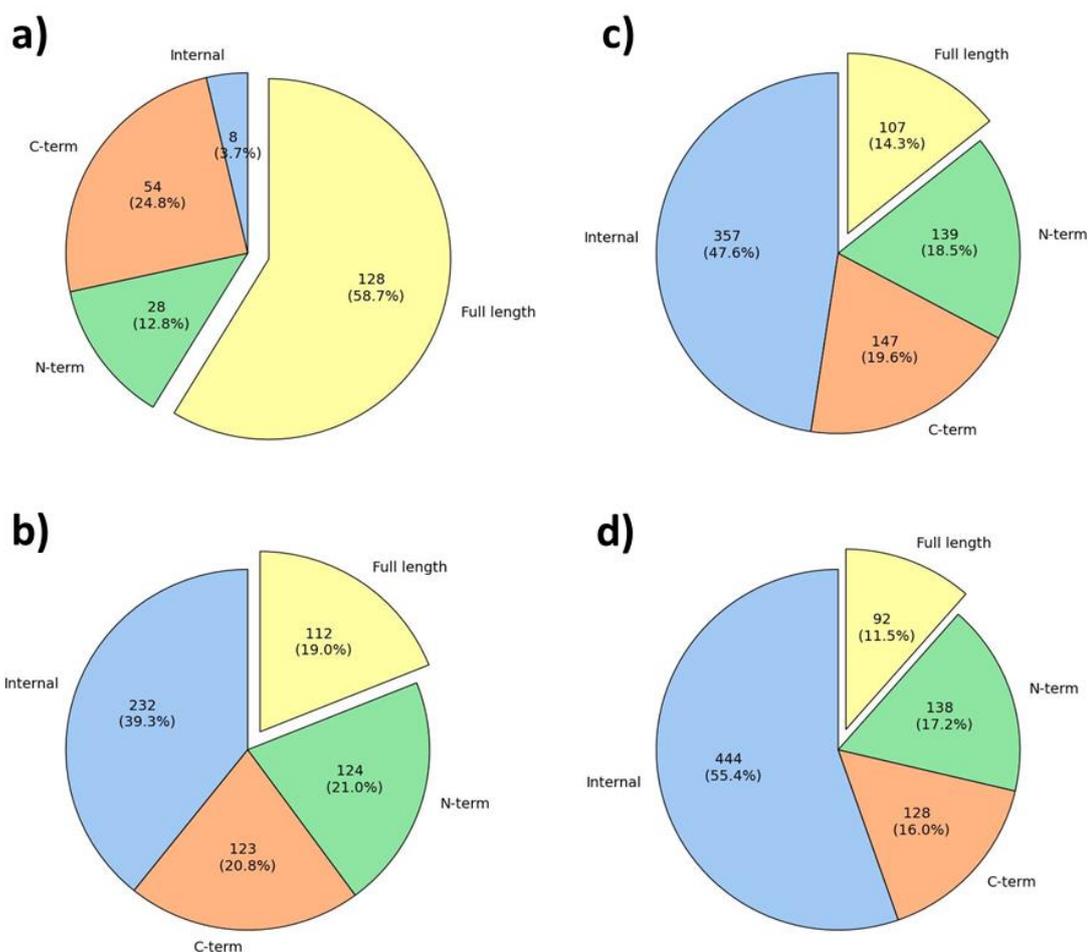


Figure 6-3: Positions of the identified analytes in relation to the protein sequence, observed in a Glu-C digest time series involving samples of ~1000 CaCo-2 cells. Charts shown are for no digest at all (a), 30 min (b), 60 min (c) and 120 min (d).

Based on these results, it was considered feasible to limit the Glu-C digestion time to 1 h, as well as increase the sample protein/protease ratio to approximately replicate that used for the trypsin/Lys-C mix during the DMF-BUP experiments described in chapter 4.

Another important factor was the final sample volume, which had the potential to vary greatly depending on whether the digest or eluate fractions were to be processed, and whether or not the two were to be merged. Across the entire sample preparation protocol, the step contributing the largest additional volume was the dilution of FA used for eluting intact proteoforms and large Middle-Down peptides from magnetic beads. It was therefore investigated whether the elution process could also be successful at lower FA concentrations, ideally making it possible to substantially decrease the final sample

volume by lowering the necessary dilution. Results were compared in particular with respect to the number of annotated (i.e. undigested) proteoforms in each sample measured (Table 6-4).

Table 6-4: Numbers of proteins, peptides, PrSMs and annotated (i.e. undigested) proteoforms, as well as the average theoretical peptides masses, in a series testing elution with different concentrations of FA (in % v/v) and using ~1000 CaCo-2 cells as starting material.

Sample	Rep.	Proteins	Peptides	PrSMs	m^{Theo} (\emptyset) in kDa	Annotated proteoforms
0% FA	1	332	584	1197	5.0	30
	2	302	493	932	4.5	9
10% FA	1	351	751	2160	6.0	88
	2	316	654	1913	6.3	103
20% FA	1	318	640	1994	6.4	100
	2	404	946	2513	5.8	96
30% FA	1	309	665	1987	6.5	111
	2	361	772	2262	6.0	87
40% FA	1	382	871	2661	6.1	108
	2	338	745	2317	6.3	112

While annotated proteoforms were observed even after elution attempts using water only (0% FA), any elution with 10% FA or more caused their numbers to rise sharply by comparison, highlighting the elution potential of FA already demonstrated in previous reports. Most notably, numbers of annotated proteoforms per sample appeared to remain almost stable after elution with 10 to 30% FA, and rise only slightly when going from 30 to 40% FA. Based on these results, the elution step with 40% FA previously used for DMF-TDP (Leipert et al., 2023) was adapted, changing the default FA concentration to 30%.

A third vital aspect to consider was how digest and eluate fractions were handled after obtainment. The LC equipment available could accommodate volumes of up to 20 μ L in the sample loop, making it possible to inject entire samples even after combining fractions. With respect to the limited amount of material available from the outset, the preferred procedure for low-input samples meant for LC-MS/MS therefore was to merge the fractions, sacrificing resolution by working steps in favor of a higher material amount per injection. It was therefore investigated how maintaining the original sample fractions would affect the outcome of measurements in terms of sensitivity, peptide size range and overlap between fractions (Table 6-5 and Figure 6-4).

Table 6-5: Numbers of proteins, peptides, PrSMs, and annotated (i.e. undigested) proteoforms, as well as the incidence of glutamate residues at peptide C termini and the average theoretical peptide masses, in a series using ~1000 CaCo-2 cells as starting material with fractionation maintained. Cells shaded grey represent digest fractions, cells shaded white eluate fractions.

Rep.	Proteins	Peptides	PrSMs	m_{Theo} (kDa)	# Glu at C term. (%)	Annotated proteoforms
1	218	358	660	4.5	75.4	3
2	271	473	1002	4.9	73.4	16
3	338	614	1373	5.0	74.1	24
4	245	390	820	4.9	73.1	16
1	221	407	1421	7.1	46.7	90
2	250	495	1631	6.7	53.0	100
3	193	384	1642	7.7	38.8	114
4	239	471	1838	7.0	48.0	107

As expected, qualitative evaluation indeed suggested the majority of the analytes in the digest fraction to be peptides resulting from hydrolysis by Glu-C, with a relatively high prevalence of C-terminal glutamate (~74% on average), average theoretical peptide masses between ~4.5 and 5 kDa, and a rare occurrence of annotated proteoforms. Accordingly, a different picture was observed with the eluate fractions, where the number of annotated proteoforms was, on average, about seven-fold higher when compared to the digest fraction. A similar trend was observed for the theoretical peptide masses, which averaged at ~7 kDa in the eluate fraction.

To investigate overlap between fractions, the ProForma annotations (LeDuc et al., 2018, 2022) of Middle-Down peptides from digest and eluate fractions were pooled over four replicates before removing duplicate entries and comparing the annotations from both fractions against each other (Figure 6-4).

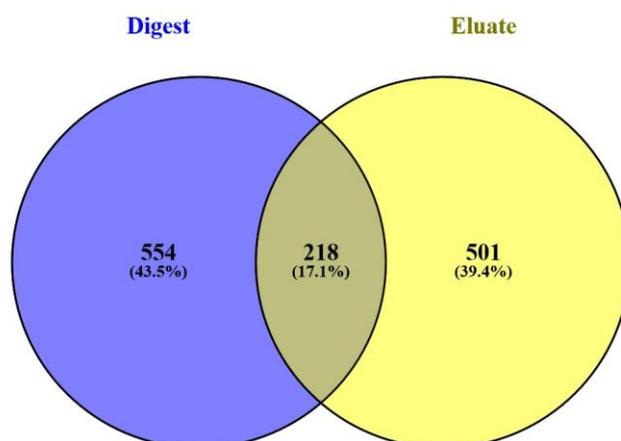


Figure 6-4: Overlap of Middle-Down peptides from CaCo-2 digests found in separate digest and eluate fractions, combined from four replicates. Venn diagram created using the Venny 2.1 online tool (Oliveros, 2007).

Data evaluation via this approach again suggested a strong separation of Middle-Down peptides by fraction, with only ~17% of ProForma annotations being shared across both. This overlap may be partially explained by small, non-recoverable remnants of the digest fraction leading to carryover of Middle-Down peptides into the eluate fraction despite removal of the digest. Furthermore, larger peptides not completely resolubilized during the digestion step could have been eluted into the subsequent fraction, again contributing to the percentage of analytes shared among both of these.

In any case, these findings highlighted the need for a deliberate elution step following the Glu-C digestion, with up to 39% of peptides or even undigested proteoforms otherwise presumably retained on the beads after the digestion step, and thus lost in the process.

6.4.4. Optimization of LC-MS/MS conditions

In terms of maximizing the yield of detected Middle-Down peptides, the LC-MS/MS process offered a variety of parameters to consider for adjustments.

An essential aspect of the proposed measurement method was the use of FAIMS with internal CV stepping. Building upon the TDP work presented in chapter 5, the total number of CVs was kept at four to cover a broad ion mobility range for Middle-Down peptides without losing out on scan cycles excessively. Anticipating a generally smaller population of analyte ions due to the incomplete proteolysis step, CVs were set to lower values

compared to the previously applied CV stepping for TDP, introducing a CV of -60 V to replace that of -20 V.

The second important parameter to optimize was the number of microscans performed both on MS and MS² level. While a higher number of microscans is associated with increasing the S/N ratio, this also causes scans to take more time to be completed, leading to less MS/MS spectra recorded, and thus a negative impact on the proteome coverage (Kalli & Hess, 2012). Accordingly, the consistently highest numbers of peptides during MDP method development were obtained using a method where MS and ddMS² microscans were reduced to one (variant 2; see Table 6-6), the only exception being the CV of -60 V (2 ddMS² microscans).

Compared to the default method, shortening or lengthening the LC gradient notably both proved detrimental on their own in terms of all four parameters evaluated, with the 60-minute gradient method (variant 4) in particular yielding the lowest values. Out of all LC-MS methods tested, variant 3 (involving a 180-minute gradient) appeared to display the highest inter-replicate variability. This was likely a result of peak broadening and associated lower signal intensity, which would have made the detection of ions, especially those of low abundance, inconsistent across replicates.

Another parameter considered for adjustment was the activation mode for data-dependent MS². CID was the default mode prior to method development, having already been used in the previously discussed BUP and TDP work involving single *C. elegans*. For comparison, HCD was considered as an alternative activation mode.

The qualitative results of the LC-MS/MS method comparison are summarized in Table 6-6, where methods tested have been assigned numbers corresponding to those used in the description in chapter 6.3.6.

Table 6-6: Numbers of proteins, peptides and PrSMs, as well as the average theoretical peptides masses, in a series comparing different LC-MS/MS methods and using ~1000 CaCo-2 cells as starting material. Method numbers correspond to those introduced in chapter 6.3.6.

Method	Replicate	Proteins	Peptides	PrSMs
Default	1	430	935	2667
	2	387	832	2408
1	1	409	897	3325
	2	446	1040	3551
2	1	460	1070	3563
	2	433	1023	3304
3	1	400	823	2570
	2	355	714	2037
4	1	358	787	2223
	2	361	790	2178
5	1	404	894	2659
	2	433	966	2775

Based on the results obtained with these six LC-MS/MS method, variant 2 was chosen as the new default method for measurements of single *C. elegans* for later LFQ evaluation.

6.4.5. *C. elegans*: Middle-Down label-free quantification

6.4.5.1. Challenges & workarounds

Due to spatial constraints, the replicates necessary for the LFQ measurement sequence could only be generated in multiple sample preparations involving several DMF chips. Accordingly, the risk of introducing batch effects through this practice needed to be considered. To alleviate this, replicates prepared per chip were always made up evenly of *C. elegans* from the control and heat-treated group; however, loss of individual samples during experiments would occasionally shift this distribution one way or the other.

On-chip incubation of samples always needed to be planned with regard to keeping evaporation under control. Early experiments suggested that incubation times of 1 h or less at approximately 20 °C led to a tolerable extent of evaporation, even with the chip left in the open, while also ensuring a favorable degree of Glu-C digestion activity. These aspects made it possible to forego the use of humidified chambers as a means to control

evaporation almost entirely. The only exception to this was the SP3 binding step, where an atmosphere saturated with ACN was imperative to maintain the solvent composition within the sample droplets necessary to induce protein binding.

Throughout the DMF-MDP protocol, manual pipetting to retrieve samples was considered a detrimental factor for multiple reasons. Aside from the risk of losing Middle-Down peptides or proteoforms due to adsorption, the possibility of damage to the hydrophobic coating of chip top plates due to contact with pipette tips needed to be taken into account. Since eventual merging of fractions was included even in earlier versions of the protocol, the final version therefore saw the initial fractionation omitted entirely, relying instead on one pipetting interaction with the top plate to retrieve the entire sample upon completion of the elution step.

Despite the workflow adjustments described here, DMF-MDP sample preparation remained relatively error-prone compared to the BUP and TDP variants used earlier. Most notably, complications were the likeliest to arise during the dilution step. In trials, drops of 0.01% w/v MNA-C12, frozen during the elution step, showed severely impaired mobility even after thawing, as had already been observed for the TDP method. Unlike the latter protocol, though, the bead-containing mixture of eluate and digest could not be pulled into the diluent drops by the magnetic lens to circumvent this problem, causing substantial delays of the dilution process. The procedure was therefore changed to adding MNA-C12 solution to the samples via the reservoir electrodes instead, immediately after the elution.

Loading of MNA-C12 solution would remain unreliable, however, since drops were frequently observed to cling to the edge of the top plate, or even crawl on top, rather than be drawn onto the reservoirs due to electrode actuation. These loading difficulties of MNA-C12 solution were suspected to be aggravated, if not caused entirely, by the formation of condensation due to the sudden transition of the chip from $-20\text{ }^{\circ}\text{C}$ to room temperature after the elution step, which had already been speculated to promote electrode failure during TDP sample preparation (see chapter 5.4.1.1).

For MDP experiments using Glu-C as described here, AmBiC-based buffers were required in order to ensure C-terminal hydrolysis specificity of Glu-C at glutamate residues only. This buffer system, however, has previously been observed to impair droplet mobility when used on DMF chips, something hypothesized to be caused by the adsorption of carbonate ions to the chip surface (Chatterjee et al., 2010; Leipert & Tholey, 2019). While reduced droplet mobility of AmBiC-containing mixtures on the electrode array was not observed, possibly due to the concentration range employed (25 mmol/L or less),

repeated loading of reagents containing AmBiC via the hydrophobically coated reservoir edges might also have been a contributor to the problems observed when loading MNA-C12 solutions.

6.4.5.2. Qualitative data evaluation and LFQ

The use of six replicates per condition in combination with feature mapping by ProSight led to the assignment of 2975 Middle-Down peptides; however, only 511 (~17%) of these were found to have valid abundance ratios and thus be quantifiable. This meant that the majority of the Middle-Down peptides from the dataset could only be evaluated qualitatively.

As had already been the case with the TDP data discussed previously, missing abundance values within the quantifiable peptides constituted a significant portion of the data: Out of 6132 total possible values (from 511 peptides), 1345 (~22%) were missing. This incidence of missing values made stringent filtering necessary to avoid compromising subsequent statistical tests. To qualify for the latter, peptides were therefore required to be present with at least four valid values per group (out of six), leaving 324 peptides in a reduced dataset consisting to 6.4% of missing values. Of these peptides, five were considered to have significant abundance differences between the two groups (Figure 6-5). One of these was more abundant after the application of heat stress and was assigned to the heat shock protein 16.2 (Hsp-16.2). The other four peptides were less abundant after heat stress and were assigned to protein disulfide isomerase, the 60S ribosomal protein L38 and myosin-4 (two peptides).

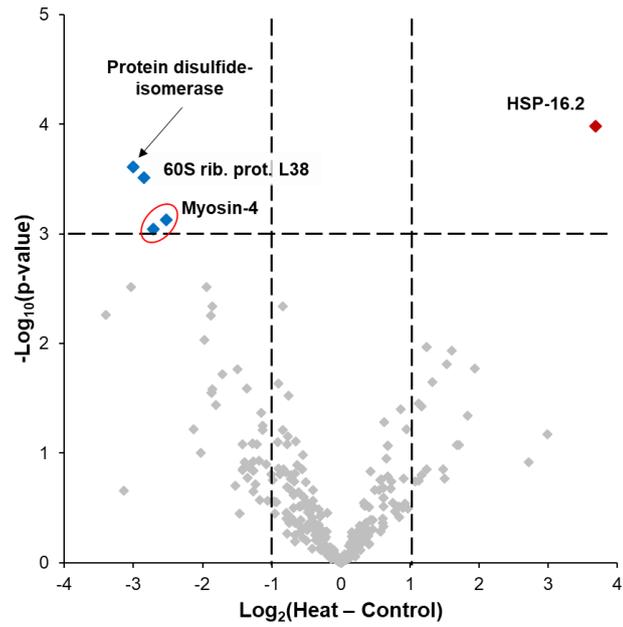


Figure 6-5: Volcano plot of Middle-Down peptides from single *C. elegans* in two biological conditions, following filtering to only include those peptides present with at least 4 valid values per group. Imputation of missing values was not applied.

Due to the loss of a substantial proportion of quantified Middle-Down peptides to missing value filtering, a second, non-statistical evaluation was performed to investigate those peptides with a particularly one-sided distribution of valid values. To be considered, peptides would have to be represented by at least four (out of six possible) valid values in one group, while being completely absent from the other.

As expected, this second evaluation showed peptides originating from multiple heat shock proteins as only having been detected in heat-treated *C. elegans*. These included Hsp-16.1/16.11, Hsp-16.2, and Hsp-70. This confirmed the previously reported findings of the related BUP study (Steinbach et al., 2022) once more and simultaneously explained the absence of the associated protein accessions from the statistical test employed.

One of the main arguments in favor of MDP is the ability to generate subsequences from proteoforms too large for identification by TDP, allowing one to deduce information on modifications, especially whenever multiple of these exist on the same subsequence. To verify this argument, the MDP dataset discussed here was compared to the data that contributed to previously published TDP-LFQ work (Leipert et al., 2023) regarding the size of proteins inferred and proteoforms identified. For the TDP data, this comparison involved all four replicates per group (“Control” and “Heat”). Since more replicates per group were available for the MDP dataset, the latter was reduced to the four per group that

contributed the highest numbers of peptides (Table 6-2). The isoform readout column was then used to compare the molecular weight of proteins identified in TDP, or inferred in MDP, as well as the number of amino acid residues making them up. For this purpose, only accessions exclusive to either dataset were taken into consideration.

Proteins exclusively identified in the control group via TDP were observed to have a molecular weight of ~27.7 kDa on average, corresponding to an average number of ~246 amino acid residues. Conversely, proteins exclusively inferred in the same group via Middle-Down peptides had an average molecular weight of ~68.5 kDa and consisted of ~616 residues on average. A similar trend was observed for the data obtained from heat-treated nematodes, with TDP-identified proteins averaging at ~30.6 kDa and ~273 residues, while MDP-inferred proteins averaged at ~62.5 kDa and ~563 residues. In combination with the results from method development with CaCo-2 cells, where Glu-C digestion under the same conditions was shown to generate an abundance of peptides above what is considered the BUP peptide mass range, these findings were seen as proof of the MDP principle as outlined above.

Another known strength of MDP is the ability to distinguish different PTMs on the same peptide, or multiple peptides containing the same sequence portion. Despite this promise, modifications were generally found to be scarce in the given dataset, which again may be explained by the low material input used and long cycle times of the multi-CV FAIMS method, as well as the (at the time) error-proneness of the sample preparation being detrimental to repeatability. Of the 2975 Middle-Down peptides inferred, only 1161 (~39%) were observed to bear modifications of any kind. A very high proportion of these could be attributed to artificially induced carbamidomethylation (introduced during the alkylation process), which was present at least once in 982 of 1161 (84.5%) modified peptides, while 274 (23.6%) were found to feature N-terminal acetylation. One example that stood out was histone H2A (P09588), where two peptides were found to feature N6-acetyl-L-lysine. While both of these peptides were quantifiable, they could not be taken into account for statistical testing due to not passing the missing value filter. Another interesting example was the elongation factor 1-alpha, which yielded 15 peptides featuring L-glutamyl-5-glycerolphosphorylethanolamine in various positions. Of these peptides, three were quantifiable and two in turn available for statistical testing, albeit not with a significant abundance difference. Additionally, across the three quantifiable peptides, the modification in question was in the same location.

6.5. Conclusion

The adaptation of the previously used DMF sample preparation methods for MDP proved to be particularly challenging for a multitude of reasons. The combined necessity of alkylation and digestion steps on the one hand, and elution at $-20\text{ }^{\circ}\text{C}$ on the other hand, likely strained the durability of the DMF chips used more than either of the BUP or TDP protocols did. This would lead to a high risk of complications during sample retrieval, or even complete loss of samples, and in turn poorer reproducibility. The latter factor in particular was considered a major detriment to LFQ, with only $\sim 16\%$ of Middle-Down peptides assigned actually quantifiable, and many of these excluded from statistical testing due to high numbers of missing values. Despite the complications observed, qualitative evaluation of Middle-Down peptides did serve to confirm quantitative trends observed on BUP and TDP level. Furthermore, the DMF-MDP results presented here clearly indicated the inference of proteins well above the mass ranged considered accessible to TDP, and therefore complementary to the latter approach. DMF-based sample preparation for quantitative low-input MDP can therefore be expected to become a valuable asset following further method development.

7. Discussion

7.1. Considerations for LC-MS/MS

All experiments described in the previous chapters relied on samples containing entire proteomes, and the resulting high sample complexity therefore needed to be taken into account accordingly. Regardless of the proteomics approach pursued, separation by nanoLC therefore was a decisive means to reduce sample complexity. This made it essential to first perform a series of benchmarking LC-MS/MS measurements on different column designs prior to attempting quantification. Columns tested within this series were compared for the yield of inferred proteins and peptide groups at different gradient lengths and input amounts.

While these results are not discussed here in detail with respect to the scope of the thesis, an essential finding was that the EIC build appeared to show not only the highest yield, but also the most consistent performance across different gradient lengths. For these reasons, the Aurora series 25 cm column by IonOpticks was chosen as the default column for BUP LFQ. Most notably, an Aurora-like prototype designed primarily for TDP, where the C_{18} stationary phase was replaced by a C_4 version, would display the same favorable performance compared to the default C_4 column used at the time for TDP. This served to demonstrate in particular the versatility of the EIC design in low-input proteomics, regardless of the approach.

7.2. Evaluation of DMF approaches

Aiming to not only minimize interaction with samples by manual pipetting, but also limit the contact of samples with other surfaces to which proteins or peptides could potentially adsorb, DMF was adopted as a preferred means of sample handling. Two inherently different builds of DMF devices and respective chips were available, one being oil-, the other air-based.

An oil-based DMF approach was expected to come with several advantages making it suitable for the proposed proteomics method development. The most important of these was considered highly limited evaporation of samples even during prolonged on-chip handling and at elevated temperatures, due to the immiscible liquid filler medium being tantamount to a closed compartment. Recurring BUP working steps requiring different

elevated temperatures, e.g. reduction (50 °C) and digestion (37 °C), were indeed found to benefit in the expected way from the oil filling, with liquid volumes not appearing to decrease significantly based on the electrode area covered. Conversely, liquid samples on air-filled DMF chips were subject to substantial evaporation even at temperatures of approximately 20 °C, during incubation steps of comparable duration, which was observed in own experiments and had also been reported earlier (Leipert & Tholey, 2019). Additionally, oil is known to facilitate different microfluidics operations themselves (Brassard et al., 2008); therefore, another expected benefit of this design was enhanced droplet mobility. It was hypothesized that this would enable substantially reducing detergent concentrations considered indispensable in air-based DMF, while maintaining the same degree of mobility.

While advantageous in some regards, the presence of a liquid filler medium also brought with it several challenges to consider, some of which were impossible to circumvent. Oil miscibility of solvents otherwise considered ideal for SP3 (e.g. ethanol, and to some extent ACN), and, more importantly, the ease of accidentally introducing oil contamination to samples during retrieval from oil-filled chips, made it imperative to perform SP3 exclusively off-chip. While SP3 thus still served the purpose of a highly potent removal option, simultaneously eliminating both silicone oil and LC-MS-incompatible reagents, the off-chip variety of the method significantly increased the degree of interaction by manual pipetting. Additionally, the lack of on-chip SP3 also meant that sample preparation had to be planned around proteases being exposed to all additives and reagents used throughout the method. The aforementioned positive influence of the oil on droplet mobility indeed made it possible to somewhat alleviate this problem by strongly limiting the introduction of detergents, albeit only beyond the lysis step. Accordingly, in an oil-based DMF-BUP experiment, the latter could be assumed to be a factor decisive for the outcome, not only making sample proteins accessible for processing, but also dictating the conditions for the eventual enzymatic proteolysis. This essential nature of the lysis step could indeed be confirmed in a series of comparative experiments, paying particular attention to the digestion efficiency. Here, the incidence of missed cleavage events varied significantly depending solely on the composition of the lysis buffer (see also Figure 4-9a and b). Aside from these limitations directly influencing the sample preparation, the oil itself was found to complicate handling of chips during experiments. The cohesive properties and low viscosity of the oil in particular added to the risk of oil leaks. These were the most likely to occur whenever chips needed to be moved during sample preparation, which was inevitable for working steps at elevated temperatures. In the end,

these combined limitations were considered to outweigh the potential benefit of oil-based DMF for low-input proteomics, at least in the form available at the time.

Due to the complete absence of a liquid filler medium, air-based DMF came with the promise of a higher degree of freedom regarding the choice of solvents, leading to a greater flexibility during method development. The only restriction in this regard was that detergents played a larger role, compared to oil-based DMF, in maintaining droplet mobility and limiting biofouling. The dependence on detergents, however, was alleviated in different ways. Most importantly, the possibility to perform SP3 protocols on-chip at any time during sample preparation enabled removal of potentially problematic detergents whenever required, whereas the exclusive use of LC- and protease-compatible detergents post-SP3 was an excellent way of limiting the number of removal steps to only one. Both of these factors increased the versatility of DMF-based sample preparation protocols for all three approaches to proteomics (BUP, TDP and MDP). Simultaneously, the absence of DMF-associated contamination sources (oil) removed the necessity of off-chip interaction with samples for cleaning purposes. Samples could thus be processed directly by LC-MS once retrieved from the chip, reducing both handling time and potential for sample loss.

Despite these considerable advantages, however, the transition to air-based DMF did not eliminate all challenges. A major concern instead would be evaporation of liquid samples and reagents, and air-based DMF workflows were hence designed or modified almost exclusively with respect to this problem. This involved conducting the majority of incubation steps at room temperature (approximately 20 °C), thereby foregoing working steps at elevated temperatures entirely. Most notably, this also included digestion steps previously reported (Takemori et al., 2024) or shown by own evidence (chapter 4.4.2, Figure 4-9) to benefit from such conditions. The primary argument in favor of this approach was that it enabled all incubation steps to be carried out without the need for the chips to be enclosed in a humidified chamber, which would have come with the risk of allowing the dielectric to be infiltrated by moisture, eventually promoting localized electric breakdown (Eydelnant et al., 2012; Leipert & Tholey, 2019). The only exception to this was incubation for protein binding during SP3, where enclosure as described was essential to maintain the necessary high ratio of organic to aqueous solvent. While limited evaporation did still occur even under the temperature restrictions described, reagent volumes for air-based DMF were always chosen so that electrode overlap would be guaranteed, ensuring droplets would stay mobile throughout the entire sample preparation.

Aside from the inherent advantages and limitations of oil- vs. air-based DMF, the properties of the chips themselves were also a factor to be considered during method development. Being from different manufacturers and intended for different variants of DMF, the two chip models were highly different concerning the number and distribution patterns of electrodes, as well as materials comprising the chips. Interestingly, despite these differences in chip geometry, the number of replicates processible per chip was similar for both designs. In oil-based DMF, the accumulation of reagents and lack of evaporation increased the spatial demands per sample, whereas the electrode distribution effectively divided the chip in two halves (see Figure 3-2), thereby providing only limited options for mixing routes. This in turn limited the number of replicates that could be handled without risking cross-contamination. Conversely, the air-based chips comprised a lower number of actuation electrodes; however, the compact arrangement of these allowed more flexibility when moving samples to and from mixing areas, and a moderate degree of evaporation was in fact somewhat beneficial in reducing spatial demands of samples. Furthermore, liquid volumes that could be handled per electrode were similar, if not identical, across both chip designs, making it possible to maintain the same scale when adapting workflows. All of these factors ensured comparable sample handling even when using two inherently different DMF devices.

Compared to both DMF principles discussed here, the primary strength of the in-tube approach was that sample preparation was not tied to a confined area, eliminating spatial constraints. This made the method highly flexible in terms of replicate numbers that could be prepared simultaneously, while also granting more freedom regarding solvent/sample ratios, which was considered particularly important for the SP3 procedure. On the other hand, numerous weaknesses could be attributed to the in-tube method. One of these was poor visibility of particulate sample remains, which was relevant in all work involving single *C. elegans*, due to the risk of accidentally aspirating non-lysed sample in the early stages of preparation. Most importantly, the notoriously high degree of manual interaction and surface contact of samples inevitably makes in-tube preparation costly in terms of sample material, especially when starting amounts are low to begin with and/or sample preparation vessels are changed over the course of the protocol.

7.3. Comparison of proteomics approaches

While pursuing inherently different approaches to proteomics (BUP, TDP and MDP), the studies described here were supposed to be designed in a similar way, making use of samples identical regarding the species of origin, pre-treatment of the latter, and subsequent input amount. Sample preparation workflows were likewise kept as simple and similar to each other as possible, making only those adjustments necessary to support the respective proteomics approach. Temperature variations were notably kept to a minimum as well across all workflows presented, with all working steps performed at room temperature or $-20\text{ }^{\circ}\text{C}$ (elution step in TDP and MDP only). One point of concern in this regard was that the temperature in the sample preparation room could not be regulated to maintain a desired value. This made any working steps at room temperature somewhat dependent on the ambient temperature, which bore the potential to introduce variance between experiments, especially whenever those made use of proteases (BUP and MDP).

Representing the most popular approach to proteomics at the time of writing, the BUP work undertaken bore several characteristics to be taken into account. Most notably, having been acquired following complete digestion of the sample proteome into peptides, the BUP dataset only allowed inference of canonical proteins, rather than identification. Although a positive identification, as well as distinguishing between individual proteoforms, therefore was not possible, the BUP approach profited from the abundance of peptides generated, leading to numbers of inferred proteins otherwise unreachable. This was considered a major advantage during LFQ following preparation of single *C. elegans* nematodes, since the amount of ~ 5000 quantifiable inferred proteins, both from DMF- and in-tube sample preparation, provided a sufficient data pool for reliable statistical analysis despite the low original material amount. Most notably, all proteins inferred were represented by valid values in at least two replicates per group ("Control" or "Heat"), making it possible to evaluate the dataset even without filtering out missing values, and thus allowing abundance differences to be quantified for the entirety of the BUP dataset. The quality of the dataset presented here was therefore found encouraging for future low-input BUP experiments, suggesting the combination of multi-CV FAIMS ion fractionation and feature mapping during the database search to alleviate one of the main drawbacks of LFQ sufficiently, namely the problem of missing values.

In contrast to BUP, the TDP approach entirely forewent any use of proteases, making this the only approach tested capable of direct identification of proteoforms. The focus of TDP on proteoforms rather than peptides notably benefitted sample throughput indirectly, since

digestion steps requiring at least one hour (MDP), or even extending into the following day (BUP), could be omitted entirely. More importantly, the ability to identify proteoforms directly, as opposed to inferring proteins, provided a different angle on the response of *C. elegans* to heat stress. This was especially true in cases where multiple proteoforms corresponded to the same protein accession. Conflicting abundance differences of such proteoforms as a result of heat stress could be assumed to paint a distorted picture in BUP, where these proteoforms could not be distinguished, and this could be confirmed indeed by a direct comparison of the two datasets. This observation highlighted a decisive weakness of the BUP approach: Whenever peptides originate exclusively from sequence regions shared across multiple proteoforms, the abundance differences of individual proteoforms may easily be masked.

One factor notably confounding the TDP experiments described was the inability to process multi-CV data directly. This inevitably made it necessary to filter files by CV ("CV slicing") prior to the database search, leading to several options as workarounds. The most straightforward one was to keep the subsets separated and conduct four searches to then evaluate independently from each other. This approach showed the complementarity of the different subsets, with theoretical mass ranges of proteoforms being dependent on the respective CV. Conversely, the CV slicing could also be reverted in different ways: During the database search, by treating the CV-sliced subsets as offline fractions, or afterwards, by summing up abundance values of each proteoform per replicate. Furthermore, in the latter case, crucial steps of data processing could be performed in a different order. All approaches tested that involved reunification of data subsets in any way notably yielded different outcomes regarding the number of quantifiable proteoforms and those displaying a significant fold change.

Aside from these considerations concerning data evaluation, limitations of the TDP approach also presented themselves during the analytical stage. The presence of intact proteoforms meant that analytes spanned a broad size range from ~5 to 15 kDa, leading to challenges regarding different aspects. Separation by nanoLC always relied on a long gradient to account for the high variability in terms of hydrophobicity. FAIMS CVs likewise needed to be chosen carefully to fully cover the expected analyte size range, while also avoiding strong sensitivity losses due to the required cycle times. Lastly, due to the proteoforms still being intact, a multitude of charge states could be expected to be present at once, adding to the complexity and reducing the intensity per individual charge state. These factors conspired to limit the number of proteoforms ultimately quantifiable. Additionally, missing abundance values proved to be highly detrimental with the relatively

low number of analytes available, especially when compared to the previously acquired BUP data. Despite these inherent challenges of TDP, the ability to directly identify proteoforms rather than infer canonical proteins proved invaluable especially from a quantitative point of view.

The primary goal of the MDP approach is to use incomplete proteolysis on proteoforms too large to be identified via a TDP analysis, in order to generate subsequences not only accessible for MS/MS measurement, but also allowing the investigation of multiple modifications on the same subsequence. Regarding interpretation on the proteoform level, the MDP data was therefore expected to provide a complementary addition to the information gained from the TDP experiment, where identified proteoforms had been in a mass range from ~5 to 15 kDa.

During MDP method development, numerous questions needed to be addressed, regarding both conceptualization of experiments and data evaluation. A unique feature of the concept is the digestion step, which is only set apart from a typical BUP workflow in that proteolysis is incomplete, deliberately leaving a considerable number of missed cleavage sites and thus leading to longer peptides. A particularly critical aspect of MDP method development for this project therefore was to establish digestion conditions with a suitable balance between generating peptides, especially of larger proteoforms, and retaining as much proteoform information as possible by not allowing peptides to become too short, effectively leading to a scenario tantamount to BUP. This could be achieved by adjusting three different parameters: Using Glu-C as the protease of choice, hydrolysis specificity was limited to glutamate residues by means of an AmBiC buffer system, thereby substantially reducing the number of theoretical hydrolysis sites when compared to other proteases like trypsin. Digestion efficiency was also controlled via the temperature and reaction time (limited to 20 °C and 1 h, respectively).

Sample preparation under the given conditions was found to meet essential criteria of MDP, with peptides mostly being in a mass range from ~4-5 kDa. More importantly, proteins inferred from these peptides exceeded the mass range of those corresponding to proteoforms identified in TDP for similar samples by far, with average molecular masses of more than 60 kDa for both the control and heat-treated group. Conversely, annotated proteoforms within the MDP dataset, which effectively constituted a data subpopulation on TDP level, did not exceed masses of ~18 kDa.

A second important point of consideration was the way in which data evaluation was to be conducted. The partial digestion step employed during sample preparation set the

approach well apart from TDP, bearing instead more resemblance to a BUP workflow with a higher number of missed cleavages. Therefore, it could be reasoned that a SequestHT search, as performed for proper BUP experiments (chapters 3 and 4), would be the method of choice. Conversely, the presence of undigested proteoforms, as well as a relatively high degree of proteoform information in Middle-Down peptides, pointed towards the use of a ProSightPD search as performed for TDP data evaluation (chapter 5). The latter approach was subsequently taken, a side effect of which being that all analytes, including Middle-Down peptides, were automatically treated as proteoforms during data evaluation. This also meant that multi-CV data could not be readily processed, instead requiring workarounds as presented in chapter 5.3.7.

Out of the three datasets discussed here, the MDP dataset notably turned out to be the one least suited for LFQ evaluation, one reason being that a surprisingly high proportion of the combined Middle-Down peptides and annotated proteoforms were not quantifiable. Additionally, similar to the TDP dataset, a missing value prevalence of ~20% within the quantifiable population made stringent data filtering necessary, reducing the final dataset even further. Both of these shortcomings were certainly due to the low starting material amount, as well as an indirect result of (at the time) existing flaws in the sample preparation workflow, where difficulties such as incomplete sample retrieval would have been associated with a high variance between replicates.

8. Conclusion

During this work, progress towards reliable miniaturized proteomics was made in different ways. In particular, different solutions for three essential aspects of the analytical process – sample handling, chromatographic separation, and measurement – were profoundly investigated. This included comparing two different builds of DMF devices (air- and oil-based), various types and designs of nanoLC columns, and finally parameters for both MS/MS as well as FAIMS, allowing aspects of both sample preparation and downstream processing to be optimized. As a result, comprehensive proteomic datasets of the model organism *C. elegans* were generated for BUP, TDP and MDP, using newly developed or adapted DMF sample preparation- and LC-MS/MS methods optimized for the respective task. This eventually allowed LFQ based on three inherently different approaches to proteomics, making it possible to directly compare these for their inherent strengths and limitations.

Overall, the three datasets provided different kinds of information, depending on the underlying proteomic approach. A highly important advantage of the BUP approach was seen in the number of canonical proteins inferred from the available peptides, as well as the sequence coverage per inferred protein. Making use of an emitter-integrated column (EIC) setup for nanoLC separation, FAIMS ion fractionation with internal CV stepping, and Orbitrap-ion trap mode (OTIT), the DMF-BUP workflow led to the inference of ~5000 proteins from single *C. elegans*. Furthermore, the evaluation of 45 proteins (from DMF and in-tube experiments combined) with significant abundance differences provided deeper insight into how this particular model organism responds to heat stress.

Starting with the same input material amount, TDP sample preparation using DMF was successfully used to quantify proteoforms of higher or lower abundance following heat-treatment of live *C. elegans*. The key strength of TDP – distinguishing multiple proteoforms associated with the same canonical sequence – could be verified in the study, with several significant proteoforms frequently assigned to the same accession, and occasionally with conflictive abundance differences. Conversely, TDP-specific challenges still hindering this proteomics approach, especially in the low-input area, could likewise be demonstrated. These comprised the inherent problem of limited sensitivity, and more importantly, processing of multi-CV data.

The DMF-MDP experiments described in this work saw the expectations towards the MDP concept only partially fulfilled. On average, *C. elegans* proteins inferred from Middle-Down peptides were found to consist of considerably more amino acid residues, and have a

higher molecular mass, than proteoforms identified in single nematodes during TDP experiments. This served to prove the ability of MDP to facilitate measurements of particularly large proteoforms. Despite this promising note, however, numerous practical challenges (inconsistent sample preparation, frequency of losses) limited the quantitative value of the study, while data evaluation and experiment design were confounded by a still present lack of standardization of MDP procedures.

Although MDP is a promising concept *per se* with a number of potential applications, the difficulties and limitations encountered during execution of the sample preparation, and especially during data evaluation, therefore led to the assumption that MDP, at least in a miniaturized DMF-based workflow, cannot compete with TDP or BUP at this time.

In conclusion, DMF was observed to perform favorably across the three different proteomics approaches, thereby demonstrating the versatility of the technology, presenting a viable alternative to in-tube sample preparation and bearing the promise of reliable quantitative low-input proteomics. Despite these encouraging results, limitations to each DMF-based solution presented here still need to be acknowledged. Future work building upon the results described should therefore focus on investigating the remaining unresolved weaknesses of the DMF workflows, with the goal of making these methods not only more repeatable, but also even more efficient.

9. Outlook

In the work presented here, the potential of the DMF technology to support various approaches to proteomics could be demonstrated. However, as the technology for DMF continues to diversify, the range of possible coupling options and applications can be expected to follow the same trend. Examples from the recent years to highlight this are numerous.

Oil-based DMF is notorious for the associated risk of sample contamination and/or the need for reliable removal. A potential way to circumvent this problem may be the use of volatile media, as demonstrated by the recent water droplet-in-oil (WinO) method (Masuda et al., 2022). While not implemented on DMF systems so far, this method was successfully used for single-cell BUP by submerging aqueous droplets in ethyl acetate, effectively creating droplets as isolated microreactors akin to those in oil-based DMF.

In 2020, the introduction of digital microfluidic isolation of single cells for –omics (DISCO) simultaneously represented a major step forward concerning isolation of material from single cells for DMF applications and demonstrated the potential of DMF chips as culture vessels. The original report described the use of laser cell lysis to selectively lyse a single target cell within an adherent population grown on-chip without collateral damage. During lysis, the cell contents were released into a droplet of liquid already present at the lysis site, ensuring that the sample could be removed for processing while leaving the remainder of the culture intact, with the possibility keeping the latter viable if required (Lamanna et al., 2020). A later adaptation of the concept, termed tissue-DISCO (tDisco), was successfully used for spatial proteomics and transcriptomics from single cells isolated on-chip from tissue sections (Scott et al., 2024).

More recently, a method was presented that allowed in-line injection of from a DMF chip onto an MS/MS system, thereby not only avoiding consumption of the entire material, but also keeping the remainder of the sample mobile via EWOD for any downstream steps desired. This was achieved by means of a designated transfer electrode capable of switching between different modes. For movement of material to and from this location, the electrode was operated in the EWOD mode, whereas a switch to electrostatic spray ionization (ESTASI) mode was made for injections (Das et al., 2022). A later DMF workflow would also allow direct interfacing with HPLC by means of an exit orifice in the top plate of a DMF chip. This enabled an all-in-one pipeline for DMF-based TMT assays, which was used for the differentiation of human breast cancer (Peng et al., 2023).

Spatial constraints of DMF chips limiting the number of replicates that can be handled per chip, as well as the importance of controlling the reaction temperature (e.g. to ensure cooling), have been recurring topics throughout this work. New chip designs with finer electrode arrays, or inherently different DMF systems, based, for instance, on thin-film transistor technology (S. Hu et al., 2023; Yang et al., 2024), will likely address these limitations in the near future. Furthermore, a DMF platform incorporating controlled cooling was recently used to study protein-protein interactions, showing the potential of DMF for functional proteomics (J. Shen et al., 2023).

The ability of handling minute amounts of sample with little loss also makes DMF a promising asset for laboratory testing in a clinical environment. Such applications notably have already been shown to be possible in the field, an example being a novel on-chip immunoassay, designed specifically for use in remote locations (Ng et al., 2018), which was used in a recent field study to assess the incidence of measles and rubella infection (Knipes et al., 2022). Accessibility of DMF to the clinical field is notably still limited for reasons such as the specialized production conditions of DMF chips and an associated high cost per assay. Recently, the production of low-cost disposable chips for semi-automated sample handling was reported (Leirs et al., 2022).

While the analytes associated with the various approaches to proteomics are highly diverse in their own right, this diversity increases even further when considering other -omics. Method development for -omics therefore needs to take into account the properties of the analyte group sought after, as well as the challenges in processing material from which these analytes are sourced. The traditional and most straightforward approach to address these challenges has always been to focus on single analyte groups. Examples for this include a variety of extraction methods to aid in lipidomics (Bligh & Dyer, 1959; Folch et al., 1957; Matyash et al., 2008), as well as the more recently established SP3 popular in proteomics (Hughes et al., 2014). However, the recent years have seen increasing efforts in establishing workflows for the analysis of multiple analyte groups from the same biological sample (“multiomics”). Examples are recently described methods for proteomics and lipidomics (He et al., 2021), lipidomics and metabolomics (Sarvin et al., 2020), or even all three of these (Muehlbauer et al., 2023), from the same material source. Due to the rapid advancement of the available technologies, DMF can be expected to be a viable solution for low-input multiomics in the years to come.

DMF remains a rapidly developing and diverse field, with advancements constantly made to push the technological boundaries, and the spectrum of applications is increasing

accordingly, both in proteomics and beyond. With the potential of DMF having been demonstrated by many examples, these developments should encourage the trend towards a more widespread adoption of DMF into the diverse fields of bio- and clinical chemistry in the near future.

10. References

- Abdelgawad, M., Freire, S. L. S., Yang, H., & Wheeler, A. R. (2008). All-terrain droplet actuation. *Lab on a Chip*, 8(5), 672–677. <https://doi.org/10.1039/B801516C>
- Aebersold, R., Agar, J. N., Amster, I. J., Baker, M. S., Bertozzi, C. R., Boja, E. S., Costello, C. E., Cravatt, B. F., Fenselau, C., Garcia, B. A., Ge, Y., Gunawardena, J., Hendrickson, R. C., Hergenrother, P. J., Huber, C. G., Ivanov, A. R., Jensen, O. N., Jewett, M. C., Kelleher, N. L., ... Zhang, B. (2018). How many human proteoforms are there? *Nature Chemical Biology*, 14(3), 206–214. <https://doi.org/10.1038/nchembio.2576>
- Altun, Z. F., Herndon, L. A., Wolkow, C. A., Crocker, C., Lints, R., & Hall, D. H. (2002). *WormAtlas*. <https://www.wormatlas.org/index.html>
- Andreadis, A., Gallego, M. E., & Nadal-Ginard, B. (1987). Generation of protein isoform diversity by alternative splicing: mechanistic and biological implications. *Annual Review of Cell Biology*, 3, 207–242. <https://doi.org/10.1146/annurev.cb.03.110187.001231>
- Arshady, R. (1991). Beaded polymer supports and gels: I. Manufacturing techniques. *Journal of Chromatography A*, 586(2), 181–197. [https://doi.org/https://doi.org/10.1016/0021-9673\(91\)85124-X](https://doi.org/https://doi.org/10.1016/0021-9673(91)85124-X)
- Attygalle, A. B., Pavlov, J., & Ruzicka, J. (2022). Monoisotopic Mass? *Journal of the American Society for Mass Spectrometry*, 33(1), 5–10. <https://doi.org/10.1021/jasms.1c00110>
- Au, S. H., Kumar, P., & Wheeler, A. R. (2011). A New Angle on Pluronic Additives: Advancing Droplets and Understanding in Digital Microfluidics. *Langmuir*, 27(13), 8586–8594. <https://doi.org/10.1021/la201185c>
- Banerjee, S., & Mazumdar, S. (2012). Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *International Journal of Analytical Chemistry*, 2012, 1–40. <https://doi.org/10.1155/2012/282574>
- Barbulovic-Nad, I., Yang, H., Park, P. S., & Wheeler, A. R. (2008). Digital microfluidics for cell-based assays. *Lab on a Chip*, 8(4), 519–526. <https://doi.org/10.1039/B717759C>

- Baret, J.-C. (2012). Surfactants in droplet-based microfluidics. *Lab on a Chip*, 12(3), 422–433. <https://doi.org/10.1039/C1LC20582J>
- Barrière, A., & Félix, M. A. (2014). Isolation of *C. elegans* and related nematodes. *WormBook : The Online Review of C. Elegans Biology*, 1–19. <https://doi.org/10.1895/wormbook.1.115.2>
- Beavis, R. C., & Chait, B. T. (1990). Rapid, sensitive analysis of protein mixtures by mass spectrometry. *Proceedings of the National Academy of Sciences*, 87(17), 6873–6877. <https://doi.org/10.1073/pnas.87.17.6873>
- Bennett, D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health*, 25(5), 464–469. <https://doi.org/https://doi.org/10.1111/j.1467-842X.2001.tb00294.x>
- Bensaddek, D., Narayan, V., Nicolas, A., Brenes Murillo, A., Gartner, A., Kenyon, C. J., & Lamond, A. I. (2016). Micro-proteomics with iterative data analysis: Proteome analysis in *C. elegans* at the single worm level. *PROTEOMICS*, 16(3), 381–392. <https://doi.org/https://doi.org/10.1002/pmic.201500264>
- Berthod, A. (1991). Silica: backbone material of liquid chromatographic column packings. *Journal of Chromatography A*, 549(C), 1–28. [https://doi.org/10.1016/S0021-9673\(00\)91415-8](https://doi.org/10.1016/S0021-9673(00)91415-8)
- Bier, M. E., Park, M., & Syka, J. E. P. (1995). *Ion trap mass spectrometer system and method* (Patent No. 5,420,425).
- Blackburn, K., Mbeunkui, F., Mitra, S. K., Mentzel, T., & Goshe, M. B. (2010). Improving Protein and Proteome Coverage through Data-Independent Multiplexed Peptide Fragmentation. *Journal of Proteome Research*, 9(7), 3621–3637. <https://doi.org/10.1021/pr100144z>
- Bligh, E. G., & Dyer, W. J. (1959). A RAPID METHOD OF TOTAL LIPID EXTRACTION AND PURIFICATION. *Canadian Journal of Biochemistry and Physiology*, 37(8), 911–917. <https://doi.org/10.1139/o59-099>
- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72(1), 248–254. [https://doi.org/https://doi.org/10.1016/0003-2697\(76\)90527-3](https://doi.org/https://doi.org/10.1016/0003-2697(76)90527-3)

- Brassard, D., Malic, L., Normandin, F., Tabrizian, M., & Veres, T. (2008). Water-oil core-shell droplets for electrowetting-based digital microfluidic devices. *Lab on a Chip*, 8(8), 1342–1349. <https://doi.org/10.1039/B803827A>
- Brenner, S. (1974). THE GENETICS OF CAENORHABDITIS ELEGANS. *Genetics*, 77(1), 71 LP – 94. <http://www.genetics.org/content/77/1/71.abstract>
- Brodbelt, J. S. (2016). Ion Activation Methods for Peptides and Proteins. *Analytical Chemistry*, 88(1), 30–51. <https://doi.org/10.1021/acs.analchem.5b04563>
- Budnik, B., Levy, E., Harmange, G., & Slavov, N. (2018). SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biology*, 19(1), 161. <https://doi.org/10.1186/s13059-018-1547-5>
- Cassidy, L., Petersen, C., Treitz, C., Dierking, K., Schulenburg, H., Leippe, M., & Tholey, A. (2018). The Caenorhabditis elegans Proteome Response to Naturally Associated Microbiome Members of the Genus Ochrobactrum. *PROTEOMICS*, 18(8), 1700426. <https://doi.org/https://doi.org/10.1002/pmic.201700426>
- Chan, C., Peng, J., Rajesh, V., Scott, E. Y., Sklavounos, A. A., Faiz, M., & Wheeler, A. R. (2023). Digital Microfluidics for Microproteomic Analysis of Minute Mammalian Tissue Samples Enabled by a Photocleavable Surfactant. *Journal of Proteome Research*, 22(10), 3242–3253. <https://doi.org/10.1021/acs.jproteome.3c00281>
- Chapman, J. D., Goodlett, D. R., & Masselon, C. D. (2014). Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. *Mass Spectrometry Reviews*, 33(6), 452–470. <https://doi.org/https://doi.org/10.1002/mas.21400>
- Chatterjee, D., Ytterberg, A. J., Son, S. U., Loo, J. A., & Garrell, R. L. (2010). Integration of Protein Processing Steps on a Droplet Microfluidics Platform for MALDI-MS Analysis. *Analytical Chemistry*, 82(5), 2095–2101. <https://doi.org/10.1021/ac9029373>
- Chen, D., Geis-Asteggiante, L., Gomes, F. P., Ostrand-Rosenberg, S., & Fenselau, C. (2019). Top-Down Proteomic Characterization of Truncated Proteoforms. *Journal of Proteome Research*, 18(11), 4013–4019. <https://doi.org/10.1021/acs.jproteome.9b00487>

- Chen, E. I., Cociorva, D., Norris, J. L., & Yates, J. R. (2007). Optimization of Mass Spectrometry-Compatible Surfactants for Shotgun Proteomics. *Journal of Proteome Research*, 6(7), 2529–2538. <https://doi.org/10.1021/pr060682a>
- Chen, S., He, Z., Choi, S., & Novosselov, I. V. (2021). Characterization of Inkjet-Printed Digital Microfluidics Devices. In *Sensors* (Vol. 21, Issue 9). <https://doi.org/10.3390/s21093064>
- Choi, K., Ng, A. H. C., Fobel, R., Chang-Yen, D. A., Yarnell, L. E., Pearson, E. L., Oleksak, C. M., Fischer, A. T., Luoma, R. P., Robinson, J. M., Audet, J., & Wheeler, A. R. (2013). Automated Digital Microfluidic Platform for Magnetic-Particle-Based Immunoassays with Optimization by Design of Experiments. *Analytical Chemistry*, 85(20), 9638–9646. <https://doi.org/10.1021/ac401847x>
- Choi, K., Ng, A. H. C., Fobel, R., & Wheeler, A. R. (2012). Digital Microfluidics. *Annual Review of Analytical Chemistry*, 5(1), 413–440. <https://doi.org/10.1146/annurev-anchem-062011-143028>
- Cong, Y., Motamedchaboki, K., Misal, S. A., Liang, Y., Guise, A. J., Truong, T., Huguet, R., Plowey, E. D., Zhu, Y., Lopez-Ferrer, D., & Kelly, R. T. (2021). Ultrasensitive single-cell proteomics workflow identifies >1000 protein groups per mammalian cell. *Chemical Science*, 12(3), 1001–1006. <https://doi.org/10.1039/D0SC03636F>
- Cooper, H. J. (2016). To What Extent is FAIMS Beneficial in the Analysis of Proteins? *Journal of the American Society for Mass Spectrometry*, 27(4), 566–577. <https://doi.org/10.1007/s13361-015-1326-4>
- Corsi, A. K., Wightman, B., & Chalfie, M. (2015). A Transparent Window into Biology: A Primer on *Caenorhabditis elegans*. *Genetics*, 200(2), 387–407. <https://doi.org/10.1534/genetics.115.176099>
- Couvillion, S. P., Zhu, Y., Nagy, G., Adkins, J. N., Ansong, C., Renslow, R. S., Piehowski, P. D., Ibrahim, Y. M., Kelly, R. T., & Metz, T. O. (2019). New mass spectrometry technologies contributing towards comprehensive and high throughput omics analyses of single cells. *The Analyst*, 144(3), 794–807. <https://doi.org/10.1039/c8an01574k>
- Creese, A. J., & Cooper, H. J. (2012). Separation and Identification of Isomeric Glycopeptides by High Field Asymmetric Waveform Ion Mobility Spectrometry. *Analytical Chemistry*, 84(5), 2597–2601. <https://doi.org/10.1021/ac203321y>

- Cristobal, A., Marino, F., Post, H., van den Toorn, H. W. P., Mohammed, S., & Heck, A. J. R. (2017). Toward an Optimized Workflow for Middle-Down Proteomics. *Analytical Chemistry*, *89*(6), 3318–3325. <https://doi.org/10.1021/acs.analchem.6b03756>
- Ctortecka, C., Hartlmayr, D., Seth, A., Mendjan, S., Tourniaire, G., Udeshi, N. D., Carr, S. A., & Mechtler, K. (2023). An automated nanowell-array workflow for quantitative multiplexed single-cell proteomics sample preparation at high sensitivity. *Molecular & Cellular Proteomics*, *22*(12), 100665. <https://doi.org/https://doi.org/10.1016/j.mcpro.2023.100665>
- Dagley, L. F., Infusini, G., Larsen, R. H., Sandow, J. J., & Webb, A. I. (2019). Universal Solid-Phase Protein Preparation (USP3) for Bottom-up and Top-down Proteomics. *Journal of Proteome Research*, *18*(7), 2915–2924. <https://doi.org/10.1021/acs.jproteome.9b00217>
- Daly, M., Ansari, P., Häubl, G., Rogers, A., & Brunner, K. (2018). Assessing Almond and Peanut Allergens Using Commercially Available Immunoanalytical Kits and LC-MS/MS: A Case Study. *Journal of AOAC INTERNATIONAL*, *101*(1), 96–101. <https://doi.org/10.5740/jaoacint.17-0398>
- Das, A., Weise, C., Polack, M., Urban, R. D., Krafft, B., Hasan, S., Westphal, H., Warias, R., Schmidt, S., Gulder, T., & Belder, D. (2022). On-the-Fly Mass Spectrometry in Digital Microfluidics Enabled by a Microspray Hole: Toward Multidimensional Reaction Monitoring in Automated Synthesis Platforms. *Journal of the American Chemical Society*, *144*(23), 10353–10360. <https://doi.org/10.1021/jacs.2c01651>
- Davies, V., Wandy, J., Weidt, S., van der Hooft, J. J. J., Miller, A., Daly, R., & Rogers, S. (2021). Rapid Development of Improved Data-Dependent Acquisition Strategies. *Analytical Chemistry*, *93*(14), 5676–5683. <https://doi.org/10.1021/acs.analchem.0c03895>
- Denisov, E., Damoc, E., Lange, O., & Makarov, A. (2012). Orbitrap mass spectrometry with resolving powers above 1,000,000. *International Journal of Mass Spectrometry*, *325–327*, 80–85. <https://doi.org/https://doi.org/10.1016/j.ijms.2012.06.009>
- Dietz, L. A. (1965). Basic Properties of Electron Multiplier Ion Detection and Pulse Counting Methods in Mass Spectrometry. *Review of Scientific Instruments*, *36*(12), 1763–1770. <https://doi.org/10.1063/1.1719460>

- Ding, X., Li, P., Lin, S.-C. S., Stratton, Z. S., Nama, N., Guo, F., Slotcavage, D., Mao, X., Shi, J., Costanzo, F., & Huang, T. J. (2013). Surface acoustic wave microfluidics. *Lab on a Chip*, 13(18), 3626–3649. <https://doi.org/10.1039/C3LC50361E>
- Dixon, C., Ng, A. H. C., Fobel, R., Miltenburg, M. B., & Wheeler, A. R. (2016). An inkjet printed, roll-coated digital microfluidic device for inexpensive, miniaturized diagnostic assays. *Lab on a Chip*, 16(23), 4560–4568. <https://doi.org/10.1039/C6LC01064D>
- Dong, X., Xiong, L., Jiang, X., & Wang, Y. (2010). Quantitative Proteomic Analysis Reveals the Perturbation of Multiple Cellular Pathways in Jurkat-T Cells Induced by Doxorubicin. *Journal of Proteome Research*, 9(11), 5943–5951. <https://doi.org/10.1021/pr1007043>
- Dongré, A. R., Jones, J. L., Somogyi, Á., & Wysocki, V. H. (1996). Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *Journal of the American Chemical Society*, 118(35), 8365–8374. <https://doi.org/10.1021/ja9542193>
- Doucette, A. A., Vieira, D. B., Orton, D. J., & Wall, M. J. (2014). Resolubilization of Precipitated Intact Membrane Proteins with Cold Formic Acid for Analysis by Mass Spectrometry. *Journal of Proteome Research*, 13(12), 6001–6012. <https://doi.org/10.1021/pr500864a>
- Douglas, D. J., Frank, A. J., & Mao, D. (2005). Linear ion traps in mass spectrometry. *Mass Spectrometry Reviews*, 24(1), 1–29. <https://doi.org/https://doi.org/10.1002/mas.20004>
- Drapeau, G. R., Boily, Y., & Houmard, J. (1972). Purification and Properties of an Extracellular Protease of *Staphylococcus aureus*. *Journal of Biological Chemistry*, 247(20), 6720–6726. [https://doi.org/https://doi.org/10.1016/S0021-9258\(19\)44749-2](https://doi.org/https://doi.org/10.1016/S0021-9258(19)44749-2)
- Eshghi, A., Xie, X., Hardie, D., Chen, M. X., Izaguirre, F., Newman, R., Zhu, Y., Kelly, R. T., & Goodlett, D. R. (2023). Sample Preparation Methods for Targeted Single-Cell Proteomics. *Journal of Proteome Research*, 22(6), 1589–1602. <https://doi.org/10.1021/acs.jproteome.2c00429>
- Eydelnant, I. A., Uddayasankar, U., Li, B., 'Betty,' Liao, M. W., & Wheeler, A. R. (2012). Virtual microwells for digital microfluidic reagent dispensing and cell culture. *Lab on a Chip*, 12(4), 750–757. <https://doi.org/10.1039/C2LC21004E>

- Félix, M.-A., & Braendle, C. (2010). The natural history of *Caenorhabditis elegans*. *Current Biology*, 20(22), R965–R969. <https://doi.org/https://doi.org/10.1016/j.cub.2010.09.050>
- Fenn, J. B., Mann, M., Meng Chin, K., Wong Shek, F., & Whitehouse, C. M. (1989). Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science*, 246(4926), 64–71. <https://doi.org/10.1126/science.2675315>
- Fobel, R., Kirby, A. E., Ng, A. H. C., Farnood, R. R., & Wheeler, A. R. (2014). Paper Microfluidics Goes Digital. *Advanced Materials*, 26(18), 2838–2843. <https://doi.org/https://doi.org/10.1002/adma.201305168>
- Folch, J., Lees, M., & Stanley, G. H. S. (1957). A SIMPLE METHOD FOR THE ISOLATION AND PURIFICATION OF TOTAL LIPIDES FROM ANIMAL TISSUES. *Journal of Biological Chemistry*, 226(1), 497–509. [https://doi.org/https://doi.org/10.1016/S0021-9258\(18\)64849-5](https://doi.org/https://doi.org/10.1016/S0021-9258(18)64849-5)
- Fulcher, J. M., Makaju, A., Moore, R. J., Zhou, M., Bennett, D. A., De Jager, P. L., Qian, W.-J., Paša-Tolić, L., & Petyuk, V. A. (2021). Enhancing Top-Down Proteomics of Brain Tissue with FAIMS. *Journal of Proteome Research*, 20(5), 2780–2795. <https://doi.org/10.1021/acs.jproteome.1c00049>
- Garcia, B. A. (2010). What does the future hold for top down mass spectrometry? *Journal of the American Society for Mass Spectrometry*, 21(2), 193–202. <https://doi.org/10.1016/j.jasms.2009.10.014>
- 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. (2022). Streamlined single-cell proteomics by an integrated microfluidic chip and data-independent acquisition mass spectrometry. *Nature Communications*, 13(1), 37. <https://doi.org/10.1038/s41467-021-27778-4>
- Geiszler, D. J., Kong, A. T., Avtonomov, D. M., Yu, F., da Veiga Leprevost, F., & Nesvizhskii, A. I. (2021). PTM-shepherd: Analysis and summarization of post-translational and chemical modifications from open search results. *Molecular and Cellular Proteomics*, 20, 100018. <https://doi.org/10.1074/MCP.TIR120.002216>
- Geromanos, S. J., Vissers, J. P. C., Silva, J. C., Dorschel, C. A., Li, G.-Z., Gorenstein, M. V., Bateman, R. H., & Langridge, J. I. (2009). The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *PROTEOMICS*, 9(6), 1683–1695. <https://doi.org/https://doi.org/10.1002/pmic.200800562>

- Giansanti, P., Tsiatsiani, L., Low, T. Y., & Heck, A. J. R. (2016). Six alternative proteases for mass spectrometry–based proteomics beyond trypsin. *Nature Protocols*, *11*(5), 993–1006. <https://doi.org/10.1038/nprot.2016.057>
- Golden, N. L., Plagens, R. N., Kim Guisbert, K. S., & Guisbert, E. (2020). Standardized methods for measuring induction of the heat shock response in *Caenorhabditis elegans*. *Journal of Visualized Experiments*, *2020*(161), 1–13. <https://doi.org/10.3791/61030>
- González-Benito, J., & Koenig, J. L. (2006). Nature of PMMA dissolution process by mixtures of acetonitrile/alcohol (poor solvent/nonsolvent) monitored by FTIR-imaging. *Polymer*, *47*(9), 3065–3072. <https://doi.org/https://doi.org/10.1016/j.polymer.2006.02.086>
- Groele, R. J., Krasicky, P. D., Chun, S.-W., Sullivan, J., & Rodriguez, F. (1991). Dissolution rates of poly(methyl methacrylate) in mixtures of nonsolvents. *Journal of Applied Polymer Science*, *42*(1), 3–8. <https://doi.org/https://doi.org/10.1002/app.1991.070420102>
- GuhaThakurta, D., Palomar, L., Stormo, G. D., Tedesco, P., Johnson, T. E., Walker, D. W., Lithgow, G., Kim, S., & Link, C. D. (2002). Identification of a novel cis-regulatory element involved in the heat shock response in *Caenorhabditis elegans* using microarray gene expression and computational methods. *Genome Research*, *12*(5), 701–712. <https://doi.org/10.1101/gr.228902>
- Hale, J. E., Butler, J. P., Gelfanova, V., You, J. S., & Knierman, M. D. (2004). A simplified procedure for the reduction and alkylation of cysteine residues in proteins prior to proteolytic digestion and mass spectral analysis. *Analytical Biochemistry*, *333*(1), 174–181. <https://doi.org/10.1016/j.ab.2004.04.013>
- Hamid, Z., Zimmerman, K. D., Guillen-Ahlers, H., Li, C., Nathanielsz, P., Cox, L. A., & Olivier, M. (2022). Assessment of label-free quantification and missing value imputation for proteomics in non-human primates. *BMC Genomics*, *23*(1), 496. <https://doi.org/10.1186/s12864-022-08723-1>
- He, Y., Rshan, E. H., Linke, V., Shishkova, E., Hebert, A. S., Jochem, A., Westphall, M. S., Pagliarini, D. J., Overmyer, K. A., & Coon, J. J. (2021). Multi-Omic Single-Shot Technology for Integrated Proteome and Lipidome Analysis. *Analytical Chemistry*, *93*(9), 4217–4222. <https://doi.org/10.1021/acs.analchem.0c04764>

- Hebert, A. S., Prasad, S., Belford, M. W., Bailey, D. J., McAlister, G. C., Abbatiello, S. E., Huguet, R., Wouters, E. R., Dunyach, J.-J., Brademan, D. R., Westphall, M. S., & Coon, J. J. (2018). Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer. *Analytical Chemistry*, *90*(15), 9529–9537. <https://doi.org/10.1021/acs.analchem.8b02233>
- Ho, M., Au, A., Flick, R., Vuong, T. V., Sklavounos, A. A., Swyer, I., Yip, C. M., & Wheeler, A. R. (2023). Antifouling Properties of Pluronic and Tetronic Surfactants in Digital Microfluidics. *ACS Applied Materials & Interfaces*, *15*(5), 6326–6337. <https://doi.org/10.1021/acsami.2c17317>
- Hodne, K., & Weltzien, F. A. (2015). Single-cell isolation and gene analysis: Pitfalls and possibilities. In *International Journal of Molecular Sciences* (Vol. 16, Issue 11, pp. 26832–26849). <https://doi.org/10.3390/ijms161125996>
- Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M., & Graham Cooks, R. (2005). The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry*, *40*(4), 430–443. <https://doi.org/https://doi.org/10.1002/jms.856>
- Hu, S., Ye, J., Shi, S., Yang, C., Jin, K., Hu, C., Wang, D., & Ma, H. (2023). Large-Area Electronics-Enabled High-Resolution Digital Microfluidics for Parallel Single-Cell Manipulation. *Analytical Chemistry*, *95*(17), 6905–6914. <https://doi.org/10.1021/acs.analchem.3c00150>
- Huang, Y., Triscari, J. M., Tseng, G. C., Pasa-Tolic, L., Lipton, M. S., Smith, R. D., & Wysocki, V. H. (2005). Statistical Characterization of the Charge State and Residue Dependence of Low-Energy CID Peptide Dissociation Patterns. *Analytical Chemistry*, *77*(18), 5800–5813. <https://doi.org/10.1021/ac0480949>
- Hug, H. (2015). *Instrumentelle Analytik: Theorie und Praxis* (1st ed.). Verlag Europa Lehrmittel, Vollmer GmbH & Co KG.
- Hughes, C. S., Foehr, S., Garfield, D. A., Furlong, E. E., Steinmetz, L. M., & Krijgsveld, J. (2014). Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular Systems Biology*, *10*(10), 757. <https://doi.org/10.15252/msb.20145625>
- Hughes, C. S., Moggridge, S., Müller, T., Sorensen, P. H., Morin, G. B., & Krijgsveld, J. (2019). Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nature Protocols*, *14*(1), 68–85. <https://doi.org/10.1038/s41596-018-0082-x>

- Jakoby, T., van den Berg, B. H. J., & Tholey, A. (2012). Quantitative Protease Cleavage Site Profiling using Tandem-Mass-Tag Labeling and LC–MALDI-TOF/TOF MS/MS Analysis. *Journal of Proteome Research*, *11*(3), 1812–1820.
<https://doi.org/10.1021/pr201051e>
- Jebrail, M. J., Luk, V. N., Shih, S. C. C., Fobel, R., Ng, A. H. C., Yang, H., Freire, S. L. S., & Wheeler, A. R. (2010). Digital microfluidics for automated proteomic processing. *Journal of Visualized Experiments*, *33*, 1–5. <https://doi.org/10.3791/1603>
- Jia, Y., Narayanan, J., Liu, X.-Y., & Liu, Y. (2005). Investigation on the Mechanism of Crystallization of Soluble Protein in the Presence of Nonionic Surfactant. *Biophysical Journal*, *89*(6), 4245–4251.
<https://doi.org/https://doi.org/10.1529/biophysj.105.066449>
- Jovic, K., Sterken, M. G., Grilli, J., Bevers, R. P. J., Rodriguez, M., Riksen, J. A. G., Allesina, S., Kammenga, J. E., & Snoek, L. B. (2017). Temporal dynamics of gene expression in heat-stressed *Caenorhabditis elegans*. *PLoS ONE*, *12*(12), e0189445.
<https://doi.org/10.1371/journal.pone.0189445>
- Jungblut, P. R., Holzhütter, H. G., Apweiler, R., & Schlüter, H. (2008). The speciation of the proteome. *Chemistry Central Journal*, *2*(1), 16. <https://doi.org/10.1186/1752-153X-2-16>
- Kalli, A., & Hess, S. (2012). Effect of mass spectrometric parameters on peptide and protein identification rates for shotgun proteomic experiments on an LTQ-orbitrap mass analyzer. *PROTEOMICS*, *12*(1), 21–31.
<https://doi.org/https://doi.org/10.1002/pmic.201100464>
- Kanithamniyom, P., Zhou, A., Feng, S., Liu, A., Vasoo, S., & Zhang, Y. (2020). A 3D-printed modular magnetic digital microfluidic architecture for on-demand bioanalysis. *Microsystems & Nanoengineering*, *6*(1), 48. <https://doi.org/10.1038/s41378-020-0152-4>
- Karas, M., Bachmann, D., & Hillenkamp, F. (1985). Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Analytical Chemistry*, *57*(14), 2935–2939. <https://doi.org/10.1021/ac00291a042>

- Karas, M., Glückmann, M., & Schäfer, J. (2000). Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *Journal of Mass Spectrometry*, *35*(1), 1–12. [https://doi.org/https://doi.org/10.1002/\(SICI\)1096-9888\(200001\)35:1<1::AID-JMS904>3.0.CO;2-0](https://doi.org/https://doi.org/10.1002/(SICI)1096-9888(200001)35:1<1::AID-JMS904>3.0.CO;2-0)
- Kasuga, K., Katoh, Y., Nagase, K., & Igarashi, K. (2017). Microproteomics with microfluidic-based cell sorting: Application to 1000 and 100 immune cells. *PROTEOMICS*, *17*(13–14), 1600420. <https://doi.org/https://doi.org/10.1002/pmic.201600420>
- Kaulich, P. T., Cassidy, L., Bartel, J., Schmitz, R. A., & Tholey, A. (2021). Multi-protease Approach for the Improved Identification and Molecular Characterization of Small Proteins and Short Open Reading Frame-Encoded Peptides. *Journal of Proteome Research*, *20*(5), 2895–2903. <https://doi.org/10.1021/acs.jproteome.1c00115>
- Kaulich, P. T., Cassidy, L., Winkels, K., & Tholey, A. (2022). Improved Identification of Proteoforms in Top-Down Proteomics Using FAIMS with Internal CV Stepping. *Analytical Chemistry*, *94*(8), 3600–3607. <https://doi.org/10.1021/acs.analchem.1c05123>
- Kimble, J., & Hirsh, D. (1979). The postembryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*. *Developmental Biology*, *70*(2), 396–417. [https://doi.org/https://doi.org/10.1016/0012-1606\(79\)90035-6](https://doi.org/https://doi.org/10.1016/0012-1606(79)90035-6)
- Kingdon, K. H. (1923). A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Physical Review*, *21*(4), 408–418. <https://doi.org/10.1103/PhysRev.21.408>
- Kirienko, N. V., Cezairliyan, B. O., Ausubel, F. M., & Powell, J. R. (2014). *Pseudomonas aeruginosa* PA14 pathogenesis in *Caenorhabditis elegans*. In A. Filloux & J.-L. Ramos (Eds.), *Methods in Molecular Biology* (Vol. 1149, pp. 653–669). Springer New York. https://doi.org/10.1007/978-1-4939-0473-0_50
- Klunk, W. E., & Pettegrew, J. W. (1990). Alzheimer's β -Amyloid Protein Is Covalently Modified when Dissolved in Formic Acid. *Journal of Neurochemistry*, *54*(6), 2050–2056. <https://doi.org/https://doi.org/10.1111/j.1471-4159.1990.tb04910.x>
- Knight, R. D. (1981). Storage of ions from laser-produced plasmas. *Applied Physics Letters*, *38*(4), 221–223. <https://doi.org/10.1063/1.92315>

- Knipes, A. K., Summers, A., Sklavounos, A. A., Lamanna, J., de Campos, R. P. S., Narahari, T., Dixon, C., Fobel, R., Ndjakani, Y. D., Lubula, L., Magazani, A., Muyembe, J. J., Lay, Y., Pukuta, E., Waku-Kouomou, D., Hao, L., Kayembe, J. K., Fobel, C., Dahmer, J., ... Scobie, H. M. (2022). Use of a rapid digital microfluidics-powered immunoassay for assessing measles and rubella infection and immunity in outbreak settings in the Democratic Republic of the Congo. *PLOS ONE*, *17*(12), e0278749. <https://doi.org/10.1371/journal.pone.0278749>
- Kohl, K., Fleming, T., Acunman, K., Hammes, H. P., Morcos, M., & Schlotterer, A. (2018). Plate-based large-scale cultivation of *Caenorhabditis elegans*: Sample preparation for the study of metabolic alterations in diabetes. *Journal of Visualized Experiments*, *2018*(138). <https://doi.org/10.3791/58117>
- Kothamachu, V. B., Zaini, S., & Muffatto, F. (2020). Role of Digital Microfluidics in Enabling Access to Laboratory Automation and Making Biology Programmable. *SLAS TECHNOLOGY: Translating Life Sciences Innovation*, *25*(5), 411–426. <https://doi.org/10.1177/2472630320931794>
- Lamanna, J., Scott, E. Y., Edwards, H. S., Chamberlain, M. D., Dryden, M. D. M., Peng, J., Mair, B., Lee, A., Chan, C., Sklavounos, A. A., Heffernan, A., Abbas, F., Lam, C., Olson, M. E., Moffat, J., & Wheeler, A. R. (2020). Digital microfluidic isolation of single cells for -Omics. *Nature Communications*, *11*(1), 5632. <https://doi.org/10.1038/s41467-020-19394-5>
- Lazar, C., Gatto, L., Ferro, M., Bruley, C., & Burger, T. (2016). Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. *Journal of Proteome Research*, *15*(4), 1116–1125. <https://doi.org/10.1021/acs.jproteome.5b00981>
- Leduc, A., Huffman, R. G., Cantlon, J., Khan, S., & Slavov, N. (2022). Exploring functional protein covariation across single cells using nPOP. *Genome Biology*, *23*(1), 261. <https://doi.org/10.1186/s13059-022-02817-5>
- Leduc, A., Huffman, R. G., & Slavov, N. (2021). Droplet sample preparation for single-cell proteomics applied to the cell cycle. *BioRxiv*, *Cdc*, 2021.04.24.441211. <https://doi.org/10.1101/2021.04.24.441211>

- LeDuc, R. D., Deutsch, E. W., Binz, P.-A., Fellers, R. T., Cesnik, A. J., Klein, J. A., Van Den Bossche, T., Gabriels, R., Yalavarthi, A., Perez-Riverol, Y., Carver, J., Bittremieux, W., Kawano, S., Pullman, B., Bandeira, N., Kelleher, N. L., Thomas, P. M., & Vizcaíno, J. A. (2022). Proteomics Standards Initiative's ProForma 2.0: Unifying the Encoding of Proteoforms and Peptidoforms. *Journal of Proteome Research*, 21(4), 1189–1195. <https://doi.org/10.1021/acs.jproteome.1c00771>
- LeDuc, R. D., Schwämmle, V., Shortreed, M. R., Cesnik, A. J., Solntsev, S. K., Shaw, J. B., Martin, M. J., Vizcaino, J. A., Alpi, E., Danis, P., Kelleher, N. L., Smith, L. M., Ge, Y., Agar, J. N., Chamot-Rooke, J., Loo, J. A., Pasa-Tolic, L., & Tsybin, Y. O. (2018). ProForma: A Standard Proteoform Notation. *Journal of Proteome Research*, 17(3), 1321–1325. <https://doi.org/10.1021/acs.jproteome.7b00851>
- Lee, H. J., McAuley, A., Schilke, K. F., & McGuire, J. (2011). Molecular origins of surfactant-mediated stabilization of protein drugs. *Advanced Drug Delivery Reviews*, 63(13), 1160–1171. <https://doi.org/https://doi.org/10.1016/j.addr.2011.06.015>
- Leipert, J., Kaulich, P. T., Steinbach, M. K., Steer, B., Winkels, K., Blurton, C., Leippe, M., & Tholey, A. (2023). Digital Microfluidics and Magnetic Bead-Based Intact Proteoform Elution for Quantitative Top-down Nanoproteomics of Single *C. elegans* Nematodes. *Angewandte Chemie - International Edition*, 62(28), e202301969. <https://doi.org/10.1002/anie.202301969>
- Leipert, J., Steinbach, M. K., & Tholey, A. (2021). Isobaric Peptide Labeling on Digital Microfluidics for Quantitative Low Cell Number Proteomics. *Analytical Chemistry*, 93(15), 6278–6286. <https://doi.org/10.1021/acs.analchem.1c01205>
- Leipert, J., & Tholey, A. (2019). 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 on a Chip*, 19(20), 3490–3498. <https://doi.org/10.1039/c9lc00715f>
- Leirs, K., Dal Dosso, F., Perez-Ruiz, E., Decrop, D., Cops, R., Huff, J., Hayden, M., Collier, N., Yu, K. X. Z., Brown, S., & Lammertyn, J. (2022). Bridging the Gap between Digital Assays and Point-of-Care Testing: Automated, Low Cost, and Ultrasensitive Detection of Thyroid Stimulating Hormone. *Analytical Chemistry*, 94(25), 8919–8927. <https://doi.org/10.1021/acs.analchem.2c00480>

- Lenčo, J., Khalikova, M. A., & Švec, F. (2020). Dissolving Peptides in 0.1% Formic Acid Brings Risk of Artificial Formylation. *Journal of Proteome Research*, *19*(3), 993–999. <https://doi.org/10.1021/acs.jproteome.9b00823>
- Levin, Y., Hradetzky, E., & Bahn, S. (2011). Quantification of proteins using data-independent analysis (MSE) in simple and complex samples: A systematic evaluation. *PROTEOMICS*, *11*(16), 3273–3287. <https://doi.org/https://doi.org/10.1002/pmic.201000661>
- Levy, M. J., Washburn, M. P., & Florens, L. (2018). Probing the Sensitivity of the Orbitrap Lumos Mass Spectrometer Using a Standard Reference Protein in a Complex Background. *Journal of Proteome Research*, *17*(10), 3586–3592. <https://doi.org/10.1021/acs.jproteome.8b00269>
- Li, Z.-Y., Huang, M., Wang, X.-K., Zhu, Y., Li, J.-S., Wong, C. C. L., & Fang, Q. (2018). Nanoliter-Scale Oil-Air-Droplet Chip-Based Single Cell Proteomic Analysis. *Analytical Chemistry*, *90*(8), 5430–5438. <https://doi.org/10.1021/acs.analchem.8b00661>
- Liang, Y., Acor, H., McCown, M. A., Nwosu, A. J., Boekweg, H., Axtell, N. B., Truong, T., Cong, Y., Payne, S. H., & Kelly, R. T. (2021). Fully Automated Sample Processing and Analysis Workflow for Low-Input Proteome Profiling. *Analytical Chemistry*, *93*(3), 1658–1666. <https://doi.org/10.1021/acs.analchem.0c04240>
- Liu, L., Chen, D., Wang, J., & Chen, J. (2020). Advances of Single-Cell Protein Analysis. *Cells*, *9*(5), 1271. <https://doi.org/10.3390/cells9051271>
- Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B. C., & Aebersold, R. (2018). Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Molecular Systems Biology*, *14*(8), e8126. <https://doi.org/https://doi.org/10.15252/msb.20178126>
- Luk, V. N., Mo, G. C. H., & Wheeler, A. R. (2008). Pluronic Additives: A Solution to Sticky Problems in Digital Microfluidics. *Langmuir*, *24*(12), 6382–6389. <https://doi.org/10.1021/la7039509>
- Luk, V. N., & Wheeler, A. R. (2009). A Digital Microfluidic Approach to Proteomic Sample Processing. *Analytical Chemistry*, *81*(11), 4524–4530. <https://doi.org/10.1021/ac900522a>

- Mádi, A., Mikkat, S., Ringel, B., Ulbrich, M., Thiesen, H.-J., & Glocker, M. O. (2003). Mass spectrometric proteome analysis for profiling temperature-dependent changes of protein expression in wild-type *Caenorhabditis elegans*. *PROTEOMICS*, 3(8), 1526–1534. <https://doi.org/https://doi.org/10.1002/pmic.200300490>
- Makarov, A. (2000). Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Analytical Chemistry*, 72(6), 1156–1162. <https://doi.org/10.1021/ac991131p>
- Marty, M. T. (2022). Fundamentals: How Do We Calculate Mass, Error, and Uncertainty in Native Mass Spectrometry? *Journal of the American Society for Mass Spectrometry*, 33(10), 1807–1812. <https://doi.org/10.1021/jasms.2c00218>
- Masuda, T., Inamori, Y., Furukawa, A., Yamahiro, M., Momosaki, K., Chang, C.-H., Kobayashi, D., Ohguchi, H., Kawano, Y., Ito, S., Araki, N., Ong, S.-E., & Ohtsuki, S. (2022). Water Droplet-in-Oil Digestion Method for Single-Cell Proteomics. *Analytical Chemistry*, 94(29), 10329–10336. <https://doi.org/10.1021/acs.analchem.1c05487>
- Matyash, V., Liebisch, G., Kurzchalia, T. V., Shevchenko, A., & Schwudke, D. (2008). Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics[†]. *Journal of Lipid Research*, 49(5), 1137–1146. <https://doi.org/https://doi.org/10.1194/jlr.D700041-JLR200>
- McLuckey, S. A. (1992). Principles of collisional activation in analytical mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 3(6), 599–614. [https://doi.org/10.1016/1044-0305\(92\)85001-Z](https://doi.org/10.1016/1044-0305(92)85001-Z)
- Moggridge, S., Sorensen, P. H., Morin, G. B., & Hughes, C. S. (2018). Extending the Compatibility of the SP3 Paramagnetic Bead Processing Approach for Proteomics. *Journal of Proteome Research*, 17(4), 1730–1740. <https://doi.org/10.1021/acs.jproteome.7b00913>
- Muehlbauer, L. K., Jen, A., Zhu, Y., He, Y., Shishkova, E., Overmyer, K. A., & Coon, J. J. (2023). Rapid Multi-Omics Sample Preparation for Mass Spectrometry. *Analytical Chemistry*, 95(2), 659–667. <https://doi.org/10.1021/acs.analchem.2c02042>
- Napier, J. A., Hey, S. J., Lacey, D. J., & Shewry, P. R. (1998). Identification of a *Caenorhabditis elegans* Delta6-fatty-acid-desaturase by heterologous expression in *Saccharomyces cerevisiae*. *The Biochemical Journal*, 330 (Pt 2(Pt 2)), 611–614. <https://doi.org/10.1042/bj3300611>

- Nazario, C. E. D., & Lanças, F. M. (2013). Suportes cromatográficos e fases estacionárias para cromatografia líquida: preparo, evolução e tendências [Chromatographic supports and stationary phases for liquid chromatography: Preparation, evolution and trends] (in Portuguese). *Scientia Chromatographica*, 5(2), 111–135.
<https://doi.org/10.4322/sc.2013.009>
- Nesvizhskii, A. I., & Aebersold, R. (2005). Interpretation of Shotgun Proteomic Data. *Molecular & Cellular Proteomics*, 4(10), 1419–1440.
<https://doi.org/https://doi.org/10.1074/mcp.R500012-MCP200>
- Nesvizhskii, A. I., Keller, A., Kolker, E., & Aebersold, R. (2003). A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. *Analytical Chemistry*, 75(17), 4646–4658. <https://doi.org/10.1021/ac0341261>
- Neukomm, L. J., Nicot, A.-S. S., Kinchen, J. M., Almendinger, J., Pinto, S. M., Zeng, S., Doukometzidis, K., Tronchère, H., Payrastra, B., Laporte, J. F., & Hengartner, M. O. (2011). The phosphoinositide phosphatase MTM-1 regulates apoptotic cell corpse clearance through CED-5–CED-12 in *C. elegans*. *Development*, 138(10), 2003–2014. <https://doi.org/10.1242/dev.060012>
- Neve, I. A. A., Sowa, J. N., Lin, C.-C. J., Sivaramakrishnan, P., Herman, C., Ye, Y., Han, L., & Wang, M. C. (2020). *Escherichia coli* Metabolite Profiling Leads to the Development of an RNA Interference Strain for *Caenorhabditis elegans*. *G3: Genes/Genomes/Genetics*, 10(1), 189 LP – 198.
<https://doi.org/10.1534/g3.119.400741>
- Ng, A. H. C., Fobel, R., Fobel, C., Lamanna, J., Rackus, D. G., Summers, A., Dixon, C., Dryden, M. D. M., Lam, C., Ho, M., Mufti, N. S., Lee, V., Asri, M. A. M., Sykes, E. A., Chamberlain, M. D., Joseph, R., Ope, M., Scobie, H. M., Knipes, A., ... Wheeler, A. R. (2018). A digital microfluidic system for serological immunoassays in remote settings. *Science Translational Medicine*, 10(438), eaar6076.
<https://doi.org/10.1126/scitranslmed.aar6076>
- Oliveros, J. C. (2007). *Venny. An interactive tool for comparing lists with Venn's diagrams*. <https://bioinfogp.cnb.csic.es/tools/venny/index.html>

- Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., & Mann, M. (2002). Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics*. *Molecular & Cellular Proteomics*, 1(5), 376–386. <https://doi.org/https://doi.org/10.1074/mcp.M200025-MCP200>
- Pamula, V. K., Srinivasan, V., Chakrapani, H., Fair, R. B., & Toone, E. J. (2005). A droplet-based lab-on-a-chip for colorimetric detection of nitroaromatic explosives. *18th IEEE International Conference on Micro Electro Mechanical Systems, 2005. MEMS 2005.*, 722–725. <https://doi.org/10.1109/MEMSYS.2005.1454031>
- Pan, W., Huang, X., Guo, Z., Nagarajan, R., & Mylonakis, E. (2021). Identification and Functional Analysis of Cytokine-Like Protein CLEC-47 in *Caenorhabditis elegans*. *MBio*, 12(5), e0257921–e0257921. <https://doi.org/10.1128/mBio.02579-21>
- Panek, J., Gang, S. S., Reddy, K. C., Luallen, R. J., Fulzele, A., Bennett, E. J., & Troemel, E. R. (2020). A cullin-RING ubiquitin ligase promotes thermotolerance as part of the intracellular pathogen response in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences*, 117(14), 7950–7960. <https://doi.org/10.1073/pnas.1918417117>
- Peng, J., Chan, C., Zhang, S., Sklavounos, A. A., Olson, M. E., Scott, E. Y., Hu, Y., Rajesh, V., Li, B. B., Chamberlain, M. D., Zhang, S., Peng, H., & Wheeler, A. R. (2023). All-in-One digital microfluidics pipeline for proteomic sample preparation and analysis. *Chemical Science*, 14(11), 2887–2900. <https://doi.org/10.1039/d3sc00560g>
- Petelski, A. A., Emmott, E., Leduc, A., Huffman, R. G., Specht, H., Perlman, D. H., & Slavov, N. (2021). Multiplexed single-cell proteomics using SCoPE2. *Nature Protocols*, 16(12), 5398–5425. <https://doi.org/10.1038/s41596-021-00616-z>
- Rea, S. L., Wu, D., Cypser, J. R., Vaupel, J. W., & Johnson, T. E. (2005). A stress-sensitive reporter predicts longevity in isogenic populations of *Caenorhabditis elegans*. *Nature Genetics*, 37(8), 894–898. <https://doi.org/10.1038/ng1608>
- Roeder, T., Stanisak, M., Gelhaus, C., Bruchhaus, I., Grötzinger, J., & Leippe, M. (2010). Caenopores are antimicrobial peptides in the nematode *Caenorhabditis elegans* instrumental in nutrition and immunity. *Developmental & Comparative Immunology*, 34(2), 203–209. <https://doi.org/https://doi.org/10.1016/j.dci.2009.09.010>

- Samiei, E., Tabrizian, M., & Hoorfar, M. (2016). A review of digital microfluidics as portable platforms for lab-on a-chip applications. *Lab on a Chip*, *16*(13), 2376–2396. <https://doi.org/10.1039/C6LC00387G>
- Sarvin, B., Lagziel, S., Sarvin, N., Mukha, D., Kumar, P., Aizenshtein, E., & Shlomi, T. (2020). Fast and sensitive flow-injection mass spectrometry metabolomics by analyzing sample-specific ion distributions. *Nature Communications*, *11*(1), 3186. <https://doi.org/10.1038/s41467-020-17026-6>
- Savaryn, J. P., Toby, T. K., & Kelleher, N. L. (2016). A researcher's guide to mass spectrometry-based proteomics. *Proteomics*, *16*(18), 2435–2443. <https://doi.org/https://doi.org/10.1002/pmic.201600113>
- Schlüter, H., Apweiler, R., Holzhütter, H.-G., & Jungblut, P. R. (2009). Finding one's way in proteomics: a protein species nomenclature. *Chemistry Central Journal*, *3*(1), 11. <https://doi.org/10.1186/1752-153X-3-11>
- Schwartz, J. C., Senko, M. W., & Syka, J. E. P. (2002). A two-dimensional quadrupole ion trap mass spectrometer. *Journal of the American Society for Mass Spectrometry*, *13*(6), 659–669. [https://doi.org/https://doi.org/10.1016/S1044-0305\(02\)00384-7](https://doi.org/https://doi.org/10.1016/S1044-0305(02)00384-7)
- Scott, E. Y., Safarian, N., Casasbuenas, D. L., Dryden, M., Tockovska, T., Ali, S., Peng, J., Daniele, E., Nie Xin Lim, I., Bang, K. W. A., Tripathy, S., Yuzwa, S. A., Wheeler, A. R., & Faiz, M. (2024). Integrating single-cell and spatially resolved transcriptomic strategies to survey the astrocyte response to stroke in male mice. *Nature Communications*, *15*(1), 1584. <https://doi.org/10.1038/s41467-024-45821-y>
- Shamsi, M. H., Choi, K., Ng, A. H. C., & Wheeler, A. R. (2014). A digital microfluidic electrochemical immunoassay. *Lab on a Chip*, *14*(3), 547–554. <https://doi.org/10.1039/C3LC51063H>
- Shaye, D. D., & Greenwald, I. (2011). OrthoList: A Compendium of *C. elegans* Genes with Human Orthologs. *PLOS ONE*, *6*(5), e20085. <https://doi.org/10.1371/journal.pone.0020085>
- Shen, J., Liao, J., Liu, H., Liu, C., Li, C., Cheng, H., Yang, H., & Chen, H. (2023). A low-temperature digital microfluidic system used for protein–protein interaction detection. *Lab on a Chip*, *23*(20), 4390–4399. <https://doi.org/10.1039/D3LC00386H>

- Shen, Y., & Smith, R. D. (2005). Advanced nanoscale separations and mass spectrometry for sensitive high-throughput proteomics. *Expert Review of Proteomics*, 2(3), 431–447. <https://doi.org/10.1586/14789450.2.3.431>
- Sielaff, M., Kuharev, J., Bohn, T., Hahlbrock, J., Bopp, T., Tenzer, S., & Distler, U. (2017). Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range. *Journal of Proteome Research*, 16(11), 4060–4072. <https://doi.org/10.1021/acs.jproteome.7b00433>
- Smith, L. M., & Kelleher, N. L. (2018). Proteoforms as the next proteomics currency. *Science*, 359(6380), 1106–1107. <https://doi.org/10.1126/science.aat1884>
- Smith, L. M., Kelleher, N. L., & Proteomics, T. C. for T. D. (2013). Proteoform: a single term describing protein complexity. *Nature Methods*, 10(3), 186–187. <https://doi.org/10.1038/nmeth.2369>
- Sok, J., Calton, M., Lu, J., Lichtlen, P., Clark, S. G., & Ron, D. (2001). Arsenite-inducible RNA-associated protein (AIRAP) protects cells from arsenite toxicity. *Cell Stress & Chaperones*, 6(1), 6–15. [https://doi.org/10.1379/1466-1268\(2001\)006<0006:airapa>2.0.co;2](https://doi.org/10.1379/1466-1268(2001)006<0006:airapa>2.0.co;2)
- Song, Y., Tian, T., Shi, Y., Liu, W., Zou, Y., Khajvand, T., Wang, S., Zhu, Z., & Yang, C. (2017). Enrichment and single-cell analysis of circulating tumor cells. *Chemical Science*, 8(3), 1736–1751. <https://doi.org/10.1039/C6SC04671A>
- Specht, H., & Slavov, N. (2018). Transformative Opportunities for Single-Cell Proteomics. *Journal of Proteome Research*, 17(8), 2565–2571. <https://doi.org/10.1021/acs.jproteome.8b00257>
- Srinivasan, V., Pamula, V. K., & Fair, R. B. (2004). Droplet-based microfluidic lab-on-a-chip for glucose detection. *Analytica Chimica Acta*, 507(1), 145–150. <https://doi.org/https://doi.org/10.1016/j.aca.2003.12.030>
- Steinbach, M. K., Leipert, J., Blurton, C., Leippe, M., & Tholey, A. (2022). Digital Microfluidics Supported Microproteomics for Quantitative Proteome Analysis of Single *Caenorhabditis elegans* Nematodes. *Journal of Proteome Research*, 21(8), 1986–1996. <https://doi.org/10.1021/acs.jproteome.2c00274>

- Steinbach, M. K., Leipert, J., Matzanke, T., & Tholey, A. (2024). Digital Microfluidics for Sample Preparation in Low-Input Proteomics. *Small Methods*, *n/a(n/a)*, 2400495. <https://doi.org/https://doi.org/10.1002/smt.202400495>
- Sulston, J. E., & Horvitz, H. R. (1977). Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Developmental Biology*, *56*(1), 110–156. [https://doi.org/https://doi.org/10.1016/0012-1606\(77\)90158-0](https://doi.org/https://doi.org/10.1016/0012-1606(77)90158-0)
- Sulston, J. E., Schierenberg, E., White, J. G., & Thomson, J. N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Developmental Biology*, *100*(1), 64–119. [https://doi.org/https://doi.org/10.1016/0012-1606\(83\)90201-4](https://doi.org/https://doi.org/10.1016/0012-1606(83)90201-4)
- Swaney, D. L., Wenger, C. D., & Coon, J. J. (2010). Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics. *Journal of Proteome Research*, *9*(3), 1323–1329. <https://doi.org/10.1021/pr900863u>
- Takayama, S., Reed, J. C., & Homma, S. (2003). Heat-shock proteins as regulators of apoptosis. *Oncogene*, *22*(56), 9041–9047. <https://doi.org/10.1038/sj.onc.1207114>
- Takemori, A., Kaulich, P. T., Konno, R., Kawashima, Y., Hamazaki, Y., Hoshino, A., Tholey, A., & Takemori, N. (2024). GeLC-FAIMS-MS workflow for in-depth middle-down proteomics. *PROTEOMICS*, *24*(3–4), 2200431. <https://doi.org/https://doi.org/10.1002/pmic.202200431>
- Taniguchi, Y., Choi, P. J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., & Sunney Xie, X. (2010). Quantifying *E. coli* Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science*, *329*(5991), 533–538. <https://doi.org/10.1126/science.1188308>
- Taylor, G. I. (1964). Disintegration of water drops in an electric field. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *280*(1382), 383–397. <https://doi.org/10.1098/rspa.1964.0151>
- Tholey, A., & Becker, A. (2017). Top-down proteomics for the analysis of proteolytic events - Methods, applications and perspectives. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1864*(11, Part B), 2191–2199. <https://doi.org/https://doi.org/10.1016/j.bbamcr.2017.07.002>

- Thompson, A., Schaefer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R. A. W., Mohammed, A. K. A., & Hamon, C. (2006). Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Analytical Chemistry*, *78*(12), 4235. <https://doi.org/10.1021/ac060310l>
- Treitz, C., Cassidy, L., Höckendorf, A., Leippe, M., & Tholey, A. (2015). Quantitative proteome analysis of *Caenorhabditis elegans* upon exposure to nematocidal *Bacillus thuringiensis*. *Journal of Proteomics*, *113*, 337–350. <https://doi.org/10.1016/j.jprot.2014.09.027>
- Tyanova, S., Temu, T., & Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols*, *11*(12), 2301–2319. <https://doi.org/10.1038/nprot.2016.136>
- Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M., & Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*, *13*(9), 731–740. <https://doi.org/10.1038/nmeth.3901>
- Venable, J. D., Dong, M.-Q., Wohlschlegel, J., Dillin, A., & Yates, J. R. (2004). Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nature Methods*, *1*(1), 39–45. <https://doi.org/10.1038/nmeth705>
- Venne, K., Bonneil, E., Eng, K., & Thibault, P. (2005). Improvement in Peptide Detection for Proteomics Analyses Using NanoLC–MS and High-Field Asymmetry Waveform Ion Mobility Mass Spectrometry. *Analytical Chemistry*, *77*(7), 2176–2186. <https://doi.org/10.1021/ac048410j>
- Vissers, J. P. C., Blackburn, R. K., & Moseley, M. A. (2002). A novel Interface for variable flow nanoscale LC/MS/MS for improved proteome coverage. *Journal of the American Society for Mass Spectrometry*, *13*(7), 760–771. [https://doi.org/10.1016/S1044-0305\(02\)00418-X](https://doi.org/10.1016/S1044-0305(02)00418-X)
- Wang, M., You, J., Bemis, K. G., Tegeler, T. J., & Brown, D. P. G. (2008). Label-free mass spectrometry-based protein quantification technologies in proteomic analysis. *Briefings in Functional Genomics*, *7*(5), 329–339. <https://doi.org/10.1093/bfpg/eln031>

- Wasinger, V. C., Cordwell, S. J., Cerpa-Poljak, A., Yan, J. X., Gooley, A. A., Wilkins, M. R., Duncan, M. W., Harris, R., Williams, K. L., & Humphery-Smith, I. (1995). Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *ELECTROPHORESIS*, *16*(1), 1090–1094.
<https://doi.org/https://doi.org/10.1002/elps.11501601185>
- Watts, J. L., & Browse, J. (2002). Genetic dissection of polyunsaturated fatty acid synthesis in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences*, *99*(9), 5854–5859. <https://doi.org/10.1073/pnas.092064799>
- Weke, K., Singh, A., Uwugiaren, N., Alfaro, J. A., Wang, T., Hupp, T. R., O'Neill, J. R., Vojtesek, B., Goodlett, D. R., Williams, S. M., Zhou, M., Kelly, R. T., Zhu, Y., & Dapic, I. (2021). MicroPOTS Analysis of Barrett's Esophageal Cell Line Models Identifies Proteomic Changes after Physiologic and Radiation Stress. *Journal of Proteome Research*, *20*(5), 2195–2205. <https://doi.org/10.1021/acs.jproteome.0c00629>
- Wheeler, A. R. (2008). Putting Electrowetting to Work. *Science*, *322*(5901), 539–540.
<https://doi.org/10.1126/science.1165719>
- Wiese, S., Reidegeld, K. A., Meyer, H. E., & Warscheid, B. (2007). Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. *PROTEOMICS*, *7*(3), 340–350.
<https://doi.org/https://doi.org/10.1002/pmic.200600422>
- Wilm, M., & Mann, M. (1996). Analytical Properties of the Nanoelectrospray Ion Source. *Analytical Chemistry*, *68*(1), 1–8. <https://doi.org/10.1021/ac9509519>
- Winkels, K., Koudelka, T., & Tholey, A. (2021). Quantitative Top-Down Proteomics by Isobaric Labeling with Thiol-Directed Tandem Mass Tags. *Journal of Proteome Research*, *20*(9), 4495–4506. <https://doi.org/10.1021/acs.jproteome.1c00460>
- Yamashita, M., & Fenn, J. B. (1984). Electrospray ion source. Another variation on the free-jet theme. *The Journal of Physical Chemistry*, *88*(20), 4451–4459.
<https://doi.org/10.1021/j150664a002>
- Yang, Z., Jin, K., Chen, Y., Liu, Q., Chen, H., Hu, S., Wang, Y., Pan, Z., Feng, F., Shi, M., Xie, H., Ma, H., & Zhou, H. (2024). AM-DMF-SCP: Integrated Single-Cell Proteomics Analysis on an Active Matrix Digital Microfluidic Chip. *JACS Au*, *4*(5), 1811–1823.
<https://doi.org/10.1021/jacsau.4c00027>

- Yin, Y., Donlevy, S., & Smolikove, S. (2016). Coordination of Recombination with Meiotic Progression in the *Caenorhabditis elegans* Germline by KIN-18, a TAO Kinase That Regulates the Timing of MPK-1 Signaling. *Genetics*, *202*(1), 45–59.
<https://doi.org/10.1534/genetics.115.177295>
- Yun, C., Stanhill, A., Yang, Y., Zhang, Y., Haynes, C. M., Xu, C.-F., Neubert, T. A., Mor, A., Philips, M. R., & Ron, D. (2008). Proteasomal adaptation to environmental stress links resistance to proteotoxicity with longevity in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences*, *105*(19), 7094 LP – 7099. <https://doi.org/10.1073/pnas.0707025105>
- Zhang, Y., & Nguyen, N.-T. (2017). Magnetic digital microfluidics – a review. *Lab on a Chip*, *17*(6), 994–1008. <https://doi.org/10.1039/C7LC00025A>
- Zhou, L., Wang, L., Bai, S., Xing, S., Li, W., Ma, J., & Fu, X. (2018). Knockdown of LMW-PTP enhances stress resistance in *Caenorhabditis elegans*. *International Journal of Biological Macromolecules*, *113*, 1015–1023.
<https://doi.org/10.1016/j.ijbiomac.2018.03.014>
- Zhou, M., Uwugiaren, N., Williams, S. M., Moore, R. J., Zhao, R., Goodlett, D., Dapic, I., Paša-Tolić, L., & Zhu, Y. (2020). Sensitive Top-Down Proteomics Analysis of a Low Number of Mammalian Cells Using a Nanodroplet Sample Processing Platform. *Analytical Chemistry*, *92*(10), 7087–7095.
<https://doi.org/10.1021/acs.analchem.0c00467>
- Zhu, Y., Piehowski, P. D., Kelly, R. T., & Qian, W.-J. (2018). Nanoproteomics comes of age. *Expert Review of Proteomics*, *15*(11), 865–871.
<https://doi.org/10.1080/14789450.2018.1537787>
- Zhu, Y., Piehowski, P. D., Zhao, R., Chen, J., Shen, Y., Moore, R. J., Shukla, A. K., Petyuk, V. A., Campbell-Thompson, M., Mathews, C. E., Smith, R. D., Qian, W.-J., & Kelly, R. T. (2018). Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nature Communications*, *9*(1), 882.
<https://doi.org/10.1038/s41467-018-03367-w>
- Zou, W., Lu, Q., Zhao, D., Li, W., Mapes, J., Xie, Y., & Wang, X. (2009). *Caenorhabditis elegans* Myotubularin MTM-1 Negatively Regulates the Engulfment of Apoptotic Cells. *PLOS Genetics*, *5*(10), e1000679.
<https://doi.org/10.1371/journal.pgen.1000679>

Declaration according to § 9 of the Doctoral Degree Regulations

I, Max Konstantin Steinbach, hereby declare the following:

- (i) This thesis, save the guidance provided by my supervisor, is my own work in terms of conceptualization and content. Furthermore, any external sources used, regardless of whether quoted directly or indirectly, have been cited, and any preparations or tools contributive to this work that were not conducted or developed by myself have been explicitly identified as such.
- (ii) This thesis, in part or its entirety, has not previously been subjected to a doctoral examination procedure by another examining body, nor has it previously been published or submitted for publication.
- (iii) This thesis has been prepared in compliance with the “Rules of Good Scientific Practice” of the German Research Foundation.
- (iv) I have never had an academic degree withdrawn, nor failed a doctoral examination procedure.

Kiel, _____

Max Konstantin Steinbach