

„Etablierung eines ArgC-analogen Verdaues für Bottom-Up-Ansätze in der Proteomforschung“

Dissertation

zur Erlangung des Grades

„Doktor der Naturwissenschaften“

Vahid Golghalyani

aus Berlin

Frankfurt 2018

(D30)

vom Fachbereich 14 Biochemie, Chemie und Pharmazie der
Johann Wolfgang Goethe – Universität als Dissertation angenommen.

Dekan: Prof. Dr. Clemens Glaubitz

Gutachter: Prof. Dr. Michael Karas

Prof. Dr. Rolf Marschalek

Datum der Disputation: 11.10.2018

Für meine Eltern, Sahra und Adel.

Danksagung

An dieser Stelle möchte ich meinen besonderen Dank all den Personen übermitteln, ohne deren Mithilfe diese Arbeit nicht zustande gekommen wäre:

Mein Dank gilt

Herrn Prof. Dr. Michael Karas, meinem Doktorvater, der mir vorurteilsfrei die Möglichkeit gegeben hat diese Arbeit zu verfassen. Ich danke ihm für seine Betreuung, die freundliche Hilfe und die Möglichkeit, an nationalen und internationalen Konferenzen teilnehmen zu können. Er gewährte mir Raum zum freien forschen, sodass ich mich in verschiedenen Bereichen weiterentwickeln konnte. Ich habe unsere Dialoge stets als Inspirationsquelle wahrgenommen, die zudem oft ermutigend und motivierend waren.

Frau Dr. Ute Bahr für ihre unschätzbare Hilfe bei der Zusammenstellung meiner Publikation, ihre wohlwollende Art und für die gute Zusammenarbeit mit ihr als Praktikumsleiterin.

Herrn Prof. Dr. Rolf Marschalek für die freundliche Übernahme des Zweitgutachtens.

Frau Dr. Ilka Wittig dafür, dass sie sehr großzügiger Art und Weise Geräte und Methoden zur Vermessung unserer Proteomproben zur Verfügung gestellt hat.

Meinem Bürokollegen Moritz Neupärtl danke ich für die Entwicklung seiner Bead-basierten Aufreinigungsmethode, die die Grundlage der hier entwickelten Methode darstellt, des Weiteren danke ich ihm für die gute Zusammenarbeit und die vielen Gespräche, die stets aufbauend und inspirierend waren.

Kathrin Carlsburg für ihre immer freundliche und zuverlässige Unterstützung.

Meinen Praktikumskollegen Dr. Jan Gorka, Tobias Schorge, Dr. Anne Arnold, Vera Jüngst, Yannik Lewin, Katharina Schlag für die stets unterhaltsame Zusammenarbeit.

Dr. Marion Bäumlisberger danke ich für ihre umfangreiche Einarbeitung in das Themengebiet.

Michael Rühl dafür, dass er mich in zwei Publikationen als Mitautor berücksichtigt hat.

Allen meinen Kollegen der Arbeitsgruppe namentlich Matthias Brandl, David Fabacher, Andreas Münch, Benjamin Müller und Stefan Schönborn.

Ferner gilt mein Dank Omar Barhoum, mich in dieser Zeit weit über das normale Maß einer gewöhnlichen Freundschaft unterstützt hat.

Zusammenfassung

Massenspektrometrie-basierte Proteomuntersuchungen erfolgen auch heute überwiegend nach dem sogenannten *Bottom-Up-Ansatz*, d.h. die Identifizierung von Proteinen erfolgt auf der Basis von Peptiden, die chromatographisch gut voneinander getrennt werden können und massenspektrometrisch leichter zu analysieren sind als Proteine. Nach Identifikation der Peptide kann rekonstruiert werden, welche Proteine ursprünglich in der Probe vorgelegen haben. Zentraler Arbeitsschritt der Probenvorbereitung ist daher die Zerlegung des Proteins, die entweder chemisch oder - wie in den meisten Fällen - enzymatisch erfolgt. Trypsin ist das mit Abstand am häufigsten genutzte Enzym, da es eine hohe Schnittspezifität aufweist und sehr effizient ist. Der Trypsin-Verdau ist darüber hinaus sehr robust, d.h. er zeigt eine hohe Toleranz gegenüber Verunreinigungen, und zudem werden Peptide erzeugt, die sowohl gute Ionisations- als auch gute Fragmentierungseigenschaften aufweisen. Die durch Trypsin gebildeten Peptide enthalten neben dem basischen N-Terminus eine weitere basische Aminosäure am C-Terminus, so dass sie leicht zumindest doppelt-geladene Ionen bilden können und sehr häufig aussagekräftige C-terminale Fragmentationserien liefern.

Neben den zahlreichen Vorteilen gibt es allerdings auch Nachteile. So können nach einem tryptischen Verdau in Abhängigkeit von der Verteilung der Schnittstellen Peptide entstehen, die entweder zu klein sind, um eine verlässliche Zuordnung zu einem Protein zu erlauben oder die zu groß sind für den Massenbereich des gewählten Massenanalysators. Eine vielversprechende Alternative zu Trypsin wäre ArgC, welches C-Terminus zu Argininen schneidet und somit im Durchschnitt größere Peptide mit Ionisations- und Fragmentierungseigenschaften ähnlich zu tryptischen Peptiden erzeugt. Das Enzym ArgC weist jedoch nur eine geringe Schnittspezifität auf und sein Trypsin-ähnliches Verhalten - also das Schneiden auch hinter Lysin - wurde öfters beobachtet und wird auch vom Hersteller angegeben. Ziel dieser Arbeit war die Entwicklung einer Verdauemethode, die Peptide erzeugt, die ausschließlich auf Argininen enden.

Das Resultat der zu entwickelnden Verdauemethode sollte somit dem eines idealen enzymatischen ArgC-Verdaues entsprechen. Realisiert wurde der ArgC-ähnliche-Verdau durch den Einsatz von Trypsin, dessen enzymatischer Schnitt durch die chemische Derivatizierung der Substrat-Lysine auf Arginine reduziert wurde. Neben dem weiteren Einsatz von Trypsin sollte dieser „Quasi-Arg-C-Verdau“ weitere systematische Vorteile für Proteomanalysen realisieren: Zum Ersten sollte die Anzahl von Fehlschnitt-Peptiden, die sich bei Trypsin insbesondere an Lysinen mit saurer chemischer Umgebung ergeben, reduziert werden, zum Zweiten sollten die Arg-C-Peptide sowohl durch ihre gewachsene Größe, als auch durch das mit dem C-terminalen Arginin verbesserte Fragmentierungsverhalten höhere Score-Werte bei der bioinformatischen Auswertung der MS-Daten ergeben.

Im ersten Teil wurden zunächst bioinformatische Werkzeuge entwi-

ckelt, die MALDI-MS-Dateien automatisiert prozessierten. Die entwickelten Programme umfassen die Identifizierung und relative Quantifizierung von Proteinen aus diesen Dateien. Des Weiteren wurde ein Programm zur Analyse von MALDI-MS-Dateien entwickelt. Automatisierte Auswertungen gelangen durch die Erstellung von *Workflows* in der Datenanalyseplattform KNIME. Diese *Workflows* kombinieren in Python geschriebene Skripte und Funktionalitäten frei verfügbarer Programme wie „MSConvert“ und „mMass“.

Nach Erstellung der bioinformatischen Werkzeuge wurde die Methodenentwicklung zur Modifizierung der Lysine für verschiedene Reagenzien durchgeführt. Die Auswahl fiel auf vier Substanzen, von denen bekannt ist, dass sie unter milden Reaktionsbedingungen im quantitativen Ausmaß mit Aminogruppen reagieren. Diese waren Sulfo-NHS-Acetat, Propionsäureanhydrid, Diethylpyrocarbonat und die reduktive Methylierung mit Formaldehyd und Picolin-Boran. Die Reaktionsbedingungen mussten zunächst für Proteine optimiert werden, da die publizierten Protokolle hauptsächlich zur Derivatierung von Peptiden verwendet worden waren. Anschließend wurden die optimierten Protokolle für eine Protein- und Proteomprobe eingesetzt und die Resultate miteinander verglichen. Die Untersuchungen führten zu dem Ergebnis, dass sowohl auf Protein- als auch auf Proteomebene die Propionierung des Lysins die besten Resultate zeigte. Insbesondere ist hervorzuheben, dass alle ArgC-ähnlichen Ansätze unabhängig vom eingesetzten Reagenz zu besseren Ergebnissen in jeder der Untersuchungen führte als der klassische enzymatische ArgC-Verdau.

Im nächsten Abschnitt wurden die Resultate eines ArgC-ähnlichen (mit propionylierten Aminogruppen) und eines konventionellen tryptischen Verdauens der gleichen Proteomprobe tiefergehend miteinander verglichen. Der Vergleich ergab, dass der ArgC-ähnliche Verdau, trotz einer geringeren Anzahl identifizierter Peptide, zu einer höheren Anzahl identifizierter Proteine führte als der konventionell tryptische Ansatz. Des Weiteren war der Anteil an fehlgeschnittenen Peptiden im ArgC-ähnlichen Ansatz deutlich geringer und die erzeugten Peptide wiesen eine Ionenscore-Verteilung auf, die im Vergleich zum tryptischen Ansatz in höhere Bereiche verschoben war. Die höhere Anzahl identifizierter Proteine und der geringere Anteil fehlgeschnittener Peptide implizieren eine höhere *Proteome Coverage* und eine höhere Schnitteleffizienz des ArgC-ähnlichen Verdauens. Eine Verschiebung der Ionenscore-Verteilung wird durch ein besser vorhersagbares Fragmentierungsverhalten begründet und erlaubt Identifikationen höherer Sicherheit von ArgC-ähnlich im Vergleich zu tryptisch erzeugten Peptiden.

Speziell die höhere Anzahl identifizierter Proteine konnte bis zum heutigen Tag nicht durch den Wechsel der Verdauemethode erreicht werden. In vergleichenden Untersuchungen führte Trypsin zu deutlich mehr Identifikationen, so dass der Einsatz anderer Proteasen häufig auf Spezialanwendungen limitiert ist. Eine höhere Proteom-Coverage konnte bis lang nur durch eine Vorfraktionierung der Probe erreicht werden, die entweder online oder offline erfolgt, wobei dieses Vorgehen zu

einer dramatischen Erhöhung der Messzeit führt und Quantifizierungen durch das Vorkommen von Peptiden in mehreren Fraktionen erschwert werden. Dem ArgC-ähnlichen Verdau gelang es somit als erste Verdaumethode, durch eine sinnvolle Reduktion der Probenkomplexität, zu einer höheren Anzahl identifizierter Proteine zu führen als der bisherige Goldstandard Trypsin.

Im letzten Teil dieser Arbeit wurde der ArgC-ähnliche Verdau auf bestimmte konkrete Anwendungsmöglichkeiten untersucht. Hier konnte dargelegt werden, dass diese Methode für die Analyse humaner Proteomproben geeignet ist. Zusätzlich wurde gezeigt, dass Quantifizierungen mit gleich hoher Präzision erfolgen können wie bei Standardmethoden. Zum Abschluss der Arbeit wurde die entwickelte Methode erfolgreich und mit wiederum besseren Ergebnissen zur Charakterisierung eines Proteins verwendet.

Inhaltsverzeichnis

| | |
|---|-------------|
| Inhaltsverzeichnis | xiii |
| Abbildungsverzeichnis | xvi |
| Tabellenverzeichnis | xix |
| 1 Einleitung | 1 |
| 1.1 Proteine, Proteom und Proteomik | 1 |
| 1.1.1 Begriffe | 1 |
| 1.1.2 Physikochemische Eigenschaften von Proteinen und Peptiden | 3 |
| 1.1.3 Ziele der Proteomik | 5 |
| 1.1.4 Arbeitsabläufe in der Proteomik | 6 |
| 1.2 Probenvorbereitung (Sample Preparation) | 7 |
| 1.2.1 Zellyse | 7 |
| 1.2.2 Protein-Aufreinigung | 7 |
| 1.2.3 Proteolyse | 8 |
| 1.2.4 Methoden zur Peptidtrennung | 9 |
| 1.3 Chemische Derivatisierung von Proteinen und Peptiden | 10 |
| 1.3.1 Derivatisierung der Cysteine | 10 |
| 1.3.2 Derivatisierung der Aminogruppen | 11 |
| 1.4 Massenspektrometrie | 13 |
| 1.4.1 Begriffe | 13 |
| 1.4.2 Ionisationsmethoden | 14 |
| 1.4.3 Massenanalysatoren | 16 |
| 1.4.4 Peptidfragmentierung | 20 |
| 1.5 Massenspektrometrische Datenanalyse | 23 |
| 1.5.1 Konvertierung der MS-Dateien | 24 |
| 1.5.2 Signalprozessierung | 24 |
| 1.5.3 Proteinidentifizierung | 25 |

| | | |
|----------|---|-----------|
| 1.5.4 | Peptididentifizierung | 25 |
| 1.6 | Massenspektrometrische Quantifizierung von Proteinen und Peptiden | 27 |
| 1.6.1 | Quantifizierung mit TMT | 28 |
| 1.6.2 | Derivatisierungsfreie Quantifizierung (<i>Label-free</i>) | 29 |
| 2 | Materialien und Methoden | 31 |
| 2.1 | Materialien | 31 |
| 2.1.1 | Modellproteine und -proteome | 31 |
| 2.1.2 | Reagenzien und Verbrauchsmaterialien | 32 |
| 2.1.3 | Software | 32 |
| 2.1.4 | Verwendete Geräte | 33 |
| 2.2 | Methoden | 38 |
| 2.2.1 | Zelllyse | 38 |
| 2.2.2 | Acetonfällung | 38 |
| 2.2.3 | Binden von Proteinen an magnetische Beads | 39 |
| 2.2.4 | Proteinquantifizierung | 40 |
| 2.2.5 | Chemische Modifizierung auf Proteinebene | 41 |
| 2.2.6 | Verdauprotokolle | 44 |
| 2.2.7 | Peptidaufreinigung | 45 |
| 2.2.8 | MALDI-Messungen | 45 |
| 2.2.9 | LC-ESI-Messungen | 45 |
| 2.2.10 | Mascot-Suchparameter | 47 |
| 3 | Fragestellung und Ziele | 49 |
| 4 | Ergebnisse und Diskussion | 53 |
| 4.1 | Entwicklung bioinformatischer Werkzeuge zur Analyse von MS-Daten | 53 |
| 4.1.1 | Konvertierung der MALDI-MS-Dateien | 54 |
| 4.1.2 | Signalprozessierung von Massenspektren mit mMass | 55 |
| 4.1.3 | Parsen von Mascot dat-Dateien | 57 |
| 4.1.4 | Integration der Auswerteprogramme in KNIME | 57 |
| 4.1.5 | Workflows zur Auswertung von MALDI-ISD-Spektren | 60 |
| 4.1.6 | Diskussion | 62 |
| 4.2 | Chemische Modifikation der Aminogruppen auf Proteinebene | 63 |
| 4.2.1 | Verifizierung der Modifizierungsprotokolle auf Peptid- und Proteinebene | 63 |
| 4.2.2 | Statistische Versuchsplanung (Design of Experiment, DOE) | 65 |
| 4.2.3 | Diskussion | 79 |
| 4.3 | Vergleich der Modifizierungsprotokolle auf Protein- und Proteomebene | 81 |
| 4.3.1 | Vergleich auf Proteinebene | 82 |

| | | |
|---|---|----------------|
| 4.3.2 | Vergleich auf Proteomebene | 90 |
| 4.3.3 | Diskussion | 100 |
| 4.4 | Vergleich mit Trypsin | 101 |
| 4.4.1 | Anzahl der Proteinidentifikationen | 102 |
| 4.4.2 | Verdauereffizienz | 103 |
| 4.4.3 | Fragmentierungsverhalten | 109 |
| 4.4.4 | Diskussion | 116 |
| 4.5 | Anwendbarkeit ArgC-ähnlicher Verdaue | 118 |
| 4.5.1 | Anwendbarkeit auf das humane Proteom | 118 |
| 4.5.2 | Anwendbarkeit für Quantifizierungsansätze | 123 |
| 4.5.3 | Anwendbarkeit für Multi-Enzym-Ansätze | 126 |
| 4.5.4 | Diskussion | 129 |
| A Reagenzien und Verbrauchsmaterialien | | 131 |
| B Software | | 133 |
| C Mascot Suchparameter | | 135 |
| Literaturverzeichnis | | 141 |
| Abkürzungsverzeichnis | | 155 |
| Eidesstattliche Versicherung | | 157 |

Abbildungsverzeichnis

| | | |
|------|--|----|
| 1.1 | Proteinogene Aminosäuren | 2 |
| 1.2 | Übersicht über <i>Top-Down</i> - und <i>Bottom-Up</i> -Ansätze | 6 |
| 1.3 | Alkylierung von Cysteinresten mit Iodacetamid | 11 |
| 1.4 | Reaktion der Aminogruppen mit Carbonsäureanhydriden | 12 |
| 1.5 | Reduktive Methylierung von Aminen | 12 |
| 1.6 | Acylierung von Aminogruppen über NHS-Ester | 12 |
| 1.7 | Funktionseinheiten eines Massenspektrometers | 13 |
| 1.8 | Elektrospray-Ionisation (ESI) | 15 |
| 1.9 | Matrix-unterstützte Laser-Desorptions/Ionisation (MALDI) ¹ | 16 |
| 1.10 | Aufbau eines Flugzeitanalysators mit Reflektron. | 19 |
| 1.11 | Schematische Darstellung zur Aufnahme von MS und MS/MS-Spektren | 22 |
| 1.12 | Nomenklatur nach Roepstorff und Fohlmann. | 23 |
| 1.13 | Schmetischer Aufbau einer Datenbank-basierten Suchmaschine | 26 |
| 1.14 | Quantifizierungsverfahren in der Proteomforschung | 28 |
| 1.15 | TMT-Reagenz und sein funktioneller Aufbau | 29 |
| 2.1 | Voyager-DE TM STR Workstation (ABSciex) | 34 |
| 2.2 | 4800 MALDI-TOF/TOF TM Analyser (ABSciex) | 35 |
| 2.3 | MALDI LTQ-Orbitrap TM XL (Thermo Fisher Scientific) | 36 |
| 2.4 | Q Exactive Orbitrap TM XL (Thermo Fisher Scientific) | 38 |
| 2.5 | Strukturformel des violetten Cu ⁺ -Bicinchoninsäure-Komplexes | 41 |
| 4.1 | KNIME- <i>Workflow</i> zur Prozessierung von MALDI-MS-Dateien | 58 |
| 4.2 | Aufbau eines Metaknoten | 59 |
| 4.3 | <i>Workflow</i> zur Auswertung von MALDI-MS-Dateien | 61 |
| 4.4 | MALDI-MS Spektrum von Substanz P und seinen derivatisierten Analoga | 64 |
| 4.5 | Übersicht über die Qualitätsmerkmale nach Optimierung der Reaktionsbedingungen der reduktiven Dimethylierung | 67 |

| | | |
|------|---|----|
| 4.6 | <i>Contour Plot</i> für die Optimierung der Reaktionsbedingungen der reduktiven Dimethylierung | 68 |
| 4.7 | MALDI-MS-Spektren vor und nach Optimierung der Reaktionsbedingungen der reduktiven Methylierung | 69 |
| 4.8 | Übersicht über die Qualitätsmerkmale nach Optimierung der Reaktionsbedingungen der Carbethoxylierung der Amino-Gruppen | 71 |
| 4.9 | <i>Contour Plot</i> für die Optimierung der Reaktionsbedingungen der Carbethoxylierung der Aminogruppen | 72 |
| 4.10 | MALDI-MS-Spektren vor und nach Optimierung der Reaktionsbedingungen der Carbethoxylierung der Aminogruppen | 73 |
| 4.11 | Übersicht über die Qualitätsmerkmale nach Optimierung der Reaktionsbedingungen der Acetylierung der Amino-Gruppen | 75 |
| 4.12 | <i>Contour Plot</i> für die Optimierung der Reaktionsbedingungen der Acetylierung der Aminogruppen | 76 |
| 4.13 | MALDI-MS-Spektren vor und nach Optimierung der Reaktionsbedingungen der Acetylierung der Aminogruppen | 76 |
| 4.14 | Übersicht über die Qualitätsmerkmale nach Optimierung der Reaktionsbedingungen der Propionylierung der Amino-Gruppen | 78 |
| 4.15 | <i>Contour Plot</i> für die Optimierung der Reaktionsbedingungen der Propionylierung der Aminogruppen | 78 |
| 4.16 | MALDI-MS-Spektren vor und nach Optimierung der Reaktionsbedingungen der Propionylierung der Aminogruppen | 79 |
| 4.17 | Übersicht über die verschiedenen Verdaustrategien. | 81 |
| 4.18 | MALDI-MS-Spektrum von Serotransferrin nach Propionylierung | 84 |
| 4.19 | MALDI-MS-Spektrum von Serotransferrin nach Acetylierung | 85 |
| 4.20 | MALDI-MS-Spektrum von Serotransferrin nach Carbethoxylierung | 86 |
| 4.21 | MALDI-MS-Spektrum von Serotransferrin nach reduktiver Methylierung | 87 |
| 4.22 | MALDI-MS-Spektrum nach ArgC-Verdau von Serotransferrin | 88 |
| 4.23 | Lage der tryptischen Schnittstellen und Übersicht über die detektierten Sequenzteile des Serotransferrin nach Anwendung der verschiedenen Protokolle. | 89 |
| 4.24 | Übersicht verschiedener Kriterien zum Vergleich der Massenspektren von unterschiedlich modifizierten Serotransferrin mit konventionellen Verdaustrategien (ArgC und Trypsin). | 91 |
| 4.25 | Gesamtzahl der aufgenommenen MS/MS-Spektren und deren nicht zu Peptiden zuweisbaren Anteile | 93 |
| 4.26 | Anzahl und Anteil N-terminal modifizierter Peptide. | 94 |
| 4.27 | Anzahl und Anteil unvollständig modifizierter Peptide. | 95 |
| 4.28 | Boxplots zur Visualisierung der Verteilung der Sequenzabdeckungen der verschiedenen Ansätze. | 98 |
| 4.29 | Boxplots zur Visualisierung der Verteilung der Sequenzabdeckungen nach Zusammenführen der Resultate mit dem tryptischen Verdau. | 99 |

| | | |
|------|--|-----|
| 4.30 | Venn-Diagramme der identifizierten Proteine mit Trypsin als Referenz. | 101 |
| 4.31 | Scatterplot mit der Anzahl zugewiesener PSM des ArgC-ähnlichen gegen den tryptischen Ansatz | 103 |
| 4.32 | Prozentualer Anteil der Peptide mit Fehlschnittstellen an der Gesamtzahl der PSMs und der Anteil an der gemessenen Gesamtintensität. | 104 |
| 4.33 | <i>Heatmap</i> zur Auszählung der Aminosäuren an der Fehlschnittregion für den tryptischen Ansatz | 106 |
| 4.34 | <i>Heatmap</i> zur Auszählung der Aminosäuren an der Fehlschnittregion für den ArgC-ähnlichen Ansatz | 107 |
| 4.35 | iceLogo der fehlgeschnittenen Peptide vom Trypsin-Protokoll. . . | 108 |
| 4.36 | iceLogo der fehlgeschnittenen Peptide vom PA-Protokoll. | 109 |
| 4.37 | Verteilung der Percolator-adjustierten Ionenscores für den tryptischen und den ArgC-ähnlichen Ansatz | 110 |
| 4.38 | Verteilung der Peptidmassen für den tryptischen und den ArgC-ähnlichen Ansatz | 111 |
| 4.39 | Violinplot der Ionenscore-Verteilungen in Abhängigkeit der Peptidmassenbereiche | 113 |
| 4.40 | Boxplots für die <i>Intensity Coverage</i> und den y -Fragmentüberschuss für das jeweilige Protokoll | 114 |
| 4.41 | Nach Massenbereichen kategorisierte Boxplots für die <i>Intensity Coverage</i> | 115 |
| 4.42 | Anteil fehlgeschnittener Peptide aufgeteilt in Massenbereiche für das jeweilige Protokoll | 116 |
| 4.43 | Scatterplot mit der Anzahl zugewiesener PSM des ArgC-ähnlichen gegen den tryptischen Ansatz nach Anwendung an dem humanen Proteom | 120 |
| 4.44 | Venn-Diagramme zur Untersuchung der Komplementarität der verschiedenen Verdauansätze auf Ebene der identifizierten Proteine | 121 |
| 4.45 | <i>Boxplot</i> zur Visualisierung der Sequenzabdeckungsverteilung und des Zugewinns an Information nach Zusammenführen der Ergebnisse | 122 |
| 4.46 | Anteil fehlgeschnittener Peptida in den verschiedenen Verdau-Ansätzen | 123 |
| 4.47 | <i>Normal-QQ-Plot</i> für die Proteinverhältnisse des ArgC-ähnlichen und des tryptischen Ansatzes | 125 |
| 4.48 | Übersicht zur erhaltenen Sequenzabdeckung mit verschiedenen Verdauansätzen an verschiedenen Geräten | 127 |
| 4.49 | Übersicht zur erhaltenen Sequenzabdeckung mit verschiedenen Verdauansätzen nach Zusammenführen der Ergebnisse an den verschiedenen Geräten | 128 |

Tabellenverzeichnis

| | | |
|------|---|-----|
| 1.1 | Physikochemische Eigenschaften von Aminosäuren. | 4 |
| 1.2 | Vergleichende Übersicht der verschiedenen Massenanalytoren. | 21 |
| 2.1 | Python Module | 33 |
| 4.1 | Versuchsplan zur Optimierung der reduktiven Methylierung von ADH | 66 |
| 4.2 | Versuchsplan zur Optimierung von DEPC | 70 |
| 4.3 | Versuchsplan zur Optimierung der Acetylierung | 74 |
| 4.4 | Versuchsplan zur Optimierung der Propionylierung | 77 |
| 4.5 | Peptidliste (PA-Protokoll) | 84 |
| 4.6 | Peptidliste (NHS-Protokoll) | 85 |
| 4.7 | Peptidliste (DEPC-Protokoll) | 86 |
| 4.8 | Peptidliste DM-Protokoll | 87 |
| 4.9 | Peptidliste ArgC-Protokoll | 88 |
| 4.10 | Anzahl der Peptid- und Proteinidentifikationen | 96 |
| 4.11 | Prozentuale Häufigkeit der Aminosäuren in dem untersuchten Proteom. | 108 |
| 4.12 | Anzahl der Peptid- und Proteinidentifikationen | 119 |
| A.1 | Reagenzien | 131 |
| A.2 | Verbrauchsmaterialien | 132 |
| B.1 | Programme | 133 |

Kapitel 1

Einleitung

Im April des Jahres 2003 wurde verkündet, dass das Humangenom-Projekt das menschliche Erbgut in praktisch allen Details entziffert habe.² Entgegen der Erwartungen stellte sich heraus, dass das menschliche Genom aus lediglich 20'000 Genen aufgebaut ist³ und somit gerade mal doppelt so viele Gene aufweist wie die DNA von Würmern oder Fliegen.⁴ Zu Beginn des Projektes wurde das Vorhandensein von mindestens 100'000 Genen postuliert, um alle Merkmale und Funktionen des humanen Körpers kodieren zu können. Die niedrige Anzahl der Gene widersprach der These vom genetischen Determinismus und die anfangs gehegten Hoffnungen, anhand von Genen auf bestimmte Merkmale zu schließen wurde zu Nichte gemacht.

Die Vorstellung von 100'000 Genen richtete sich nach der Anzahl humaner Proteine,⁵ welche entscheidend für die Funktionen und den Aufbau eines Organismus sind. Die großen Unterschiede in der Anzahl von Proteinen und Genen weisen darauf hin, dass die Vorgänge ausgehend vom Ablesen der DNA bis zum gebildeten Protein deutlich komplexer sein müssen. Diese Ernüchterung nach der Entzifferung des Genoms führte zu einer Verlagerung der Anstrengungen auf andere Bereiche wie die Analyse der Gesamtheit aller Proteine, welches Gegenstand der Proteomforschung ist. In den folgenden Abschnitten wird dieses Forschungsfeld eingeführt, wobei der Schwerpunkt auf der Massenspektrometrie-basierten Proteomforschung liegt.

1.1 Proteine, Proteom und Proteomik

1.1.1 Begriffe

Proteine stellen neben den Nukleinsäuren und Kohlenhydraten die dritte große Gruppe der biologischen Makromoleküle dar. Sie sind aufgebaut aus 21 unterschiedlichen natürlich vorkommenden (biogenen) Aminosäuren, die

1. EINLEITUNG

über Peptidbindungen miteinander verbunden sind. Neben den proteinogenen Aminosäuren sind zur Zeit 400 nicht proteinogene Aminosäuren bekannt,⁶ die biologische Funktionen haben. Die Carboxylgruppe ist bei proteinogenen Aminosäuren über ein Kohlenstoffatom mit der Aminogruppe verbunden, wobei das Kohlenstoffatom zwischen den beiden funktionellen Gruppen eine für die Aminosäure spezifische Seitenkette trägt. Die Einteilung der Aminosäuren in Gruppen erfolgt nach der Polarität, Basizität oder

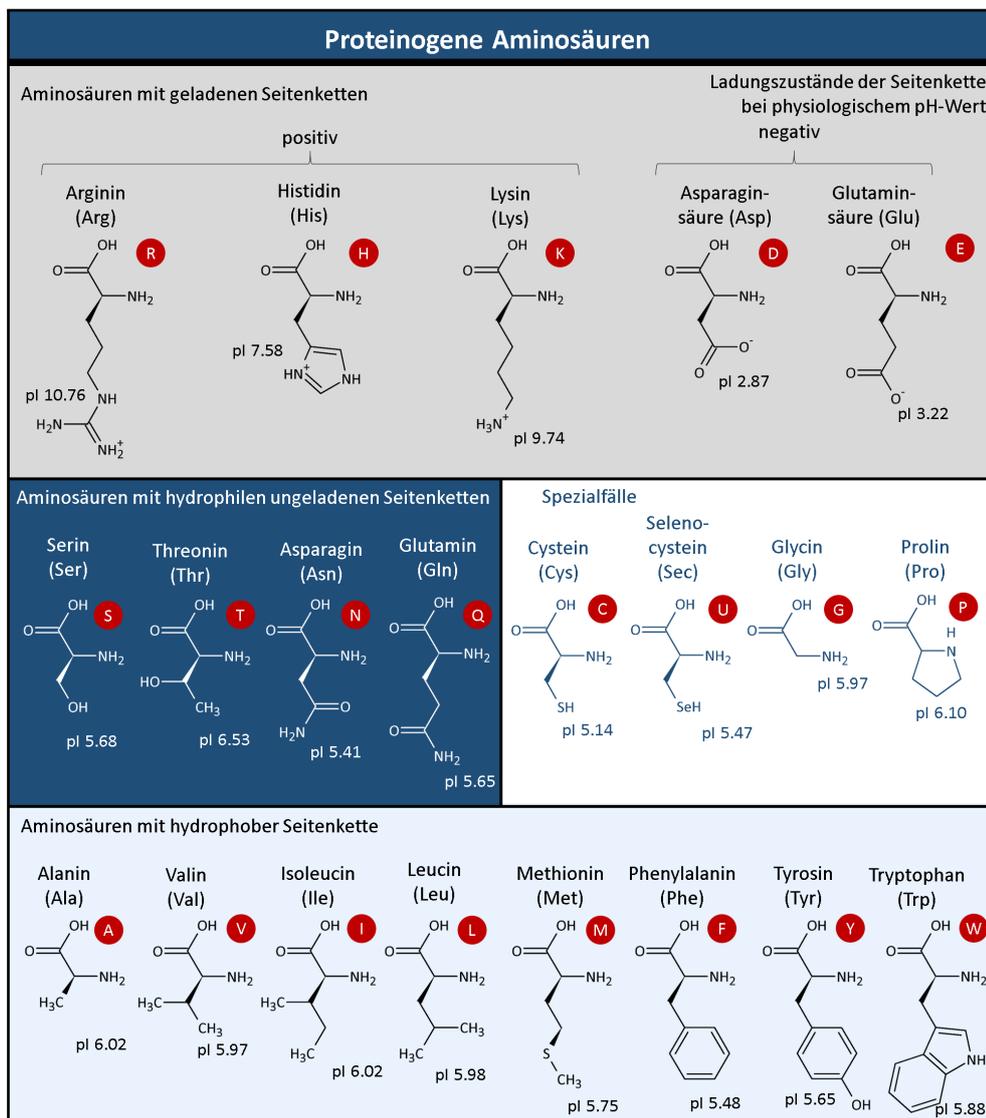


Abbildung 1.1: Proteinogene Aminosäuren unterteilt nach den Eigenschaften ihrer Seitenkette mit dem Drei- und Einbuchstabencode (roter Kreis). pI ist die Abkürzung für isoelektrischer Punkt.

dem Ladungszustand der Seitenkette (siehe Abbildung 1.1 auf der vorherigen Seite).

Proteine haben aufgrund der hohen Anzahl unterschiedlicher Bausteine (z.B. im Vergleich zu Nukleinsäuren mit nur vier verschiedenen Monomeren) sehr unterschiedliche Funktionen in der Physiologie eines Organismus. Neben der Funktion als Energieträger und Strukturgeber stellen sie „molekulare Maschinen“ dar, die Zellbewegungen ermöglichen, Metabolite transportieren, Ionen pumpen, chemische Reaktionen katalysieren und Signalstoffe erkennen können.

Die Gesamtheit aller Proteine in einem Lebewesen, einem Gewebe, einer Zelle oder einem Zellkompartiment, unter exakt definierten Bedingungen und zu einem bestimmten Zeitpunkt, wird als *Proteom* bezeichnet. Schätzungen gehen davon aus, dass das menschliche Proteom aus 100'000 Proteinen aufgebaut ist, wobei ca. 70'000 Proteine Spleißvarianten darstellen. Unter Berücksichtigung vorkommender posttranslationaler Modifikationen (PTM) erhöht sich die Komplexität auf über eine Million Proteine.⁵ Die Abundanzunterschiede der Proteine können in Zellen über sechs, im Plasma sogar über zehn Größenordnungen erreichen.⁷

Neben der Komplexität des Proteoms und den Abundanzunterschieden einzelner Proteine wird die Erforschung des Proteoms, die *Proteomik* (im englischen *Proteomics*), durch das Nichtvorhandensein einer Amplifizierungsmethode wie die Polymerase-Kettenreaktion für die DNA erschwert. Tyers und Mann erweiterten den Begriff Proteomics zu allen Vorgängen, die postgenomisch einen Bezug zu Proteinen haben.⁸ Heutzutage gehören die Charakterisierung aller Isoformen und Modifikationen, die Interaktionen und die Strukturaufklärung der Proteine ebenfalls zum Feld der Proteomik. Methoden innerhalb der Proteomik lassen sich im wesentlichen in zwei Gruppen aufteilen: Methoden, in denen Antikörper eingesetzt werden sowie Methoden, die auf Massenspektrometrie (MS) basieren.

1.1.2 Physikochemische Eigenschaften von Proteinen und Peptiden

Die physikochemischen Eigenschaften von Peptiden und Proteinen werden im wesentlichen durch die Effekte der Seitenketten der Aminosäuren determiniert, aus denen sie aufgebaut sind. Vorhersagen über die im Anschluss beschriebenen Eigenschaften sind möglich, wenn die Aminosäuresequenz bekannt ist. Experimentelle Einzelbestimmungen können somit umgangen werden, welche durch die hohe Anzahl vorkommender Proteine und Peptide geradezu unmöglich sind. Abweichungen der gemessenen von den vorhergesagten Werten werden durch intramolekulare Wechselwirkungen verursacht. Tabelle 1.1 auf der nächsten Seite gibt eine Übersicht über die wichtigsten Eigenschaften.

Tabelle 1.1: Physikochemische Eigenschaften von Aminosäuren.

| Aminosäure | Masse der Seitenkette (Da) | | pI ⁹ | Hydrophobizität ¹⁰ | Häufigkeit* (%) |
|------------|----------------------------|------------------|-----------------|-------------------------------|-----------------|
| | mono-isotopisch | durchschnittlich | | | |
| Ala | 71,04 | 71,08 | 6,01 | 1,8 | 7,87 |
| Arg | 156,10 | 156,19 | 10,76 | -4,5 | 5,42 |
| Asn | 114,04 | 114,10 | 5,41 | -3,5 | 4,13 |
| Asp | 115,03 | 115,09 | 2,77 | -3,5 | 5,34 |
| Cys | 103,01 | 103,15 | 5,07 | 2,5 | 1,50 |
| Gln | 128,06 | 128,13 | 5,65 | -3,5 | 3,96 |
| Gly | 57,02 | 57,05 | 5,97 | -0,4 | 6,95 |
| His | 137,06 | 137,14 | 7,59 | -3,2 | 2,29 |
| Ile | 113,08 | 113,16 | 6,02 | 4,5 | 5,91 |
| Leu | 113,08 | 113,16 | 5,98 | 3,8 | 9,65 |
| Lys | 128,10 | 128,17 | 9,74 | -3,9 | 5,92 |
| Met | 131,04 | 131,20 | 5,74 | 1,9 | 2,39 |
| Phe | 147,07 | 147,18 | 5,48 | 2,8 | 3,95 |
| Pro | 97,05 | 97,12 | 6,48 | -1,6 | 4,82 |
| Ser | 87,03 | 87,08 | 5,68 | -0,8 | 6,84 |
| Thr | 101,05 | 101,11 | 5,87 | -0,7 | 5,41 |
| Trp | 186,08 | 186,21 | 5,89 | -0,9 | 1,13 |
| Tyr | 163,06 | 163,18 | 5,66 | -1,3 | 3,02 |
| Val | 99,07 | 99,13 | 5,97 | 4,2 | 6,73 |
| N-terminus | 1,01** | | | | |
| C-terminus | 17,01** | | | | |

* Ermittelt aus der Swiss-Prot Datenbank, Version 52.1 vom 20.03.2007

** Befindet sich ein Peptid am C- oder N-Terminus wird diese Masse mit der Masse der Seitenkette addiert.

Da ein Peptid wie jedes chemische Molekül aus Elementatomen aufgebaut ist, die in Isotopen auftreten können, existieren unterschiedliche Massen für ein Peptid:

Die **monoisotopische Masse** (oder exakte Masse) ist die Masse eines Moleküls, das nur aus den am häufigsten natürlich vorkommenden Isotopen zusammengesetzt ist.

Die **durchschnittliche Molmasse** (oder relative Masse) wird mittels der durchschnittlichen Atommassen der Elemente in einer Summenformel bestimmt. Die durchschnittliche Atommasse berechnet sich wiederum aus dem gewichteten arithmetischen Mittel der Atommassen der Isotope mit den natürlichen Häufigkeiten der Isotope als Gewichte.

Der **isoelektrische Punkt (pI)** ist der pH-Wert, bei dem die Zahl der posi-

tiven und negativen Ladungen eines amphoteren Moleküls im statistischen Mittel genau gleich ist. Experimentell lässt sich diese Größe über isoelektrische Fokussierung ermitteln. Durch Bildung des arithmetischen Mittels aus den Einzelwerten der vorkommenden Aminosäuren lässt sich der Wert rechnerisch abschätzen.

Die **Hydrophobizität**, welche das Ausmaß des hydrophoben Effekts bei Molekülen beschreibt, lässt sich durch eine von Kyte und Doolittle entwickelte Skala messen.¹⁰ Je hydrophober die Aminosäure ist, desto höher ist ihr Hydrophobizitätsindex auf einer Skala, die von -4,5 bis 4,5 reicht.

Der **GRAVY-Score** („GRand AVerages of hYdrophobicity“) berechnet sich aus dem arithmetischen Mittel der Hydrophobizitätsindizes der Aminosäuren, aus denen ein Protein aufgebaut ist und stellt eine gute Approximation für die Hydrophobizität des Proteins dar.

1.1.3 Ziele der Proteomik

Das übergeordnete Ziel der Proteomforschung ist die Beschreibung der Proteine bezüglich ihrer molekularen Funktion, ihrer Beteiligung an biologischen Prozessen und ihre Lokalisation innerhalb der Zelle bzw. ihrer Kompartimente. Speziell die MS-basierte Proteomforschung liefert hierzu ihren Beitrag durch vier Eckpfeiler:

Identifizierung einzelner Proteine bis hin zur parallelen Identifizierung tausender Proteine wird durch MS-basierte Analyseverfahren wie dem *Peptidmassen-Fingerprint* (PMF)¹¹⁻¹⁵ oder der *Peptidfragmentmassen-Fingerprint-Analyse* (PFF)¹⁶⁻¹⁹ ermöglicht, wie in den Abschnitten 1.5.3 auf Seite 25 und 1.5.4 auf Seite 25 näher ausgeführt wird.

Charakterisierungen von Proteinen erfolgen über die Bestimmung biophysikalischer und biochemischer Eigenschaften. Der Fokus in der MS-basierten Charakterisierung von Proteinen liegt in der Beschreibung von Modifikationen im Allgemeinen und insbesondere von *posttranslationalen Modifikationen* (PTMs).

Quantifizierungen von Proteinen erfolgen entweder mit absoluten oder relativen Werten und dienen der Bestimmung darüber, in welchen Anteilen bzw. Abundanzen Proteine in einer Probe vorliegen.

Vergleichende Proteomanalyse beschreibt die Analyse von Abweichungen bzw. Ähnlichkeiten in der Proteinzusammensetzung von verschiedenen Proben. Abweichungen können die Präsenz, die Abundanz oder den Modifizierungsgrad betreffen.

1.1.4 Arbeitsabläufe in der Proteomik

Die MS-basierte Proteomforschung lässt sich in zwei große Stränge unterteilen: die Analyse von intakten Proteinen, *Top-Down-Proteomics*,^{20,21} und die Analyse von Peptidmischungen, die mit einem enzymatischen Verdau eines Proteins erzeugt werden, *Bottom-Up-Proteomics* (siehe Abbildung 1.2).²² Die Arbeitsabläufe beider Stränge sind in ihrer Grundstruktur identisch und gliedern sich in drei Teile. Der erste Teil betrifft die **Probenvorbereitung** (Abschnitt 1.2 auf der nächsten Seite), in der die Probe dem Massenspektrometer zugänglich gemacht und die Komplexität der Probe reduziert wird. Im Anschluss folgen die **massenspektrometrischen Messungen** (Abschnitt 1.4 auf Seite 13). Hierbei muss für die entsprechende Probe und den verfolgten Ansatz ein optimales Instrument gewählt und dessen Einstellungen angepasst werden. Die **Datenanalyse** (Abschnitt 1.5 auf Seite 23) bildet den Abschluss, die gesammelten Daten werden mittels geeigneter Software ausgewertet.

Der *Top-Down*-Ansatz wird deutlich seltener angewendet, da geeignete Verfahren zur Reduzierung der Probenkomplexität fehlen. Gemische, die bereits mehr als 100 Proteine aufweisen, lassen sich mit den heutigen chromatographischen Systemen nur unzureichend trennen.

Gelelektrophoretische Methoden spielen heute in Proteomik-Experimenten

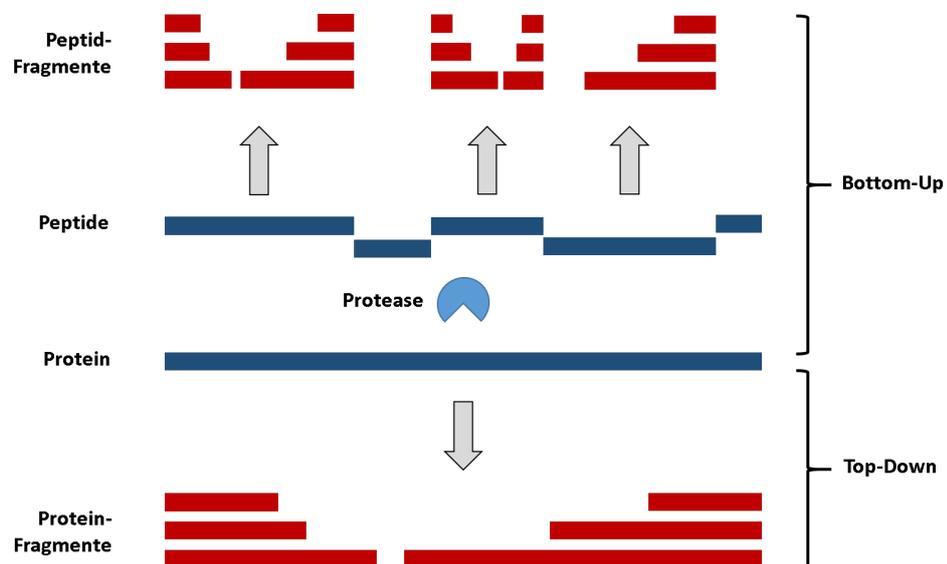


Abbildung 1.2: Übersicht über *Top-Down*- und *Bottom-Up*-Ansätze. In *Top-Down*-Ansätzen werden wie im oberen Teil der Abbildung zu sehen, die Proteine zunächst in Peptide zerkleinert, bevor diese massenspektrometrisch analysiert werden.

trotz höherer Trennkapazitäten aufgrund ihrer hohen technischen Anforderungen und des hohen Zeitaufwandes nur noch eine untergeordnete Rolle. Liegt keine komplexe Probe vor, kann diese auch direkt mit dem Massenspektrometer analysiert werden. Neben der intakten Masse des Proteins, lässt sich dieses zusätzlich in dem Massenspektrometer fragmentieren. Die Kombination aus intakter Masse und Fragmentspektrum ermöglicht die Identifizierung und eventuell auch die Charakterisierung des Proteins.

Bottom-Up- oder *Shot-Gun*-Ansätze sind trotz der Generierung einer Probe mit einem höheren Komplexitätsgrad die meist genutzten Ansätze und liefern die höchste Anzahl an Identifikationen. Gründe hierfür sind das Vorliegen von chromatographischen Methoden, die eine hohe Trenneffizienz aufweisen und so die Vermessung von zehntausenden Peptiden ermöglichen. Des Weiteren werden *Top-Down*-Ansätze durch die meist schlechtere Ionisierbarkeit und Fragmentierung der Proteine limitiert.²³ Neben den chromatographischen existieren Gel-basierte *Bottom-Up*-Ansätze. In dieser Arbeit wurden ausschließlich *Bottom-Up*-Versuche durchgeführt. Dementsprechend liegt der Fokus der nachfolgenden Abschnitte auf eben diesem Ansatz.

1.2 Probenvorbereitung (Sample Preparation)

1.2.1 Zellyse

Zur Analyse des Zellproteoms ist es zunächst notwendig, die Proteine aus dem zu untersuchenden Gewebe oder Organismus zu extrahieren. Im ersten Schritt werden hierfür die intakten Zellen lysiert, so dass der Zellinhalt mit den enthaltenen Proteinen freigesetzt wird. Es stehen eine Reihe von Verfahren zur Verfügung, die sich grob in zwei Gruppen einordnen lassen: Mechanische und nicht-mechanische Aufschlussverfahren. Zu der Gruppe der mechanischen Verfahren gehört zum Beispiel die Lyse mit Hilfe von Ultraschall oder durch wiederholtes Einfrieren und Auftauen. Diese Methoden werden noch heutzutage noch breit angewendet, haben jedoch zu den nicht mechanischen Methoden, wie zum Beispiel der Aufschluss mit Hilfe von tensidhaltigen Lyse-Puffern zunehmend an Bedeutung verloren. Gründe hierfür sind, dass diese Verfahren insgesamt schonender sind und zu Ergebnissen mit höherer Reproduzierbarkeit führen.²⁴

1.2.2 Protein-Aufreinigung

Das Homogenat, welches aus dem Zellaufschluss resultiert, weist neben den Proteinen Verunreinigungen auf, die entfernt werden müssen, da sie die darauf folgenden Prozessschritte empfindlich stören können. Dieser Schritt fällt üblicherweise weg, wenn die Proteine mithilfe eines SDS-Gel aufgetrennt werden. Aufreinigung durch Fällung der Proteine gehört zu den

ältesten Methoden. Dies erfolgt klassischerweise durch Zusatz eines organischen Lösungsmittels, so dass sich die Verunreinigungen nach Fällung im Überstand befinden. Häufig genutzte Lösungsmittel sind 5-10% Trichloressigsäure (TCA) und Aceton (Aceton und TCA-Fällung).²⁵

Aufreinigungen basierend auf Fällung können gewisse Nachteile mit sich bringen. Beispielsweise lassen sich manche Proteine, nachdem sie gefällt wurden, nur durch Zusatz von Tensiden wieder in Lösung bringen. Die Fällung kleinerer Mengen erweist sich ebenfalls häufig als problematisch.²⁴ Neuere Methoden setzen Molekularfilter ein, die Proteine zurückhalten und Verunreinigungen kleinerer Molekulargewichte bei Zentrifugation passieren lassen. Die Filter-assistierten Probenaufbereitungsmethoden (*Filter Assisted Sample Preparation*, FASP^{26,27}) gehören mittlerweile zu den am häufigsten genutzten Verfahren. Häufig wird das crude Zellysate direkt auf den Filter gegeben und Verunreinigungen werden abzentrifugiert. Nach proteolytischem Verdau der Proteine entstehen Peptide, die aufgrund ihres niedrigeren Molekulargewichtes den Filter nach Zentrifugation passieren können.

Zu den neuesten Innovationen gehören *Bead*-basierte Aufreinigungen wie sie in der Methode *SP3 (Single-Pot Solid-Phase-enhanced Sample Preparation)* verwendet werden.²⁸ Die SP3-Methode verwendet magnetische Beads, die einen ferromagnetischen Kern und eine hydrophile Oberfläche aufweisen. Eine Bindung der Proteine an die Beadoberfläche gelingt durch den Zusatz organischer Lösungsmittel. Durch Anlegen eines Magneten an der Gefäßwand werden die Beads zurückgehalten, während der Überstand abpipettiert werden kann. Durch Wiederaufnahme der Beads in einem wässrigen Lösungsmittel, liegen die Proteine wieder gelöst vor.

1.2.3 Proteolyse

Wie im Abschnitt 1.1.4 auf Seite 6 erläutert, existieren in der Proteomik zwei Hauptstrategien (*Top-Down*- und *Bottom-Up*-Ansätze). In *Bottom-Up*-Ansätzen wird ein Protein in Peptide zerlegt, welche leichter massenspektrometrisch zu analysieren sind. Anhand der Identifikation eines oder mehrerer Peptide lässt sich auf die Präsenz eines Proteins in der Probe schließen. Proteolysen können entweder chemisch (Bromcyan-Spaltung, Säurespaltung)^{29,30} oder enzymatisch mithilfe von Proteasen erfolgen. Die mit großem Abstand am häufigsten verwendete Protease ist Trypsin,³¹ welches C-terminal nach Arginin oder Lysin schneidet.

Das Verdauenzym Trypsin genießt nicht nur aufgrund seiner hohen Spezifität, seiner Robustheit gegenüber den Reaktionsbedingungen und seiner Schnitteleffizienz eine so große Beliebtheit, sondern auch wegen der Produkte, die aus der Proteolyse entstehen. Die entstandenen Peptide weisen (mindestens) zwei basische Gruppen auf - mit Ausnahme des C-terminalen Peptids und dem Sonderfall eines modifizierten N-Terminus: die Aminogruppe am

N-Terminus des Peptids und die basische Gruppe am C-Terminus. Die leichte Protonierung basischer Gruppen und die damit einhergehende positive Ladung des Peptids schaffen positive Voraussetzungen für eine empfindliche massenspektrometrische Analyse. Des Weiteren wurde gezeigt, dass die Fragmentierung von Peptiden durch basische Gruppen in der Seitenkette beeinflusst wird.³² Die Präsenz von basischen Gruppen an den Termini des Peptids führt zu einer gerichteten Fragmentierung des Peptids (speziell bei Stoß-induzierten Fragmentierungen, siehe Abschnitt 1.4.4 auf Seite 20).³³ Fragmentationenspektren von tryptischen Peptiden sind aus diesem Grund leichter zu interpretieren.

Neben Trypsin finden auch andere Proteasen Verwendung,³⁴ so etwa das ArgC, LysC, GluC und die weniger spezifischen Proteasen Chymotrypsin,³⁵ Pepsin³⁶ und Elastase.^{37,38} Diese werden allerdings aufgrund ihrer geringeren Spezifität und der Generierung von Peptiden, deren Fragmentspektren im Gegensatz zu denen von tryptischen Peptiden schwieriger zu interpretieren sind,³⁹ seltener und häufig nur komplementär zum Trypsinverdau eingesetzt.

1.2.4 Methoden zur Peptidtrennung

Bevor eine proteomische Probe dem Massenspektrometer zugeführt werden kann, muss in der Regel eine Reduzierung der Komplexität der Probe vorgenommen werden. Beim Vorliegen eines sehr hohen Komplexitätsgrades können zwei chromatographische Methoden, die orthogonale Eigenschaften aufweisen, miteinander verknüpft werden.⁴⁰⁻⁴³ Die erste chromatographische Methode dient hierbei meist zur Vorfraktionierung, die auch offline erfolgen kann wie zum Beispiel über *Stage-Tipping*^{44,45}, die hochaufgelöste Peptid-Trennung erfolgt dann typischerweise mit C18-Material (*Reversed Phase*, RP).

Stage-Tips sind Pipetenspitzen, die an ihrer Spitze z.B. Kationenaustauscher-Material haben. Die Peptidprobe kann durch Aufziehen mit der Pipette auf dieses Material geladen werden. Durch das anschließende Auf- und Abziehen von Elutionsmitteln steigender Konzentration ist eine Vorfraktionierung der Probe möglich.⁴⁶

Alternativ können auch zwei chromatographische Säulen mit unterschiedlichem Material direkt miteinander gekoppelt sein (*MudPIT*⁴⁷). Da Säulen mit C18-Material die höchste Trenneffizienz aufweisen, sind diese direkt mit dem Massenspektrometer gekoppelt. Die Vorfraktionierungsmethode muss demnach chromatographisch orthogonale Eigenschaften zum C18 aufweisen. Am häufigsten verwendet werden Ionenaustauscher, wobei Kationenaustauscher (*Strong-Cation-Exchange*, SCX) ein saures und Anionenaustauscher (*Strong-Anion-Exchange*, SAX) ein basisches Milieu brauchen. Orthogonale Resultate lassen sich auch mit dem selben Säulenmaterial erzielen

durch Veränderung des pH-Wertes des Elutionsmittel. Zum Beispiel zeigen die Ergebnisse mit einer RP-Säule mit einem sauren und einem basischen Elutionsmittel eine ausgeprägte Orthogonalität auf.^{48,49}

Proben mit einem geringen Grad an Komplexität werden meist nur über eine chromatographische Säule (meist RP-Material) getrennt, welche direkt mit dem Massenspektrometer gekoppelt ist. In der Proteomik werden Säulen mit einem geringen Innendurchmesser (ca. 100 μm) zur Steigerung der Sensitivität verwendet. Je kleiner der Durchmesser des Säulenmaterials ist (1,7 - 5 μm) desto schmaler werden die Peaks und desto höher ist die Peakkapazität. Eine Verringerung des Partikeldurchmessers verlangt einen erhöhten Druck zur Aufrechterhaltung der Fließmittelgeschwindigkeit. Anlagen, die bei einem Druck von bis zu 1000 bar arbeiten, werden als *UPLC*-Systeme (Ultra-Performance-Liquid-Chromatography) bezeichnet.

1.3 Chemische Derivatisierung von Proteinen und Peptiden

Peptide und Proteine werden für die verschiedensten analytischen Fragestellungen wie Quantifizierungen, Struktur- und Funktionsaufklärung chemisch derivatisiert. Besteht die Probe aus einer Mischung von Proteinen, ist eine vollständige Derivatisierung ein Hauptziel. Weiterhin muss beachtet werden, dass der Umfang der unerwünschten Seitenreaktionen so gering wie möglich gehalten wird und die Reaktionsbedingungen nicht zu harsch gewählt werden, da es sonst zum Abbau des Proteins kommen kann. Potentielle Angriffspunkte für chemische Derivatisierungen sind die funktionellen Gruppen des Proteins, wobei die reaktivsten die Thiolgruppen der Cysteine und die Aminogruppen der Lysine sind. Aufgrund der hohen Reaktivität dieser Gruppen, werden diese Gruppen bevorzugt modifiziert, da die Umsetzung unter relativ milden Bedingungen nahezu vollständig erfolgen kann. Die in der Proteomik breit genutzten Modifizierungsreaktionen sind im Folgenden aufgeführt. Einen sehr gute Übersicht über die chemische Reagenzien zur Modifizierung von Proteinen hat Roger L. Lundblad verfasst.⁵⁰

1.3.1 Derivatisierung der Cysteine

In nahezu jedem *Bottom-Up*-Ansatz werden Disulfidbrücken, die sich durch eine kovalente Bindung zwischen den Thiolen der Cysteinreste bilden, aufgebrochen. Im ersten Schritt werden hierzu die Disulfide zu Thiolgruppen reduziert und mit der anschließenden Alkylierung des Thiols wird eine Wiederherstellung der Disulfidbrücke verhindert. Insbesondere Proteine mit vielen Disulfidbrücken sind nach Reduktion und Alkylierung für einen trypti-

schen Verdau besser zugänglich und liefern mehr Signale, die zudem eine höhere Intensität aufweisen.

Der zu Grunde liegende Reaktionsmechanismus der Alkylierung ist entweder eine nukleophile Additions- oder Substitutionsreaktion. Die Reaktion mit der Halogen-keto-Verbindung Iodacetamid (siehe Abbildung 1.3) ist ein Beispiel für eine nukleophile Substitutionsreaktion zweiter Ordnung (S_N2 -Reaktion). Die am häufigsten eingesetzten Reduktionsmittel sind 1,4-Dithiothreitol (DTT) und Tris(2-carboxyethyl)phosphin (TCEP).⁵¹

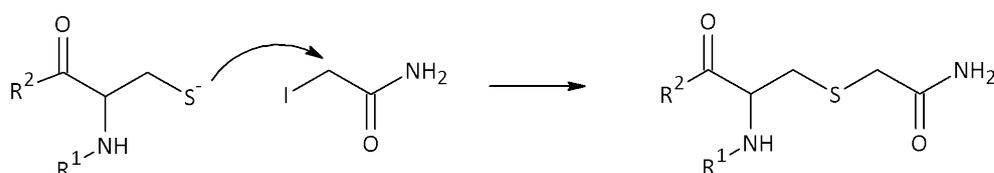


Abbildung 1.3: Alkylierung von Cysteinresten mit Iodacetamid

1.3.2 Derivatisierung der Aminogruppen

Aminogruppen sind an der ϵ -Position der Lysin-Seitenkette und am N-Terminus von Proteinen und Peptiden aufzufinden. Somit verfügt beinahe ausnahmslos jedes Peptid oder Protein über mindestens eine Aminogruppe. Das hohe Vorkommen und die hohe Reaktivität sind die Hauptgründe für die hohe Anzahl an Methoden, die einen Derivatisierungsschritt der Aminogruppen enthalten. Die drei am häufigsten genutzten Derivatisierungsstrategien werden im folgenden gezeigt.

Derivatisierungen mit Carbonsäureanhydriden

Die Derivatisierung mit Carbonsäureanhydriden ist eine schon seit längerem bekannte Möglichkeit Aminogruppen zu acylieren (siehe Abbildung 1.4 auf der nächsten Seite). Unerwünschte Seitenreaktionen treten bei phenolischen Hydroxylgruppen, Carbonsäuren und dem Imidazolring der Histidine auf. Diese sind jedoch im Gegensatz zur Modifikation von Aminogruppen labil und lassen sich durch den Zusatz von Hydroxylamin quantitativ umkehren. Eine der bekanntesten Aminomodifikationen mit einem Carbonsäureanhydrid in der Proteomik ist die Propionylierung der Histon-Aminogruppen mit Propionsäureanhydrid.⁵²

Reduktive Methylierung

Die reduktive Methylierung von Aminogruppen besteht aus zwei Schritten. Im ersten Schritt wird Formaldehyd zugesetzt, welches vom Stickstoff des Amins nucleophil unter Bildung einer Schiff'schen Base angegriffen

tid oder Protein zu binden (siehe Abbildug 1.6 auf der vorherigen Seite). Aus diesem Grund werden sie z. B. verwendet, um das Protein zu biotinylieren. *TMT* (Tandem Mass Tag, Thermo Scientific) ist einer der bekanntesten Quantifizierungsmethoden,⁵⁴ in der die Aminogruppen ebenfalls durch NHS-Ester acyliert werden.

1.4 Massenspektrometrie

Die Massenspektrometrie ist ein Verfahren zur Bestimmung der Masse von Atomen und Molekülen und geht auf die Arbeiten von J.J. Thomson im Jahre 1913 zurück, in denen er zeigte, dass das Edelgas Neon aus einer Mischung von zwei Isotopen der Massen 20 und 22 besteht. Heutzutage existieren unterschiedlichste Apparaturen für massenspektrometrische Messungen, die jedoch alle aus den drei selben Funktionseinheiten aufgebaut sind. Die erste Einheit erzeugt Ionen in der Gasphase und stellt somit eine Ionenquelle dar. Die ionisierte Probe gelangt in den Massenanalysator, der die Ionen nach ihrem Masse-zu-Ladungs-Verhältnis trennt, bevor die „Menge der Ionen“ oder die „Intensität“ vom Detektor erfasst werden.⁵⁵ Die hohe Anzahl der unterschiedlichen Massenspektrometer ist begründet durch die hohe Auswahl an Ionisationsmethoden und Massenanalysatoren, die je nach analytischer Fragestellung zum Einsatz kommen. In den folgenden Abschnitten werden nur die Massenanalysatoren vorgestellt, die in dieser Arbeit verwendet wurden.



Abbildung 1.7: Funktionseinheiten eines Massenspektrometers

1.4.1 Begriffe

Das **Masse-zu-Ladungs-Verhältnis** - genauer die relative Masse dividiert durch die Ladungszahl - (m/z -Verhältnis) ist eine dimensionslose physikalische Größe, die in verschiedenen physikalischen Bereichen von Bedeutung ist. Zwei Teilchen mit dem gleichen m/z -Verhältnis legen innerhalb eines Vakuums den gleichen Weg zurück, wenn das selbe elektrische Feld angelegt wird. Folglich können unterschiedliche Teilchen mit dem selben m/z -Verhältnis massenspektrometrisch nicht auseinander gehalten werden. Die Einheit dieser Größe ist das *Thomson* (Th).

Ein **Massenspektrum** ist eine zweidimensionale Darstellung der Signalintensität gegen das m/z -Verhältnis der Ionen.

Der **Basispeak** ist das höchste Signal in einem Massenspektrum, auf dessen Intensität die Intensitäten der anderen Signale (*Peaks*) normiert werden können (*relative Intensität*).

Eine **Peakliste** ist eine tabellarische Wiedergabe eines Massenspektrums. Sie besteht aus mindestens zwei Spalten, wobei die erste Spalte das m/z -Verhältnis und die zweite Spalte die Intensität wiedergibt.

1.4.2 Ionisationsmethoden

Ionisation beschreibt den Übergang eines neutralen Teilchens zu einem geladenen Teilchen. Die Bildung von positiv geladenen Teilchen (Kationen) geht einher mit dem Verlust von Elektronen oder erfolgt durch die Anlagerung positiv geladener Teilchen wie Protonen oder anderer Kationen.

Die *Elektronenstoßionisation* (EI) ist das klassische Ionisationsverfahren. Hierzu werden Elektronen, emittiert von der Oberfläche eines Filaments, durch ein elektrisches Feld beschleunigt, bei der Kollision mit einem Atom oder Molekül kommt es zur Freisetzung eines Elektrons aus dem Analyten, so dass dieses positiv geladen vorliegt. Diese Art der Ionisation führt wegen der damit verbundenen Energieübertragung in der Regel zum Zerfall des Moleküls.⁵⁵

In den späten achtziger Jahren wurden zwei neue Methoden eingeführt, bei denen der Analyt während des Ionisationsvorganges nicht zerfällt. Diese Art der Ionisierung wird als „sanfte“ Ionisation bezeichnet und fand 1984 mit der Entwicklung der *Elektrospray-Ionisation* (ESI) ihren ersten Vertreter^{56,57}. Neben dieser Sprühhmethode wurde praktisch gleichzeitig die *Matrix-unterstützte Laser-Desorption/Ionisation* (MALDI) entwickelt⁵⁸, die zu den Photoionisationsmethoden gehört. Die Entwicklung dieser beiden Ionisationsmethoden ermöglichte die Ausweitung der massenspektrometrischen Analyse auf große Biomoleküle und stellte die Basis für die heutige Massenspektrometrie-basierte Proteomforschung dar.

Elektrospray-Ionisation (ESI)

Elektrospray-Ionisation ist ein Verfahren, mit dem Ionen aus der flüssigen in die Gasphase überführt werden. Ausgehend von einer Spritzenpumpe oder durch ein gekoppeltes HPLC-System wird die Analytlösung durch eine Kapillare geleitet, die in Relation zu einer zylindrischen Gegenelektrode auf ein Potenzial von mehreren kV gehalten wird. Die am Kapillarende austretende Flüssigkeit wird durch das elektrische Feld zu einem Taylor-Kegel verformt. An der Spitze dieses Konus wird ein filamentarti-

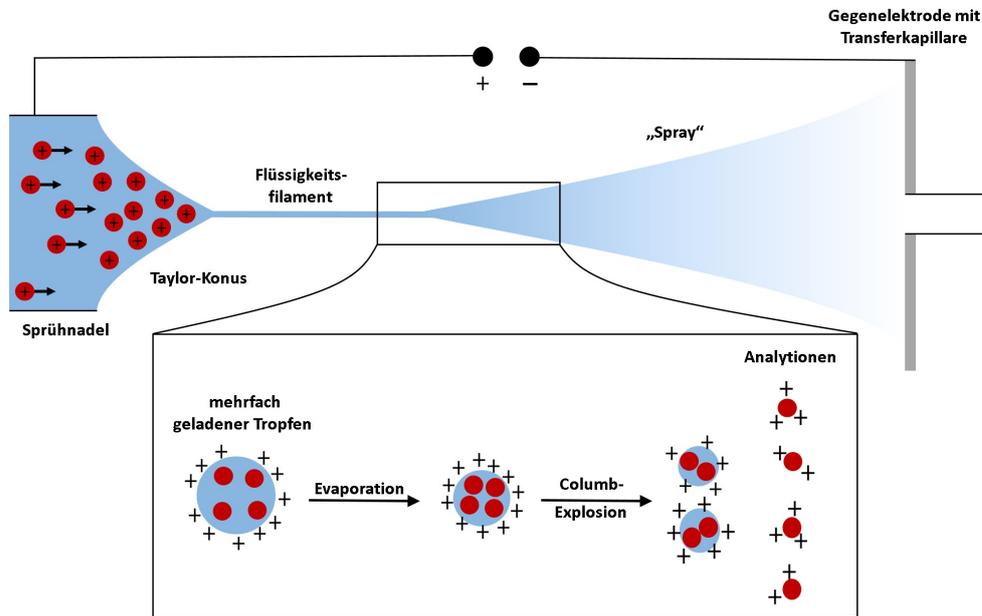


Abbildung 1.8: Elektrospray-Ionisation (ESI)

ger Flüssigkeitsstrom emittiert. Dieser Jet zerfällt unter Ausbildung kleiner Tröpfchen.

Durch die Verdampfung des Lösungsmittels, bei höheren Flußraten häufig unterstützt durch ein eingeleitetes Trockengas, schrumpfen die Primärtröpfchen, bis es bei einer kritischen Ladungsdichte in der Oberfläche zur Coulomb-Explosion kommt. Dabei werden ca. 20 Sekundärtröpfchen emittiert, die eine Anreicherung der Oberflächenladung erfahren. Der Prozeß wiederholt sich, so dass ein Aerosol feinsten und zunehmend kleiner Tröpfchen entsteht (siehe Abbildung 1.8).^{59,60} Dieser Vorgang findet unter Atmosphärendruck statt. Der ionisierte Analyt wird entlang eines Potenzial- und Druckgefälles über verschiedene Transfereinrichtungen zum Massenanalysator freigesetzt.

MALDI

Im Gegensatz zur Elektrospray-Ionisation werden die Ionen bei der Matrix-unterstützten Laser-Desorption/Ionisation direkt aus der Festphase ins Vakuum des Massenspektrometers überführt. Die Festphase ist ein Matrix-Analyt-Kokristall, wobei die Matrix eine hochreine Substanz ist, die ihr Absorptionsmaximum nahe der vom Laser emittierten Wellenlänge hat. Meist werden der Analyt und die Matrix, gelöst in einem leicht flüchtigen Lösungsmittel, zusammen auf einen flachen Metall-Proben-teller gegeben, so dass die Lösungen sich miteinander mischen, bevor das Lösungsmittel verdampft. Matrix und Analyt liegen in einem molaren Verhältnis von 1000:1 bis 10'000:1

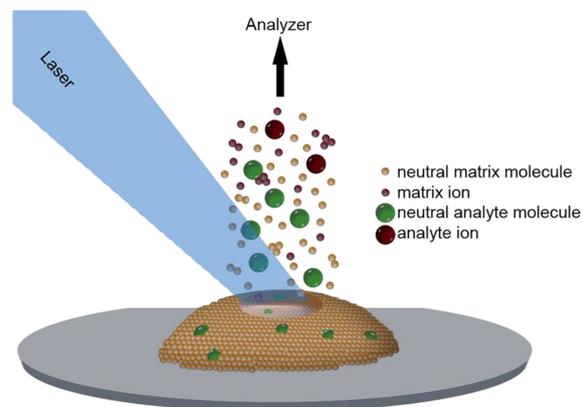


Abbildung 1.9: Matrix-unterstützte Laser-Desorption/Ionisation (MALDI)¹

in dem entstandenen Kokristall vor. Die von der Matrix absorbierte Energie aus dem Beschuss mit den hochenergetischen Laserimpulsen führt zu einer Explosion des Kokristalls und die desorbierten und ionisierten Analytmoleküle befinden sich nun in der Gasphase (siehe Abbildung 1.9). Entlang eines elektrischen Potenzialgefälles werden die Ionen in Richtung des Massenanalysators beschleunigt.

1.4.3 Massenanalysatoren

Im Massenanalysator werden die in der Ionenquelle erzeugten und beschleunigten Ionen getrennt und dann vom Detektor registriert. Orbitrap und FT-ICR-Geräte weichen in der Detektion der Ionen von den anderen Geräten ab. In diesen Geräten ist der Massenanalysator zugleich der Detektor, wohingegen die Detektion der Ionen bei den anderen Messinstrumenten üblicherweise über einen Sekundärelektronenvervielfacher erfolgt.

Bestimmte Massenanalysatoren (z.B. Quadrupole und Magnetsektorinstrumente) beruhen auf dem Prinzip der Bahnstabilität, so dass es zur Aufnahme eines Massenspektrums notwendig ist, die verschiedenen m/z -Werte nacheinander zu registrieren (*Scan-Modus*). Andere Massenanalysatoren nutzen das Prinzip einer Ionenfalle, um unterschiedliche m/z -Verhältnisse eines Ionenpaketes zu vermessen (z.B. Quadrupol-Ionenfallen, FT-ICR und Orbitrap). Für gepulste Ionenerzeugung können *Time-of-Flight* (TOF)-Geräte eingesetzt werden, die alle Ionen eines Ionisierungsereignisses registrieren können. Neben diesen Unterscheidungsmerkmalen existieren noch eine Reihe weiterer verschiedener Parameter, die ein Massenspektrometer charakterisieren:

Der zugängliche **Massenbereich** (angegeben in m/z).

Massenauflösung oder **Auflösungsvermögen (R)**:

Das Auflösungsvermögen R bzw. die Massenauflösung ist das wichtigste Leistungskriterium für ein Massenspektrometer und wird bestimmt durch:

$$R = \frac{m}{\Delta m} = \frac{m/z}{\Delta m/z} \quad (1.1)$$

Es existieren im wesentlichen zwei unterschiedliche Methoden zur Bestimmung der Auflösung. Die *Peakbreiten-Definition* setzt für Δm die Breite eines Peaks bei einer bestimmten Peakhöhe ein. Üblich sind Breitenbestimmungen bei Höhenanteilen von 0,5, 5, 10 und 50%, wobei der Wert bei 50% auch als *Halbwertsbreite (Full Width at Half Maximum, FWHM)* bezeichnet wird. Die *10%-Taldefinition der Auflösung ($R_{10\%}$)* definiert zwei benachbarte Peaks mit gleicher Intensität als hinreichend getrennt, wenn das Tal zwischen den Peaks nicht oberhalb von 10% ihrer Intensität liegt.

Die **Massenbestimmungsgenauigkeit** gibt an, wie genau die Masse des Teilchens bestimmt werden kann, welches über die Bildung der Differenz zwischen dem experimentell und theoretischem m/z -Verhältnis bestimmt wird. Die Angabe erfolgt meist in ppm als Verhältnis zum theoretischen Wert.

Quadrupol

Ein Quadrupol besteht aus vier hyperbolisch geformten Elektroden, die zu jeweils zwei Paaren entlang einer Symmetrieachse im selben Abstand quer gegenüberliegen. Die gegenüberliegenden Elektroden befinden sich auf dem gleichen Potenzial. Zwischen den Elektrodenpaaren wird eine Hochfrequenzspannung mit überlagerter Gleichspannung angelegt. Ionen, die in dieses Spannungsfeld gelangen, werden in ihrer Flugbahn beeinflusst und nur Ionen mit einem bestimmten m/z -Verhältnis gelangen auf eine stabile Flugbahn und somit zum Detektor.

In bestimmten Versuchsansätzen liegt das Interesse auf der Detektion eines Analyten mit einem bestimmten m/z -Verhältnis. In diesen Fällen werden die angelegten Frequenzen und Spannungen konstant gehalten. Geräte, die hierzu zum Einsatz kommen, haben einen Massenanalysator, der aus drei hintereinander geschalteten Quadrupolen aufgebaut ist. Der erste Quadrupol selektiert ein Vorläufermolekül mit einem bestimmten m/z -Verhältnis, welches im nächsten Quadrupol fragmentiert wird und aus den Fragmentionen selektiert wiederum der dritte Quadrupol ein Ion mit einem vordefinierten m/z -Verhältnis. Dieses Verfahren wird als *Single Ion Monitoring (SIM)* bezeichnet und zeichnet sich durch eine hohe Empfindlichkeit aus.

Deutlich häufiger werden Quadrupole jedoch für die Aufnahme von Massenspektren mit größeren Massenbereichen verwendet. Dies erfolgt durch

die zeitlich festgelegte Variation der angelegten Hochfrequenzen und Gleichspannungen. Quadrupole werden häufig mit anderen Massenanalysatoren kombiniert und dienen diesen als vorgeschaltetem Massenfilter.

Orbitrap

Die Orbitrap^{61,62} gehört neben den *FT-ICR* (Fourier-Transform-Ionencyclotronresonanz) -Analysatoren zu den Analysatoren mit dem höchsten Auflösungsvermögen und ermöglicht sehr exakte Massenbestimmungen. *FT-ICR*-Geräte, auf die hier nicht näher eingegangen wird, benötigen Magnetfelder hoher Flussdichte (bis zu 21 Tesla werden eingesetzt), so dass diese Geräte zum Einen groß und durch die Kryomagnete auch sehr teuer sind.

Die Orbitrap ist aufgebaut aus einer spindelförmigen Zentralelektrode, die umgeben ist von einer zweigeteilten fassförmigen äußeren Elektrode. Ein Ionenpaket kann auf einer Umlaufbahn entlang der spindelförmigen Zentralelektrode eingefangen werden und somit gehört die Orbitrap zur Gruppe der Ionenfallen, die Ursprünge sind auf K. H. Kingdons⁶³ Arbeiten zurückzuführen. Beide Elektroden sind mit Spannungsquellen verbunden und erzeugen ein elektrostatisches Feld, welches von außen injizierte Ionen auf einer Umlaufbahn entlang der Spindelelektrode fängt. Die elektrostatische Anziehung in Richtung der zentralen Elektrode wird durch eine Zentrifugalkraft kompensiert, die von der anfänglichen Tangentialgeschwindigkeit der Ionen herrührt. Neben der kreisförmigen Rotation schwingen die Ionen auch axial entlang der Spindel. Die Frequenz dieser axialen Schwingung (ω) ist allein Abhängig von dem Verhältnis der Ionenladung (z) und der Ionenmasse (m), sowie einer vom Analyten unabhängigen Größe (Feldkrümmung, k):

$$\omega = \sqrt{k \frac{z}{m}} \quad (1.2)$$

Die oszillierenden Ionenpakete erzeugen an den Außenelektroden der Orbitrap einen abklingenden Induktionsstrom, der über eine gewisse Zeit gemessen wird. Das Massenspektrum wird durch Fouriertransformation des Transienten erzeugt.

Flugzeitanalysatoren

Mit einem Flugzeitanalysator (Time-of-Flight, TOF)⁶⁴ kann die Zeit gemessen werden, die Ionen zur Passage einer definierten Strecke benötigen. Die Ionen werden hierzu durch Anlegung eines elektrischen Feldes bekannter Stärke zunächst beschleunigt, sodass gleich geladene Teilchen dieselbe kinetische Energie erlangen. Die Ionen werden in einen feldfreien Raum hin beschleunigt, die Driftstrecke (L). Linear aufgebaute Flugzeitanalysatoren detektieren die Ionen unmittelbar nach der Driftstrecke. Die benötigte Zeit (t) zur Passage der Driftstrecke ist allein abhängig von dem m/z -Verhältnis

der Ionen und einer Konstanten (K , entspricht der kinetischen Energie der Ionen):

$$\frac{m}{z} = 2K \frac{t^2}{L^2} \quad (1.3)$$

Der Einsatz von TOF-MS-Systemen setzt eine gepulste Ionenerzeugung voraus und ist damit direkt für eine MALDI-Ionenquelle nutzbar. Aus diesem Grund wird die Kombination von MALDI-Quellen mit Flugzeitanalysatoren häufig genutzt.

Die beim MALDI-Prozess erzeugten Ionen haben durch den Ablationsprozess eine Geschwindigkeit (und Geschwindigkeitsverteilung) senkrecht zur Oberfläche des metallischen Probetellers, erfahren aber auf Grund der hohen Teilchendichte bei der Auflösung des Festkörpers unmittelbar nach dem Laserbeschuss zahlreiche Stöße, was bei angelegter Beschleunigungsspannung zu einer komplexen und massenabhängigen Verteilung von Startort, -geschwindigkeit und -energie führt. Damit sind die in einem einfach TOF-Analysator erreichbaren Leistungsdaten extrem limitiert. Um insbesondere den Auflösungsverlust möglichst gering zu halten, wurden die Instrumente in Ihrem Aufbau kontinuierlich weiterentwickelt, wobei zwei Weiterentwicklungen die größten Effekte gezeigt haben.

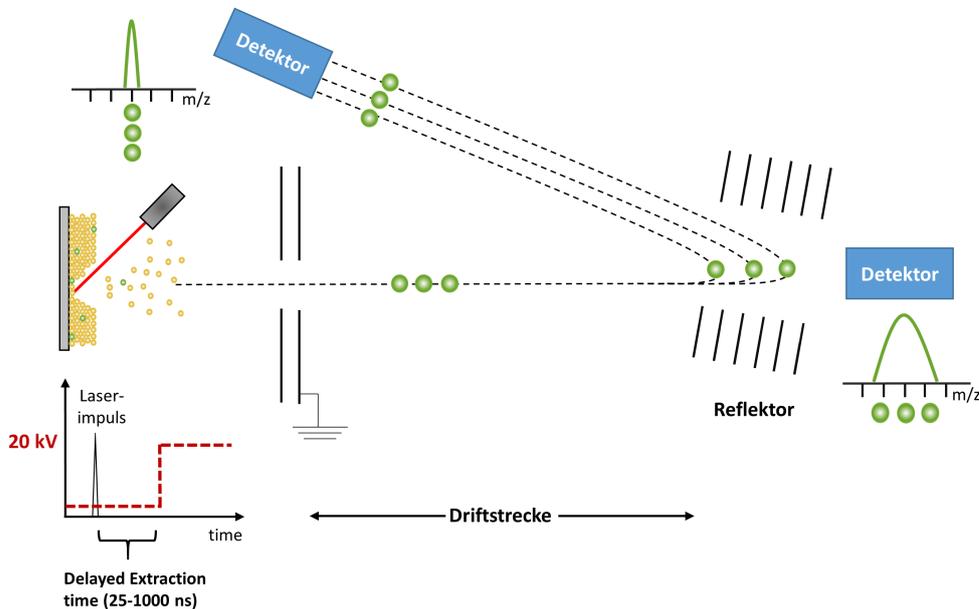


Abbildung 1.10: Aufbau eines Flugzeitanalysators mit Reflektion.

Werden Ionen nicht unmittelbar nach dem Laserbeschuss zur Driftstrecke hin beschleunigt, sondern verzögert im Zeitbereich von ca. 100 ns, wird

aus der ursprünglichen Geschwindigkeitsverteilung eine Ortsverteilung, die nach der Beschleunigung der Ionen zu einer Energieverteilung der Ionen führt, die bei geeigneter Konfiguration der Ionenquelle (zweistufig) und Einstellung der experimentellen Parameter (Spannungsverteilung und Verzögerungszeit) genutzt werden kann, so dass die Flugzeitfehler für den Ort des Detektors zumindest partiell kompensiert werden. Gleichzeitig erfolgt in der Verzögerungszeit eine Verdünnung der Dichte der Ablationswolke, so dass die Stöße und ihre abträglichen Effekte reduziert sind. Dieses Vorgehen der verzögerten Extraktion der Ionen wird als *Delayed-Extraction* (DE) bezeichnet.^{65,66}

Eine weitere Erhöhung der Massenauflösung wurde durch den Einbau eines Reflektrons⁶⁷ erreicht, welches die Flugrichtung der Ionen umkehrt. In einem ein- oder zweistufigen elektrischen Gegenfeld dringen die Ionen mit höherer kinetischer Energie tiefer ein, so dass sie mehr Zeit im Reflektron verbringen, was wiederum für den Ort des Detektors eine Kompensation der Laufzeitunterschiede in der feldfreien Driftstrecke ermöglicht (siehe Abbildung 1.10 auf der vorherigen Seite). Wird der Zeitfokus der verzögerten Extraktion nahe an die MALDI-Ionenquelle gezogen und der unterschiedlichen kinetischen Energien der Ionen durch ein nachfolgendes Reflektron-TOF Rechnung getragen, sind trotz der ursprünglich extrem anspruchsvollen Ioneneigenschaften sehr gute Leistungsdaten des MS-Systems möglich.

Vergleich der Massenanalytoren

Neben den in dieser Arbeit vorgestellten Massenanalytoren existieren wie bereits oben erwähnt noch eine Reihe weiterer, die je nach analytischer Fragestellung ihre Vorzüge aufweisen und in bestimmten Bereichen gehäuft eingesetzt werden. In der Tabelle 1.2 auf der nächsten Seite werden die Analytoren, die in dieser Arbeit Verwendung fanden, miteinander verglichen. Die Charakteristika wurden einleitend erklärt (siehe den Beginn dieses Abschnittes). Hervorzuheben sind beim TOF der hohe Massenbereich, beim Quadrupol die Robustheit und bei der Orbitrap das hohe Auflösungsvermögen.⁶⁸

1.4.4 Peptidfragmentierung

Fragmentierung beschreibt ganz allgemein die Spaltung von Molekülonen, die bei den meisten Ionisationsverfahren schon während dieses Prozesses eintritt (Ausnahme: ESI und MALDI). Bei Reinsubstanzen ist die Fragmentierung der Probe während der Ionisation nicht weiter problematisch. Das Signal mit dem größten m/z -Verhältnis lässt sich in den meisten Fällen als unfragmentierter Analyt identifizieren, das zugehörige Signal wird als *Molpeak* bzw. *Quasi-Molpeak* bezeichnet. Die übrigen im Massenspektrum zu beobachtenden Signale lassen sich auf Fragmentionen zurückführen, die

Tabelle 1.2: Vergleichende Übersicht der verschiedenen Massenanalysatoren.

| | TOF | Quadrupol | Orbitrap* |
|---------------------------|--|--|--|
| Auflösungsvermögen (FWHM) | Bis über 30'000, aber typischerweise 15'000 | 100 - 10'000 | 100'000 |
| Massengenauigkeit (ppm) | 2 -50, häufig 20 | 50 - 100 | 1-2 |
| Massenbereich (m/z) | über 100'000 | bis zu 4000, üblicherweise unter 2000 | bis 4000 |
| Vorteile | <ul style="list-style-type: none"> • schnelle Aufnahmege- schwindigkeit • höchster Massenbereich • gut kombinierbar mit MALDI | <ul style="list-style-type: none"> • toleriert hohe Drücke • geringe Größe der Apparatur • gut geeignet für Tandem-MS | <ul style="list-style-type: none"> • hohe Massengenauigkeit und Auflösung • sehr sensitiv (sub-femtomol) |
| Nachteile | <ul style="list-style-type: none"> • hohe Anforderung bzgl. der initialen Ionen-Energie • geringe Auflösung im linearen Modus • Hochvakuum ist erforderlich | <ul style="list-style-type: none"> • geringe Auflösung • schwer zu kombinieren mit MALDI • geringe Scan-Geschwindigkeit | <ul style="list-style-type: none"> • Hochvakuum ist erforderlich • Hohe Preis |
| Kosten | €€ | € | €€€ |

* Gilt für die LTQ-Orbitrap

häufig zur Identifizierung und Strukturaufklärung des Analyten herangezogen werden.

Aufgrund der hohen Komplexität der proteomischen Proben, kommen nur „sanfte“ Ionisationsverfahren wie ESI und MALDI zum Einsatz. Peptide, die aus einem *Bottom-Up*-Ansatz resultieren, werden in zwei Schritten massenspektrometrisch analysiert. Im ersten Schritt werden die Molpeaks der Peptide aufgenommen. Aufgrund des gewählten Ionisationsverfahrens werden keine Fragmentationen erzeugt und detektiert. Im zweiten Schritt werden die einzelnen Molpeaks selektiert und fragmentiert. Die massenspektrometrische Analyse dieser Ionen liefert das *MS/MS-Spektrum* des selektierten

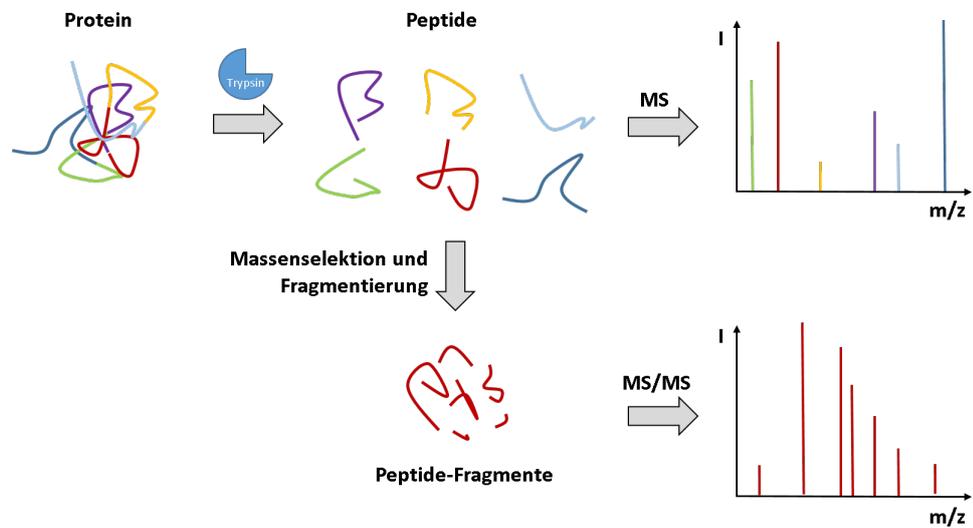


Abbildung 1.11: Schematische Darstellung zur Aufnahme von MS und MS/MS-Spektren

Vorläuferions. Diese Technik, bei der massenselektierte Ionen einer zweiten massenspektrometrischen Analyse unterzogen werden, wird als MS/MS-Verfahren bezeichnet (siehe Abbildung 1.11),^{69,70} je nach verwendeten Analysator als *Tandem-* (zwei gleiche) oder *Hybrid-MS/MS* (zwei unterschiedlichen Analysatoren).

Durch den ähnlichen Aufbau der Peptide und das Vorhandensein von leicht zu spaltenden Bindungen innerhalb des Peptid-Rückgrats ist ein Peptid-Fragment-Spektrum zu gewissen Teilen vorhersagbar und somit leicht zu interpretieren. Die Kombination aus Bestimmung der exakten Masse eines Peptides und das Vorliegen seines Fragment-Spektrums ermöglicht häufig eine Identifizierung des Peptides mit hoher Sicherheit (*De-novo-Sequenzierung*). Erleichtert wird der Prozeß, wenn die möglichen Peptide und ihre Sequenzen bekannt sind, wie im Falle der Proteomics-Analytik durch das bekannte Genom. Damit kann die Spektreninterpretation und die Identifizierung der Peptide und der zugehörigen Proteine mit bioinformatischen Techniken durchgeführt werden (siehe Abschnitt 1.5.4 auf Seite 25).

Roepstorff und Fohlmann führten eine Nomenklatur ein,⁷¹ die es ermöglicht die entstehenden Fragmentionen zu unterscheiden. Fragmente können nur identifiziert werden, wenn sie mindestens eine positive Ladung haben. Verbleibt die Ladung nach dem Fragmentieren auf der N-terminalen Seite wird von a,b oder c-Ionen gesprochen. Wenn sich die Ladung auf der C-terminalen Seite befindet, spricht man von x,y oder z-Ionen. Innerhalb des Peptidrückgrates liegen drei mögliche Bruchstellen vor, sodass jeder Buchstabe eine

davon beschreibt. Eine zusätzlich Indexierung der Buchstaben mit Zahlen gibt die Position der Bruchstelle an (siehe Abbildung 1.12).

Fragmentspektren werden häufig komplexer durch Abspaltungsreaktionen, Umlagerungsreaktionen und Ausbildung interner Fragmentationen.⁷² Bestimmte Aminosäuren wie Prolin führen meist zu starken Signalen und zu Ionenserien ausgehend von der internen Spaltungstelle in Richtung des C-Terminus.⁷³

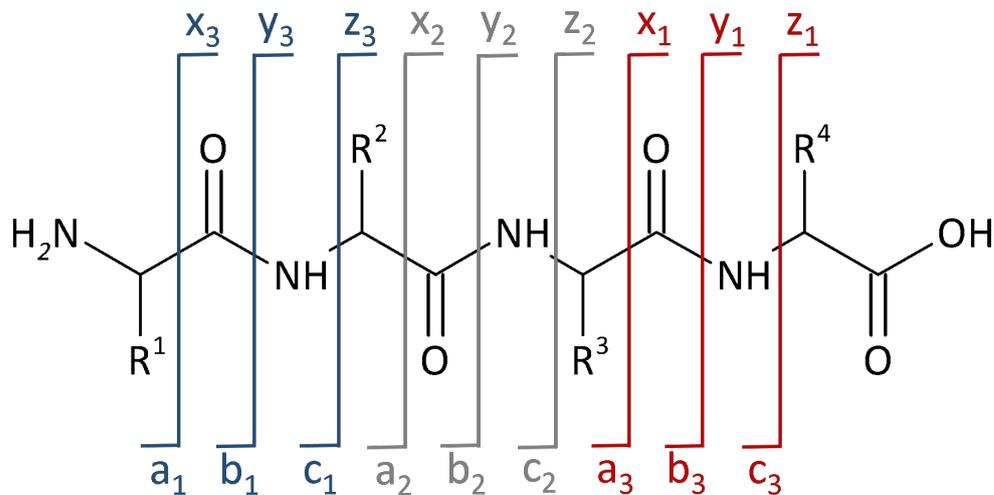


Abbildung 1.12: Nomenklatur nach Roepstorff und Fohlmann.

1.5 Massenspektrometrische Datenanalyse

Neben der Probenvorbereitung und der massenspektrometrischen Vermessung der Probe ist das dritte große Feld in der Proteomik die Datenanalyse. Aufgrund der großen Datenmengen ist hier eine manuelle Interpretation der Daten undenkbar. Erst die Bereitstellung von Software seitens kommerzieller Unternehmen und Forschungsinstituten ermöglichte die Auswertung der Daten, wobei die jetzigen angebotenen Lösungen noch entwicklungsbedürftig sind und vermutlich noch für einige Zeit als Forschungsschwerpunkte erhalten bleiben. Die von den Forschungsinstituten entwickelten Software-Lösungen stehen üblicherweise zur freien Nutzung zur Verfügung (*Freeware*) und in den meisten Fällen ist auch der Quellcode offen (*Open-Source*). Es ist ein zunehmender Trend zu beobachten, dass auch kommerzielle Anbieter Schnittstellen für Freeware-Programme zur Verfügung stellen

und somit dem Benutzer eine flexiblere Auswertung ermöglichen (z.B. ProteomdiscovererTM von Thermo Fisher).

Der grundlegende Arbeitsablauf in der Datenanalyse von proteomischen Daten unterteilt sich in die Konvertierung der Dateien, die Signalprozessierung der Massenspektren und die Identifizierung von Peptiden oder Proteinen mit Hilfe einer geeigneten Suchmaschine. Je nach analytischer Fragestellung können sich dem weitere Schritte anschließen, die beispielsweise Aussagen zu Proteinabundanz zulassen. Im weiteren Verlauf dieses Abschnitts werden die datenanalytischen Einzelschritte kurz erklärt.

1.5.1 Konvertierung der MS-Dateien

Die Grundlage jeglicher datenanalytischen Auswertung ist der Zugriff zu den Rohdaten, die während der massenspektrometrischen Messung üblicherweise in einem herstellerspezifischen Format gespeichert werden. *Open-Source* Anwendungen versuchen hingegen herstellerspezifische Lösungen anzubieten, welches durch die Fülle verschiedener Rohdaten-Formate erschwert wurde. Aus diesem Grund erfolgte auf Initiative der HUPO (*Human Proteome Organization*) und des SPC (*Seattle Proteome Center*) die Festlegung auf ein einheitliches Datenformat namens *mzML*.⁷⁴ Wie jedoch anfangs beschrieben, erfolgt die Speicherung zunächst in einem herstellerspezifischen Format, so dass eine Konvertierung zu *mzML* für eine weitere Datenprozessierung mit *Open-Source*-Programmen notwendig ist. Konvertierungsprogramme wie *MSConvert* von Proteowizard ermöglichen dies.^{75,76}

1.5.2 Signalprozessierung

Ein unprozessiertes Massenspektrum ist anfangs nur eine Kurve, auf deren Abszisse das m/z -Verhältnis und auf der Ordinate die Intensität abzulesen ist. Ziel der Signalprozessierung ist das Auffinden von Signalen (Peaks) im Rauschen. Üblicherweise setzt sich die Signalprozessierung von Massenspektren aus drei Schritten zusammen. Im ersten Schritt wird das Spektrum geglättet. Gängige Glättungsverfahren sind Savitzky-Golay- und Gauß-Filter. Im Anschluss wird die Basislinie bestimmt und sollte diese einen *Drift* aufweisen, kann eine Basislinienkorrektur vorgenommen werden. Auch hierzu gibt es verschiedene Verfahren wie die *Lineare Interpolation* und die *Continuous-Wavelet-Transformation* (CWT). Im Anschluss erfolgt das *Peak-Picking* mit dem Ziel die exakte Position, die Höhe und die Breite bei halber Höhe (FWHM) eines Peaks aus dem Spektrum zu extrahieren. Eine vollständigere Auflistung und Vergleiche der möglichen Methoden wurden von Yang *et al.* durchgeführt.⁷⁷

Das Resultat der erfolgreichen Signalprozessierung ist eine Peakliste, die bei MS-Spektren minimal aus zwei Spalten aufgebaut ist. In der ersten Spalte

sind die m/z -Verhältnisse und in der zweiten Spalte die dazugehörigen Signalintensitäten aufgeführt. Peaklisten von MS/MS-Spektren enthalten in der ersten Zeile zusätzliche Informationen über das m/z -Verhältnis des Vorläuferions und die dazugehörige gemessene Intensität.

1.5.3 Proteinidentifizierung

Wie im Abschnitt 1.4.4 auf Seite 20 gezeigt, können in *Bottom-Up*-Ansätzen aus einem Protein erzeugte Peptide über MS oder MS/MS analysiert werden. Wird ein einzelnes Protein tryptisch verdaut, reicht in vielen Fällen die Aufnahme eines MS-Spektrums. Das Signalmuster in diesem Spektrum weist eine derartig hohe Spezifität auf, dass es eindeutig einem Protein zugewiesen werden kann, ähnlich wie ein Fingerabdruck zu einer Person. In Anlehnung zu dieser Analogie wird dieses Verfahren als *Peptide-Mass-Fingerprint* (PMF) bezeichnet. Diese Technik wurde 1993 von drei Gruppen unabhängig voneinander entwickelt und basierte auf der Isolation von Proteinen mithilfe von SDS-Gelen. Die Gelbanden wurden tryptisch verdaut und über MALDI-TOF analysiert.¹¹⁻¹⁵

Zur Identifizierung muss die Sequenz des Proteins bekannt und in einer Datenbank abgelegt sein (Textdatei im FASTA-Format). In PMF-Suchen berechnet die Suchmaschine zunächst die theoretischen Peptidmassen nach *In-silico*-Verdau der in der Datenbank hinterlegten Proteine. Es folgt ein Abgleich der theoretischen mit den gemessenen Massen und je höher die Anzahl der Zuweisungen, ist desto sicherer ist die Proteinidentifizierung. Es stehen verschiedene Suchmaschinen für PMF-Suchen zur Verfügung:

- Mascot (www.matrix-science.com)
- Profound (www.prowl.com)
- MS-FIT (www.prospector.ucsf.edu/)

In der vorliegenden Arbeit wurde ausschließlich Mascot eingesetzt.⁷⁸

1.5.4 Peptididentifizierung

Wird ein Gemisch aus Proteinen tryptisch verdaut, erfolgt die Identifizierung über die Aufnahme von MS/MS-Spektren der Peptide, die zuvor mit Hilfe von Flüssigchromatographie separiert werden. Die Identifizierung der Peptide anhand ihrer Fragmentationenspektren nennt sich *Peptide-Fragment-Fingerprint* (PFF). Je mehr Signale in einem Massenspektrum Peptidfragmenten zugewiesen werden können, desto sicherer ist die Identifikation. Mittels der identifizierten Peptide lässt sich wiederum rekonstruieren, welches Protein vorlag. Es existieren hauptsächlich drei verschiedene Wege aus Fragmentationenspektren Peptide zu identifizieren, die im Nachfolgenden aufgezeigt werden.

Protein-Datenbank-basierte Suchmaschinen

Ähnlich wie bei der PMF-Suche werden auch hier die in der Datenbank hinterlegten Proteine zunächst in-silico verdaut. Auch hier werden die theoretischen mit den gemessenen Massen der Vorläuferionen abgeglichen. Im Anschluss werden nur diejenigen Peptide, deren Massen innerhalb eines Toleranzfensters liegen, in-silico fragmentiert und es folgt der Abgleich der theoretischen mit den gemessenen Fragmentspektren (siehe Abbildung 1.13).⁷⁹ Dieser Abgleich erfolgt in der Suchmaschine von einem Funktionsmodul na-

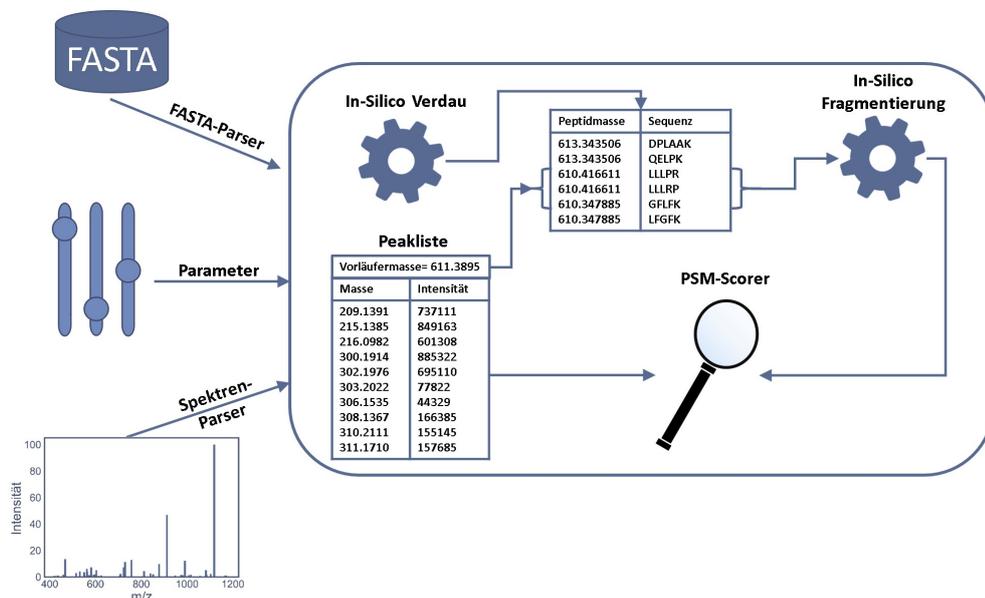


Abbildung 1.13: Schematischer Aufbau einer Datenbank-basierten Suchmaschine

mens *PSM-Scorer* (Peptide-Spectrum-Matches). Es gibt heute eine breite Anzahl verschiedener Suchmaschinen für PFF-Suchen (Mascot,⁷⁸ Sequest,^{78,80} OMSSA,⁸¹ X! Tandem,^{82,83} Myrimatch,⁸⁴ MS-GF+,^{85,86} MS Amanda,⁸⁷ Andromeda⁸⁸), die im Aufbau sehr ähnlich sind und sich hauptsächlich in ihrem *Scoring* unterscheiden. Der PSM-Scorer ist somit das Charakteristikum einer Suchmaschine und führt eventuell zu unterschiedlichen Ergebnissen.

De-Novo-Sequenzierung

De-Novo-Sequenzier-Algorithmen interpretieren ohne zugrunde legende Datenbank MS/MS-Spektren und nutzen die Signalabstände in einem Spektrum zur Vorhersage von Ionenserien aus. Der große Vorteil dieser Methode ist, dass die Sequenzen der Proteine nicht bekannt sein müssen und somit eventuelle Spleißvarianten identifiziert werden können, deren Vorhanden-

sein aus dem Genom nicht ermittelt werden kann. Ein weiterer Vorteil ist die geringere Empfindlichkeit gegenüber post-translationalen Modifikationen. Abspaltungen und das Fehlen von Peaks erschweren die Interpretation. Des Weiteren sind die Algorithmen sehr rechenintensiv und die Suchen dauern wesentlich länger als bei den anderen Verfahren. Aus diesem Grund werden De-Novo-Sequenzier-Algorithmen häufig nur komplementär zu den Datenbank-basierten Suchmaschinen genutzt.

Spektren-Datenbank-basierte Suchmaschinen

Eine Spektren-Datenbank enthält Fragmentationenspektren, die bereits Peptidsequenzen erfolgreich zugeordnet wurden. Die Zuordnung erfolgte im Vorfeld mit Hilfe von Protein-Datenbank-basierten Suchmaschinen oder mittels De-Novo-Sequenzierung. Auch hier kann ein Massenfenster definiert werden, um die potenziellen Zuordnungen sinnvoll zu reduzieren. Das in der Datenbank enthaltene Spektrum wird mit dem zu analysierenden Spektrum verglichen und der ermittelte Score ist eine Maßzahl für die Ähnlichkeit der beiden Spektren. Diese Art der Identifizierung gehört zu den schnellsten und sensitivsten Verfahren. Der große Nachteil dieser Methode ist, dass nur Peptide identifiziert werden können, die zuvor massenspektrometrisch analysiert und erfolgreich identifiziert wurden.

1.6 Massenspektrometrische Quantifizierung von Proteinen und Peptiden

Das Proteom ist im Gegensatz zum statischen Genom ein dynamisches System, welches sich in seiner qualitativen und quantitativen Zusammensetzung aufgrund unterschiedlicher Reize verändern kann. Es stellte sich auch zunehmend heraus, dass die qualitative Zusammensetzung der Proteome aus verschiedenem Gewebe beim Menschen sehr ähnlich sind und die großen Unterschiede in den unterschiedlichen Abundanzen der exprimierten Proteine liegen. Dies sind die Hauptgründe zusätzlich zur qualitativen auch die quantitative Zusammensetzung eines Proteoms zu analysieren. In der Anfangszeit der Proteomik waren beinahe nur qualitative Aussagen möglich.

Die ersten Ansätze zur Formulierung von quantitativen Aussagen waren gelbasiert. Hierzu wurde jeweils ein 2D-Gel für eine Probe angefertigt und die Größe der Protein-Spots wurden miteinander verglichen.⁸⁹ Seitdem haben sich diverse Methoden etabliert, die aussagekräftiger und zuverlässiger sind. Die Abbildung 1.14 auf der nächsten Seite gibt eine Übersicht über die gängigsten Methoden und verdeutlicht zudem, an welcher Stelle des Arbeitsablaufes ein zusätzlicher Schritt inkorporiert werden muss.⁹⁰ Einige der Methoden wie QconCAT^{91,92} und die Nutzung von AQUA-Peptiden⁹³ erlauben die absolute Quantifizierung von Proteinen. In diesen Ansätzen

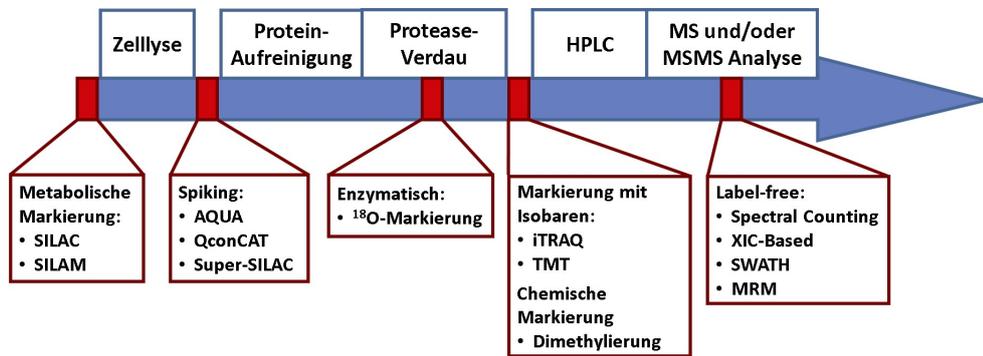


Abbildung 1.14: Übersicht über die am häufigsten genutzten Quantifizierungsverfahren in der Proteomforschung

wird mit einem internem Standard gearbeitet, d.h. ein Modelprotein oder -peptid wird dem Ansatz hinzugegeben (*Spiking*). Die anderen Methoden ermöglichen eine relative Quantifizierung und somit Aussagen über Abundanzunterschiede von Proteinen in mindestens zwei unterschiedlichen Proben. Im Folgenden werden nur die Methoden vorgestellt, die in dieser Arbeit Verwendung gefunden haben.

1.6.1 Quantifizierung mit TMT

Die relative Quantifizierung mit TMTTM (*Tandem-Mass-Tags*, Thermo Fisher Scientific)⁵⁴ gehört zu den Methoden mit der höchsten Sensitivität. Des Weiteren ermöglicht sie die zeitgleiche Aufnahme von bis zu zehn unterschiedlichen Proben. Die Peptide der unterschiedlichen Proben werden hierzu nach erfolgreicher Proteolyse der Proteine chemisch markiert. Die eingesetzten Reagenzien sind isobar und weisen drei strukturelle Funktionseinheiten auf: NHS-Ester, Balance- und Reporter-Gruppe (siehe Abbildung 1.15). Der NHS-Ester ist die reaktive Gruppe und überträgt den Acylrest auf die Aminogruppen der Peptide (siehe Abschnitt 1.3.2 auf Seite 12). Die Balance- und die Reportergruppe der einzelnen Reagenzien enthalten in der Summe die gleiche Anzahl an Isotopen, die bei den unterschiedlichen Reagenzien im unterschiedlichen Ausmaß zwischen diesen beiden Gruppen verteilt sind. TMT-markierte Peptide aus unterschiedlichen Proben sind folglich über MS nicht unterscheidbar. Erst die Fragmentierung der Peptide und die damit einhergehende Abspaltung der unterschiedlich schweren Reportergruppen ermöglicht eine massenspektrometrische Diskriminierung. Die Intensitäten der verschiedenen Reporter-Gruppen werden zueinander ins Verhältnis gesetzt und ermöglichen so die relative Quantifizierung der Proteine.

Im oberen Teil der Abbildung 1.15 ist das TMT-Reagenz mit den unterschiedlichen Gruppen zu sehen. Die Positionen, an denen schwere Isotope

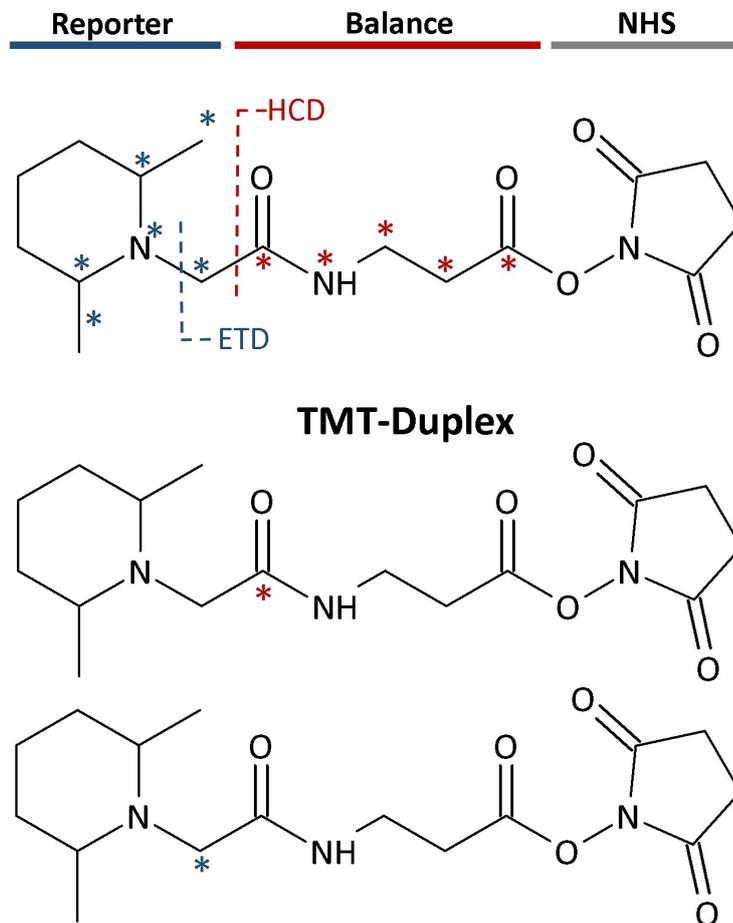


Abbildung 1.15: TMT-Reagenz und sein funktioneller Aufbau

eingefügt werden können, sind markiert. Die einfachste Realisierung dieser Methode ist im unteren Teil der Abbildung gezeigt mit jeweils einem schweren Isotop (TMT-Duplex). Dieser Ansatz erlaubt somit nur die gleichzeitige Messung von zwei Proben.

1.6.2 Derivatisierungsfreie Quantifizierung (Label-free)

Label-free Quantifizierungsverfahren erfordern keinen zusätzlichen Derivatisierungsschritt im Arbeitsablauf und lassen sich in zwei Gruppen unterteilen.⁹⁴ Die erste Gruppe leitet quantitative Aussagen anhand der gemessenen Intensitäten der Vorläuferionen ab und die zweite Gruppe anhand der Anzahl gemessener MS/MS-Spektren. Es gibt Methoden, die Mischformen darstellen oder zusätzlich die Intensitäten der Fragmentionen mit einbeziehen. In dieser Arbeit wurde lediglich eine Methode verwendet die zur zweiten

Gruppe gehört und als NSAF (*Normalized Spectral Abundance Factor*)⁹⁵ bezeichnet wird. Zur Berechnung des NSAF-Wertes eines Proteins k müssen die Anzahl der gemessenen Spektren S , die dem Protein zugewiesen werden konnten, die zugehörige Proteinlänge L und die Anzahl aller identifizierter Proteine N bekannt sein. Der Wert errechnet sich dann folgendermaßen:

$$NSAF_k = \left(\frac{S}{L}\right)_k * \frac{1}{\sum_{i=1}^N \left(\frac{S}{L}\right)_i} \quad (1.4)$$

Je größer der NSAF-Wert eines Proteins desto höher ist seine Abundanz in dem Gemisch. In vergleichenden Untersuchungen wurde gezeigt, dass der NSAF-Wert zu den zuverlässigsten Algorithmen gehört.⁹⁶

Materialien und Methoden

2.1 Materialien

2.1.1 Modellproteine und -proteome

In der vorliegenden Arbeit steht die Entwicklung einer Methode im Vordergrund. Aus diesem Grund wurden die biologischen Proben nach Möglichkeit käuflich erworben. Das Anlegen einer Zellkultur war lediglich für einen abschließenden Versuch notwendig, um beurteilen zu können, ob die entwickelte Methode sich in die gängigen Arbeitsabläufe integrieren lässt. In der Proteomik stellen lebende Zellen in der Regel den Ausgangspunkt dar.

Bis zum Abschnitt 4.3 auf Seite 81 kamen folgende Substanzen zum Einsatz:

- Substanz P, erworben als Acetat-Salz bei Sigma-Aldrich Chemie GmbH in München, UniProt-Nr.: P20366.
- Alkohol-Dehydrogenase (ADH), *Saccharomyces cerevisiae*, Sigma-Aldrich Chemie GmbH in München, UniProt-Nr.: P00330
- Apo-Serotransferrin, *Homo sapiens*, deglykolisiert, Sigma-Aldrich Chemie GmbH in München, UniProt-Nr.: P02787

Das verwendete *Escherichia coli*-Lysat (E.coli) aus Abschnitt 4.3 auf Seite 81 wurde von Bio-rad (Hercules, Kalifornien, Produktnummer 1632110) erworben.

Das humane Proteom aus Abschnitt 4.5 auf Seite 118 wurde aus HEK-239T-Zellen extrahiert. HEK-Zellen (*Human Embryonic Kidney*) sind menschliche embryonale Nierenzellen, die als ein Transformationsprodukt mit DNA-Teilen des menschlichen Adenovirus-5 geschaffen wurden. Die verwendete Variante dieser Zellen ist mit einem „T“ gekennzeichnet, da sie ein zusätzliches Protein exprimieren.

Die Zellen wurden in dem HyClone™ Dulbecco's *Modified Eagles Medium* (DMEM) *High Glucose* (GE Healthcare Life Sciences) unter Zusatz von Glutamin, Penicillin/Streptomycin und fetalem Kälberserum auf Sarstedt™ Standard-Platten kultiviert. Die Wachstumskontrolle erfolgte mit Hilfe eines Mikroskops.

Das Protein *Arachidonat-5-Lipoxygenase* (5-LO) (siehe Abschnitt 4.5.3 auf Seite 126) wurde vom Arbeitskreis Steinhilber (Goethe-Universität-Frankfurt) zur Verfügung gestellt. Das Protein wurde rekombinant mithilfe von *E. coli* Zellen hergestellt.

2.1.2 Reagenzien und Verbrauchsmaterialien

Die in dieser Arbeit verwendeten chemischen Reagenzien, Proteasen, MALDI-Matrizes und Lösungsmitteln sind unter Angabe des Herstellers in Anhang A tabellarisch aufgeführt. Des Weiteren können aus dem Anhang A die verwendeten Verbrauchsmaterialien entnommen werden.

2.1.3 Software

Eine vollständige Liste der verwendeten Programme ist im Anhang B zu finden. Auf die meisten Programme wird in den jeweiligen Abschnitten ausführlich eingegangen, so dass in diesem Abschnitt der Fokus auf die Programmiersprache *Python* gelegt wird.

Python ist eine interpretierte höhere Programmiersprache, die von Guido van Rossum entwickelt und zum ersten Mal 1991 veröffentlicht wurde. Python zeichnet sich durch eine hohe Lesbarkeit und durch eine Syntax aus, die eine geringere Anzahl Zeilencodes benötigt im Vergleich zu anderen Programmiersprachen wie C++ und Java. Diese Eigenschaften und die hohe Anzahl an Modulen machten Python in den letzten Jahrzehnten zu einer der am häufigsten benutzten Programmiersprache im Bereich der Naturwissenschaften.

Ein Modul ist eine Datei, in der Definitionen (z.B. von Funktionen und Klassen) hinterlegt werden können. Die Definitionen eines Moduls können innerhalb eines Python-Skriptes importiert und verwendet werden, so dass es nicht notwendig ist, innerhalb eines jeden neuen Skriptes häufig verwendete Definitionen immer wieder neu aufzustellen. In den letzten Jahren erlangte Python auch im Feld der Datenanalyse (*Data Sciences*) zunehmende Bedeutung und verdrängt immer mehr die in diesem Feld etablierte Programmiersprache R.

Für die Auswertung von Massenspektren ist es zum einen erforderlich eine Signalprozessierung durchzuführen, die in eine Peakliste resultiert. Nach erfolgter Datenbanksuche, erhält man speziell bei *Bottom-Up*-Ansätzen Peptidlisten, die eine sehr hohe Anzahl an Zeilen aufweisen und eine manuelle

Interpretation somit nahezu unmöglich ist. Python bietet für diese verschiedenen Aufgabenstellungen Module an, die in den nächsten Abschnitten je nach Fragestellung verwendet wurden und so eine Auswertung der großen Datenmengen ermöglichten. Zusätzlich existieren Module zur Erstellung von Grafiken zur Visualisierung von Resultaten. Die in dieser Arbeit verwendeten Module (*Packages*) sind in Tabelle 2.1 aufgelistet.

Tabelle 2.1: Python Module

| Modul | Version | Funktion |
|------------|---------|--|
| biopython | 1.68 | Berechnung von Gravy-Score und isoelektrischem Punkt von Aminosäuresequenzen |
| lxml | 3.7.2 | Parsen von XML-Dateien |
| matplotlib | 2.0.0 | Tool zum Visualisieren von Ergebnissen, z. B. mit Hilfe von Boxplots und Säulendiagrammen |
| mspartner | 2.4 | Entwickelt von Matrix Science zum Parsen von Mascot Ergebnis Dateien (dat-files) |
| numpy | 1.11.3 | Stellt schnelle vorkompilierte Funktionen für mathematische und numerische Routinen bereit |
| pandas | 0.19.2 | Stellt Datenstrukturen zur Verfügung für Tabellenkalkulationen |
| seaborn | 0.7.1 | Basiert auf matplotlib und enthält vorkonfigurierte Grafiken |

2.1.4 Verwendete Geräte

Voyager-DE™ STR Workstation (ABSciex)

Der Aufbau dieses Gerätes ist in Abbildung 2.1 auf der nächsten Seite zu sehen. Es handelt sich hierbei um einen Flugzeitanalysator (siehe Abschnitt 1.4.3 auf Seite 18), welcher gekoppelt ist an eine MALDI-Ionenquelle. Aus dem Aufbau des Gerätes lässt sich entnehmen, dass dieses Gerät eine *Delayed Extraction* (DE) ermöglicht. Durch den zusätzlichen Einsatz des Reflektrons erlangt diese Apparatur ein Auflösungsvermögen von bis zu 20'000 FWHM. Bei dem eingesetzten Laser handelt es sich um einen 20-Hz-Stickstofflaser, der 3ns-Lichtimpulse mit einer Wellenlänge von 337 nm erzeugt. Im Linearmodus betrieben, weist das Gerät ein sehr geringes Auflösungsvermögen auf (ungefähr 2000 FWHM), erreicht jedoch einen Massenbereich von bis zu 300'000 Da. Im Linearmodus werden die einzelnen Isotope der Ionen nicht aufgelöst, so dass die durchschnittliche Masse für die Auswertung herangezogen werden muss. Dieses Gerät wurde in dieser Arbeit für Proben

verwendet, die eine monoisotopische Masse größer als 4000 Da (siehe Abschnitt 4.5.3 auf Seite 126) haben.

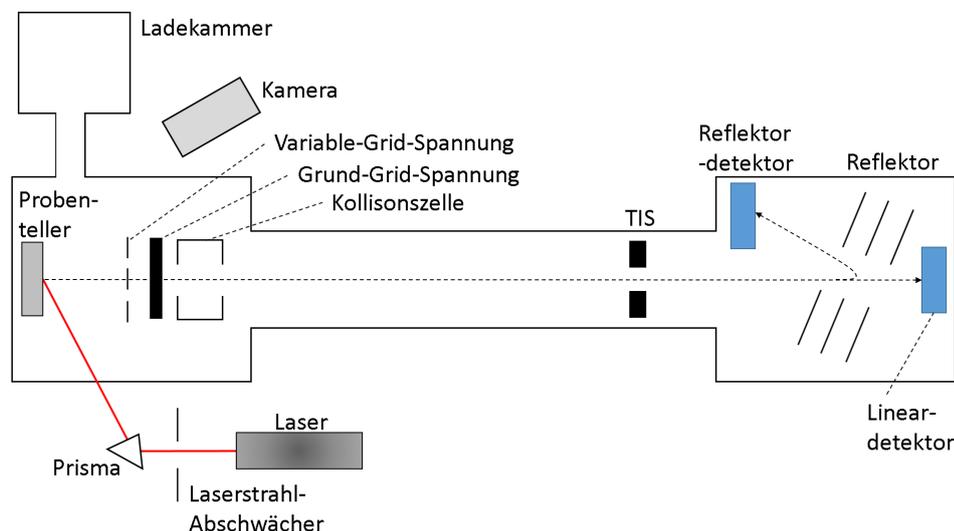


Abbildung 2.1: Voyager-DE™ STR Workstation (ABSciex)

4800 MALDI-TOF/TOF™ Analyser (ABSciex)

Dieses Gerät ist eines der Nachfolge-Modelle des Voyager STR und zeigt einen recht ähnlichen Aufbau wie aus der Abbildung 2.2 auf der nächsten Seite entnommen werden kann. Die Abbildung stammt von der Firma ABSciex und wurde für diese Arbeit modifiziert. Auch hierbei handelt es sich um einen Flugzeitanalysator, welcher mit einer MALDI-Quelle verbunden ist.

Der wesentliche Unterschied zum vorherigen Gerät ist das Vorhandensein einer zweiten Ionenquelle. Diese dient der Aufnahme von Fragmentionenspektren und wird folglich auch nur zu deren Aufnahme aktiviert. Die Fragmentierung beruht auf der mit dem MALDI-Prozess verbundenen Anregung der Ionen, die bereits in der Quelle oder aber auch in der feldfreien Driftstrecke zum metastabilem Zerfall führen können. Da diese Fragmentierung nach der Ionenquelle erfolgt, wird sie als *Post-Source Decay* (PSD) bezeichnet. Die kinetische Energie der Fragmente (PSD-Ionen) unterscheidet sich deutlich von denen der Vorläuferionen, jedoch ist die Geschwindigkeitsverteilung klein, sodass beide Ionenspezies in einem linearen TOF nahezu zeitgleich zum Detektor gelangen und nicht voneinander unterschieden werden können. Die erzeugten Ionen werden zunächst mit einer vergleichsweise geringen Spannung zur zweiten Ionenquelle beschleunigt. Die gerin-

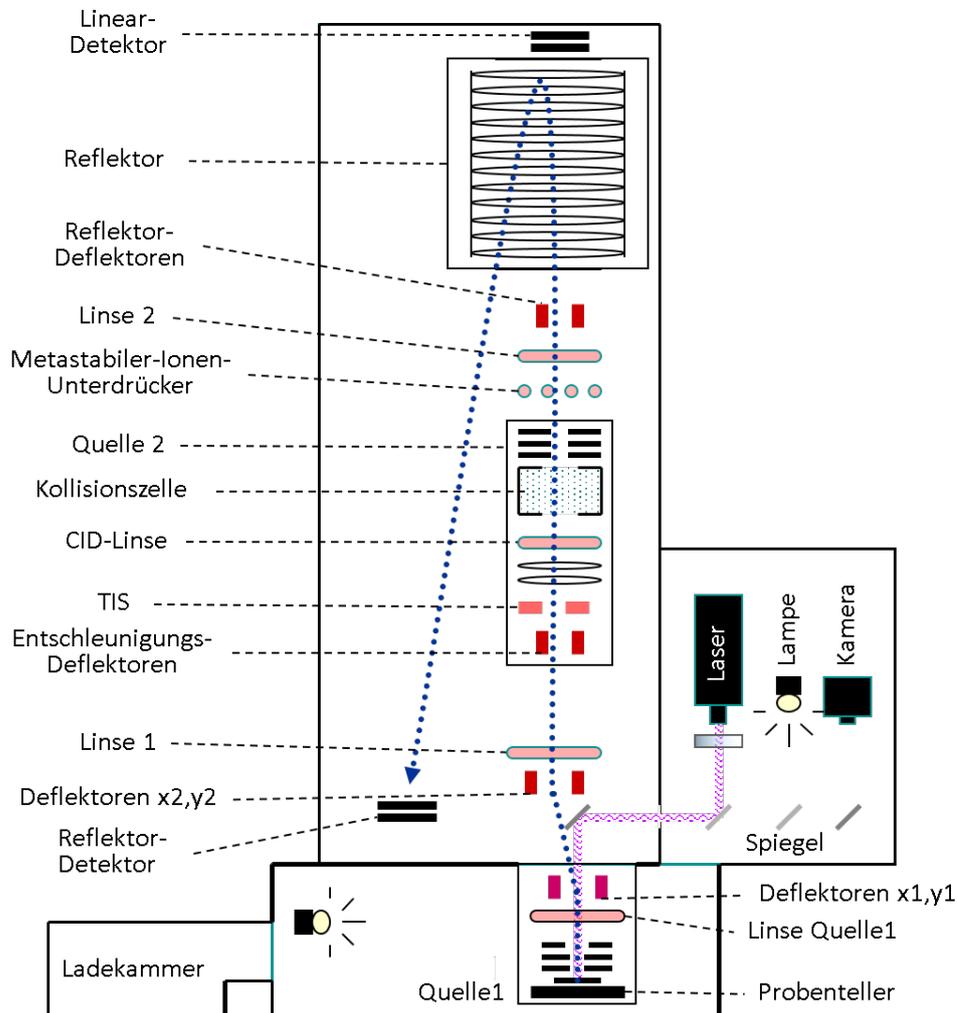


Abbildung 2.2: 4800 MALDI-TOF/TOF™ Analyser (ABSciex)

ge Geschwindigkeit der Ionen vor Eintritt in die zweite Ionenquelle vereinfacht zum einen die Selektion der Vorläuferionen über den *Timed-Ion-Selector* (TIS) innerhalb eines vordefinierten Massenbereiches aus dem Ionenstrom. Des Weiteren ergibt sich somit ein größeres Zeitfenster für den metastabilen Zerfall der Ionen. Die Ionen werden zunächst auf 1-2 keV abgebremst. Die Fragmentierung kann durch eine Kollisionszelle verstärkt und je nach Konzentration des Stoßgases reguliert werden. Im Anschluss zur Stoßkammer werden die Ionen auf 20 keV beschleunigt. Dieser Vorgang bewirkt eine Verringerung der Verteilung der kinetischen Energie, die vom Reflektor toleriert werden kann. Durch diese Anordnung aus Stoßkammer und Beschleunigungslinsen lassen sich PSD-Ionen hoher Auflösung aufnehmen.

Ein weiteres Unterscheidungsmerkmal zum Voyager-DE™ STR ist der Eintrittswinkel des Laserstrahles auf den Probenhalter. Durch Reflektion verschiedener Spiegel tritt der Laserstrahl orthogonal auf dem Probenhalter auf, so dass sich die Desorptionswolke (*MALDI-plume*) in Richtung der Beschleunigungsstrecke ausbildet. Durch diese Anordnung werden höhere Sensitivitäten erreicht.

Dieses Gerät wurde ausschließlich im Reflektormodus genutzt für Messungen innerhalb eines Massenbereiches von 800-4000 m/z. Aufgrund der hohen Messgeschwindigkeiten wurde dieses Gerät bevorzugt für Quantifizierungen über einen relativen Standard genutzt. Quantifizierungen erwiesen sich als präziser bei hohen Scanraten. Das Gerät verfügt weiterhin über einen 200-Hz-Nd:YAG-Feststofflaser mit einer emittierten Wellenlänge von 355 nm.

MALDI LTQ-Orbitrap™ XL (Thermo Fisher Scientific)

Die MALDI LTQ-Orbitrap™ XL ist ein Hybridinstrument aufgebaut aus einer linearen Ionenfalle und einem Orbitrap-Massenanalysator (siehe Abbildung 2.3, Quelle: Thermo Fisher Scientific). Die Ionenfalle kam in dieser Arbeit nicht zum Einsatz und fand deshalb auch keine Erwähnung im Abschnitt 1.4.3 auf Seite 16.

Die Ionen, die in der MALDI-Quelle erzeugt werden, gelangen über fokussierende Linsen und Oktopole in die lineare Ionenfalle. Die Ionen könnten bereits hier je nach Gerätekonfiguration mit geringer Auflösung analysiert werden. Ist die Ionenfalle inaktiviert, passieren die Ionen diese und gelangen über einen weiteren Oktopol in die sogenannte *C-trap*. Diese speichert die Ionen, um sie anschließend fokussiert und in diskreten Paketen in die Orbitrap zu injizieren. Dieses Gerät bietet mehrere Möglichkeiten der Frag-

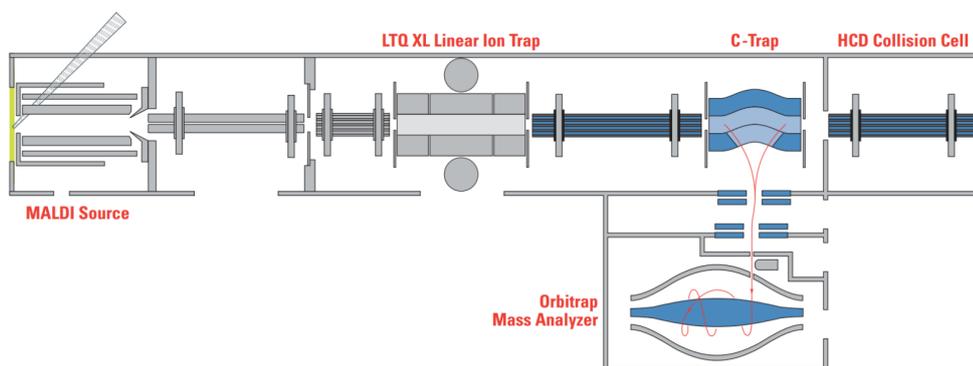


Abbildung 2.3: MALDI LTQ-Orbitrap™ XL (Thermo Fisher Scientific)

mentierung an, die jedoch in dieser Arbeit nicht angewendet wurden und somit an dieser Stelle nicht weiter darauf eingegangen wird. Die MALDI-LTQ Orbitrap XL erreicht ein Auflösungsvermögen von bis zu 100'000 FWHM bei $m/z = 400$. Bei dem eingesetzten Laser handelt es sich um einen 20-Hz-Stickstofflaser, der Strahlen mit einer Wellenlänge von 337 nm erzeugt.

Innerhalb dieser Arbeit wurde dieses Gerät ausschließlich zur Aufnahme von MS-Spektren in einem Massenbereich von 800-4000 m/z verwendet. Die geringen Massenabweichungen (i.d. R. <5 ppm) vereinfachten die anschließende Auswertung.

Q Exactive Orbitrap™ XL (Thermo Fisher Scientific)

Proben höherer Komplexität, wie aus einem Proteom-Verdau erzeugte Peptidmischungen, wurden mit diesem Gerät vermessen. Es handelt sich hierbei erneut um ein Gerät von Thermo Fisher Scientific, welches jedoch eine ESI-Quelle aufweist und so direkt mit einer HPLC-Anlage gekoppelt werden kann (siehe Abbildung 2.4 auf der nächsten Seite, Quelle: Thermo Fisher Scientific). Die Ionen müssen in dieser Geräteanordnung zunächst ein Quadrupol passieren, bevor sie in die *C-trap* gelangen. Von der *C-trap* ausgehend gelangen die Ionen entweder in die Kollisionszelle oder in den Orbitrap-Detektor.

Die *C-trap* eröffnet in dieser vorliegenden Anordnung zusätzlich die Möglichkeit spezieller Analyseverfahren wie dem sogenannten *Multiplexing*. Dabei können Ionen aus bis zu zehn verschiedenen engen Massenfenstern nacheinander mithilfe des Quadrupolfilters selektiert, in der *C-trap* gesammelt und zeitgleich in der Orbitrap analysiert werden. In ähnlicher Weise können auch Fragmentspektren von bis zu zehn verschiedenen Vorläuferionen generiert werden. Die Vorläuferionen werden nacheinander im Quadrupol isoliert und in der *C-trap* gesammelt bevor diese die Kollisionszelle geleitet und dort fragmentiert werden. Die Fragmentierung in der Kollisionszelle erfolgt über HCD (*Higher-Energy-Collisional-Dissociation*), welches zu den kollisionsinduzierten Dissoziationen (CID, *collision-induced dissociation*) gehört. Hierbei werden die Ionen nach Beschleunigung in einem elektrischen Feld durch Kollision mit neutralen Gasmolekülen wie Helium oder Stickstoff fragmentiert. HCD findet bislang nur in Orbitrap-Geräten Anwendung. HCD erfolgt mit höheren Aktivierungsenergien bei kürzeren Aktivierungszeiten im Vergleich zur CID in der Ionenfalle.

Durch den Austausch der Ionenfalle mit dem Quadrupol und der damit einhergehenden schnelleren Auswahl von Vorläuferionen werden Messgeschwindigkeiten erreicht, die bis zu drei mal höher sind. Die Vermessung komplexer Proben liefert somit in der gleichen Messzeit mehr Fragmentionenspektren und somit eine höhere Anzahl an Identifikationen. Diese Ge-

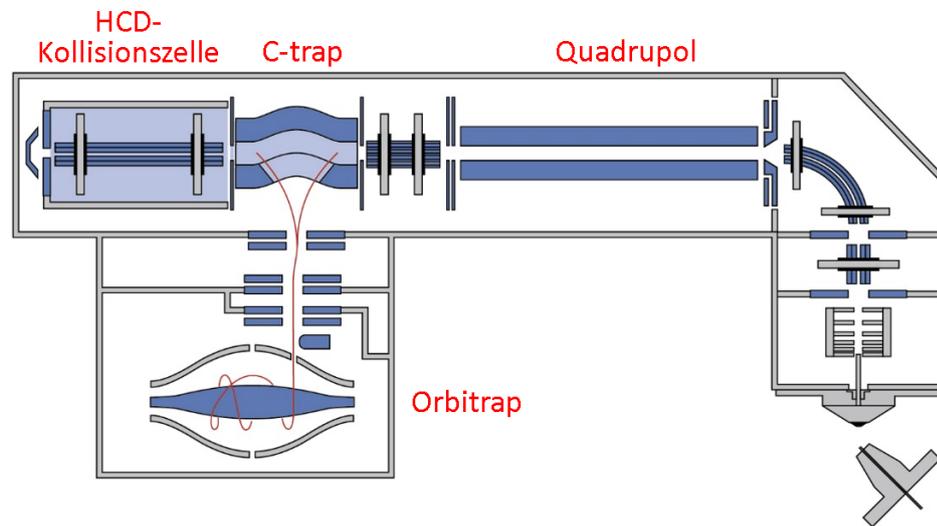


Abbildung 2.4: Q Exactive Orbitrap™ XL (Thermo Fisher Scientific)

räteeigenschaften sind die Hauptursachen dafür, dass die Q Exactive zu einem der meist benutzten Geräte für proteomische Fragestellungen gehört.⁹⁷

2.2 Methoden

2.2.1 Zellyse

Die Lyse von den Humanzellen erfolgte chemisch. Die Zellsuspension wurde zunächst zentrifugiert (4000g, 4°C, 15 min) und der Überstand dekantiert. Auf 1 g feuchtem Zellpellet wurden 1,5 mL vom Lysepuffer hinzugegeben. Bei dem eingesetzten Lysepuffer handelte es sich um den Bugbuster Mastermix (Merck KGaA, Darmstadt), dessen Zusammensetzung im Abschnitt 4.5.1 auf Seite 118 diskutiert wird. Nach 15 minütiger Inkubation wurde das Lysat erneut zentrifugiert (14'000 g, 4°C, 15 min) und im Überstand befanden sich die gelösten Proteine, die sofort weiter prozessiert oder bei -20 °C gelagert wurden.

2.2.2 Acetonfällung

Proteine, die über chemische Lyse aus Zellen extrahiert worden sind, mussten für die weitere Prozessierung zunächst aufgereinigt werden. Als Aufreinigungsmethode wurde hier die Acetonfällung gewählt. Hierzu wird ungefähr das vierfache Volumen einer bei -20°C vorgekühlten reinen Acetonlösung zum Zellysate hinzugegeben. Die Lösung wird dann anschließend für mindestens zwei Stunden bei -20°C gelagert. Nach Zentrifugation (14'000g,

4°C, 15 min) wird der Überstand verworfen. Das Proteinpellet wird durch Zugabe eines geeigneten Puffer resuspendiert. Die im Puffer unlöslichen Bestandteilen werden durch Zentrifugation abgetrennt.

2.2.3 Binden von Proteinen an magnetische Beads

Der ArgC-ähnliche Verdau (siehe Fragestellung und Ziele) beruht auf der schrittweisen Modifizierung der funktionellen Gruppen der Proteine. Im ersten Schritt werden die Cysteine und im zweiten Schritt die Lysine chemisch modifiziert. Zwischen diesen Schritten ist es notwendig, die Proteine von den Reagenzien aufzureinigen, um eine vollständige Modifikation der funktionelle Gruppen zu erhalten und den nachfolgenden Verdau durch eine eventuell eintretende Inaktivierung des Trypsins nicht einzuschränken oder gar zu verhindern.

Als Aufreinigungsmethode wurde eine im Arbeitskreis von Moritz Neupärtl entwickelte Methode verwendet, die die Proteine kovalent an magnetische Beads bindet. Die Beadoberfläche weist NHS-Ester auf, die mit den Aminen der Proteine (siehe Abschnitt 1.3.2 auf Seite 12) reagieren und sie somit kovalent binden. Mit Hilfe eines Magneten können die Beads mit den gebundenen Proteinen im Gefäß zurückgehalten werden, während die überstehende Lösung mit Hilfe einer Pipette entnommen werden kann. So war es möglich, mit wenig Aufwand an immobilisierten Proteinen konsekutive Reaktionsschritte durchzuführen.

Im Laufe der Arbeit wurden die NHS-aktivierten Beads eines anderen Herstellers genutzt, da diese eine höhere Bindungskapazität aufwiesen. Die beiden Bindungsprotokolle sind im folgenden für die jeweiligen Beads aufgeführt.

NHS Mag Sepharose Beads (GE Healthcare GmbH, Solingen)

Diese Beads wurden bis zum Abschnitt 4.3 verwendet. Die Bindung der Beads erfolgte in starker Anlehnung an das Herstellerprotokoll und ist bezogen auf 100 µg Proteinmenge:

1. Resuspendieren der Stocklösung und Entnahme von 40 µL.
2. Entnahme des Überstandes und waschen der Beads mit 100 µL kalter HCl-Lösung (1 mM).
3. Hinzugabe der Proteine, welche in 100 µL TEAB (100 mM) gelöst vorliegen und Inkubation für mindestens 45 Minuten bei Raumtemperatur.
4. Entfernen des Überstandes und Deaktivierung der Beadoberfläche durch Hinzugabe von Ambic-Puffer (160 mM).

BcMag™ NHS-Beads (1µm, Biclone Inc, San Diego, USA)

Im Laufe der Arbeit stellte sich heraus, dass diese Beads eine höhere Bindungskapazität aufweisen. Aus diesem Grund wurden sie für die Arbeiten ab Abschnitt 4.5 eingesetzt. Die Beads wurden als Feststoff geliefert und das Protokoll für 100 µg Proteinmenge sah folgendermaßen aus:

1. Abwaage von 10 mg Bead-Pulver.
2. Hinzugabe und Entnahme von 100 µL kalter HCl-Lösung (1 mM) zur Aktivierung der Beadoberfläche.
3. Hinzugabe der Proteine, welche in 100 µL TEAB (100mM) gelöst vorliegen, und Inkubation für mindestens 45 Minuten bei Raumtemperatur.
4. Hinzugabe von 100 µL ACN und weitere Inkubation für mindestens 45 Minuten bei Raumtemperatur.
5. Entfernen des Überstandes und Deaktivierung der Beadoberfläche durch Hinzugabe von Ambic-Puffer (160 mM).

Beads mit immobilisiertem Protein sollten nicht trocken gelagert und zeitnah weiter prozessiert werden.

2.2.4 Proteinquantifizierung

Zur Wahl der Reagenzmenge und des Proteasezusatzes ist die Ermittlung der vorliegenden Proteinmenge nach Zellaufschluss bzw. nach Bindung an die Beads erforderlich. Die quantitative Bestimmung der Proteine erfolgte mit Hilfe einer *BCA-Reaktion*. Hierbei reagieren zunächst zugesetzte zweiwertige Kupferionen quantitativ mit Proteinen zu einwertigen Kupferionen. Diese bilden mit Bicinchoninsäure (BCA) einen violetten Komplex (siehe Abbildung 2.5 auf der nächsten Seite), der Licht mit einer Wellenlänge von 562 nm absorbiert und somit photometrisch quantifiziert werden kann.⁹⁸

Zur Quantifizierung wurde das gebrauchsfertige „Pierce™ BCA Protein Assay Kit“ (Thermo Fisher Scientific) genutzt. Die Kalibriergerade wurde mit dem mitgelieferten BSA-Standard erstellt. Das weitere Vorgehen erfolgte strikt nach Herstelleranweisung. Hierzu wurden 200 µL der frisch hergestellten Reagenzlösung in Mikrotiterplatten mit 96 Wells und flachem Boden (Greiner) vorgelegt. Anschließend wurden 10 µL Probelösung hinzugegeben und die Lösung wurde 30 Minuten bei 37°C im Schüttler inkubiert. Die photometrische Messung erfolgte zeitnah mit Hilfe eines „Tecan Readers“. Die magnetischen Beads wirken mit ihrem Eisenkern bereits ohne die Anwesenheit von Proteinen reduzierend auf das Kupfer. Bei der Quantifizierung von Bead-gebundenem Protein ist somit die Messung eines zusätzlichen Blindwertes notwendig, der nur die Beads ohne zugesetztes Protein enthält. Des Weiteren muss bei der Messung von immobilisierten

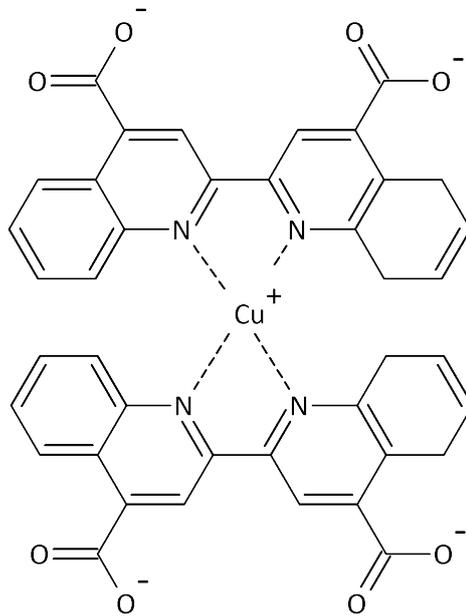


Abbildung 2.5: Strukturformel des violetten Cu^+ -Bicinchoninsäure-Komplexes

Proteinen berücksichtigt werden, dass die Beads aufgrund ihrer turbidometrischen Eigenschaften die photometrische Messung stören. Die Inkubation der Beads mit der BCA-Lösung erfolgte somit zunächst in einem Mikroreaktionsgefäß und die überstehende Bead-freie Lösung wurde anschließend in die Mikrotiterplatten überführt.

2.2.5 Chemische Modifizierung auf Proteinebene

Die verschiedenen Möglichkeiten, funktionelle Gruppen eines Proteins zu modifizieren wurden bereits im Abschnitt 1.3 gezeigt. Wie schon mehrfach erwähnt, ist es für den ArgC-ähnlichen Verdau sowohl notwendig die Thiol- als auch die Amin-Gruppen zu modifizieren. Die Modifikation der Amingruppen erfolgte zunächst mit publizierten Protokollen. Diese erwiesen sich, wie später im Detail aufgeführt, als unzureichend, so dass die Reaktionsbedingungen optimiert wurden. Beide Protokolle für das jeweilige Reagenz sind im folgenden aufgeführt. Die hier beschriebenen Protokolle beziehen sich erneut auf 100 μg Protein.

Modifizierung der Amin-Gruppen vor Optimierung der Reaktionsbedingungen

- **Propionylierung mit Propionsäureanhydrid (PA)⁵²**

1. Löse die Proteinprobe in destilliertem Wasser zu einer Konzentration von $1\mu\text{g}/\mu\text{L}$ auf.
2. Entnehme einen Aliquot von $5\mu\text{L}$ und gebe $5\mu\text{L}$ 100 mM -Ambic-Puffer hinzu.
3. Füge $1\mu\text{L}$ konzentrierte Ammoniumhydroxid-Lsg. der Probe hinzu.
4. Stelle das Propionylierungsreagenz her, in dem $75\mu\text{L}$ Propionsäureanhydrid mit $25\mu\text{L}$ Methanol gemischt werden.
5. Gebe zügig $20\mu\text{L}$ des Propionylierungsreagenz der Probe hinzu und kontrolliere den pH-Wert mit Indikatorpapier. Sollte der pH-Wert kleiner als 8 sein, füge tropfenweise konzentrierte Ammoniumhydroxid-Lösung hinzu bis wieder ein pH-Wert von ca. 8,0 erreicht ist.
6. Inkubiere für 20 min bei Raumtemperatur.
7. Dampfe die Probe bis zur Trockenheit in der Vakuumzentrifuge ein.
8. Nehme die Probe in $5\mu\text{L}$ 100 mM -Ambic-Puffer auf.
9. Wiederhole die Schritte 3-7.
10. Verdünne die Probe mit $50\mu\text{L}$ des 100-mM -Ambic-Puffers.

- **Acetylierung mit Sulfo-NHS-Acetat (NHS)**

1. Löse deine Peptid- oder Proteinprobe in 100 mM TEAB-Puffer (pH=8,5).
2. Löse das Sulfo-NHS-Acetat in ultrareinem Wasser zu einer Konzentration von 10 mM ($2,6\text{ g/ml}$). Gebe einen 25-fachen molaren Überschuss des Acetylierungsreagenzes zu der Probe hinzu. Sollte die Anzahl der Aminogruppen nicht bekannt sein, gebe die selbe Massenanteile an Acetylierungsreagenz hinzu.
3. Inkubiere für eine Stunde bei Raumtemperatur.
4. Stoppe die Reaktion durch Zugabe eines aminhaltigen Puffers.

- **Carbomethoxylierung mit Diethylpyrocarbonat (DEPC)⁹⁹**

1. Löse die Probe in 50 mM Phosphat-Puffer (pH=6,0).
2. Gebe einen 10-fachen molaren Überschuss an Diethyl-Pyrocarbonat hinzu.

3. Inkubiere für 20 min bei 35°C.

• **Reduktive Dimethylierung (DM)**¹⁰⁰

1. Löse die Probe (25 µg Protein oder Peptid) in 100 µL 100 mM-TEAB-Puffer (pH =8,5).
2. Füge 4 µL 4 % Formaldehyd-Lösung hinzu.
3. Füge 4 µL 0,6 M % NaBH₃CN-Lösung hinzu (gelöst in ACN).
4. Inkubiere für eine Stunde bei Raumtemperatur.
5. Stoppe die Reaktion durch Zugabe eines aminhaltigen Puffers.
6. Säure die Probe durch die Zugabe von Ameisensäure an.

Modifizierung der Amin-Gruppen nach Optimierung der Reaktionsbedingungen

Die Modifizierungen mit optimierten Reaktionsbedingungen erfolgten an Proteinen, die zuvor an magnetische Beads gebunden wurden (siehe Abschnitt 2.2.3). Die folgenden Protokolle beziehen sich auf einer Ausgangsmenge von 100 µg Protein. Die Reaktionen wurden in jedem Fall durch die Entnahme des Überstandes und der Hinzugabe eines aminhaltigen Puffers gestoppt.

• **Propionylierung mit Propionsäureanhydrid (PA)**

1. Vermische 480 µL Methanol mit 120 µL TEAB-Puffer (1,0 M; pH=8,5) und gebe die resultierende Lösung zu den immobilisierten Proteinen.
2. Kühle die Suspension auf 4°C herunter und gebe nach Erreichen der Temperatur 6 µL Propionsäureanhydrid hinzu.
3. Inkubiere die Lösung bei 4°C für 2 Stunden.

• **Acetylierung mit Sulfo-NHS-Acetat (NHS)**

1. Löse Sulfo-NHS-Acetat in einer Mischung aus 50% DMSO und 50% TEAB-Puffer (200 mM; pH=8,5), so dass eine Konzentration von 2 mg/mL resultiert.
2. Gebe 300 µL des Acetylierungsreagenzes zu den Beads und inkubiere bei Raumtemperatur für eine Stunde.

- **Carbethoxylierung mit Diethylpyrocarbonat (DEPC)**

1. Stelle eine 10%ige DEPC-Lösung (Lösungsmittel: ACN) her.
2. Resuspendiere die Beads in 480 μL TEAB-Puffer (1 M; pH=8,5) und kühle den Ansatz auf 4°C ab.
3. Geben 120 μL des Carbethoxylierungsreagenzes zu der Beadsuspension hinzu.
4. Inkubiere für eine Stunde bei 4°C.

- **Reduktive Dimethylierung (DM)**

1. Resuspendiere die Beads in 200 μL TEAB-Puffer (100 mM; pH=8,5).
2. Füge 200 μL 2-Picolin-Boran-Lösung hinzu (28 mg/mL, Lösungsmittel: ACN).
3. Füge 200 μL Formaldehyd-Lösung (36%) hinzu und inkubiere für eine Stunde bei Raumtemperatur.

2.2.6 Verdauprotokolle

Der Verdau der Proteine erfolgte entweder auf den Beads oder in Lösung unter den gleichen Reaktionsbedingungen. In den meisten Fällen wurde ein Enzym-Substrat-Verhältnis von 1:50 gewählt. Der Verdau erfolgte unabhängig von der Protease über Nacht (ca. 18 h) bei 37°C. Das Abstoppen der Verdaue erfolgte durch die Zugabe von 1 μL TFA (100%).

Verdau mit Trypsin

Das Trypsin-Lyophilisat wurde in Wasser/1mM HCl zu einer Konzentration von 0,1 $\mu\text{g}/\mu\text{L}$ gelöst. Als Verdau-puffer diente der TEAB-Puffer (50mM; pH=8,5).

Verdau mit ArgC

Die Proteinprobe mit einer Menge von 100 μg wurde in 80 μL Verdau-Puffer (50 mM Tris/HCl; 10 mM CaCl_2 ; pH=7,6) gelöst. 2 μg ArgC-Lyophilisat wurden in 10 μL des Verdau-puffers gelöst und der Proteinprobe hinzugegeben. Durch Zugabe einer Aktivierungslösung (50mM Tris/HCl; 10mM CaCl_2 ; 50mM DTT; pH=7,6) wurde der Verdau gestartet.

Verdau mit LysC

Das Verdauprotokoll mit LysC ist analog zu dem mit Trypsin (siehe oben).

2.2.7 Peptidaufreinigung

Die aus dem proteolytischen Verdau entstandenden Peptide müssen für chromatographische Anwendungen zunächst aufgereinigt werden. Aufgrund der eingesetzten Proteinmenge wurden in der Regel Spin-Säulen mit C18 Material verwendet, die eine Bindungskapazität von 33 µg aufweisen. Die Aufreinigung erfolgte strikt nach Herstellerprotokoll (Pierce® C 18 Spin Columns, Thermo Scientific). Das Eluat der Aufreinigung wurde zur Trockne vakuumzentrifugiert und anschließend in einem geeigneten Puffer wieder in Lösung gebracht.

2.2.8 MALDI-Messungen

Die MALDI-MS-Messungen an der MALDI LTQ-Orbitrap XL (Thermo Scientific) erfolgten im Fourier-Transformations- und Positiv-Modus mit einem Auflösungsvermögen 30'000 in einem Massenbereich von 800 - 4000 m/z. Messungen am MALDI TOF/TOF Analyzer (ABI 4800, Applied Biosystems) erfolgten im selben Massenbereich bei der gleichen Polarität im Reflektor-Modus. Für Identifizierungsversuche an diesem Gerät wurden 1000 und für Quantifizierungen 5000 Spektren aufgenommen und aufsummiert. Bei beiden Geräten wurde α -Cyano-4-hydroxymethylsäure als Matrix verwendet (3 mg/mL in 70% ACN; 0,1% TFA).

Für Messungen am Voyager-DE STR erwies sich eine Mischung der Matrices 5-Dihydroxybenzoesäure und 5-Methoxysalicylsäure als nützlich, welche unter der Abkürzung sDHB bekannt ist.¹⁰¹ Diese Matrix wurde für die Messungen aus Abschnitt 4.5.3 verwendet (10 mg/mL in 30% ACN; 0,1% TFA). Die Messungen erfolgten auch hier im Positiv-Modus, jedoch in einem Massenbereich von 2000 - 10'000 m/z. Die Kalibrierung erfolgte extern mit dem Protein Cytochrom C. Für die Vermessung jeder Probe wurden 1000 Spektren aufsummiert.

2.2.9 LC-ESI-Messungen

Die LC-ESI-Messungen erfolgten an der Q Exactive von Thermo Scientific (siehe Abschnitt 2.1.4 auf Seite 37), die direkt mit der UHPLC *Ultimate 3000* (Thermo Scientific) gekoppelt wurde und als Ionenquelle wurde die „Nanospray Flex Ion-Source“ (Thermo Scientific) verwendet. Die Peptidproben wurden zunächst auf einer C18-Vorsäule (Thermo Scientific) geladen, bevor sie auf die Hauptsäule gelangen. Die Säule wurde mit *Reprosil C18* gepackt, dessen Partikel einen durchschnittlichen Durchmesser von 2,4µm aufweisen (Dr. Maisch HPLC GmbH, Ammerbruch-Entringen). Die Säule hat einen Innendurchmesser von 100 µm und ist 15 cm lang. Der Gradient begann mit 95% mobiler Phase A (4% ACN, 0,1% Ameisensäure) und verlief über 70 Minuten zu einem Anteil von 30% mobiler Phase B (80% ACN, 0,1% Amei-

sensäure). Ein zweiter Gradient schloss sich nach 70 Minuten an, der bis zu einem Anteil von 60% mobiler Phase B innerhalb von 30 Minuten anstieg. Abschließend wurde die Säule mit 99% B für fünf Minuten gespült, bevor sie für den nächsten Lauf mit 1% B requalibriert wurde. Die eingestellte Fließmittelgeschwindigkeit betrug 300 nL/min.

Die zehn intensivsten Signale wurden zur Aufnahme eines MS/MS-Spektrums im Positiv-Modus über HCD-Fragmentierung ausgewählt. Die MS-Aufnahme erfolgte in einem Massenbereich von 300 bis 2000 m/z mit einem Auflösungsvermögen von 70'000. Der *Automatic-Gain-Control* wurde auf $3 \cdot 10^6$ eingestellt mit einer Injektionszeit von 160 ms. Ausschließlich Signale mit einer Ladung größer oder gleich zwei wurden zur Aufnahme von MS/MS-Spektren selektiert. Die Aufnahme der MS/MS-Spektren erfolgte mit einem Auflösungsvermögen von 17'500 und einem Isolationsfenster von 2 m/z. Der *Automatic-Gain-Control* wurde hier auf 10^5 gesetzt mit einer Injektionszeit von 150 ms. Die zur MS/MS-Aufnahme gewählten Signale wurden für 30 s von einer erneuten Auswahl ausgeschlossen. Die MS-Spektren wurden im Profil-Modus von der Xcalibur-Software (Thermo Scientific) aufgenommen.

2.2.10 Mascot-Suchparameter

In den Abschnitten 1.5.3 und 1.5.4 wurde gezeigt, wie sich Proteine oder Peptide anhand verschiedener Suchalgorithmen identifizieren lassen. Die Suchmaschine Mascot (Matrix Science, London, UK) bietet zwei dieser Möglichkeiten an: PMF- und PFF-Suchen. Beide Suchoptionen laufen in Mascot Protein-Datenbank-basiert. Zusätzlich bietet Mascot die Option an, einen Algorithmus namens *Percolator*¹⁰² über die Suchergebnisse laufen zu lassen. Percolator benutzt ein semi-überwachtes maschinelles Lernverfahren zur besseren Diskriminierung von korrekten und inkorrekten Peptid-Spektrum-Zuweisungen (PSM). Das Suchverfahren mit den einzelnen Suchparametern ist im Anhang C angefügt.

Zwingende Voraussetzung zur Durchführung des Percolator ist eine *Target-Decoy-Suche*. Bei einer derartigen Suche finden zwei Läufe statt. Im ersten Lauf erfolgt die Suche gegen die eingepflegte Protein-Datenbank (*Target-Suche*). Im zweiten Lauf werden zunächst die Einträge in der Datenbank modifiziert, indem beispielsweise die Aminosäuresequenzen der Proteine in ihrer Reihenfolge umgekehrt werden. Eine Suche gegen eine solche Datenbank wird als *Decoy-Suche* bezeichnet und die identifizierten Peptide als *Decoys*. Je geringer die Anzahl der *Decoys* ist, desto akkurater arbeitet die Suchmaschine und umso geringer ist die Wahrscheinlichkeit, dass die Suchmaschine falsch positive Ergebnisse liefert. Als Maßzahl dient hierbei die *False Discovery Rate* (FDR), die sich im einfachsten Fall so ausrechnen lässt:

$$\text{FDR} = \frac{\text{Anzahl der Decoys}}{\text{Anzahl der Targets}} \quad (2.1)$$

Kapitel 3

Fragestellung und Ziele

In der Proteomik gehören die in der Einleitung vorgestellten Bottom-Up-Ansätze zu den am meisten genutzten Verfahren zur Identifizierung und Charakterisierung von Proteinen mit Hilfe von Massenspektrometrie. In diesen Verfahren wird in der Regel das proteolytische Enzym Trypsin eingesetzt,³¹ was unter anderem begründet ist durch seine hohe Schnittspezifität und -effizienz. Seine Robustheit gegenüber den Reaktionsbedingungen und die Erzeugung von Peptiden, die im Schnitt ein gutes Fragmentierungsverhalten aufweisen, sind weitere Argumente für seine Beliebtheit. Bei Nennung all der Vorteile müssen auch gewisse Limitierungen des Trypsineinsatzes berücksichtigt werden. Das unregelmäßige Vorkommen von Trypsinschnittstellen entlang von Proteinen kann zu Peptiden führen, die entweder zu groß oder zu klein für den Massenbereich des gewählten Gerätes sind und somit nicht erfasst werden können. Speziell für Fragestellungen die zur Charakterisierung eines Proteines abzielen, ist dies problematisch und kann durch den kombinierten Einsatz von Protease mit verschiedenen Schnittspezifitäten umgangen werden. Alternative Proteasen, die in diesen Ansätzen in Kombination mit Trypsin zum Einsatz kommen, sind ArgC, GluC, Chymotrypsin, LysC, LysN und AspN.¹⁰³⁻¹⁰⁵

In dieser Arbeit lag ein besonderes Augenmerk auf der Endopeptidase ArgC, welche selektiv Proteine C-Terminal nach Argininen schneidet. Das restriktivere Schnittverhalten dieses Enzyms im Vergleich zu Trypsin, welches zusätzlich auch nach Lysinen schneidet, sollte zu Peptiden führen, die im Durchschnitt größer sind und die Anzahl der generierten Peptide sollte niedriger sein. Hinsichtlich größerer Peptide ist die Wahrscheinlichkeit, dass sie nur einem Protein zugewiesen werden können, höher, so dass es zu weniger Überschneidungen gemeinsamer Peptide kommt. Das Vorkommen von Peptiden, die zu mehreren Proteinen zugewiesen werden können, stellt sowohl in Fragen der Identifizierung, als auch der Quantifizierung der Proteine eine große Herausforderung dar. Die geringere Anzahl der generierten Peptide

weist den Vorteil auf, dass die Probe weniger komplex ist und somit die Zeitanforderungen bei der Generierung der MS/MS-Daten reduziert sind. Damit kann der Informationsgehalt erhöht werden. Der Vorteil der Protease ArgC gegenüber Trypsin ist folglich, dass ein höherer Anteil spezifischer Peptide gebildet wird und dass die Probenkomplexität abnimmt, beides sollte theoretisch zu einer höheren Anzahl an Proteinidentifikationen führen.

In der Praxis wird ArgC jedoch vergleichsweise selten genutzt. Einer der Gründe mag in den höheren Kosten des Enzyms liegen, obwohl dieser Kostenfaktor bei Betrachtung der Gesamtkosten proteomischer Versuche relativ gering ausfällt. Der Hauptgrund ist sicherlich das Fehlen einer ausgeprägten Enzymspezifität, ArgC schneidet zwar bevorzugt nach Arginen, Schnitte nach Lysinen können ebenfalls beobachtet werden. Das Ausmaß der Schnitte nach Lysinen mag zwar nicht so hoch sein wie nach einem tryptischen Verdau, dennoch ist die Häufigkeit so hoch, dass die Auswertung erschwert wird. Des Weiteren sind die Anforderungen an den Reaktionsbedingungen deutlich höher als bei einem tryptischen Verdau. Zuletzt sei genannt, dass Peptide, die über ArgC generiert werden, Lysine innerhalb der Peptidkette enthalten können, welches die Kollision-induzierte Fragmentierung der Peptide negativ beeinflussen kann.³³

Die oben aufgezählten Nachteile der Protease ArgC sind gravierend, daher wurde in verschiedenen zuvor veröffentlichten Arbeiten der Einsatz vermieden, auch wenn auf ein ArgC entsprechendes Resultat hingearbeitet wurde. Garcia et al vermieden zum Beispiel den Einsatz von ArgC durch die Entwicklung eines Protokolls, wobei die Lysine im ersten Schritt auf Proteinebene unter Einsatz von Propionsäureanhydrid propionyliert werden und anschließend das Protein mit Trypsin verdaut wird.⁵² Die chemische Modifizierung der Lysine bewirkte deren Funktionsverlust als Schnittstelle und Trypsin schneidet folglich beinahe exklusiv nur noch nach Argininen. Das erhaltene Resultat ist dabei dem Ergebnis eines ArgC-Verdaus ähnlich. Diese Methode wurde zur Charakterisierung von Histonen verwendet, welche hohe Anteile an Lysinen aufweisen und für die der tryptische Verdau somit nur zu sehr kleinen und wenig aussagekräftigen Peptiden führen würde.

Ähnliche Vorgehensweisen für andere Fragestellungen erfolgten mit anderen Reagenzien. Eine dieser Methoden diente zur Identifizierung des C-Terminus von Proteinen.¹⁰⁶ Das dazugehörige Protokoll enthielt die Dimethylierung der Protein-Aminogruppen. Ein anderes Verfahren ist die Acetylierung der Aminogruppen des Proteins unter Einsatz von Sulfo-NHS-Acetat mit dem Ziel der Anreicherung der N-Termini der Proteine nach einem tryptischen Verdau.¹⁰⁷ Ein weiteres potenzielles Reagenz, welches bislang noch nicht für Methoden benutzt wurde, die zu einem ArgC-ähnlichen Resultat führen, ist das Diethylpyrocabonat. Es ist ein sehr reaktives Reagenz, welches mit den verschiedenen Nukleophilen eines Proteins reagieren

kann. Durch die Wahl bestimmter Reaktionsbedingungen und den Einsatz von Hydroxylamin kann die Selektivität der Reaktion auf die Modifizierung der Aminogruppen gelenkt werden.^{99,108,109}

Trotz der positiven Erfahrungen, einen ArgC-Verdau durch Modifizierung der Proteinaminogruppen mit anschließendem tryptischen Verdau zu umgehen, wurde diese Vorgehensweise bislang nicht auf ihre Eignung zur Analyse eines Proteoms untersucht. Gründe hierfür könnten die Schwierigkeiten sein, die damit verbunden sind, Proteine chemisch zu modifizieren. Erfolgt die Umsetzung beispielsweise nicht quantitativ, erhöht sich der Komplexitätsgrad einer an sich schon sehr komplexen Probe in einem Maß, dass mit Einbußen in den Resultaten zu rechnen ist. Zusätzlich wird die Aufreinigung der Probe schwieriger, da der gesamte Arbeitsablauf unter Berücksichtigung der Reduzierung und Alkylierung der Cysteine nun zwei Modifizierungsschritte enthält. Die Reagenzien müssen nach Ablauf der Reaktion jeweils entfernt werden, um Interaktionen unter den Modifizierungen zu vermeiden sowie um Aktivitätsverluste des Trypsins zu vermeiden.

In den vergangenen Jahren wurden neue Aufreinigungsmethoden entwickelt, die nützliche Alternativen zu den traditionellen auf Proteinfällungsbasierten Methoden darstellen. Der Einsatz von Molekularfiltern (FASP,^{26,27} siehe Abschnitt 1.2.2) gehört mittlerweile zu den gängigsten Methoden. Eine kürzlich eingeführte Aufreinigungsmethode basiert auf der Nutzung von paramagnetischen Beads (SP3,²⁸ siehe Abschnitt 1.2.2), die in dieser Arbeit in abgewandelter Form genutzt wurden (siehe Abschnitt 2.2.3).

Ziel dieser Arbeit ist die Entwicklung einer Methode, die zu einem ArgC-ähnlichem Resultat führt und zur Analyse von proteomischen Proben geeignet ist. Hierzu werden im ersten Schritt die zu analysierenden Proteine kovalent an magnetische Beads gebunden. Nach Reduktion und Alkylierung erfolgt die chemische Modifizierung der Aminogruppen der Proteine. Die modifizierten und immobilisierten Proteine werden auf den Beads mit Trypsin verdaut, so dass das Resultat dem von ArgC ähnlich ist.

Zu Beginn dieser Arbeit werden die bioinformatischen Werkzeuge entwickelt, die zur Analyse proteomischer Daten zwingend notwendig sind. Im Anschluss daran beginnt die Entwicklung des ArgC-ähnlichen Verdauens unter Einbeziehung von magnetischen Beads. Im ersten Schritt wurden die im oberen Teil aufgezählten Möglichkeiten zur chemischen Modifizierung von Protein-Aminogruppen auf ihren Umsetzungsgrad hin untersucht. Im einzelnen werden die Reaktionsbedingungen optimiert, so dass die Umsetzungen nahezu quantitativ erfolgen und das Ausmaß der Nebenprodukte auf ein Minimum reduziert wurden. Nach erfolgter Optimierung wurden die Resultate der ArgC-ähnlichen Ansätze mit denen von konventionellen tryptischen und ArgC-Verdauen verglichen. Das Reagenz, welches in diesen vergleichenden Analysen in der Summe zu den besten Resultaten

3. FRAGESTELLUNG UND ZIELE

geführt hat, wurde anschließend detailliert mit dem konventionell tryptischen Verdau verglichen. Ziel dieser Untersuchungen war die Beantwortung der Frage, ob die Resultate des ArgC-ähnlichen Verdaues lediglich als komplementär zu dem tryptischen Verdau betrachtet werden können oder aber diese Art des Verdaues tatsächlich zu besseren Ergebnissen führt und somit eine ernstzunehmende Alternative darstellt. Zum Schluss werden Anwendungsmöglichkeiten des ArgC-ähnlichen Verdaues für Fragestellungen wie Quantifizierungen und Charakterisierungen untersucht.

Ergebnisse und Diskussion

Die Ergebnisse der vorliegenden Untersuchungen und ihre Diskussion werden im folgenden Abschnitt dargestellt. Dazu werden zunächst die bioinformatischen Grundlagen zur Analyse massenspektrometrischer Daten beschrieben. Anschließend wird die Entwicklung eines alternativen *Bottom-Up*-Ansatzes aufgezeigt. Schließlich wird diese entwickelte Methode auf ihre Anwendbarkeit für ausgewählte analytische Fragestellungen überprüft.

4.1 Entwicklung bioinformatischer Werkzeuge zur Analyse von MS-Daten

Die von Geräteherstellern mitgelieferten Auswerteprogramme sind häufig nur für bestimmte Fragestellungen konzipiert und lassen keine Erweiterungen ihrer Funktionalitäten zu. So ist es beispielsweise nicht möglich, mit dem von Thermo Fisher mitgeliefertem Programm automatisierte Auswertungen („Batch-Analysen“) durchzuführen. Das Programm bietet lediglich die Möglichkeit, einzelne Massenspektren auszuwerten, die zuvor mit der MALDI LTQ-Orbitrap™ XL aufgenommen wurden. Auch *Open-Source* Programme (siehe Abschnitt 1.5) boten zu dem damaligen Zeitpunkt keine vorgefertigte Lösung an, MALDI-MS-Dateien automatisiert auszuwerten. Jedoch weisen sie den Vorteil auf, dass ihr Quellcode unverschlüsselt ist und somit Erweiterungen der Funktionalitäten dieser Programme möglich sind.

Die Auswertung von MALDI-MS-Dateien unterteilt sich wie im Abschnitt 1.5 auf Seite 23 beschrieben in eine optionale Konvertierung der Daten (in der Regel bei Anwendung von *Open-Source*-Programmen), einer Signalprozessierung und die sich anschließende Protein- oder Peptididentifizierung mit Hilfe einer Suchmaschine. Zur Umsetzung dieser Einzelschritte wurden verschiedene Programme verwendet. Im Folgenden werden einerseits die Einzelprogramme kurz vorgestellt und zudem beschrieben, wie die Verkettung

der Einzelprogramme gelang, um eine automatisierte Auswertung zu ermöglichen.

4.1.1 Konvertierung der MALDI-MS-Dateien

Die Signalprozessierung erfolgte mit dem *Open-Source*-Programm „mMass“,¹¹⁰ so dass es zwingend erforderlich war die aufgenommenen MALDI-MS-Dateien in ein für dieses Programm lesbares Format zu überführen. Neben dem im Abschnitt 1.5.1 auf Seite 24 eingeführten mzML-Format⁷⁴ existieren noch eine Reihe weiterer massenspektrometrischer Rohdateienformate, die mMass einlesen kann, wie etwa *mzData*,¹¹¹ *mzXML*¹¹² und *mgf*. Die Konvertierung in eines dieser Formate erfolgte instrumentenabhängig, wie in den nächsten Unterabschnitten dargelegt.

RAW-Dateien von der MALDI LTQ-Orbitrap™ XL

Die MALDI-Orbitrap nimmt eine vom Benutzer definierte Anzahl an Subspektren auf und speichert diese in eine RAW-Datei als *Scans* ab. Die RAW-Datei kann mit Hilfe von MSConvert⁷⁶ von Proteowizard¹¹³ in eine mzML-Datei konvertiert werden. In der generierten mzML-Datei liegen die Scans als Einzelspektren vor. Zur weiteren Auswertung müssen die Einzelspektren zu einem Gesamtspektrum aufsummiert werden.

t2d-Dateien von dem 4800 MALDI-TOF/TOF™ Analyser

Das Rohdaten-Format dieses Gerätes ist das *t2D*-Format, welches nicht von MSConvert erkannt wird. Es existieren aber glücklicherweise noch andere frei zugängliche Konverter wie „t2d-Convert“ von „proteomecommons“,¹¹⁴ welcher t2d-Dateien zwar nicht in mzML-Dateien, jedoch in mzXML-Dateien konvertieren kann. Die mzXML-Datei enthält nur das Gesamtspektrum, so dass eine Aufsummierung der Einzelaufnahmen nicht notwendig ist.

txt-Dateien der Voyager-DE™ STR Workstation

Massenspektren, die am Voyager-DE STR aufgenommen werden, lassen sich mit Hilfe der Aufnahme-Software in Textdateien exportieren. Diese Textdatei enthält zwei Spalten, wobei die erste Spalte das m/z-Verhältnis und die zweite Spalte die Intensität des aufgenommenen Massenspektrums wiedergibt. Auch hier war die Konvertierung problematisch, da die verfügbaren freien Konverter diesen Dateityp nicht einlesen konnten. Das Programm „mMass“, welches in den nächsten Abschnitten ausführlicher vorgestellt wird, enthält *Parser* für die verschiedenen Datenformate. Ein Parser zerlegt in der Informatik eine Eingabe und wandelt sie in ein für die Weiterverarbeitung geeignetes Format um. Nach geringfügigen Anpassungen des Quellco-

des eines dieser Parser, konnten die Textdateien von mMass zur weiteren Verarbeitung eingelesen werden.

Der Code 4.1 zeigt, dass lediglich eine Funktion verändert werden musste. Die for-Schleife in den Zeilen 13 - 17 splittet jede Zeile der Textdatei in die Teile, die von der Leerstelle Tab (\t) getrennt sind und fügt diese der Liste „data“ hinzu. Der Rückgabewert dieser Funktion ist das Listenobjekt „data“, welches von mMass weiter verarbeitet werden kann.

Code 4.1: Parsen von txt-Dateien

```
1 def _parseData(self):
2     """Parse data."""
3
4     # open document
5     try:
6         document = file(self.path)
7         rawData = document.readlines()
8         document.close()
9     except IOError:
10        return False
11    # read lines
12    data = []
13    for line in rawData:
14        if line[0]=='T': continue
15        buff=line.split('\t')
16        data.append([float(buff[0]), float(buff[1])])
17    return data
```

4.1.2 Signalprozessierung von Massenspektren mit mMass

Nach erfolgreicher Konvertierung können die in den Dateien enthaltenen Massenspektren mit Hilfe von mMass prozessiert werden. Das Programm mMass wurde von Martin Strohhalm in der Programmiersprache Python geschrieben und ist im Wesentlichen aus drei Einheiten aufgebaut:

- Das Modul **gui** (General-User-Interface) enthält den Code für die Benutzeroberfläche des Programmes.
- Das Modul **mspy** enthält sämtliche Funktionen zum Parsen von Dateien, zur Signal-Prozessierung und Objekdefinitionen.
- Der Ordner **configs** enthält die Konfigurationsdateien, in denen beispielsweise die Spezifität von Enzymen gespeichert sind.

Code 4.2: Generierung von „peaklist“ aus mzML-Dateien

```
1 def genPL(mzmlPfad="", pickingHeight=0.95, absThreshold=0.,
2           relThreshold=0.001, snThreshold=0.0,
3           baselineWindow=0.1, baselineOffset=0.,
4           smoothMethod="SG", smoothWindow=0.05,
5           smoothCycles=1):
6
7     import mMass.mspy.parser_mzml as parser
8     from AKK.tools import raw
9     mzml=parser.parseMZML(raw(mzmlPfad))
10    numScans=len(mzml.scanlist())
11    if numScans==0: return False
12
13    i=2
14    scan=mzml.scan(1)
15    while mzml.scan(i):
16        scan.combine(mzml.scan(i))
17        i+=1
18
19    scan.labelscan(pickingHeight, absThreshold,
20                  relThreshold, snThreshold,
21                  baselineWindow, baselineOffset,
22                  smoothMethod, smoothWindow,
23                  smoothCycles)
24    return scan.peaklist
```

Der Code 4.2 zeigt, wie mzML-Dateien, die aus der Konvertierung von Thermo RAW-Dateien entstanden sind, eingelesen und verarbeitet werden. In der neunten Zeile des Codes wird das Objekt „mzml“ durch Angabe des Dateipfades instanziiert. Objekte haben in Python, wie in anderen objektorientierten Programmiersprachen, Attribute und Methoden. So hat beispielsweise das Objekt mzml die Methode „scan“, die unter Angabe einer Nummer zum Aufruf eines Einzelspektrum führt. Wie im Abschnitt 4.1.1 auf Seite 54 besprochen, müssen die Einzelspektren der MALDI-Orbitrap-Aufnahmen zunächst aufsummiert werden. Dies geschieht in Zeilen 14-17 und das Gesamtspektrum wird der Variable „scan“ zugewiesen. Die Methode „labelscan“ ermittelt ab der Zeile 19 die Basislinie, glättet das Signal und das sich anschließende Peak-Picking führt zu dem Objekt „peaklist“. Dieses Objekt kann im Anschluss deisotopiert werden, so dass im Anschluss nur noch die monoisotopischen Signale im Massenspektrum verbleiben. Nach endgültiger Prozessierung kann dieses Objekt als mgf-Datei gespeichert werden, bevor es der Suchmaschine Mascot übermittelt wird.

4.1.3 Parsen von Mascot dat-Dateien

Das Ergebnis der Suchmaschine Mascot (Matrix Science) wird in einer Datei im dat-Format gespeichert. Die Entwickler von Mascot stellen gratis einen Parser zur Verfügung, der die Extrahierung der Informationen aus den dat-Dateien mithilfe verschiedener Programmiersprachen (Python, C++, Java, Perl) ermöglicht. So können die Ergebnisse eines Suchmaschinenlaufes beispielsweise als Protein oder Peptidlisten wiedergegeben werden. Der Parser, der dies ermöglicht, heißt „MSParser“¹¹⁵ und innerhalb dieser Arbeit wurde er strikt nach der von Matrix Science herausgegebenen Dokumentation verwendet. Bei der verwendeten Programmiersprache handelt es sich erneut um Python.

4.1.4 Integration der Auswerteprogramme in KNIME

Die in den vorangegangenen Abschnitten beschriebenen Einzelprogramme ließen sich mit Hilfe eines Programmes namens „KNIME“¹¹⁶ verknüpfen, so dass Batch-Analysen von MALDI-MS-Dateien ermöglicht wurden. KNIME ist ein Datenanalyse-Programm und findet breite Anwendung in verschiedenen Bereichen, in denen größere Datenmengen ausgewertet werden müssen. KNIME wurde bereits zur Auswertung massenspektrometrischer Daten genutzt, so sind beispielsweise Funktionalitäten des Projektes „OpenMS“¹¹⁷ in KNIME nutzbar.¹¹⁸ In KNIME werden *Workflows* erstellt, die aus Knoten aufgebaut sind, welche über *Ports* miteinander verknüpft werden. Ein Knoten empfängt Daten über seinen *Input-Port*, verarbeitet diese und gibt das Resultat über den *Output-Port* an den nächsten Knoten weiter. Für KNIME existieren weit über 1000 Knoten, die über die verschiedensten Funktionalitäten verfügen. Einige dieser Knoten bieten Schnittstellen zu anderen Programmiersprachen wie Python an und ermöglichen so die Integration von Python-Code in diese Workflows.

Beispielhaft ist ein *Workflow* zur Auswertung von MALDI-MS-Dateien, wie in der Abbildung 4.1 auf der nächsten Seite zu sehen ist. Der erste Knoten listet die auszuwertenden Rohdateien in einem vom Benutzer spezifizierten Verzeichnis auf. Der daran angeknüpfte Knoten konvertiert diese Dateien in das mzML-Format und darauffolgend erfolgt im dritten Knoten die Signalprozessierung und die Abspeicherung des Resultates als mgf-Datei. Der vierte Knoten übermittelt die mgf-Datei der Suchmaschine Mascot, dessen Ergebnisse in dat-Dateien gespeichert werden. Die Erstellung der Peptid- und Proteinlisten erfolgt mit Hilfe des MS-Parsers und diese Listen können anschließend nach Excel exportiert werden.

Die eckigen Knoten in den KNIME-*Workflows* sind *Metaknoten*, welche wie in Abbildung 4.42 zu sehen ist aus anderen Knoten aufgebaut sind. Das Zusammenfassen mehrerer Knoten zu einem Metaknoten führt zu übersichtlicheren *Workflows* und durch den Einsatz von *Quicknodes* lassen sich diese

auch einfacher konfigurieren. Des Weiteren ist in der Abbildung zu sehen, dass in dem Metaknoten ein Python-Script-Knoten enthalten ist, welcher als Schnittstelle zur Programmiersprache Python fungiert. Der Code 4.3 auf der nächsten Seite soll beispielhaft zeigen wie der Code in einem Python-Script-Knoten aussehen kann. In diesem Beispiel werden die eingelesenen mzML-Dateien signalprozessiert und das Resultat als mgf-Datei ausgegeben. In den Zeilen 5-13 des Codes werden die Prozessierungsparameter, die über Quicknodes eingegeben wurden, eingelesen. Innerhalb der for-Schleife erfolgt dann die Signalprozessierung, das Deisotopieren und die Ausgabe als mgf-Datei. Diese aufgezählten Schritte erfolgen mit Funktionen aus dem Programm-Code von mMass und dieses Beispiel verdeutlicht wie die Integration dieser Funktionen in KNIME gelingt. Die anderen Python-Module wie der MSParser, wurden in ähnlicher Art und Weise in KNIME integriert.

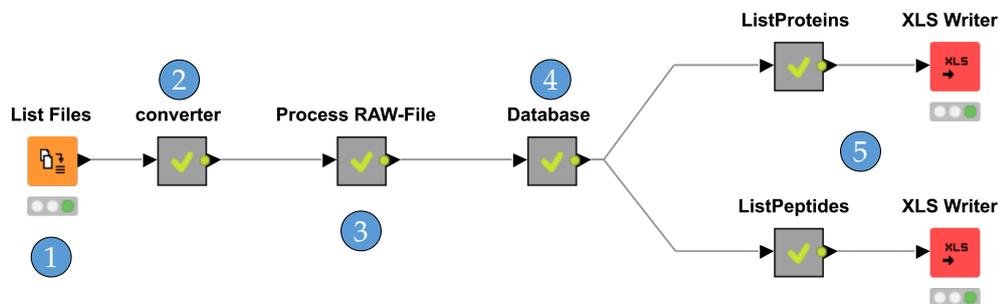


Abbildung 4.1: KNIME-Workflow zur Prozessierung von MALDI-MS-Dateien

- ① Auflistung der zuprozessierenden RAW-Dateien.
- ② Konvertierung zu mzML-Dateien.
- ③ Dieser Knoten wurde im Abschnitt 4.1.2 besprochen. Hier wird eine aus einer mzML-Datei eine deisotopierte Peakliste erzeugt, die als mgf-Datei gespeichert werden kann. Des Weiteren lässt sich hier eine Referenzmasse definieren, dessen Intensität als Maß für eine relative Quantifizierung dient.
- ④ Mascot-Datenbanksuche wird durchgeführt, dessen Resultat in einer dat-Datei gespeichert wird.
- ⑤ Mit Hilfe des MSParser können die Resultate der Datenbanksuche aus der erzeugten dat-Datei extrahiert und aufgelistet werden. Die Tabellen werden als Excel-Dateien gespeichert.

4.1. Entwicklung bioinformatischer Werkzeuge zur Analyse von MS-Daten

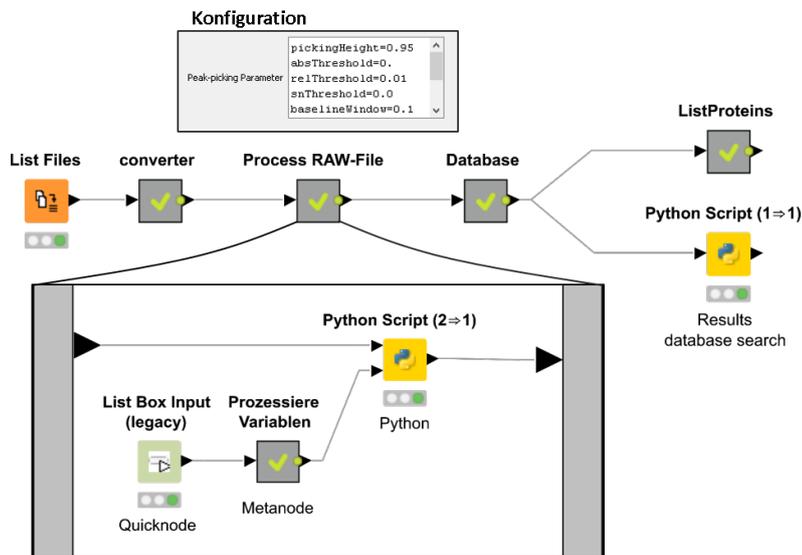


Abbildung 4.2: Aufbau eines Metaknotens

Code 4.3: Erzeugung von mgf-Dateien aus mzML-Dateien

```

1 import AKK.processing as test
2
3 l_mgfURL=[]
4 pickingHeight=float(input_table_2['Values']['pickingHeight'])
5 relThreshold=float(input_table_2['Values']['relThreshold'])
6 blWindow=float(input_table_2['Values']['baselineWindow'])
7 blOffset=float(input_table_2['Values']['baselineOffset'])
8 smoothMethod=input_table_2['Values']['smoothMethod']
9 smoothWindow=float(input_table_2['Values']['smoothWindow'])
10 smoothCycles=int(input_table_2['Values']['smoothCycles'])
11
12 for loc in input_table_1['mzml']:
13     mzml=test.mzML(mzmlPfad=loc)
14     peaklist=mzml.genPL(pickingHeight,
15                        relThreshold, blWindow,
16                        blOffset, smoothMethod,
17                        smoothWindow, smoothCycles)
18     peaklist=mzml.deisotopePL(peaklist)
19     mgf=mzml.makeMGF(peaklist)
20     l_mgfURL.append(mgf)
21
22 input_table_1['mgf']=l_mgfURL
23 output_table =input_table_1

```

4.1.5 Workflows zur Auswertung von MALDI-ISD-Spektren

Die Ionisierung über MALDI kann unter Einsatz bestimmter Matrices wie 1,5-Diaminonaphthalin (DAN-Matrix) zur Fragmentierung des Analyten schon während des Ionisierungsvorganges führen.¹¹⁹ Aufgrund des Auftretens in der Ionenquelle wird diese Art der Fragmentierung als *In-Source-Decay* (ISD) bezeichnet. Es handelt sich hierbei um keine MS/MS-Methode, da die Fragmentierung ohne Auswahl eines Vorläuferionens erfolgt. MALDI-ISD-Spektren von Proteinen können aussagekräftige Fragmentionenserien enthalten, die die Identifikation oder die Charakterisierung des Proteins ermöglichen. Die

Code 4.4: Umwandlung von mgf-Dateien zu ISD-mgf-Dateien

```
1 def makeISDmgf(mgf):
2     import os.path as path
3     peaklist=mgf.getPeaklist()
4     peaklist=collections.OrderedDict(sorted(peaklist.items(),
5         reverse=True))
6     liste_mz=[mz for mz,ai in peaklist.iteritems()]
7     liste_ai=[ai for mz,ai in peaklist.iteritems()]
8     text=""
9     while len(liste_mz)>1:
10        text+="Begin ions\n"
11        text+="Pepmass= %s\n"%(liste_mz[0])
12        for i in range(1,(len(liste_mz))):
13            text+="s\t%s\n"%(liste_mz[i],liste_ai[i])
14        text+='End ions\n'
15        del liste_mz[0]; del liste_ai[0]
16        mgfpath=path.splitext(self.mgfPfad)[0] + "_ISD_mgf.mgf"
17        fobj=open(mgfpath,'w'); fobj.write(text)
18    return mgfpath
```

Fragmentierung wird vermutlich durch den Übertrag von Wasserstoff-Radikalen von der Matrix auf den Analyten induziert¹²⁰, die darüber hinaus auch zur Spaltung von Disulfid-Brücken führen. Das letztere wird ausgenutzt für die Analyse von Antikörpern, dessen Untereinheiten über Disulfidbrücken miteinander verbunden sind.¹²¹ Die voneinander getrennten Untereinheiten können dann anschließend über eine andere Fragmentierungstechnik wie CID weiter analysiert werden.

Da die Aufnahme von MALDI-ISD-Spektren ohne eine Auswahl von Vorläuferionen erfolgt, ist das Resultat der Messung ein MS-Spektrum. Dieses Spektrum enthält jedoch Signale von Fragmentionen und muss folglich wie ein MS/MS Spektrum ausgewertet werden. Dies ist jedoch beispiels-

4.1. Entwicklung bioinformatischer Werkzeuge zur Analyse von MS-Daten

weise mit der Mascot-Suchmaschine nicht möglich, da die Angabe eines Vorläuferions zwingend notwendig ist. Der Aufbau der mgf-Dateien unterscheidet sich somit je nach Suchmodus. Erfolgt eine PMF-Suche, enthalten die mgf-Dateien lediglich eine Peak-Liste. Eine mgf-Datei für eine PFF-Suche kann die Information mehrerer MS/MS-Spektren enthalten. Für jedes einzelne MS/MS-Spektrum ist das Vorläuferion angegeben, gefolgt von der Peakliste des dazugehörigen Fragmentspektrums, welches umklammert ist von Textzeilen mit dem Inhalt „BEGIN ION“ und „END ION“.

Wie die Umwandlung von mgf-Dateien, die nur eine Peakliste enthalten, zu einer MS/MS-mgf-Datei erfolgen kann, zeigt der Code 4.4 auf der vorherigen Seite. Hierzu wird aus jeder gelisteten Masse in der Peakliste ein Vorläuferion und sein Fragmentspektrum setzt sich zusammen aus den Peaks deren Massen kleiner sind als das ausgewählte Vorläuferion.

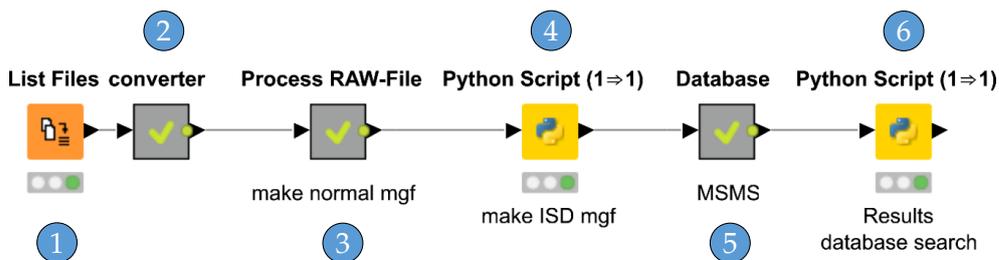


Abbildung 4.3: Workflow zur Auswertung von MALDI-MS/MS-Dateien

- ① Auflistung der zu prozessierenden RAW-Dateien.
- ② Konvertierung zu mzML-Dateien.
- ③ Signalprozessierte Spektren werden als mgf-Dateien gespeichert.
- ④ Umwandlung von zu MS/MS-mgf-Dateien
- ⑤ Mascot MS/MS-Datenbanksuche
- ⑥ Ausgabe der Ergebnis-URLs in Form einer HTML-Datei

In der Abbildung ?? ist beispielhaft ein Workflow zur Auswertung vom MALDI-MS/MS-Spektren abgebildet. Bis zum dritten Knoten ist dieser Workflow identisch zu dem zur Auswertung von MALDI-MS-Spektren, so dass auch hier mgf-Dateien entstehen. Wie die Umwandlung dieser mgf-Dateien in MS/MS-mgf-Dateien erfolgt, wurde oben gezeigt und der dazugehörige Code befin-

det sich im vierten Knoten. Im Anschluss erfolgt eine PFF-Suche und das Ergebnis dieser Suche wird exportiert.

4.1.6 Diskussion

Für die Auswertung von MALDI-MS-Dateien lagen verschiedene Software-Lösungen vor, die häufig nur einen einzelnen Aspekt bearbeiteten. Für eine vollständige Auswertung war somit die manuelle Übertragung der Ergebnisse von einem Programm zum nächsten notwendig, was zeitaufwendig und fehleranfällig ist. Durch die Integration von MSConvert, mMass und dem MSParser in KNIME wurde eine leicht zu konfigurierende Lösung entwickelt, die eine vollständige und automatisierte Auswertung der verschiedenen MALDI-MS-Dateitypen erlaubt. Des Weiteren wurden *Workflows* in KNIME generiert, welche MALDI-MS Daten auswerten können. Dies war bislang nur mit kommerziellen Programmen möglich. Ein weiterer Vorteil, der sich durch die Integration der Programme in KNIME ergibt, ist die vereinfachte anschließende Datenauswertung. Die von KNIME im Basisprogramm enthaltenen Knoten ermöglichen die Aufbereitung und Visualisierung der Daten. Es bedarf somit keines Exportes der Daten in andere Programme, um diese auszuwerten.

Die hier vorgestellten *Workflows* bedienen sich der Programmiersprache Python, welche im Vergleich zu anderen Programmiersprachen eine mangelnde Leistungsfähigkeit aufweist, die durch eine vergleichsweise lange Bearbeitungszeit spürbar wird. Eine Übersetzung des Codes von Python zu Java würde mit Sicherheit zu einer spürbaren Beschleunigung der Bearbeitungszeit führen, da KNIME selbst in Java geschrieben ist und Skripte in dieser Sprache die höchste Integrierbarkeit in KNIME aufweisen.

4.2 Chemische Modifikation der Aminogruppen auf Proteinebene

Im vorherigen Abschnitt wurden lediglich die bioinformatischen Grundlagen zur Auswertung von MALDI-MS-Dateien geschaffen, ohne konkret auf die im Abschnitt 3 auf Seite 49 gestellte Zielsetzung hinzuwirken. Die in dieser Zielsetzung beschriebene Etablierung eines ArgC-ähnlichen Verdauens beinhaltet die chemische Modifizierung der Aminogruppen eines Proteins. Die verschiedenen Möglichkeiten diese zu modifizieren wurden bereits im Abschnitt 1.3.2 gezeigt und werden hier mit den entsprechenden Abkürzungen aufgelistet.

- Propionylierung mit Propionsäureanhydrid (PA)
- Carboxylierung unter Zugabe von Diethylpyrocarbonat (DEPC)
- Reduktive Dimethylierung (DM)
- Acetylierung mit Hilfe eines NHS-Esters (NHS)

Die aufgezählten Modifizierungen wurden bislang in verschiedenen Zusammenhängen mehrfach publiziert und die dazugehörigen Protokolle sind im Abschnitt 2.2.5 aufzufinden.

4.2.1 Verifizierung der Modifizierungsprotokolle auf Peptid- und Proteinebene

Die publizierten Protokolle aus Abschnitt 2.2.5 wurden zunächst auf Ihren Umsetzungsgrad hin untersucht. Als Modellsubstanz diente das Peptid Substanz P, welches in seiner Aminosäurekette ein Lysin enthält und somit unter Berücksichtigung seines N-Terminus zwei Aminogruppen aufweist. Nach vollständiger Modifikation der Aminogruppen dieses Peptides sollte folglich eine Massenverschiebung im Spektrum zu beobachten sein, die doppelt so hoch ist wie nach einer einfachen Modifikation des Peptides.

Das erste Spektrum in Abbildung 4.4 auf der nächsten Seite zeigt das Massenspektrum des unmodifizierten Peptides. Die übrigen Massenspektren wurden nach Anwendung des jeweiligen Modifizierungsprotokolls aufgenommen. Zur besseren Darstellung wurden diese Spektren geflippt, welches durch die Vorzeichenänderung der Intensitäten realisiert wurde. In den geflippten Spektren ist die Ausgangsmasse des Peptides nicht zu sehen. Des Weiteren sind bis auf den Produktpeak keine Zwischenprodukte zu sehen. Die Modifizierung des Peptides lief somit in jedem Fall vollständig und quantitativ nach Anwendung der bereits publizierten Protokolle ab.

Im nächsten Schritt wurde der Umsetzungsgrad der Modifizierungsprotokolle auf Proteinebene überprüft. Alkohol-Dehydrogenase (ADH) wurde als

4. ERGEBNISSE UND DISKUSSION

Modellprotein gewählt, da nach erfolgreicher Modifikation und Trypsinverdau nur drei (bei Vorliegen von Methioninoxidation vier) Signale in dem

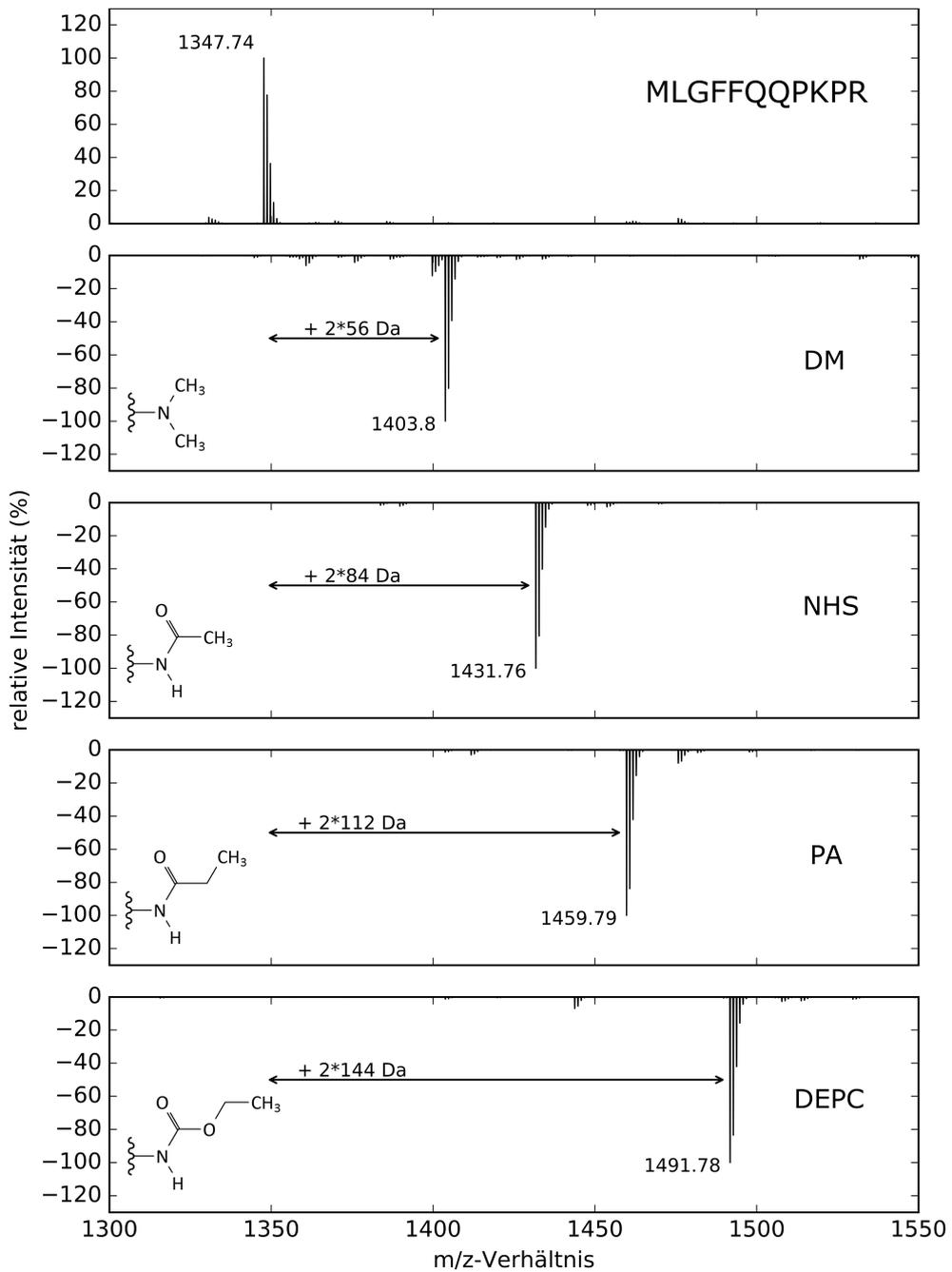


Abbildung 4.4: MALDI-MS Spektrum von Substanz P und seinen derivatisierten Analoga

gemessenen Massenbereich (800 - 4000) zu erwarten sind und somit ein leicht zu interpretierendes Massenspektrum resultieren sollte. Die gemessenen Massenspektren wiesen jedoch zusätzliche Peaks auf, welche durch eine nicht vollständige Umsetzung des ADHs erklärt werden können.

Unvollständige Modifizierungen würden auf Proteomebene zu noch komplexeren Proben führen und folglich zu einer verringerten Anzahl an Proteinidentifikationen. Des Weiteren wird die Reproduzierbarkeit der Resultate verschlechtert und Quantifizierungen wären dadurch erschwert. Es war somit zwingend erforderlich, die Reaktionsbedingungen so anzupassen, dass höhere Umsetzungsraten erzielt werden konnten. Die Optimierung der Reaktionsbedingungen erfolgte mit Hilfe der statistischen Versuchsplanung (*Design of Experiment, DOE*), welche im nächsten Unterabschnitt theoretisch eingeführt wird. In den darauffolgenden Abschnitten wird im Einzelnen gezeigt, wie die Modifizierungsprotokolle mit Hilfe eines DOEs auf Ihren Umsetzungsgrad hin optimiert werden konnten. Anhand des ersten Fallbeispiels (reduktive Methylierung) werden die statistischen Begriffe und zum Teil auch die Auswertung erklärt.

4.2.2 Statistische Versuchsplanung (Design of Experiment, DOE)

Die Statistische Versuchsplanung (Design of Experiment, DOE)¹²² ermittelt mit möglichst wenigen Versuchen den funktionalen Zusammenhang von Einflussparametern (z.B. Temperatur) und einer Zielgröße (*Response Factor*, im folgenden Beispiel die *Intensity Coverage*¹²³) und beschreibt diesen mathematisch. Dies gelingt durch eine systematische Planung und die statistischen Auswertung der Versuche. In derartigen Ansätzen können auch Wechselwirkungen von Einflussfaktoren erkannt werden, die mittels eines intuitiven Vorgehens, wie der schrittweisen Veränderung eines einzelnen Faktors, schwer zu ermitteln sind. Zur Planung und Auswertung wurde die Software MODDE von Umetrics^{124,125} herangezogen, die für die verschiedenen Bereiche eines DOE Assistenten bereitstellt.

Der „Design-Wizard“ ist ein Assistent zur Erstellung von Versuchsplänen. Hierzu werden die Einflussfaktoren und die Zielgröße definiert. Nach praktischer Durchführung des Versuchsplanes und Eintragungen der Resultate in die entsprechende Versuchsmatrix kann mit Hilfe des „Analysis-Wizard“ geschlussfolgert werden, welcher Einflussfaktor in welchem Maße das Ergebnis beeinflusst. Dies lässt letztlich Vorhersagen über optimierte Reaktionsbedingungen zu.

In jedem der unten aufgeführten Optimierungsansätze wurde als Zielgröße die *Intensity Coverage* gewählt, da sie am ehesten Aufschluss über den Umsetzungsgrad der Modifikation gibt. Sie errechnet sich folgendermaßen:

$$\text{Intensity Coverage} = \frac{\sum \text{Intensity of matched Peaks}}{\sum \text{Intensity of all Peaks}} \quad (4.1)$$

Je höher die *Intensity Coverage*, desto höher muss folglich der Umsetzungsgrad sein, da ansonsten nicht zuordenbare Peaks mit hoher Intensität auftreten, die eine Erhöhung des Wertes im Nenner bewirken und somit zu einer Verringerung der Gesamtgröße führen.

Reduktive Methylierung

Die reduktive Methylierung von ADH gehörte zu den Modifikationen mit dem höchsten Umsetzungsgrad, so dass eine Optimierung der Reaktionsbedingungen primär nicht notwendig war. Dennoch wurde ein DOE mit MODDE durchgeführt, um die Effekte der Einflußgrößen auf das Resultat zu ermitteln. So kann ein besseres Verständnis über die Modifizierungsreaktion gewonnen werden. Die Tabelle 4.1 gibt den vom Design-Wizard kreierten Versuchsplan mit den dazugehörigen Ergebniswerten an, welche aus der praktischen Durchführung dieses Plans resultierten.

Tabelle 4.1: Versuchsplan zur Optimierung der reduktiven Methylierung von ADH

| Nummer | Acetonitril (%) | Reagenzmenge (µL) | Intensity coverage (%) |
|--------|-----------------|-------------------|------------------------|
| 1 | 0 | 10 | 51,1 |
| 2 | 40 | 10 | 43,5 |
| 3 | 80 | 10 | 34,5 |
| 4 | 40 | 50 | 65,4 |
| 5 | 80 | 50 | 57,0 |
| 6 | 40 | 100 | 70,2 |
| 7 | 80 | 100 | 59,6 |
| 8 | 40 | 50 | 66,0 |
| 9 | 40 | 50 | 66,7 |
| 10 | 40 | 50 | 64,3 |

Mit Hilfe des Analysis-Wizards können anschließend die Daten ausgewertet werden. Abbildung 4.5 auf der nächsten Seite gibt eine Übersicht über die Ausprägungen der verschiedenen Qualitätsmerkmale an.

Der *Replicate Plot* oben links in der Abbildung zeigt die Resultate aller Ergebnisse für eine schnelle Rohdaten-Analyse. Experimente, die unter den gleichen Bedingungen durchgeführt wurden sind gekennzeichnet durch eine rote Einfärbung. Solche Experimente sollten in ihren Ergebnissen eine

4.2. Chemische Modifikation der Aminogruppen auf Proteinebene

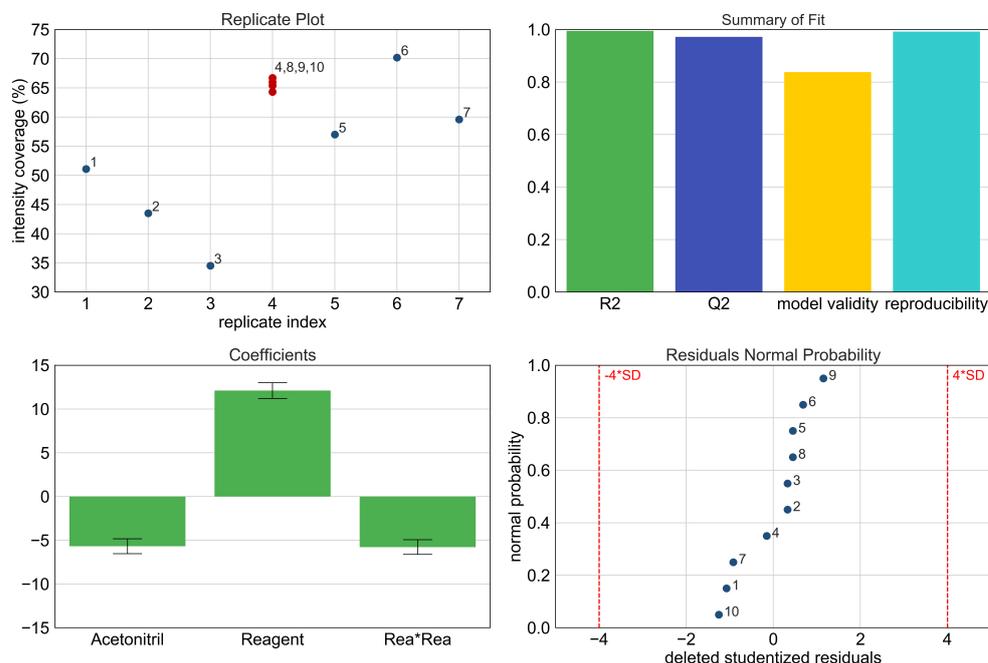


Abbildung 4.5: Übersicht über die Qualitätsmerkmale nach Optimierung der Reaktionsbedingungen der reduktiven Dimethylierung

deutlich geringere Variabilität aufweisen als die anderen Ergebnisse untereinander.

Das Koeffizientendiagramm im unteren linken Teil der Abbildung gibt an, welcher Faktor berücksichtigt werden soll und welcher keinen Einfluss auf die Zielgröße hat. Hierzu muss erwähnt werden, dass in der statistischen Versuchsplanung Auswertungen mit quantitativen (metrischen Faktoren) mit Hilfe von Regressionsanalysen erfolgen. Das Resultat der Regressionsanalyse ist ein Gleichungssystem (mathematisches Modell) mit dessen Hilfe eventuell auch Vorhersagen möglich sind. Hat ein untersuchter Faktor keinen Einfluss auf eine Zielgröße resultiert ein Koeffizient für dieses Modell, dessen Konfidenzintervall größer als der Koeffizient selbst ist. Ein derartiger Faktor ist nicht signifikant und sollte bei der Aufstellung des Modells nicht berücksichtigt werden. Im vorliegenden Fall ist aus der Abbildung zu entnehmen, dass sowohl der Zusatz von Acetonitril als auch die eingesetzte Menge an Methylierungsreagenz einen signifikanten Einfluss auf die Zielgröße haben. Der Zusatz von Acetonitril führt zu einer Erniedrigung der *Intensity Coverage*, erkennbar an der negativen Ausprägung des Koeffizienten. Des Weiteren ist aus der Abbildung zu entnehmen, dass kein linearer Zusammenhang zwischen der *Intensity Coverage* und der Menge zugesetztem

4. ERGEBNISSE UND DISKUSSION

Methylierungsreagenzes herrscht. Dem wird Rechnung getragen, in dem man das Quadrat dieser Größe in dem Modell mit einfließen lässt.

Im *Summary of Fit* im oberen rechten Teil der Abbildung sind verschiedene Größen dargestellt, die Aussagen über die Qualität des erstellten mathematischen Modells machen:

Das *Bestimmtheitsmaß* R^2 wird aus dem Quadrat des Korrelationskoeffizienten R (auch Pearson-Korrelation genannt) gebildet und gibt an, wie viel Prozent der Variation (Varianz) in den Daten durch ein vorliegendes Regressionsmodell erklärt werden können. Der Wertebereich von R^2 liegt zwischen 0 und 1. Dabei gilt, je näher R^2 an 1 liegt, umso besser lässt sich die Varianz der Variablen durch das gefundene Regressionsmodell beschreiben.

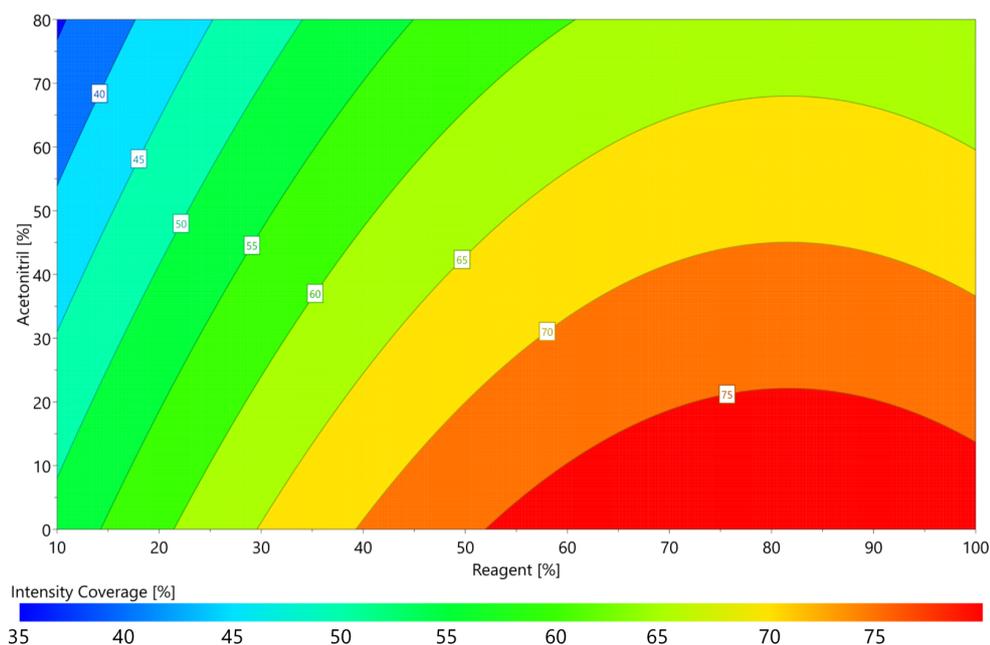


Abbildung 4.6: *Contour Plot* für die Optimierung der Reaktionsbedingungen der reduktiven Dimethylierung

Die *vorhergesagte Varianz* Q^2 gibt die Güte eines Regressionsmodelles an, Werte vorherzusagen.

Die *Modellvalidität* sollte größer 0,25 sein, anderenfalls liegt eine fehlende Anpassung (Lack of Fit, LOF) vor. Dieser liegt dann vor, wenn ein Regressionsmodell die funktionale Beziehung zwischen den experimentellen Faktoren und der Antwortvariablen (Werte der Zielgröße) nicht adäquat beschreibt. Ein LOF kann indiziert werden, wenn der „Modellfehler“ größer ist als der

4.2. Chemische Modifikation der Aminogruppen auf Proteinebene

„reine Fehler“. Replikationen geben den „reine Fehler“ wieder, da Unterschiede zwischen den beobachteten Werten und der Antwortvariablen nur durch zufällige Streuung verursacht werden können. Wohingegen der „Modellfehler“ aus einem fehlerhaften Modellansatz resultiert und kein statistisches Phänomen darstellt, da er durch eine Korrektur der Modellstruktur behoben werden kann.

Die Reproduzierbarkeit ist ein Maß für die Variation der Ergebnisse, die unter den gleichen Bedingungen erzielt wurden. Je kleiner der „reine Fehler“ bei Replikaten ist, desto höher ist die Reproduzierbarkeit.

In der letzten Grafik wird die „Verteilung der Residuen“ dargestellt. Ein „Residuum“ berechnet sich durch die Differenz des beobachteten vom vorhergesagten Wert des Regressionsmodells. Innerhalb dieser Grafik wurden die Werte „studentisiert“ (Transformation der Werte, so dass das arithmetische Mittel bei Null und die Stichprobenvarianz bei Eins liegt.) und in einem Normal-QQ-Plot dargestellt. Die Residuen sind normal verteilt, wenn die Punkte auf einer Diagonalen liegen. Punkte außerhalb der rot gestrichelten Linie sind Ausreißer und deren Berücksichtigung führt zu einem Modell schlechter Güte. Aus dem *Summary of Fit* ist ersichtlich, dass alle

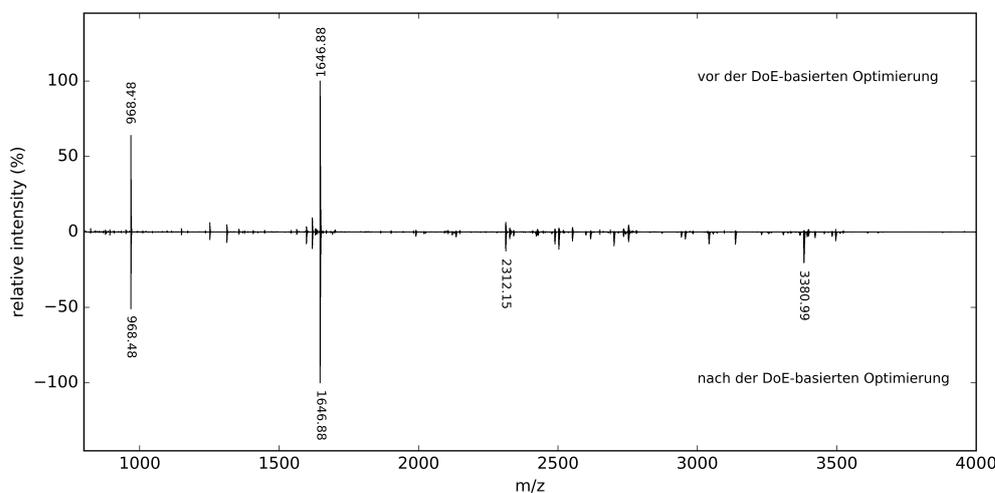


Abbildung 4.7: MALDI-MS-Spektren vor (oben) und nach (unten) Optimierung der Reaktionsbedingungen der reduktiven Methylierung

oben erwähnten Qualitätsmerkmale vorhanden sind und das vorgeschlagene Modell vertrauenswürdige Vorhersagen erlaubt. Anhand eines *Contour-Plot*, wie es in Abbildung 4.6 auf der vorherigen Seite dargestellt ist, lässt sich grafisch ermitteln, wie man die Faktoren Acetonitril-Zusatz und Menge des Methylierungsmittels wählen muss, um in einen Bereich mit einem hohen Umsetzungsgrad zu gelangen. Im vorliegenden Fall ist der hohe

Zusatz an Methylierungsreagenz ausschlaggebend für einen hohen Umsetzungsgrad, wohingegen schon geringe Zusätze von Acetonitril zu einer Verschlechterung der Ergebnisse führen.

Zum Vergleich wurden beide Methylierungsprotokolle an ADH angewendet und nach dem tryptischen Verdau resultierten die Massenspektren in Abbildung 4.7 auf der vorherigen Seite. Das obere Massenspektrum zeigt das Resultat vor und das untere nach Optimierung der Reaktionsbedingungen. Eine Verbesserung ist im hohen Massenbereich zu sehen. Das Peptid mit dem m/z-Verhältnis über 3300 ist vor Optimierung der Reaktionsbedingungen kaum detektierbar, wohingegen mit dem neuen Protokoll ein deutliches Signal zu sehen ist.

Tabelle 4.2: Versuchsplan zur Optimierung von DEPC

| Nummer | Zeit (min) | Acetonitril (%) | DEPC (µL) | Temperatur (°C) | IC (%) |
|--------|------------|-----------------|-----------|-----------------|--------|
| 1 | 15 | 0 | 1 | 4 | 16,8 |
| 2 | 120 | 0 | 1 | 4 | 27,6 |
| 3 | 15 | 80 | 1 | 4 | 28,8 |
| 4 | 120 | 80 | 1 | 4 | 53,6 |
| 5 | 15 | 0 | 20 | 4 | 67,3 |
| 6 | 120 | 0 | 20 | 4 | 60,1 |
| 7 | 15 | 80 | 20 | 4 | 64,1 |
| 8 | 120 | 80 | 20 | 4 | 60,5 |
| 9 | 15 | 0 | 1 | 25 | 26,6 |
| 10 | 120 | 0 | 1 | 25 | 59,4 |
| 11 | 15 | 80 | 1 | 25 | 55,4 |
| 12 | 120 | 80 | 1 | 25 | 60,4 |
| 13 | 15 | 0 | 20 | 25 | 60,6 |
| 14 | 120 | 0 | 20 | 25 | 58,5 |
| 15 | 15 | 80 | 20 | 25 | 53,2 |
| 16 | 120 | 80 | 20 | 25 | 54,1 |
| 17 | 60 | 40 | 10 | 25 | 63,7 |
| 18 | 60 | 40 | 10 | 25 | 54,8 |
| 19 | 60 | 40 | 10 | 25 | 56,1 |

Carbethoxylierung

DEPC ist ein sehr reaktives Reagenz, welches mit sämtlichen nukleophilen Gruppen eines Proteins reagieren kann. In vorherigen Publikationen wurden Reaktionen mit dem N-Terminus, Histidin, Tyrosin, Serin, Cysteine, Threonin, Lysin, Glutamin- und Asparaginsäure berichtet.^{99,108} Die ge-

4.2. Chemische Modifikation der Aminogruppen auf Proteinebene

nannten Reaktionen lassen sich bis auf die Reaktion mit den Aminogruppen durch den Zusatz von Hydroxylamin umkehren. Aus diesem Grund ist das Abstoppen der Reaktion mit Hydroxylamin ein zwingend notwendiger Schritt, wenn nur die Carboethoxylierung der Aminogruppen erwünscht ist. Die Umkehrung der unerwünschten Modifikationen ist nicht zwingend quantitativ und das Auftreten von Störsignalen im Massenpektrum kann hierdurch bedingt sein. Des Weiteren ist zu berücksichtigen, dass das Nebenprodukt dieser Reaktion saure Eigenschaften aufweist und dies bei nicht ausreichender Pufferung der Reaktionslösung zur Herabsetzung des pH-Wertes führen kann. Niedrige pH-Werte beschleunigen wiederum die unerwünschte Hydrolyse von Amidgruppen der Seitenkette von Proteinen, welches sich ebenfalls durch das Auftreten von Störsignalen bemerkbar machen kann.

Ähnlich wie im vorherigen Abschnitt wurden zunächst mögliche einflussnehmende Faktoren in den *Design-Wizard* eingetragen und der Versuchsplan generiert (siehe Tabelle 4.2 auf der vorherigen Seite). In der letzten Spalte dieser Tabelle sind die resultierenden Ergebnisse nach der praktischen Durchführung der Einzelversuche zu sehen. Nach Analyse der Ergebnisse mit Hilfe des *Analysis-Wizard* ergab sich die Übersichtsgrafik in der Abbildung 4.8.

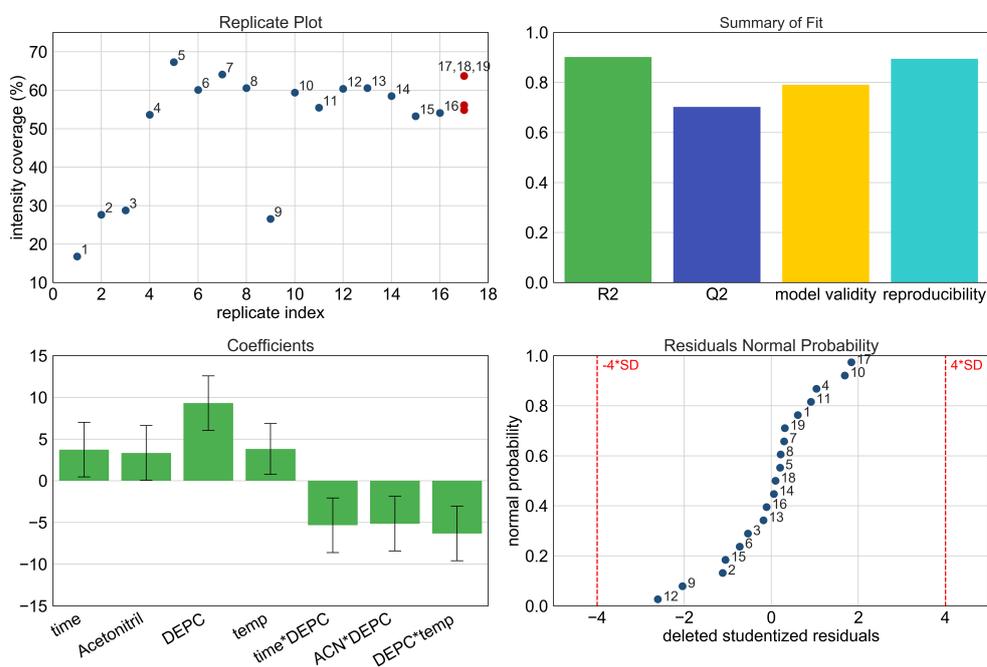


Abbildung 4.8: Übersicht über die Qualitätsmerkmale nach Optimierung der Reaktionsbedingungen der Carboethoxylierung der Amino-Gruppen

4. ERGEBNISSE UND DISKUSSION

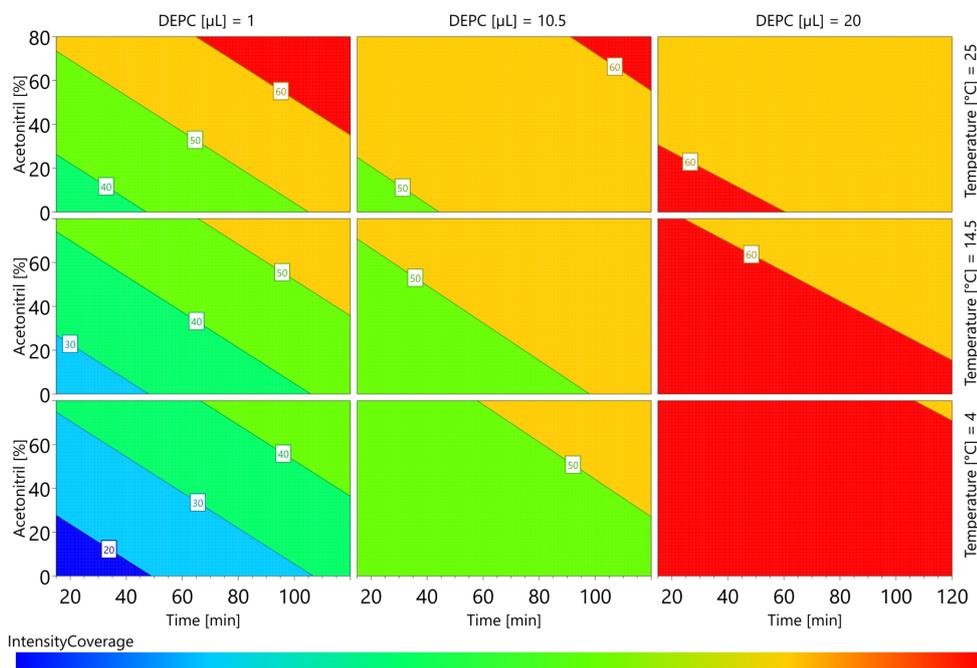


Abbildung 4.9: *Contour Plot* für die Optimierung der Reaktionsbedingungen der Carbethoxylierung der Aminogruppen

Das Modell zur Vorhersage der *Intensity Coverage* erfüllt alle Qualitätskriterien wie aus dem *Summary of Fit* im oberen rechten Teil der Übersichtsgrafik zu entnehmen ist. Die Residuen zeigen eine Normalverteilung und somit kann jedes Ergebnis zur Generierung eines vertrauenswürdigen Modells einbezogen werden (siehe unten rechts in der Übersichtsgrafik). Aus dem Konfidenzdiagramm im unteren linken Teil der Abbildung ist zu erkennen, dass vier Faktoren einen positiven Einfluss auf die *Intensity Coverage* haben: die Reaktionszeit, die Zugabe von Acetonitril, die eingesetzte Menge an Reagenz und die eingestellte Temperatur. Zusätzlich ist aus der Grafik ersichtlich, dass Interaktionen existieren, die zu einer Verschlechterung der *Intensity Coverage* führen. Hierzu gehören Interaktionen der eingesetzten Reagenzmenge mit der Reaktionszeit, der zugegebenen Menge Acetonitril und der Temperatur.

Dies bedeutet zunächst, dass die Effekte der Einzelfaktoren, die an sich positive Auswirkungen auf das Resultat haben, sich in ihrem Effekt umkehren, wenn sie zeitgleich erhöht werden. Dies liegt vermutlich daran, dass die Reagenzaktivität in einem derartigen Maß erhöht wird, dass Nebenreaktionen gehäuft vorkommen, die zu Störsignalen führen. Das gehäufte Aufkommen von Störsignalen erniedrigt wiederum die *Intensity Coverage*.

Durch die hohe Anzahl der Faktoren lassen sich die Abhängigkeiten der

4.2. Chemische Modifikation der Aminogruppen auf Proteinebene

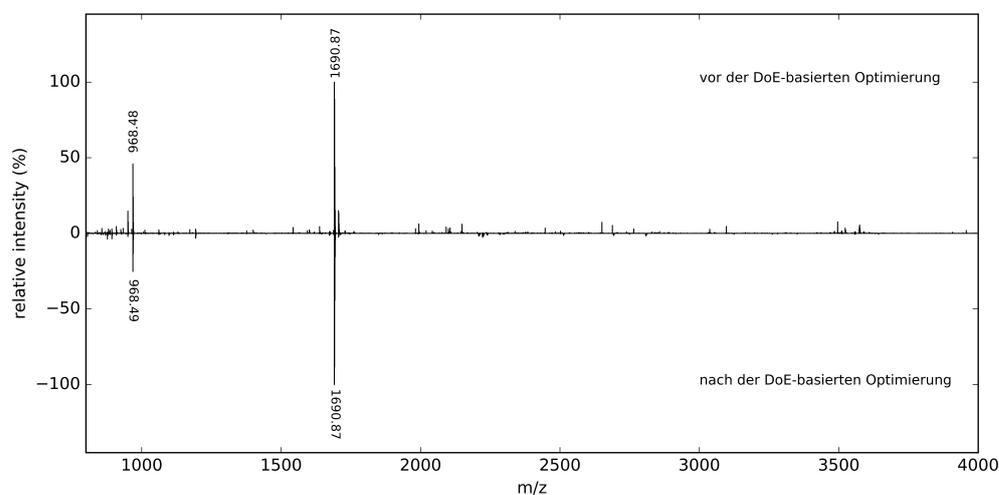


Abbildung 4.10: MALDI-MS-Spektren vor (oben) und nach (unten) Optimierung der Reaktionsbedingungen der Carbethoxylierung der Aminogruppen

Intensity Coverage von den Einflussfaktoren nicht durch einen zweidimensionalen *Contour Plot* visualisieren, wie es bei der reduktiven Methylierung der Fall war. Abhilfe schaffen hierbei vierdimensionale *Contour Plots*, die letztlich aus einzelnen 2D-*Contour Plots* mit fixen Werten für bestimmte Faktoren zusammengesetzt sind. Beispielsweise zeigt der erste *Contour Plot* in der Grafik 4.9 die Abhängigkeit der *Intensity Coverage* von der Reaktionszeit und der zugesetzten Menge Acetonitril bei 25°C und 1 µL DEPC-Zugabe.

Aus diesem Plot ist ersichtlich, dass eine niedrige Temperatur und ein relativ hoher Zusatz an DEPC zu Ergebnissen mit hoher *Intensity Coverages* führen, die weitgehend unabhängig von der Reaktionszeit und der zugesetzten Menge Acetonitril sind. Das Protokoll vor und nach Optimierung der Reaktionsbedingungen ist aus dem Methodenteil zu entnehmen (siehe Abschnitt 2.2.5). Erneut wurden beide Protokolle auf ADH angewendet und das Ergebnis ist in der Abbildung 4.10 dargestellt. Aus den Massenspektren ist ersichtlich, dass das Ausmaß an Nebenreaktionen reduziert wurde. Es sind weniger Störsignale zu sehen und die Basislinie weist ein geringeres Rauschen auf. Nachbarpeaks der Hauptsignale sind ein Indiz für Hydrolyseprodukte, deren Ausprägung ebenfalls durch die Optimierung der Reaktionsbedingungen verringert wurde.

Acetylierung

Acylierungen von Aminogruppen mit Hilfe von NHS-Ester erfreuen sich einer großen Beliebtheit in der Proteom-Forschung. Die dabei ablaufende Reaktion ist im Abschnitt 1.3.2 abgebildet und zeichnet sich durch ihren ho-

hen Umsetzungsgrad unter nativen Bedingungen aus. NHS-Ester mit einem Acetylrest weisen eine geringe Wasserlöslichkeit auf, die durch den Zusatz einer Sulfonsäuregruppe am N-Hydroxysuccinimid-Ring erhöht wurde, so dass auch Umsetzungen im wässrigen Milieu möglich sind.

Tabelle 4.3: Versuchsplan zur Optimierung der Acetylierung

| Nummer | DMSO (%) | Acetonitril (%) | IC (%) |
|--------|----------|-----------------|--------|
| 1 | 0 | 10 | 22,8 |
| 2 | 40 | 10 | 58,2 |
| 3 | 80 | 10 | 61,6 |
| 4 | 0 | 55 | 61,2 |
| 5 | 40 | 55 | 77,5 |
| 6 | 80 | 55 | 78,4 |
| 7 | 0 | 100 | 70,4 |
| 8 | 40 | 100 | 85,4 |
| 9 | 80 | 100 | 79,9 |
| 10 | 40 | 55 | 60,2 |
| 11 | 40 | 55 | 76,0 |
| 12 | 40 | 55 | 80,1 |

Nach Anwendung des vom Hersteller mitgelieferten Protokolls auf ADH wurden eine Vielzahl von Störsignalen detektiert, die auf einen unzureichenden Umsetzungsgrad hinweisen. Der generierte Versuchsplan mit den dazugehörigen Ergebnissen ist aus der Tabelle 4.3 zu entnehmen. Auch hier wurden alle Qualitätskriterien erfüllt, jedoch wie aus der Abbildung 4.11 ersichtlich ist, nicht in der Deutlichkeit wie bei den zuvor durchgeführten Ansätzen, so dass die Vorhersagekraft anhand des erstellten Modells angezweifelt werden kann. Aus der Übersichtsgrafik im unteren linken Teil kann zudem entnommen werden, dass die Menge an zugesetztem Reagenz und der Zusatz von DMSO kritische Faktoren für den Umsetzungsgrad dieser Reaktion darstellen. Im *Residual Plot* im unteren rechten Teil der Grafik ist zu sehen, dass die Residuen der Experimente mit der Nummer 1 und 9 eine starke Abweichung zeigen. Ein Ausschluss dieser Experimente führte jedoch nicht zu einer Verbesserung der Qualitätsmerkmale, so dass diese beibehalten wurden.

Aus dem *Contour Plot* in Abbildung 4.12 auf Seite 76 kann entnommen werden, dass ein hoher Zusatz von DMSO und ein hohe Konzentration des Acetylierungsreagenzes zu hohen Umsetzungsgraden führen. Die Angaben erfolgen in Prozent, wobei 100% äquivalent sind zu 1,8 mg/mL.

Trotz der schlechten Qualitätsmerkmale des mathematischen Modells konn-

4.2. Chemische Modifikation der Aminogruppen auf Proteinebene

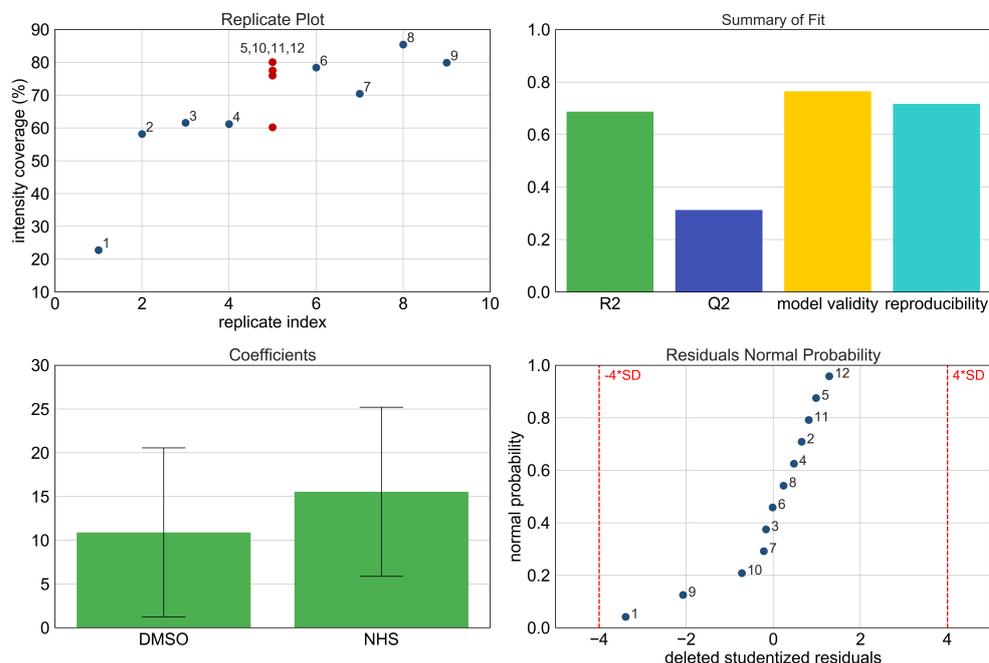


Abbildung 4.11: Übersicht über die Qualitätsmerkmale nach Optimierung der Reaktionsbedingungen der Acetylierung der Amino-Gruppen

ten mit Hilfe des *Contour Plots* Reaktionsbedingungen hervorgesagt werden, die zu einer deutlichen Verbesserung der Resultate geführt haben, wie aus den Massenspektren in Abbildung 4.13 ersichtlich ist. Das Massenspektrum nach Umsetzung unter optimierten Bedingungen weist lediglich die Signale auf, die nach einer vollständigen Umsetzung des Proteins zu erwarten sind. Die Störsignale des unteren Massenspektrums lassen sich Peptiden zuordnen, die nach einer unvollständigen Umsetzung des Proteins mit dem Acetylierungsreagenzes generiert werden.

Propionylierung

Modifizierungen mit Propionsäureanhydrid haben die gleiche Charakteristik wie die mit DEPC. Auch hier sind hohe Umsetzungen zu erwarten und das Auftreten von Störsignalen

lässt sich auf unerwünschte Seitenreaktionen und Hydrolyseprodukte zurückführen. Seitenreaktionen lassen sich auch hier durch den Zusatz von Hydroxylamin rückgängig machen und durch ausreichende Pufferung der Reaktionslösung wird die Bildung von Hydrolyseprodukten verhindert. Der Versuchsplan der Optimierung der Reaktionsbedingungen mit den dazugehörigen Ergebnissen ist in der Tabelle 4.4 abgebildet.

4. ERGEBNISSE UND DISKUSSION

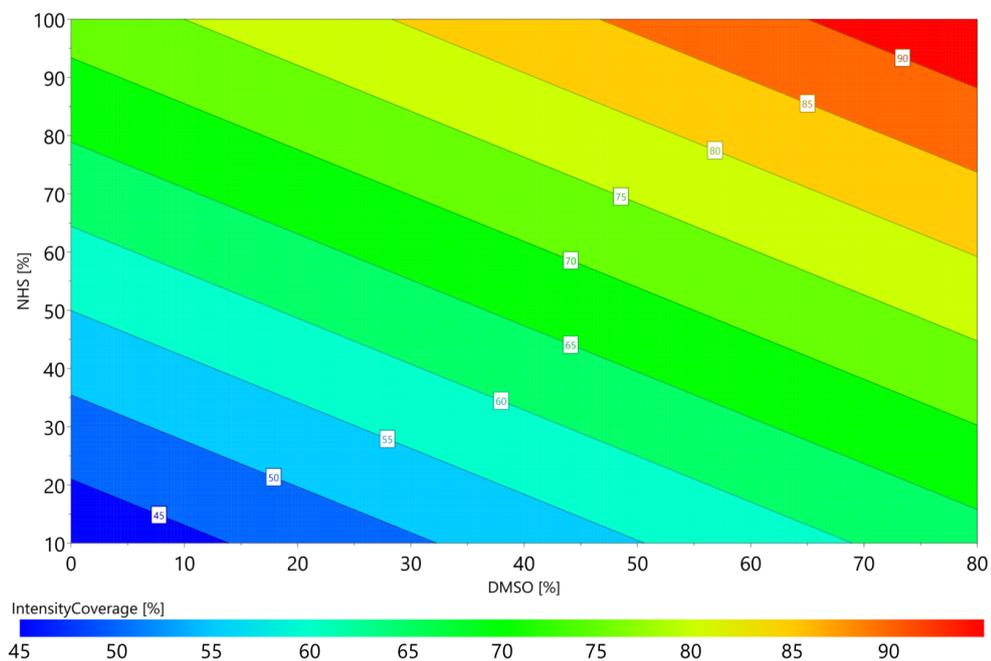


Abbildung 4.12: *Contour Plot* für die Optimierung der Reaktionsbedingungen der Acetylierung der Aminogruppen

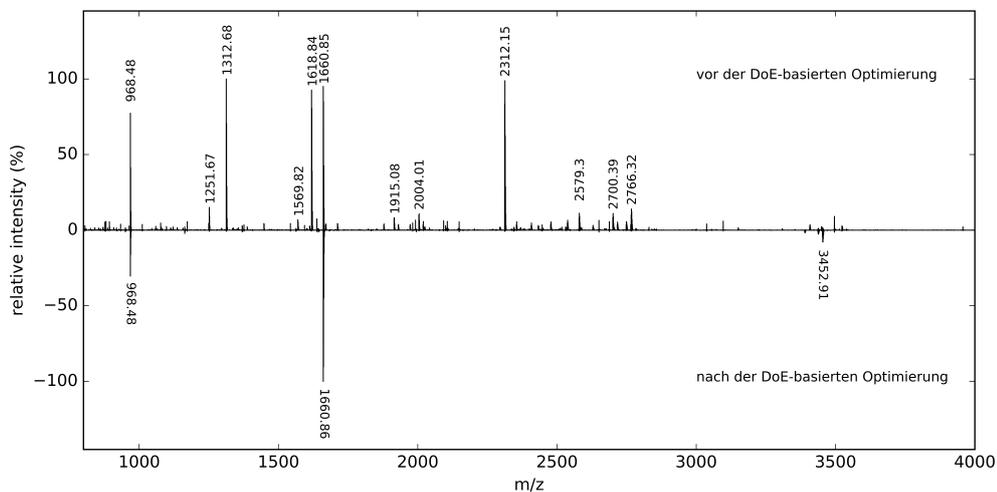


Abbildung 4.13: MALDI-MS-Spektren vor (oben) und nach (unten) Optimierung der Reaktionsbedingungen der Acetylierung der Aminogruppen

Zur Generierung eines vertrauenswürdigen Modells mussten die Experimente mit der Nummer 6 und 15 ausgeschlossen werden. Anhand des *Residual Plots* im unteren linken Teil der Abbildung 4.14 ist zu sehen, dass

4.2. Chemische Modifikation der Aminogruppen auf Proteinebene

Tabelle 4.4: Versuchsplan zur Optimierung der Propionylierung

| Nummer | Zeit (min) | Methanol (%) | PA (μL) | Temperatur ($^{\circ}\text{C}$) | IC (%) |
|--------|---------------|-----------------|-------------------------|--------------------------------------|-----------|
| 1 | 15 | 0 | 1 | 4 | 56,5 |
| 2 | 120 | 0 | 1 | 4 | 64,0 |
| 3 | 15 | 80 | 1 | 4 | 62,8 |
| 4 | 120 | 80 | 1 | 4 | 63,3 |
| 5 | 15 | 0 | 20 | 4 | 43,3 |
| 6 | 120 | 0 | 20 | 4 | 63,5 |
| 7 | 15 | 80 | 20 | 4 | 53,1 |
| 8 | 120 | 80 | 20 | 4 | 45,7 |
| 9 | 15 | 0 | 1 | 25 | 51,6 |
| 10 | 120 | 0 | 1 | 25 | 59,1 |
| 11 | 15 | 80 | 1 | 25 | 44,5 |
| 12 | 120 | 80 | 1 | 25 | 58,8 |
| 13 | 15 | 0 | 20 | 25 | 41,4 |
| 14 | 120 | 0 | 20 | 25 | 32,5 |
| 15 | 15 | 80 | 20 | 25 | 6,1 |
| 16 | 120 | 80 | 20 | 25 | 38,6 |
| 17 | 60 | 40 | 10 | 15 | 45,6 |
| 18 | 60 | 40 | 10 | 15 | 43,9 |
| 19 | 60 | 40 | 10 | 15 | 45,2 |

keine weiteren Ausreißer vorlagen und die verbleibenden Residuen normal verteilt vorliegen. Im unteren rechten Teil der Abbildung ist aus dem Koeffizientendigramm abzulesen, dass eine Erhöhung der Temperatur und der Konzentration des eingesetzten Reagenzes zu Resultaten mit niedriger *Intensity Coverage* führen. Der Faktor Zeit leistet isoliert betrachtet keinen signifikanten Beitrag, muss jedoch weiterhin berücksichtigt werden, da das Produkt aus Zeit und eingesetzter Menge an PA einen signifikanten Faktor darstellt. Ähnlich wie bei DEPC, muss auch hier die hohe Reaktivität des PA durch die Reaktionsbedingungen eingeschränkt werden.

Die geringe Modellvalidität im *Summary of Fit* im oberen rechten Teil der Abbildung ist verursacht durch die hohe Reproduzierbarkeit (letzter Balken), was gleichbedeutend ist mit einem sehr kleinen reinen Fehler. Der Modellfehler braucht somit nicht sehr ausgeprägt zu sein, um den reinen Fehler zu überragen und dies täuscht einen LOF (*Lack of Fit*) vor. Zieht man jedoch die Ausprägung der Gütefaktoren R^2 und Q^2 mit in Betracht, wird deutlich, dass das erstellte Modell zur Vorhersage herangezogen werden kann.

Da mehr als zwei Faktoren einen Einfluss auf die *Intensity Coverage* haben, ist es auch hier nicht möglich, einen zweidimensionalen *Contour Plot* aufzu-

4. ERGEBNISSE UND DISKUSSION

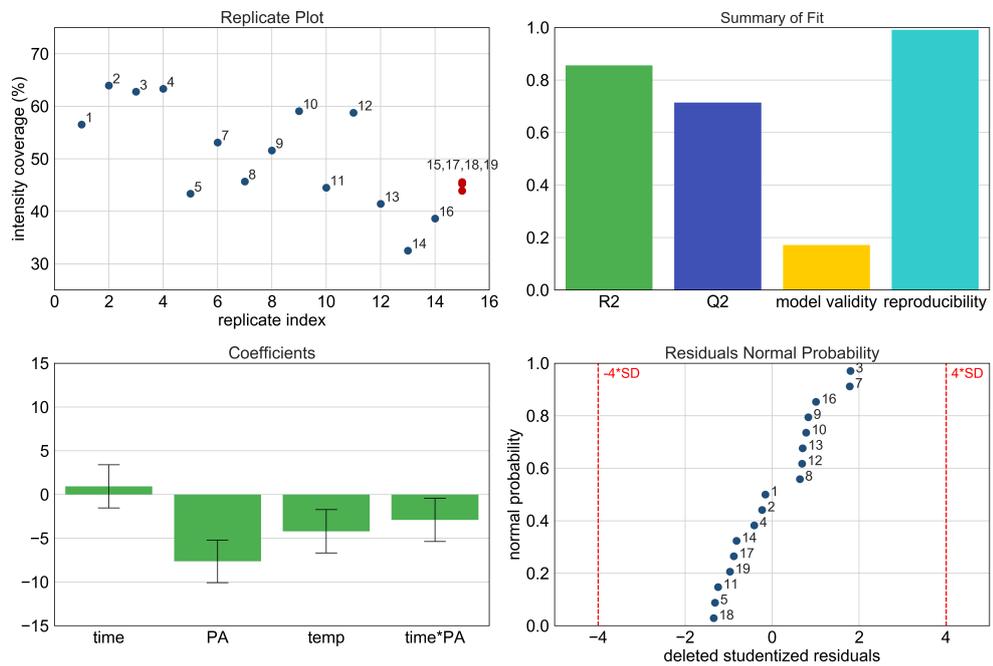


Abbildung 4.14: Übersicht über die Qualitätsmerkmale nach Optimierung der Reaktionsbedingungen der Propionylierung der Amino-Gruppen

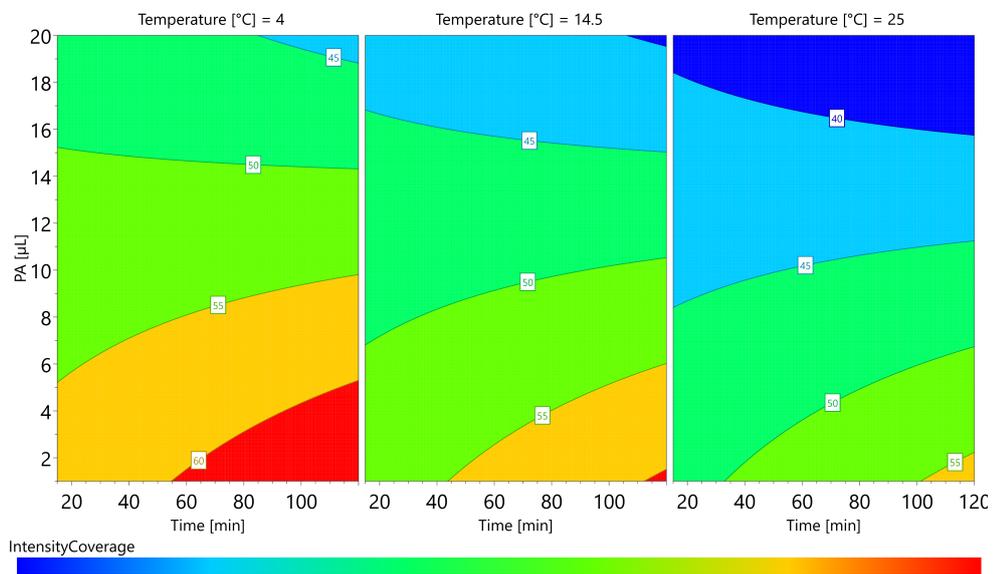


Abbildung 4.15: *Contour Plot* für die Optimierung der Reaktionsbedingungen der Propionylierung der Aminogruppen

4.2. Chemische Modifikation der Aminogruppen auf Proteinebene

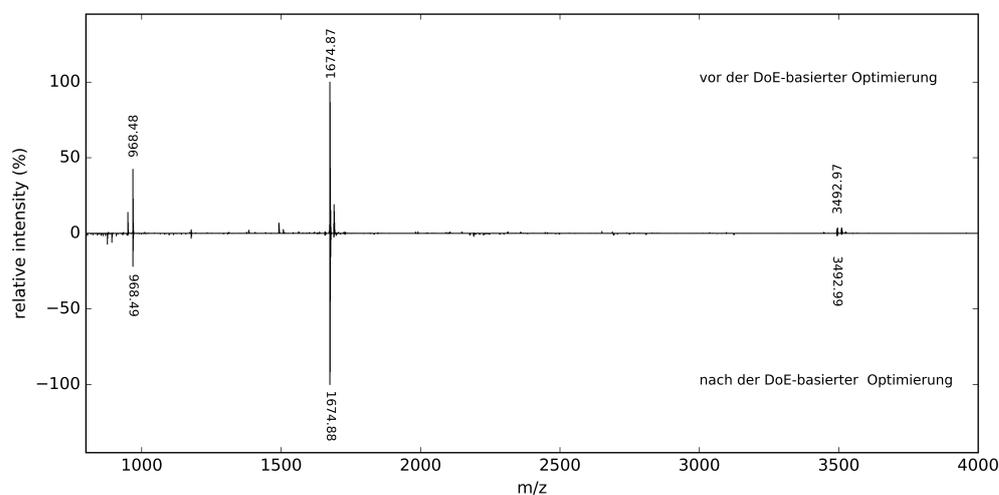


Abbildung 4.16: MALDI-MS-Spektren vor (oben) und nach (unten) Optimierung der Reaktionsbedingungen der Propionylierung der Aminogruppen

zeigen. In dem 4D-Contour Plot in der Abbildung 4.15 ist die Abhängigkeit der *Intensity Coverage* von der zugesetzten Menge PA und der Reaktionszeit jeweils für die untersuchte Reaktionstemperaturen abgebildet. Aus diesem Contour Plot ist ersichtlich, dass geringe Temperaturen, lange Reaktionszeiten und geringe Mengen eingesetzten Reagenzes zu hohen Umsetzungsgraden führen.

Nach Betrachtung der Massenspektren in der Abbildung 4.16 sind keine großen Unterschiede vor und nach Optimierung der Reaktionsbedingungen zu erkennen. Allerdings ist eine Verbesserung bezüglich des Auftretens von Seitenreaktionen zu beobachten. Störsignale, die in unmittelbarer Nachbarschaft zu den Hauptsignalen liegen, sind nach der Optimierung der Reaktionsbedingungen nicht so stark ausgeprägt, was sich wiederum durch eine höhere *Intensity Coverage* äußert.

4.2.3 Diskussion

Die in der Zielsetzung formulierte Etablierung eines ArgC-ähnlichen Verdauens, erfordert eine chemische Modifizierung der Aminogruppen der Proteine bevor diese mit Trypsin verdaut werden. Es wurden in diesem Abschnitt vier häufig in der Proteinchemie genutzte Reagenzien zur Derivatisierung von Aminogruppen auf ihren Umsetzungsgrad hin untersucht und es stellte sich nach Anwendung der publizierten Protokolle heraus, dass diese zum Teil zu unbefriedigenden Ergebnissen führten. Dies lag hauptsächlich daran, dass diese Protokolle für die Modifizierung von Peptiden oder nur kleinen Proteinen optimiert wurden. Liegen solche Proben vor, sind die Re-

4. ERGEBNISSE UND DISKUSSION

sultate - wie es im Abschnitt 4.2.1 gezeigt wurde - sehr zufriedenstellend. In diesem Abschnitt wurden die Reaktionsbedingungen der einzelnen Protokolle auf ein möglichst hohen Umsetzungsgrad für Proteine optimiert. Die Optimierung verlief DOE-basiert und die Resultate waren in allen Fällen zufriedenstellend. Nach Vergleich der Massenspektren vor und nach Optimierung der Reaktionsbedingungen waren Verbesserungen entweder in Bezug auf den Umsetzungsgrad oder auf das geringere Vorkommen von Störsignalen sichtbar.

4.3 Vergleich der Modifizierungsprotokolle auf Protein- und Proteomebene

Im vorherigen Abschnitt wurden die Reaktionsbedingungen für die verschiedenen Reagenzien bezüglich ihres Umsetzungsgrades auf Proteinebene DOE-basiert optimiert. In diesem Abschnitt gilt es nun zu überprüfen, welches der vier Reagenzien zu den besten Resultaten für Proteomproben führt. Als Referenz dienen konventionelle ArgC- und Trypsin-Verdaue, so dass insgesamt sechs verschiedene Ansätze miteinander verglichen werden (siehe Abbildung 4.17).

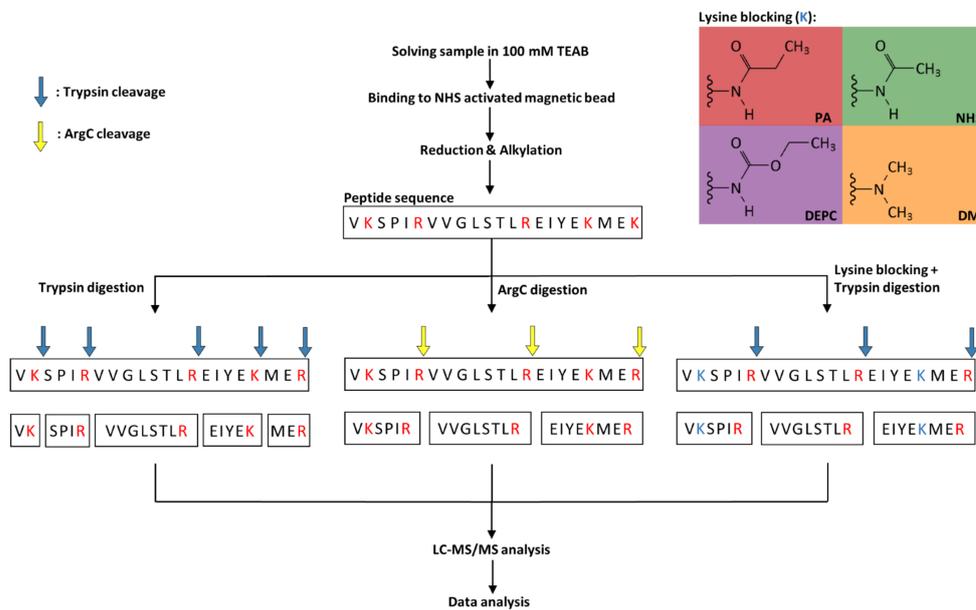


Abbildung 4.17: Übersicht über die verschiedenen Verdaustrategien.

Den Ansätzen ist gemein, dass die Proteine zunächst an magnetische Beads gebunden werden, bevor sie reduziert und alkyliert werden. Bei den konventionellen Strategien wird die Protease zugegeben und der Verdau erfolgt auf den Beads, wohingegen beim ArgC-ähnlichen Verdau (in der Abbildung ganz rechts) die Aminogruppen mit den verschiedenen Reagenzien zunächst modifiziert werden, bevor der Verdau auf den Beads durch den Zusatz von Trypsin erfolgen kann. Aus textökonomischen Gründen werden die ArgC-ähnlichen Verdaue mit dem Kürzel des eingesetzten Reagenzes abgekürzt. So steht beispielsweise PA-Protokoll für dasjenige Protokoll, welches die Aminogruppen der Proteine mit PA modifiziert, bevor diese anschließend mit Trypsin verdaut werden.

Der Vergleich der ArgC-ähnlichen Ansätze mit den konventionellen Verdaustrategien erfolgte zunächst nur auf Proteinebene. Durch die hohe Komplexität der Proteomproben lassen sich die Ergebnisse, die mit einem einzelnen Protein erzielt wurden, nicht verallgemeinern. Aus diesem Grund erfolgt im zweiten Schritt der Vergleich auf Proteomebene, so dass allgemeingültige Aussagen gemacht werden können.

4.3.1 Vergleich auf Proteinebene

Der Vergleich der Protokolle auf Proteinebene erfolgt mit Hilfe des Modellproteins Serotransferrin, welches eine sehr hohe Zahl an Lysinen hat und somit die Möglichkeit bietet, den Umsetzungsgrad der Modifizierungsprotokolle unter dem Einfluss verschiedener benachbarter Aminosäuren zu untersuchen.

Die Massenspektren wurden am 4800 MALDI-TOF/TOFTM Analyser (AB-Sciex) (siehe Abschnitt 2.1.4) aufgenommen. Von jedem Protokoll wurden sechs Replikate angefertigt, die acht Mal gespottet wurden. Die acht Massenspektren von jedem Replikat wurden aufsummiert und deren Signalintensitäten wurde auf den Basispeak normalisiert. Die Zuweisungen der Signale zu den Peptiden erfolgte mit Hilfe einer PMF-Datenbanksuche mit der Suchmaschine Mascot. Die Parameter wurden so gewählt, dass die höchste Enzymspezifität vorausgesetzt wurde, in dem die Anzahl der Fehlschnittstellen auf null gesetzt wurde (siehe Anhang C).

Die Massenspektren lassen sich auf zwei unterschiedlichen Wegen miteinander vergleichen: visuell und mit Hilfe von quantitativen Messgrößen. Wichtige visuelle Parameter sind die Verteilung der Intensitäten in Abhängigkeit vom Massenbereich, die Anzahl nicht annotierter Peaks und die Interpretierbarkeit von Signalen hoher Intensität, d.h. können diese Signale einem Peptid zugewiesen werden.

Visuelle Evaluierung

Die aufgenommenen Massenspektren mit den Resultaten der PMF-Datenbanksuche befinden sich gebündelt am Ende dieses Unterabschnittes. Signale, die nach der Datenbanksuche einem Peptid zugewiesen werden konnten, sind im Spektrum annotiert und nummeriert. In der zugehörigen Tabelle sieht man eine Auflistung der Peptidsequenzen mit deren Masse und Intensitäten. Eine mehrfache Auflistung einzelner Peptide ist verursacht durch das Auftauchen von Oxidationen am Methionin.

Das Massenspektrum des PA-Protokolls (siehe Abbildung 4.18) weist eine gleichmäßige Intensitätsverteilung und nur eine geringe Anzahl nicht annotierter Peaks auf und zudem sind die Signale hoher Intensität alle erklärbar, so dass dieses Massenspektrum im Vergleich zu den anderen auf-

4.3. Vergleich der Modifizierungsprotokolle auf Protein- und Proteomebene

genommenen Spektren den höchsten Qualitätsstandard erreicht. Das Massenspektrum für das NHS-Protokoll (siehe Abbildung 4.19) weist eine nicht so gleichmäßige Verteilung der Intensitäten auf. Des Weiteren sind Signale hoher Intensität zu sehen, die nach der Datenbanksuche keinem Peptid zugewiesen werden konnten. Das Massenspektrum des DEPC-Protokolls (siehe Abbildung 4.20 auf Seite 86) ist qualitativ noch niedriger einzustufen als das vom NHS-Protokoll, da es eine ungleichmäßigere Verteilung der Intensitäten und eine höhere Anzahl nicht annotierter Peaks aufweist. Das DM-Protokoll (siehe Abbildung 4.21 auf Seite 87) liefert ein qualitativ gleichwertig gutes Massenspektrum wie das PA-Protokoll mit dem Unterschied, dass das Intensitätsmaximum in Richtung höherer Massen verschoben ist. Durch die reduktive Methylierung erhöht sich die Basizität und somit augenscheinlich auch die Ionisierbarkeit der Peptide, die Lysine enthalten. Das Massenspektrum des ArgC-Protokoll (siehe Abbildung 4.22) ist als qualitativ niedrigstes Spektrum einzustufen, da es eine hohe Anzahl an Signalen hoher Intensität aufweist, die von der Suchmaschine keinem Peptid zugeordnet werden konnten.

4. ERGEBNISSE UND DISKUSSION

Tabelle 4.5: Peptidliste (PA-Protokoll)

| Peak-Nummer | Peptid-Sequenz | m/z-Verhältnis | Intensität |
|-------------|-----------------------------|----------------|------------|
| 1 | VPDKTVR | 926,6001 | 922 |
| 2 | APNHAVVTR | 964,6081 | 183 |
| 3 | SETKDLLFR | 1164,7066 | 530 |
| 4 | GKKSCHTGLGR | 1312,7606 | 2207 |
| 5 | DDTVCLAKLHDR | 1498,8161 | 528 |
| 6 | KPVEEYANCHLAR | 1642,8874 | 6069 |
| 7 | KDKEACVHKILR | 1665,0007 | 7570 |
| 8 | KCSTSSLLEACTFR | 1715,8947 | 4586 |
| 9 | MDAKMYLGYEYVTAIR | 1980,0505 | 3556 |
| 10 | MDAKMYLGYEYVTAIR | 1996,0336 | 1534 |
| 11 | MDAKMYLGYEYVTAIR | 2012,0316 | 278 |
| 12 | WCAVSEHEATKCQSFR | 2051,9961 | 3922 |
| 13 | SAGWNIPIGLLYCDLPEPR | 2171,1832 | 1125 |
| 14 | TAGWNIPMGLLYNKINHCR | 2314,2555 | 235 |
| 15 | TAGWNIPMGLLYNKINHCR | 2330,241 | 298 |
| 16 | NTYEKYLGEYVKAAGNLR | 2358,2874 | 4273 |
| 17 | KPVDEYKDCHLAQVPSHTVVAR | 2661,4297 | 3674 |
| 18 | EGTCPEAPTDECKPVKWCALSHHER | 3106,4585 | 1515 |
| 19 | DHMKSVIPSDGSPVACVKKASYLDCIR | 3201,6408 | 158 |
| 20 | DHMKSVIPSDGSPVACVKKASYLDCIR | 3217,6311 | 56 |

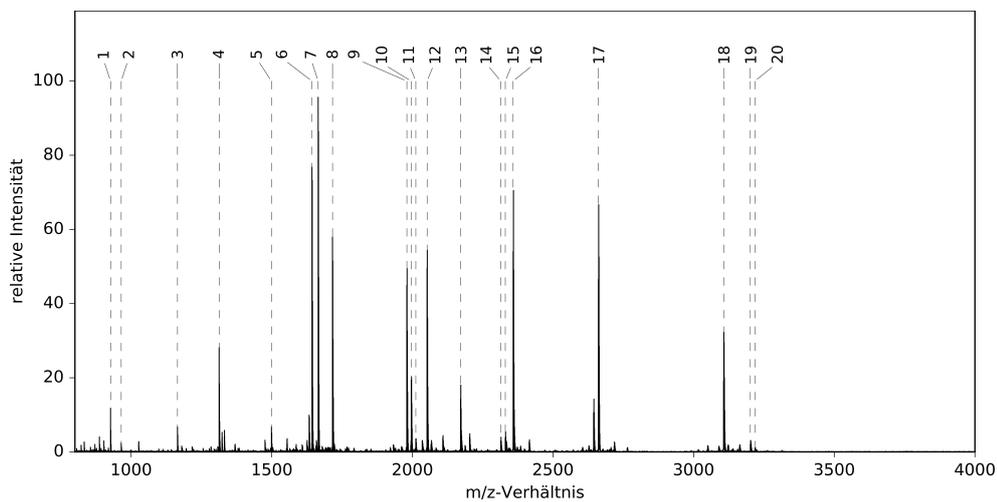


Abbildung 4.18: MALDI-MS-Spektrum von Serotransferrin nach Propionylierung

4.3. Vergleich der Modifizierungsprotokolle auf Protein- und Proteomebene

Tabelle 4.6: Peptidliste (NHS-Protokoll)

| Peak-Nummer | Peptid-Sequenz | m/z-Verhältnis | Intensität |
|-------------|-----------------------------|----------------|------------|
| 1 | VPDKTVR | 898,5545 | 1433 |
| 2 | APNHAVVTR | 964,5942 | 397 |
| 3 | SETKDLLFR | 1150,6701 | 557 |
| 4 | GKKSCHTGLGR | 1284,7144 | 2087 |
| 5 | DDTVCLAKLHDR | 1484,7811 | 462 |
| 6 | DQYELLCLDNTR | 1539,7428 | 338 |
| 7 | KDKEACVHKILR | 1622,9365 | 12311 |
| 8 | KPVEEYANCHLAR | 1628,8577 | 11830 |
| 9 | KCSTSSLLEACTFR | 1701,8636 | 8607 |
| 10 | MDAKMYLGYEYVTAIR | 1966,0199 | 3845 |
| 11 | MDAKMYLGYEYVTAIR | 1981,9982 | 2319 |
| 12 | MDAKMYLGYEYVTAIR | 1997,987 | 1004 |
| 13 | WCAVSEHEATKCQSFR | 2037,9576 | 3525 |
| 14 | SAGWNIPIGLLYCDLPEPR | 2171,1575 | 231 |
| 15 | TAGWNIPMGLLYNKINHCR | 2300,2093 | 134 |
| 16 | TAGWNIPMGLLYNKINHCR | 2316,2064 | 194 |
| 17 | NTYEKYLGEYVKAAGNLR | 2330,2334 | 3580 |
| 18 | KPVDEYKDCHLAQVPSHTVVAR | 2633,3711 | 5027 |
| 19 | EGTCPEAPTDECKPVKWCALSHHER | 3078,3983 | 2058 |
| 20 | DHMKSVIPSDGPSVACVKKASYLDCIR | 3159,5637 | 113 |

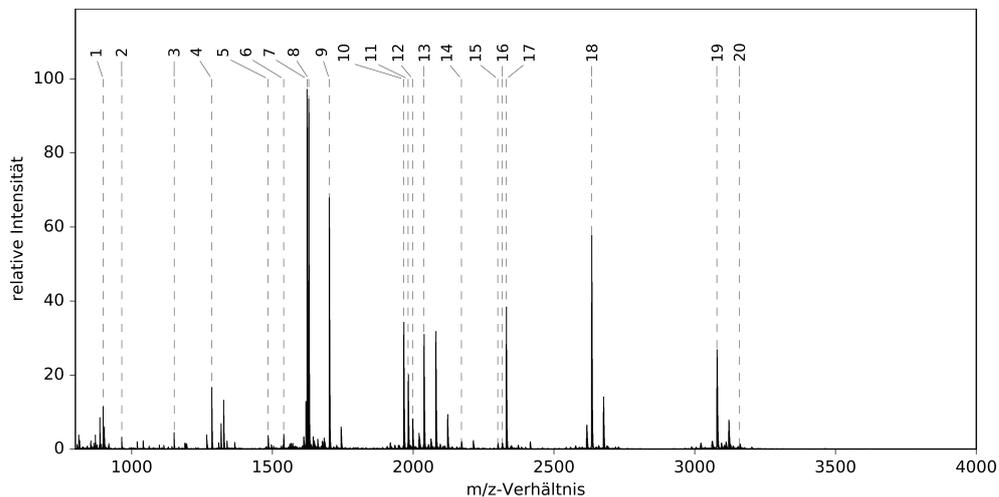


Abbildung 4.19: MALDI-MS-Spektrum von Serotransferrin nach Acetylierung

4. ERGEBNISSE UND DISKUSSION

Tabelle 4.7: Peptidliste (DEPC-Protokoll)

| Peak- Nummer | Peptid-Sequenz | m/z- Verhältnis | Inten- sität |
|-----------------|-----------------------------|--------------------|-----------------|
| 1 | VPDKTVR | 958,5591 | 2332 |
| 2 | APNHAVVTR | 964,571 | 143 |
| 3 | SETKDLLFR | 1180,6631 | 680 |
| 4 | GKKSCHTGLGR | 1344,7142 | 1876 |
| 5 | DDTVCLAKLHDR | 1514,7711 | 344 |
| 6 | DQYELLCLDNTR | 1539,7683 | 41 |
| 7 | KPVEEYANCHLAR | 1658,8389 | 3503 |
| 8 | KDKEACVHKILR | 1712,9414 | 6852 |
| 9 | KCSTSSLLEACTFR | 1731,8433 | 3880 |
| 10 | MDAKMYLGYEYVTAIR | 1995,9986 | 1167 |
| 11 | MDAKMYLGYEYVTAIR | 2011,9846 | 679 |
| 12 | MDAKMYLGYEYVTAIR | 2027,9742 | 168 |
| 13 | WCAVSEHEATKQCQSF | 2067,9512 | 1765 |
| 14 | SAGWNIPIGLLYCDLPEPR | 2171,1307 | 1780 |
| 15 | TAGWNIPMGLLYNKINHCR | 2330,1878 | 150 |
| 16 | TAGWNIPMGLLYNKINHCR | 2346,177 | 231 |
| 17 | NTYEKYLGEYVKAIGNLR | 2390,2189 | 1382 |
| 18 | KPVDEYKDCHLAQVPSHTVVAR | 2693,3584 | 1176 |
| 19 | EGTCPEAPTDECKPVKWCALSHHER | 3138,3903 | 281 |
| 20 | DHMKSVIPSDGPSVACVKKASYLDCIR | 3249,5485 | 50 |

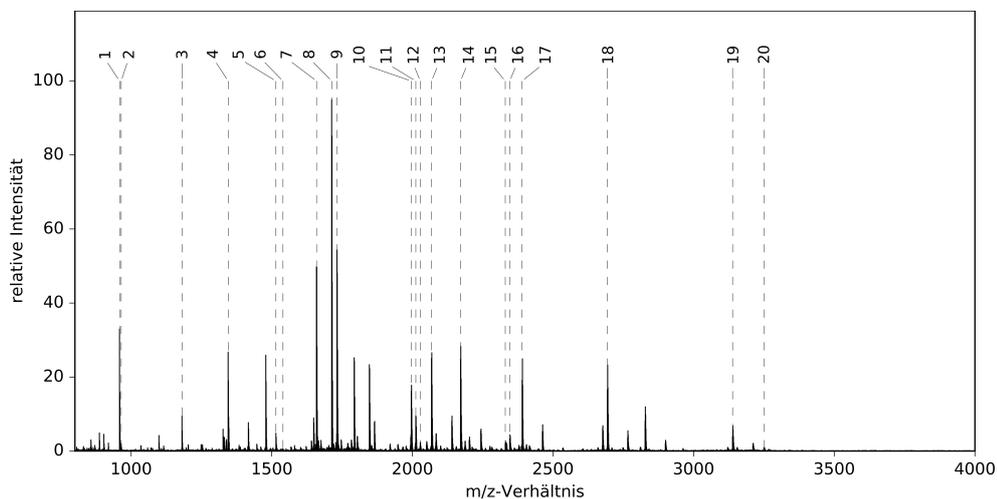


Abbildung 4.20: MALDI-MS-Spektrum von Serotransferrin nach Carbethoxylierung

4.3. Vergleich der Modifizierungsprotokolle auf Protein- und Proteomebene

Tabelle 4.8: Peptidliste DM-Protokoll

| Peak-Nummer | Peptid-Sequenz | m/z-Verhältnis | Intensität |
|-------------|-----------------------------|----------------|------------|
| 1 | APNHAVVTR | 964,5657 | 201 |
| 2 | SETKDLLFR | 1136,6677 | 521 |
| 3 | GKKSCHTGLGR | 1256,7295 | 349 |
| 4 | DDTVCLAKLHDR | 1470,7746 | 399 |
| 5 | DQYELLCLDNTR | 1539,7511 | 737 |
| 6 | KDKEACVHKILR | 1580,9696 | 851 |
| 7 | KPVEEYANCHLAR | 1614,8436 | 4040 |
| 8 | KCSTSSLLEACTFR | 1687,8498 | 3342 |
| 9 | MDAKMYLGYEYVTAIR | 1952,0003 | 5840 |
| 10 | MDAKMYLGYEYVTAIR | 1967,9868 | 1027 |
| 11 | MDAKMYLGYEYVTAIR | 1983,9835 | 179 |
| 12 | WCAVSEHEATKCQSFR | 2023,9458 | 3410 |
| 13 | SAGWNIPIGLLYCDLPEPR | 2171,1295 | 4712 |
| 14 | TAGWNIPMGLLYNKINHCR | 2286,1935 | 2105 |
| 15 | NTYEKYLGEYVKAAGNLR | 2302,2329 | 5894 |
| 16 | KPVDEYKDCHLAQVPSHTVVAR | 2605,3795 | 2544 |
| 17 | EGTCPEAPTDECKPVKWCALSHHER | 3050,4031 | 1404 |
| 18 | DHMKSVIPSDGPSVACVKKASYLDCIR | 3117,5979 | 541 |
| 19 | DHMKSVIPSDGPSVACVKKASYLDCIR | 3133,5674 | 54 |

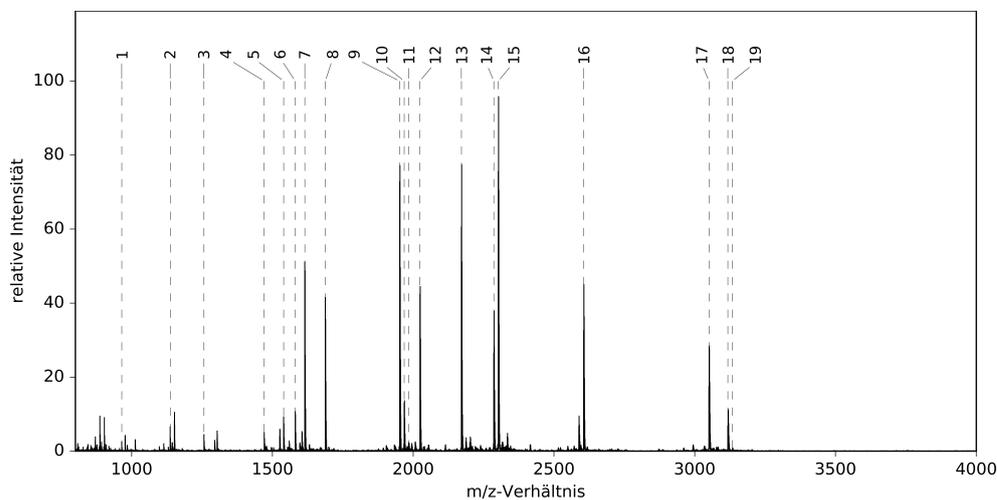


Abbildung 4.21: MALDI-MS-Spektrum von Serotransferrin nach reduktiver Methylierung

4. ERGEBNISSE UND DISKUSSION

Tabelle 4.9: Peptidliste ArgC-Protokoll

| Peak- Nummer | Peptid-Sequenz | m/z- Verhältnis | Inten- sität |
|-----------------|---------------------------|--------------------|-----------------|
| 1 | APNHAVVTR | 964,6007 | 766 |
| 2 | SETKDLLFR | 1108,6724 | 571 |
| 3 | DDTVCLAKLHDR | 1442,7863 | 181 |
| 4 | DQYELLCLDNTR | 1539,7959 | 12680 |
| 5 | KPVEEYANCHLAR | 1586,8598 | 21781 |
| 6 | KCSTSSLLEACTFR | 1659,8688 | 9800 |
| 7 | MDAKMYLGYEYVTAIR | 1940,0772 | 140 |
| 8 | WCAVSEHEATKCQSFR | 1995,9772 | 274 |
| 9 | SAGWNIPIGLLYCDLPEPR | 2171,1851 | 317 |
| 10 | TAGWNIPMGLLYNKINHCR | 2258,0641 | 897 |
| 11 | KPVDEYKDCHLAQVPSHTVVAR | 2549,3792 | 7776 |
| 12 | EGTCPEAPTDECKPVKWCALSHHER | 2994,4049 | 1315 |

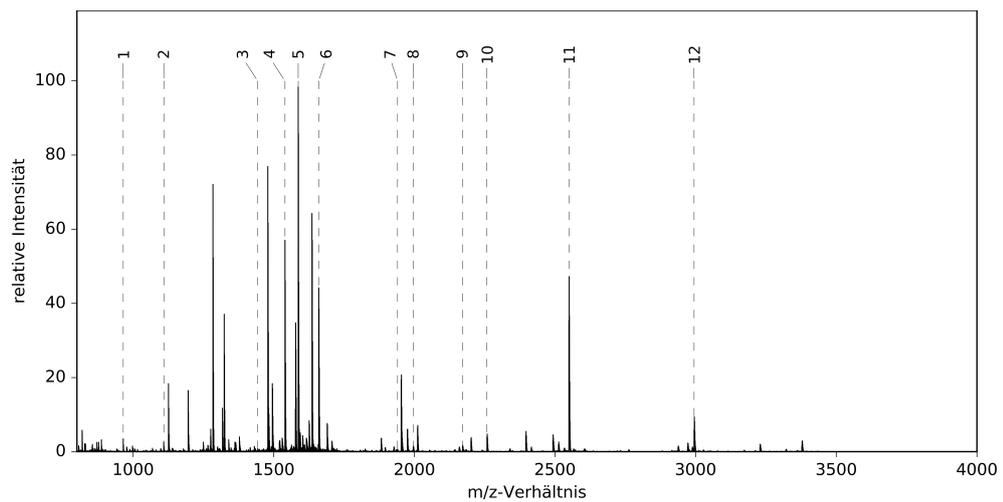


Abbildung 4.22: MALDI-MS-Spektrum nach ArgC-Verdau von Serotransferrin

Evaluierung anhand quantitativer Messgrößen

Visuelle Evaluierungen unterliegen einer gewissen Subjektivität, welche bei der Heranziehung quantitativer Messgrößen nicht gegeben ist. Im Folgenden werden die zuvor gezeigten Spektren anhand drei quantitativer Größen charakterisiert und miteinander verglichen. Dazu gehören die Anzahl zugewiesener Peptide, die Sequenzabdeckung und die *Intensity Coverage*. Trotz der starken Korrelation zwischen der Anzahl zugewiesener Peptide und der Sequenzabdeckung, ist es sinnvoll, die beiden Größen separat aufzuzeigen. Längere Peptide führen nämlich zu höheren Sequenzabdeckungen und die Länge der Peptide ist aus der bloßen Angabe der Peptidanzahl nicht erkennbar.

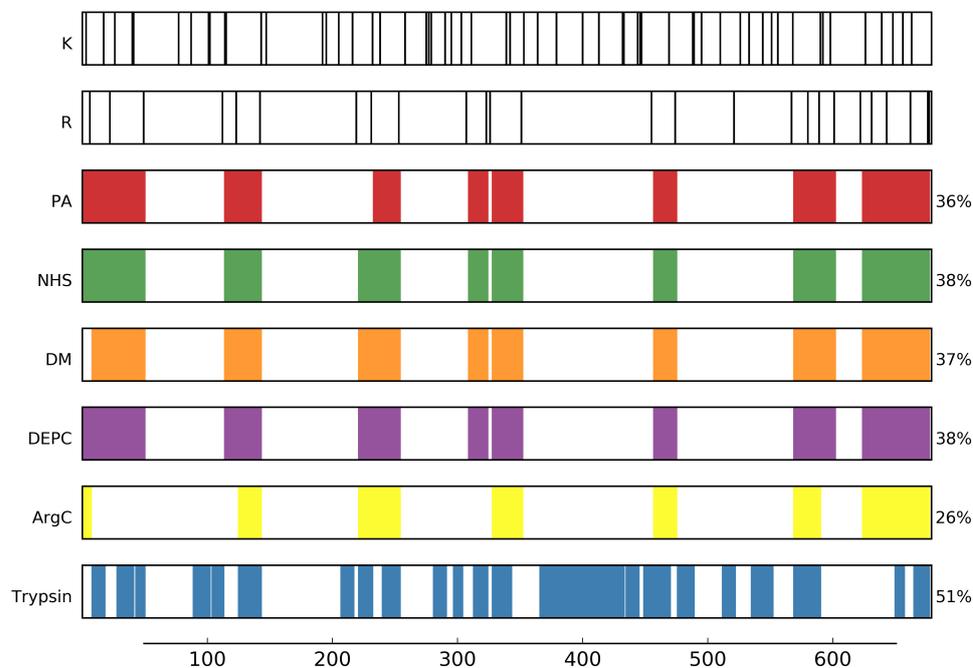


Abbildung 4.23: Lage der tryptischen Schnittstellen und Übersicht über die detektierten Sequenzteile des Serotransferrin nach Anwendung der verschiedenen Protokolle.

In der Abbildung 4.24 auf Seite 91 ist jeweils für die aufgezählten Größen ein Balkendiagramm mit den ermittelten Werten zu sehen. Aus dem Balkendiagramm für die Sequenzabdeckung (in der Mitte der Abbildung) ist zu erkennen, dass die ArgC-ähnlichen Protokolle bis auf das PA-Protokoll zu gleichen Werten führen. Das PA-Protokoll weist eine verringerte Sequenzabdeckung auf, da es im Gegensatz zu den anderen Protokollen kein Signal aufweist, das sich zu dem Peptid mit der Sequenz DQYELLCLDNTR zuord-

nen lässt. Das Fehlen dieses Peptides erklärt auch die verringerte Anzahl an zugewiesenen Peptiden im oberen Balkendiagramm. Auffällig in beiden Balkendiagrammen, ist dass ArgC gegenüber jedem anderen Protokoll schlechtere Werte aufweist.

Die hohe Anzahl der zugewiesenen Peptide und die hohe Sequenzabdeckung des Trypsin-Protokolls erklärt sich durch die Verteilung der Schnittstellen, die zu einer höheren Anzahl von Peptiden führen, die sich im gemessenen Massenbereich detektieren lassen. Die Verteilung der Schnittstellen ist in der Abbildung 4.23 auf der vorherigen Seite zu sehen. Die ersten zwei Balken zeigen jeweils die Lage von Argininen und Lysin innerhalb der Sequenz des Serotransferrins. Jeder weitere Balken visualisiert die Sequenzabdeckung für das jeweilige gekennzeichnete Protokoll. Auch hier ist die hohe Ähnlichkeit der Resultate zwischen den ArgC-ähnlichen Protokollen erkennbar. Die großen Lücken bei den ArgC-ähnlichen Protokollen erklären sich durch das Fehlen von Argininen in diesen Sequenzteilen, welches zur Generierung sehr langer Peptide führen, welche außerhalb des gemessenen Massenbereichs liegen.

Im dritten Balkendiagramm der Abbildung 4.24 auf der nächsten Seite ist die berechnete *Intensity Coverage* für das jeweilige Protokoll zu sehen. Die quantitativen Werte stimmen mit den Ergebnissen der zuvor durchgeführten visuellen Auswertung überein. Das PA-Protokoll weist hier die höchsten Werte auf, welche sich durch die geringe Anzahl nicht annotierter Peaks ergeben. Bemerkenswerterweise weist selbst das Trypsin-Protokoll geringere Werte auf, bedingt durch das höhere Vorkommen von Peptiden mit Fehlschnittstellen, die nicht berücksichtigt wurden, was die Anzahl nicht annotierter Peaks erhöht. Auch hier weist ArgC die geringsten Werte. Das generell schlechte Abschneiden von ArgC ist verursacht durch seine geringe Enzymspezifität.

4.3.2 Vergleich auf Proteomebene

Die Vergleichsmöglichkeiten nach Anwendung der verschiedenen Protokolle auf Proteomebene sind vielfältiger und zudem noch aussagekräftiger als diejenigen auf der Ebene eines einzelnen Proteins. Dies liegt an den unterschiedlichen Aufbau der Proteine, was verallgemeinerte Aussagen ausgehend von den Ergebnissen eines einzelnen Proteins nicht ermöglicht.

Für die Untersuchung wurde ein Zelllysate von *Escherichia coli* verwendet, da dessen Proteom sehr gut untersucht ist und Proteindatenbanken hoher Qualität vorliegen. Des Weiteren eignen sich im Allgemeinen Proteome von Prokaryonten für Evaluierungen dieser Art, da nur wenige posttranslationale Modifikationen und Isoformen vorliegen, was die datenanalytische Auswertung im Anschluss an der Messung stark vereinfacht.

4.3. Vergleich der Modifizierungsprotokolle auf Protein- und Proteomebene

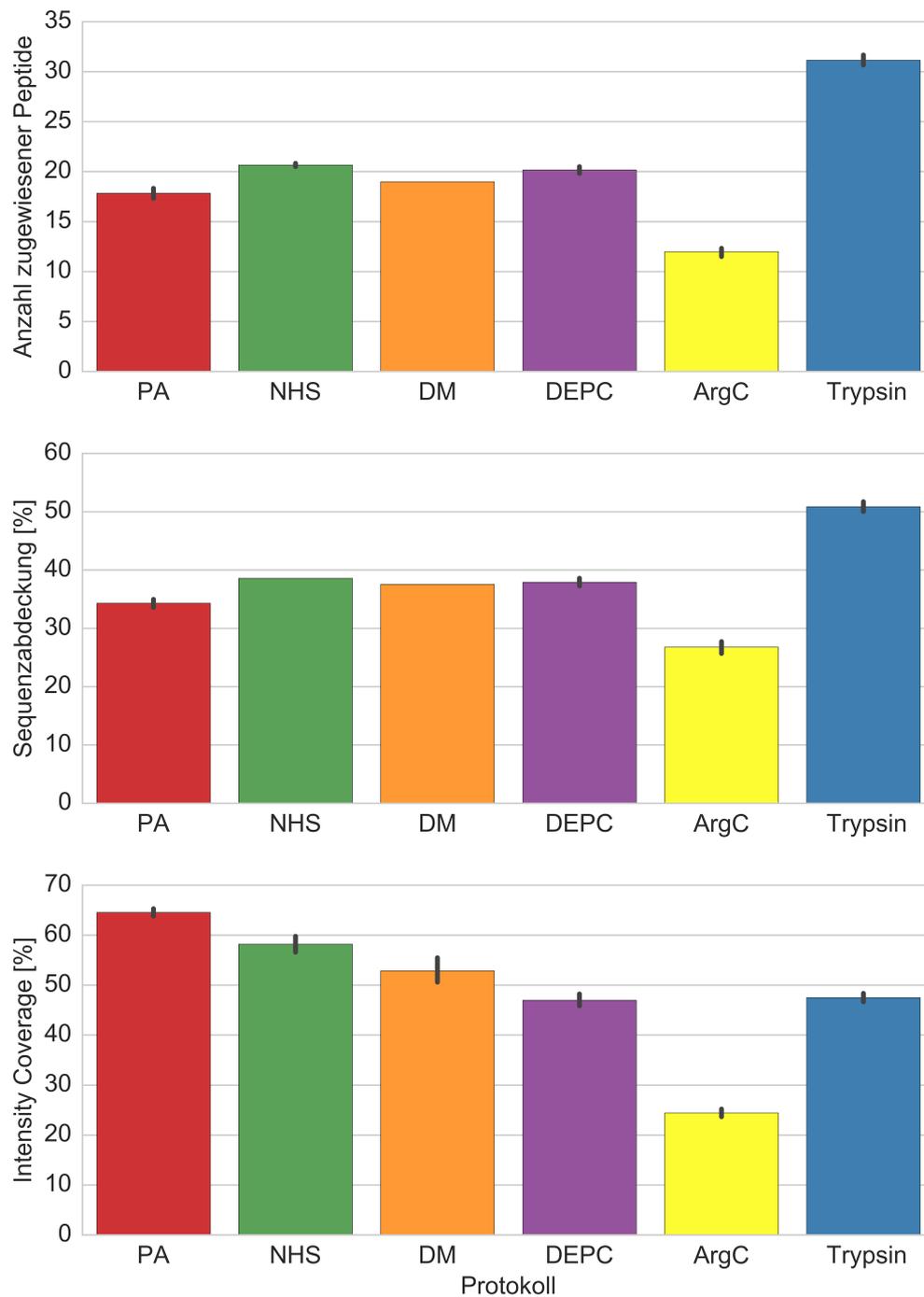


Abbildung 4.24: Übersicht verschiedener Kriterien zum Vergleich der Massenspektren von unterschiedlich modifiziertem Serotransferrin mit konventionellen Verdaustrategien (ArgC und Trypsin).

Die komplexen Peptidgemische, die aus den einzelnen Verdauen resultieren, lassen sich nur adäquat mit einer vorgeschalteten Separierung durch Hochdruckflüssigkeitschromatographie (HPLC) vermessen. Die HPLC ist direkt verknüpft mit einem Orbitrap-Massenspektrometer (Q Exactive von Thermo). Genaueres zur HPLC und den Vermessungsparametern findet sich im Methodenteil im Abschnitt 2.2.9 zu entnehmen.

Die aus der Messung resultierenden Dateien wurden mit Hilfe von MSConvert zu mgf-Dateien konvertiert. Die Datenbanksuche erfolgte für jede Datei zweimal mit unterschiedlichen Suchparametern. Im ersten Suchlauf wurden die Modifikationen als fix auftretend und die Anzahl von Fehlschnittstellen auf Null gesetzt. Des Weiteren wurde bei dem ArgC-Verdau und seinen Analoga ArgC als Enzym gewählt. Der erste Suchlauf ist somit recht spezifisch, wohingegen der Schwerpunkt im zweiten Suchlauf auf der Identifikation nicht vollständig modifizierter Aminogruppen liegt, so dass die verschiedenen Modifikationen als variabel auftretend gesetzt wurden. Des Weiteren wurde Trypsin als Enzym und eine hohe Zahl an tolerablen Fehlschnittstellen gewählt. Die genauen Suchparameter sind im Anhang C einsehbar.

Anzahl identifizierter Peptide und Proteine

Zur Bestimmung der Anzahl identifizierter Peptide und Proteine wurden nur die Resultate des ersten Suchmaschinenlaufes berücksichtigt. In der Abbildung ist ein gruppiertes Balkendiagramm zu sehen. Aus diesem Balkendiagramm ist ersichtlich, zu welchem Anteil die aufgenommenen MS/MS-Spektren zur Identifikation eines Peptides geführt haben. Der erste Balken gibt die Gesamtanzahl der aufgenommenen MS/MS-Spektren an, der zweite die nicht zuweisbaren und der dritte Balken die Anzahl der Spektren, die zu einem Peptid zugeordnet werden konnten. Beim letzten Balken spricht man auch von der Anzahl der PSMs (*Peptide to Spectra Matches*).

In der Abbildung ist zu erkennen, dass im Vergleich zum konventionellen ArgC Verdau alle ArgC-ähnlichen Protokolle höhere Anteile zuweisbarer Spektren aufweisen, wobei den höchsten Anteil innerhalb der ArgC-ähnlichen Verdaue das PA-Protokoll hat. Trypsin überragt mit seinem hohen Anteil an PSMs alle anderen Protokolle. Mögliche Ursachen hierfür können unerwünschte Seitenreaktionen, unvollständige Modifikationen und das Auftreten von Modifikationen am proteolytisch generierten N-Terminus der Peptide sein. Der letzte Fall würde eintreten, wenn das Modifizierungsreagenz bei den Aufreinigungsschritten nicht vollständig entfernt werden konnte. Der oben beschriebene zweite Suchmaschinenlauf diente der Identifikation der Peptide, bei denen einer dieser Fälle eingetreten ist. Diese Peptide wurden im ersten Suchmaschinenlauf aufgrund der gesetzten Suchparameter nicht erfasst.

Zunächst wurde der Anteil von unerwünschten Modifikationen am N-Terminus

4.3. Vergleich der Modifizierungsprotokolle auf Protein- und Proteomebene

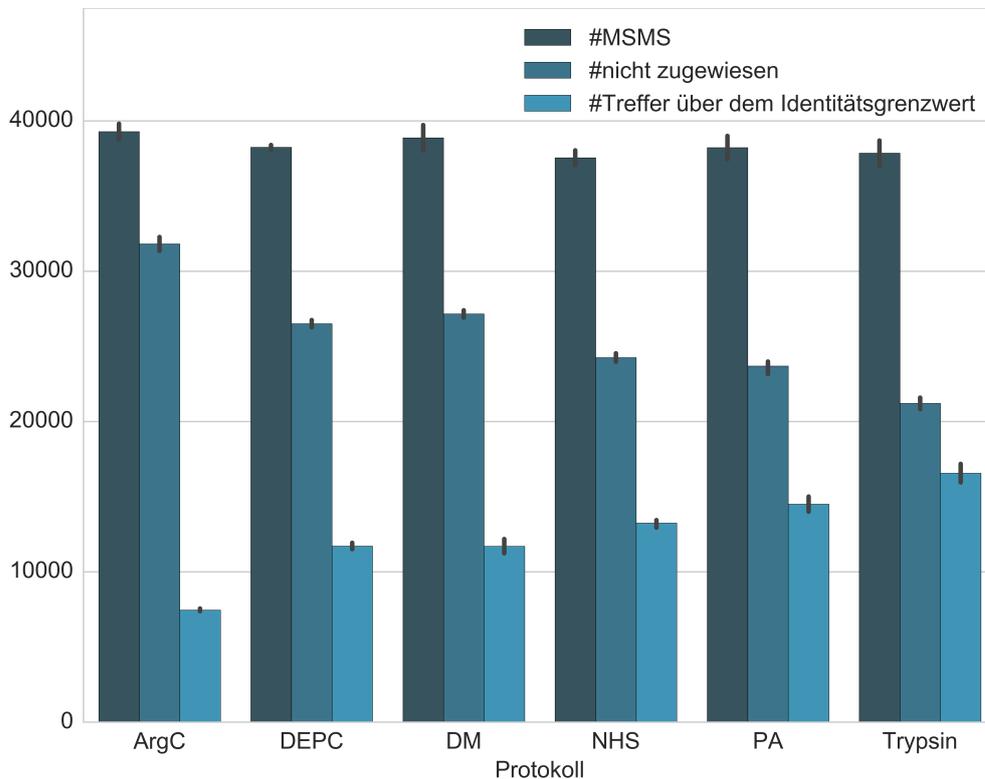


Abbildung 4.25: Gesamtzahl der aufgenommenen MS/MS-Spektren und deren nicht zu Peptiden zuweisbaren Anteile

von Peptiden untersucht. Hierbei wurde berücksichtigt, dass einige der eingesetzten Modifikationen auch natürlich vorkommen, wie die Acetylierung von Proteinen an ihren N-Termini. Das Ergebnis dieser Untersuchung ist in Abbildung 4.26 auf der nächsten Seite zu sehen. Erkennbar ist, dass die ungewollte reduktive Methylierung der Peptid-N-Termini nur extrem selten vorkommt. Dies deutet darauf hin, dass während des Verdauens kein Methylierungsreagenz vorhanden ist. Dies kann begründet sein, dass die Reagenzien effizienter entfernt werden mit der eingesetzte Aufreinigungsmethode oder der Grad der Inaktivierung - durch den Zusatz von Hydroxylamin - höher ist.

Die Acylierungsreagenzien weisen zahlenmäßig gleich häufig vorkommende N-terminal modifizierte Peptide auf, wobei im Falle des DEPC diese Peptide einen größeren Anteil an der Gesamtintensität ausmachen. PA und NHS verhalten sich in dieser Untersuchung weitgehend gleichwertig.

Eine weitere Ursache für die höhere Zahl unzuweisbarer MS/MS-Spektren bei den ArgC-ähnlichen Verdauen könnte die unvollständige Modifizierung der Aminogruppen sein. Nicht modifizierte Lysine stellen eine Schnittstelle

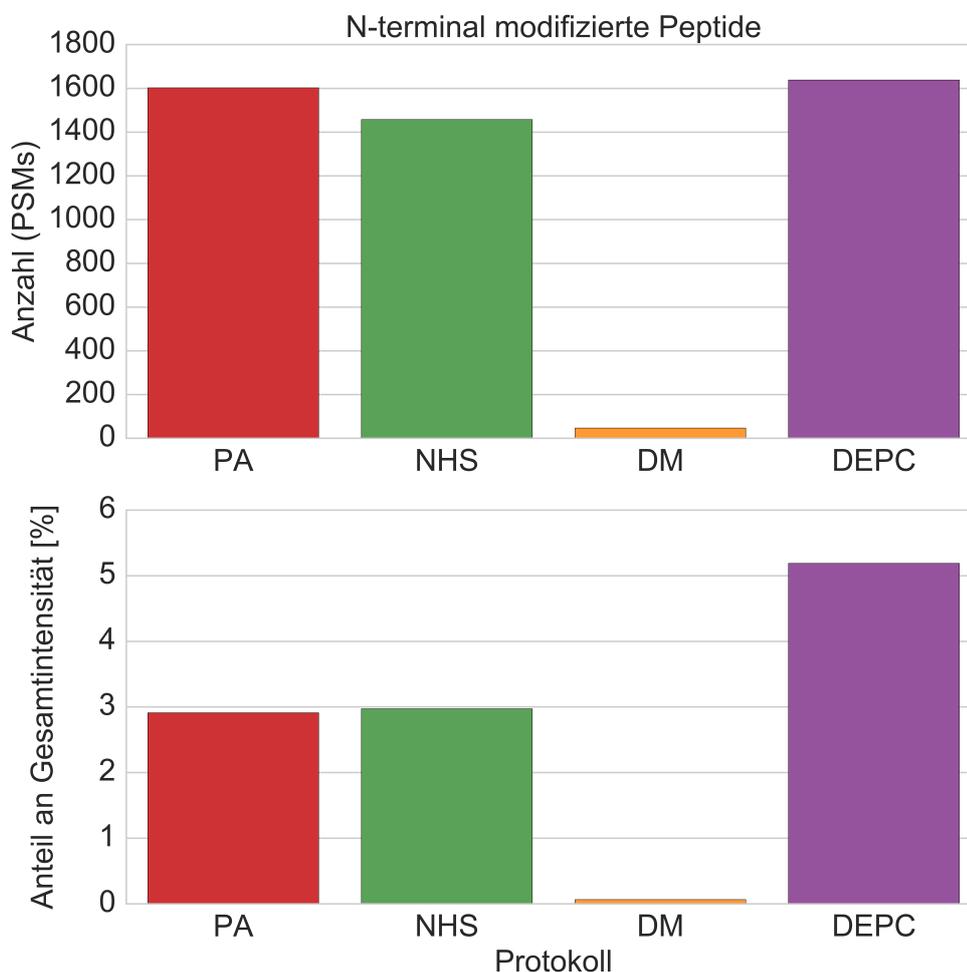


Abbildung 4.26: Anzahl und Anteil N-terminal modifizierter Peptide.

für Trypsin dar. Aus diesem Grund wurde im zweiten Suchlauf Trypsin als Protease gewählt und die Menge zulässiger Fehlschnittstellen auf 8 gesetzt, so dass sowohl die unvollständig als auch die vollständig modifizierten Peptide identifiziert werden konnten. Der Suchlauf ergab, dass keine der Modifizierungen quantitativ erfolgt ist, wobei sowohl die Anzahl als auch der Anteil der Nebenprodukte in einem erfreulich niedrigem Prozentbereich liegen (siehe Abbildung 4.27 auf der nächsten Seite). Das NHS-Protokoll schneidet dabei signifikant schlechter ab als die anderen Modifizierungsprotokolle.

Zwar weisen das PA- und das DEPC-Protokoll geringere Anteile unvollständig modifizierter Peptide an den gemessenen Gesamtintensitäten auf, dennoch lässt sich beim Zusammenführen der beiden obigen Ergebnisse schlussfolgern, dass die reduktive Methylierung am besten abschneidet. Dies

4.3. Vergleich der Modifizierungsprotokolle auf Protein- und Proteomebene

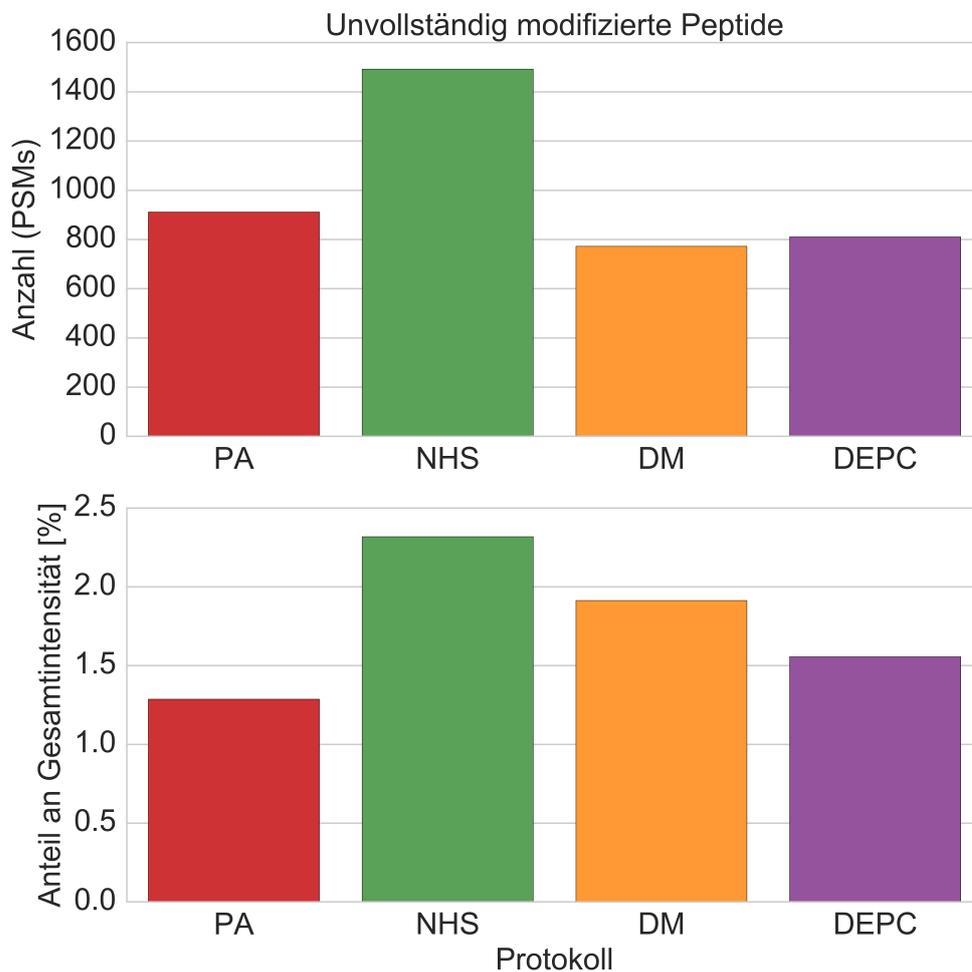


Abbildung 4.27: Anzahl und Anteil unvollständig modifizierter Peptide.

ist begründet durch die vernachlässigbare kleine Anzahl N-terminal modifizierter und die geringe Anzahl unvollständig modifizierter Peptide bei Anwendung dieses Protokolls. Von den Acylierungsreagenzien liefert das PA-Protokoll die besten Resultate nach dem Zusammenführen beider Ergebnisse. Die hohe Reaktivität von DEPC wurde bereits im letzten Abschnitt angesprochen und verursacht im vorliegenden Fall mit hoher Wahrscheinlichkeit den hohen Anteil N-terminal modifizierter Peptide. Des Weiteren wurde im vorherigen Abschnitt besprochen, dass DEPC verschiedene Reaktionen mit den nukleophilen Gruppen eines Proteins eingehen kann. Im zweiten Suchlauf wurde jedoch nur die Reaktion mit dem Imidazol-Ring des Histidins berücksichtigt. Der Suchlauf ergab, dass nach Berücksichtigung dieser Modifikation bis zu 200 zusätzliche Peptide identifiziert werden konnten, die einen Anteil von bis zu 2% an der Gesamintensität ausmachen.

Das Sulfo-NHS-Acetat, welches zur Acetylierung der Aminogruppen eingesetzt wurde, scheint die geringste Reaktivität zu haben, da die höchste Anzahl unvollständig modifizierter Peptide mit dem größten Anteil an der gemessenen Gesamtintensität bei diesem Ansatz berechnet wurde.

Abschließend werden in diesem Abschnitt auf die einmalig identifizierten Peptide und Proteine eingegangen. Bislang lag der Fokus auf den PSMs, welche jedoch häufig Duplikate enthalten. Diese entstehen durch die Mehrfachmessung eines Peptides. Peptide eluieren nämlich in einem Zeitbereich und werden somit häufig mehrfach bei der MS-Aufnahme erfasst. Gehört ihr Signal zu den stärksten im aufgenommenem MS-Spektrum, werden sie als Vorläuferion für die MS/MS-Aufnahme ausgewählt. Diese mehrfach vermessenen Peptide liefern jedoch keinen Informationsgewinn, so dass die Anzahl der einmalig identifizierten Peptide relevanter als die Anzahl der PSMs ist. In der Tabelle 4.10 können die Anzahl der einmalig identifizierten Peptide und Proteine mit der dazugehörigen *False Discovery Rates* (FDR) entnommen werden (siehe Abschnitt 2.2.9).

Tabelle 4.10: Anzahl der Peptid- und Proteinidentifikationen

| Protokoll | Einmalig identifizierte | | False Discovery Rate | |
|-----------|-------------------------|----------|----------------------|----------|
| | Peptide | Proteine | Peptide | Proteine |
| PA | 9216 | 1439 | 0,40 | 0,91 |
| NHS | 8683 | 1402 | 0,44 | 1,45 |
| DM | 7166 | 1206 | 0,40 | 1,35 |
| DEPC | 7868 | 1425 | 0,47 | 1,64 |
| ArgC | 4624 | 1164 | 0,24 | 0,78 |
| Trypsin | 11178 | 1287 | 0,52 | 2,71 |

Aus den Tabellen kann entnommen werden, dass es mit drei der ArgC-ähnlichen Ansätze möglich war, mehr Proteine zu identifizieren als mit dem tryptischen Ansatz, der den „Goldstandard“ in der proteomischen Welt darstellt. Auffällig ist hierbei, dass eine höhere Anzahl an Proteinidentifikationen erzielt wurde, obwohl der tryptische Ansatz eine höhere Anzahl identifizierter Peptide aufweist. Dieser Sachverhalt wird ausführlich im nächsten Abschnitt diskutiert. Des Weiteren ist anzumerken, dass zum wiederholten Male der enzymatische Verdau mit ArgC zu den vergleichsweise schlechtesten Ergebnissen geführt hat.

Komplementarität der ArgC-ähnlichen Verdau zu Trypsin

In diesem Unterabschnitt soll die Komplementarität der Resultate der ArgC-ähnlichen Ansätze mit dem tryptischen Ansatz verglichen werden. In die-

4.3. Vergleich der Modifizierungsprotokolle auf Protein- und Proteomebene

sem Zusammenhang kann Komplementarität auf zwei unterschiedliche Arten gedeutet werden. Zum Einen weisen zwei Resultate eine hohe Komplementarität auf, wenn deutlich höhere Sequenzabdeckungen erreicht werden nach Zusammenführen der Sequenzinformationen der Proteine, die aus den einzelnen Ansätzen gewonnen wurden. Eine andere Art der Deutung liegt auf Ebene der Proteinidentifikation. Je kleiner die Schnittmenge und umso größer die Randmengen der identifizierten Proteine zweier unterschiedlicher Ansätze sind, desto höher ist deren Komplementarität zueinander.

Zunächst wird die Komplementarität auf Ebene der Sequenzabdeckung untersucht. Zuvor ist jedoch zu sagen, dass hohe Sequenzabdeckungen speziell für Fragenstellungen interessant sind, die auf die Charakterisierung von Proteinen abzielen. Die Ermittlung der vollständigen Sequenz eines Proteins mit nur einer Protease ist in der Regel nicht möglich. Die Verteilung der Schnittstellen und der eingeschränkte Massenbereich der Geräte bedingen, dass nicht jedes generierte Peptid detektiert werden kann, da es entweder zu klein oder zu groß ist. Möchte man jedoch Proteine charakterisieren, sind hohe Sequenzabdeckungen erforderlich, wobei man heutzutage dies hauptsächlich durch den Einsatz unterschiedlicher Proteasen erreicht (*Multi-Enzyme-Approach*).¹⁰⁴ Durch die unterschiedliche Verteilung der Schnittstellen, können in der Regel mehr Sequenzinformationen generiert werden. Wie hoch der Zugewinn bei den ArgC-ähnlichen Verdauen ist und welches Reagenz die besten Resultate liefert, wird im Folgenden untersucht.

Bevor der Zugewinn der Sequenzinformationen analysiert wird, werden die Verteilungen der vorliegenden Sequenzabdeckungen der Proteine für jeden Ansatz betrachtet. Bei nicht Vorliegen einer Normalverteilung, lässt sich die Streuung der Daten durch die Angabe von Maßzahlen, wie es bei der „Fünf-Punkte-Zusammenfassung“ der Fall ist, beschreiben. Die Fünf-Punkte-Zusammenfassung einer Verteilung besteht aus x_{\min} (Minimum), $x_{0,25}$ (1. Quartil), x_{med} (Median), $x_{0,75}$ (3. Quartil) und x_{\max} (Maximum). Ein *Box-Plot* ist die Visualisierung der Fünf-Punkt-Zusammenfassung, welches sich als grafische Darstellung der Daten zum Vergleich verschiedener Verteilungen sehr gut eignet. Ein *Box-Plot* besteht aus einer Schachtel („Box“) und zwei Linien („Whiskers“). Den Anfang der Schachtel definiert die Lage des 1. Quartils und das Ende wird definiert durch das 3. Quartil. Der Querbalken innerhalb der Box gibt die Lage des Medians wieder. Die beiden Linien außerhalb der Box gehen von x_{\min} bis x_{\max} , wobei die Maximallänge der Whiskers das 1,5-fache der Boxlänge ist. Liegen Punkte außerhalb dieser Maximallänge, werden sie als Ausreißer gekennzeichnet.¹²⁶

In der Abbildung 4.28 auf der nächsten Seite ist die Verteilung der Sequenzabdeckung für jedes Protokoll jeweils durch ein *Box-Plot* visualisiert. Es ist sofort erkennbar, dass die Protokolle PA, NHS und DM ähnliche Verteilun-

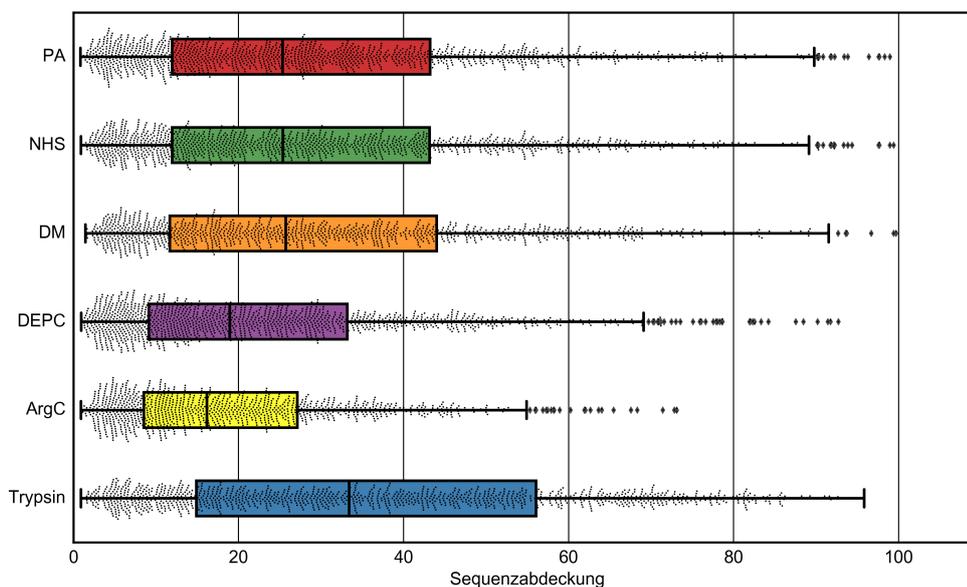


Abbildung 4.28: Boxplots zur Visualisierung der Verteilung der Sequenzabdeckungen der verschiedenen Ansätze.

gen aufweisen. Beim DEPC-Protokoll ist eine deutliche Verschiebung der Verteilung in niedrigere Bereiche zu sehen. Dennoch liefert es ein besseres Resultat, als der konventionelle ArgC-Ansatz. In dieser Abbildung wird erneut ersichtlich, aus welchem Grund Trypsin mit großem Abstand die am häufigsten genutzte Protease ist. Die Verteilung der Sequenzabdeckung ist deutlich in höhere Bereiche verschoben und somit liefert Trypsin das beste Resultat in dieser Untersuchung.

Der Zugewinn der Sequenzinformation nach Zusammenführen zweier Resultate lässt sich ebenfalls mit einem Boxplot visualisieren. Hierbei ist jedoch zu beachten, dass die resultierende Sequenzabdeckung nicht durch bloße Addition oder Subtraktion der Ausgangswerte ermittelt werden kann, da überschneidende Sequenzteile vorhanden sind. Bedauerlicherweise existierte kein Programm zur Beantwortung dieser Fragestellung. Aus diesem Grund wurde ein kurzes Python-Script geschrieben, das jedes einzelne Peptid eines Proteins zunächst in eine Liste von Zahlen transkribiert, die angibt an welcher Position im Protein die Aminosäuren des Peptides sich befinden. Befand sich zum Beispiel ein Peptid am N-Terminus eines Proteins und war dieses 6 Aminosäuren lang, wurde es in die Liste [1,2,3,4,5,6] transkribiert. Die Listen wurden abhängig vom Protein vereinigt und Doppelnennungen entfernt, sodass die Länge der Liste dividiert durch die Länge des Proteins die Sequenzabdeckung eines Proteins wiedergab. Diese resultierende

4.3. Vergleich der Modifizierungsprotokolle auf Protein- und Proteomebene

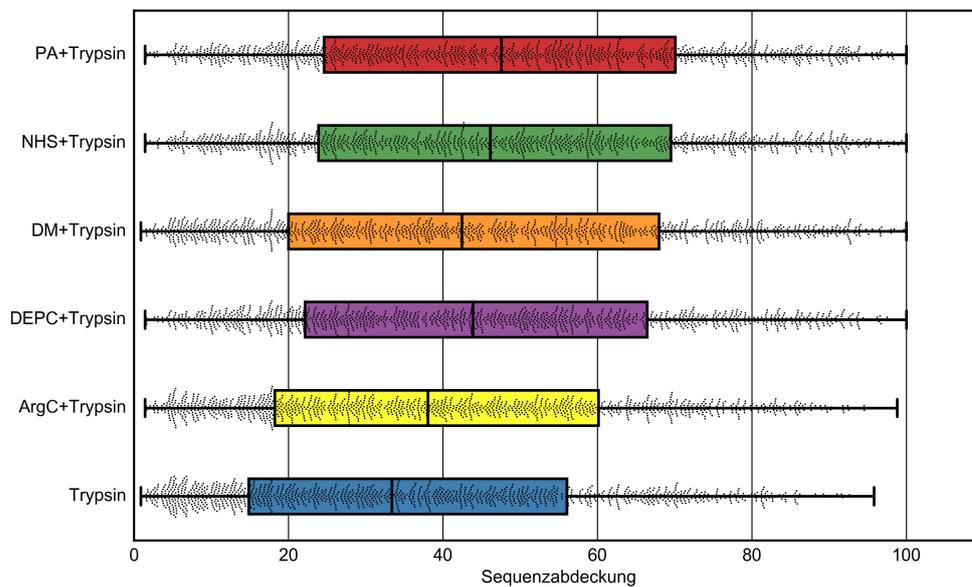


Abbildung 4.29: Boxplots zur Visualisierung der Verteilung der Sequenzabdeckungen nach Zusammenführen der Resultate mit dem tryptischen Verdau.

Sequenzabdeckung gibt also an, wieviel Prozent des Proteins beschrieben werden können, wenn die Resultate beider Protokolle vereinigt werden.

In der Abbildung 4.29 zeigt der erste Boxplot von unten die Verteilung der Sequenzabdeckungen vom tryptischen Ansatz. Alle anderen abgebildeten Boxplots zeigen die Sequenzabdeckungen, die man erhält, nachdem die Sequenzinformationen des einen Protokolls mit dem des Trypsin-Protokoll vereinigt wurden. Aus der Abbildung ist erkennbar, dass die Acylierungsprotokolle zur größten Verschiebung in höhere Bereiche geführt haben. Von diesen zeigt das PA-Protokoll zwar nur sehr knapp, aber erkennbar die größte Verschiebung und erreicht somit die größte Komplementarität im Hinblick auf den Zugewinn an Sequenzinformationen gegenüber dem tryptischen Ansatz.

Zur Analyse der Komplementarität auf Ebene der identifizierten Proteine, eignen sich Venn-Diagramme, aus denen das Ausmaß der Schnittmengen und Randmengen sehr leicht ersichtlich wird. In der Abbildung 4.30 auf Seite 101 sind fünf Venn-Diagramme gezeigt, die jeweils die Komplementarität zum tryptischen Ansatz aufzeigen. Das Venn-Diagramm des DEPC-Protokolls gibt zu erkennen, dass dieser Ansatz die höchste Komplementarität mit 262 zusätzlich identifizierten Proteinen aufweist, dicht gefolgt vom PA-Protokoll mit 252 Proteinen. Auf den dritten und vierten Rang sind das NHS- und das ArgC-Protokoll mit jeweils 219 und 123 Proteinen. Auf den

letzten Rang befindet sich das DM-Protokoll mit 109 zusätzlich identifizierten Proteinen.

4.3.3 Diskussion

Rückblickend ergab der Vergleich der unterschiedlichen Protokolle in diesem Abschnitt, dass die Methoden, die einen ArgC-Verdau nachahmen, in jedem Fall zu einem besseren Ergebnis geführt haben als der konventionelle ArgC-Verdau. Auf die Frage, welches der eingesetzten Reagenzien zu den besten Resultaten geführt hat, kann nach Zusammenführen aller Ergebnisse dies mit Propionsäure-Anhydrid beantwortet werden, auch wenn in einzelnen Untersuchungen andere Reagenzien zu besseren Ergebnissen geführt haben.

Auf Proteinebene lieferte das PA-Protokoll sowohl nach visuellen Kriterien als auch nach Einbezug quantitativer Messgrößen das „sauberste“ Spektrum. Dies ließ sich auf Proteomebene durch die geringste Anzahl unzuweisbarer MS/MS-Spektren verifizieren. Der Anteil der unvollständig modifizierten Peptide ist sehr gering und die Anzahl einmalig identifizierter Peptide ist sehr hoch, so dass dieses Protokoll die höchste Anzahl identifizierter Proteine aufwies.

Das PA-Protokoll zeigt zusätzlich hinsichtlich der Sequenzabdeckung die in höchsten Maße komplementären Resultate zu den Ergebnissen des tryptischen Verdauens. In Hinblick auf die Proteinidentifikationen weist es ähnlich hohe Werte auf wie das DEPC-Protokoll, welches die höchste Anzahl zusätzlich identifizierter Proteine zeigte. Lediglich bei der Anzahl N-terminal modifizierter Peptide wurden nach Dimethylierung der Aminogruppen bessere Resultate erhalten. Im nächsten Abschnitt werden folglich nur noch die Resultate des PA-Protokolls mit denen des konventionellen tryptischen Verdauens verglichen.

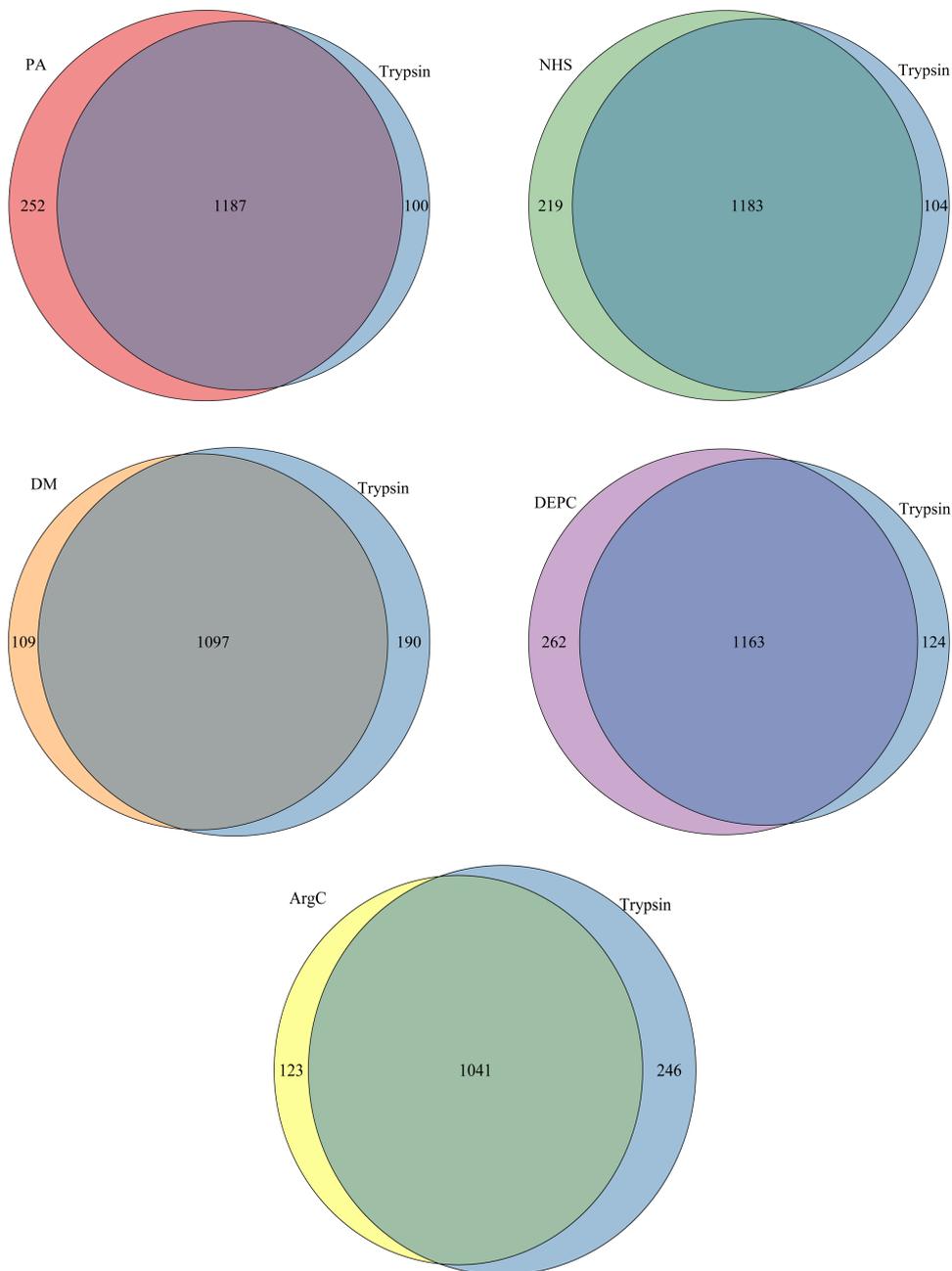


Abbildung 4.30: Venn-Diagramme der identifizierten Proteine mit Trypsin als Referenz.

4.4 Vergleich mit Trypsin

Aus den Untersuchungen des letzten Abschnittes ging nach Zusammenführen aller Ergebnisse hervor, dass das PA-Protokoll in der Summe inner-

halb der ArgC-ähnlichen Ansätze die besten Resultate liefert. Anknüpfend an den letzten Abschnitt werden nachfolgend zunächst mögliche Gründe für die höhere Anzahl identifizierter Proteine des PA-Protokolls im Vergleich zu den Resultaten des tryptischen Ansatzes genannt, der zunächst überraschend eine höhere Anzahl identifizierter Peptide aufwies. Im Anschluss werden beide Protokolle auf das Fragmentierverhalten der generierten Peptide und das Auftreten von Fehlschnittstellen untersucht.

4.4.1 Anzahl der Proteinidentifikationen

Aus dem Venn-Diagramm in der Abbildung 4.30 ging hervor, dass mit dem PA-Protokoll 252 zusätzliche Proteine identifiziert werden konnten. Aus der Tabelle 4.10 ist abzulesen, dass absolut gesehen 152 Proteine mehr identifiziert werden konnten, obwohl die Anzahl der identifizierten Peptide beim tryptischen Ansatz deutlich höher ist.

Die höhere Anzahl an Identifikationen trotz geringerer Anzahl identifizierter Peptide lässt sich nur durch eine sinnvolle Reduktion der Probenkomplexität erklären. Im Scatterplot in der Abbildung 4.31 auf der nächsten Seite wurden die durchschnittliche Anzahl zugewiesener PSMs pro Protein aus dem PA-Protokoll gegen die aus dem Trypsin-Protokoll aufgetragen. Je größer und je dunkler ein Punkt ist desto abundanter liegt dieses Protein im Proteom vor. Die Abundanzen wurden mit Hilfe des NSAF ermittelt (siehe Abschnitt 1.6.2). Nach dieser Kalkulation ist der Elongationsfaktor TU1 das höchst abundante Protein im untersuchten *Escherichia coli* Proteom. Es konnten aus den tryptischen Ansätzen im Durchschnitt 580 PSMs diesem Protein zugewiesen werden, wohingegen es beim PA-Protokoll im Durchschnitt nur 200 PSMs waren. Punkte, die auf der gestrichelten Linie liegen, gehören zu Proteinen, denen in beiden Ansätzen die gleiche Anzahl an PSMs zugewiesen werden konnten. Liegt ein Punkt unterhalb der gestrichelten Linie, weist dieses Protein im tryptischen Ansatz eine höhere Anzahl an PSMs auf, als im ArgC-ähnlichen Ansatz. Liegt der Punkt oberhalb der gestrichelten Linie, verhält es sich analog anders herum.

Aus der Abbildung wird ersichtlich, dass die höchst abundanten Proteine zu deutlich mehr PSMs führen, nachdem sie tryptisch verdaut wurden. Die Anzahl der möglichen Messpunkte sind in einem LC-MS/MS-Lauf begrenzt. Entfallen, wie beim Trypsin-Protokoll, zu viele Messpunkte auf die hochabundanten Proteine, so reicht die Kapazität der Apparatur nicht aus, um auch die niedrig abundanten Proteine zu erfassen. Dieser Sachverhalt erklärt somit, wieso das Trypsin-Protokoll trotz höherer Anzahl an Peptididentifikationen im Bereich der Proteinidentifikationen dem ArgC-ähnlichen Ansatz unterlegen ist. Durch das enge zusammenliegen der Punkte im Bereich unter 100 PSMs pro Protein ist nicht erkennbar, wie viele Punkte ober- oder unterhalb der gestrichelten Linie liegen. Kalkulationen ergaben jedoch, dass knapp

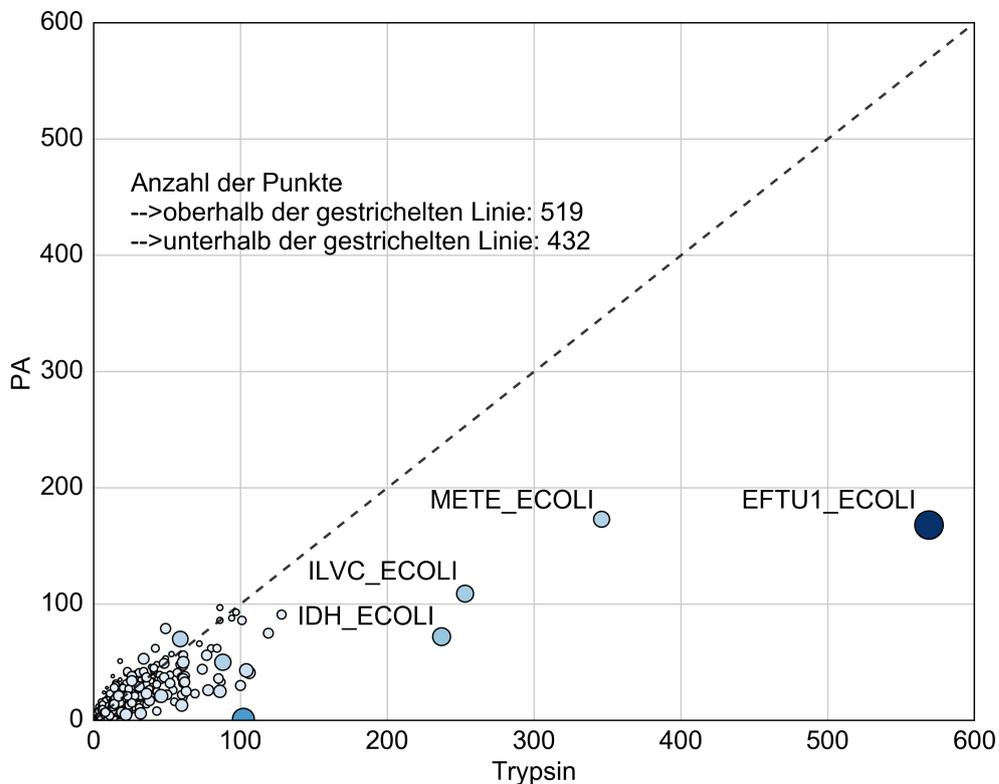


Abbildung 4.31: Scatterplot. x-Achse: Anzahl zugewiesener PSM je Protein für den tryptischen Ansatz; y-Achse: Anzahl zugewiesener PSM je Protein für den ArgC-ähnlichen Ansatz.

90 zusätzliche Punkte oberhalb der gestrichelten Linie liegen. Es konnten also im Durchschnitt mehr PSMs pro Protein durch das PA-Protokoll zugewiesen werden, was insgesamt zu Identifikationen höherer Sicherheit führt.

4.4.2 Verdaueffizienz

Wie bereits in der Einleitung erwähnt, werden bis zu 90% aller proteomischen Ansätze mit Trypsin durchgeführt.³¹ Neben seiner hohen Spezifität liegt das auch an seiner hohen Effizienz. Es schneidet somit nicht nur sehr spezifisch, es tut dies auch sehr schnell. Je höher die Verdaueffizienz einer Protease ist, desto reproduzierbarer sind seine Resultate, was wiederum für Quantifizierungsexperimente notwendig ist. Verdaueffizienzen von Proteasen lassen sich über verschiedene Wege bestimmen. Im Folgenden wird die Verdaueffizienz mit dem Vorkommen von Fehlschnittstellen korreliert. Fehlschnitte werden durch das Auftreten von Lysinen und Argininen innerhalb der Peptide indiziert. Je geringer die Anzahl der Peptide mit Fehlschnittstellen ist, desto höher ist die Verdaueffizienz. Die im Folgenden

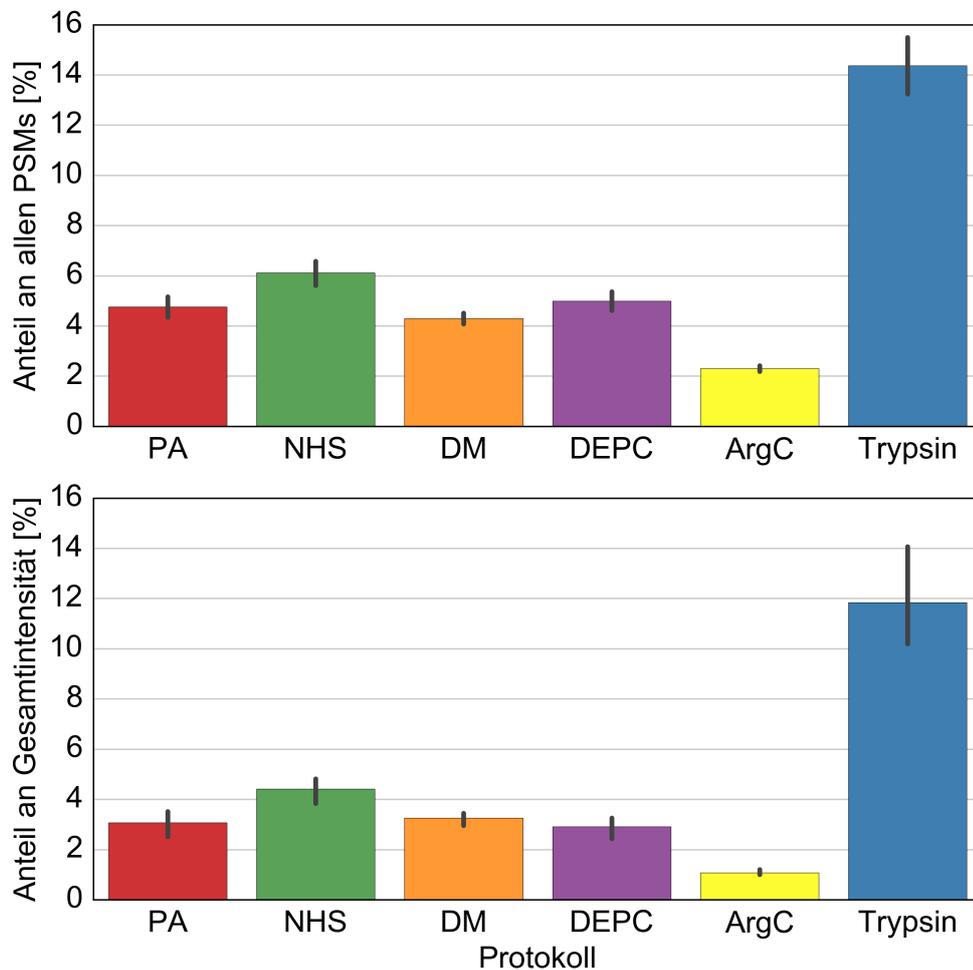


Abbildung 4.32: Prozentualer Anteil der Peptide mit Fehlschnittstellen an der Gesamtzahl der PSMs und der Anteil an der gemessenen Gesamtintensität.

dargestellten Untersuchungen erfolgten mit den Daten, die aus dem ersten Suchlauf generiert wurden. In diesem Suchlauf werden nur Peptide erfasst, die maximal einen Fehlschnitt aufweisen. Voruntersuchungen ergaben vernachlässigbar kleine Unterschiede, wenn mehr Fehlschnittstellen zugelassen wurden.

Vergleicht man nun die Verdaueffizienzen zwischen dem konventionell tryptischen und den ArgC-ähnlichen Verdauen, würde man zunächst keine Unterschiede vermuten, da beide Ansätze Trypsin verwenden. Berücksichtigt man jedoch, dass es häufiger Fehlschnitte nach Lysin gibt als nach Arginin, ist durch die Blockade des Lysins in den ArgC-ähnlichen Ansätzen sehr wohl mit einer veränderten Verdaueffizienz zu rechnen, trotz des Einsatzes

der selben Protease. In vorherigen Untersuchungen wurde bereits über die höhere Tendenz von Fehlschnittstellen nach Lysin berichtet,^{127,128} welches sich durch die hier zugrundeliegenden Daten bestätigen lässt. Die Analyse des konventionellen tryptischen Verdauers ergab, dass ungefähr 70% der Peptide mit einem Fehlschnitt diesen am Lysin aufweisen.

Im oberen Teil der Abbildung 4.32 auf der vorherigen Seite ist zu sehen, wie hoch der Anteil der fehlgeschnittenen Peptide in der Gesamtzahl der PSMs für jedes Protokoll ist. Die ArgC-ähnlichen Protokolle liegen ungefähr auf einem Niveau und rangieren zwischen 4 bis 6%. Mit Anteilen über 14% weist der konventionelle tryptische Verdau deutlich höhere Anteile an Peptiden mit einem Fehlschnitt auf. Am besten schneidet mit einem Anteil von knapp über 2% ArgC ab. Im unteren Graphen ist zu erkennen, dass der prozentuale Anteil mit dem Anteil an der Gesamtintensität korreliert. Auch hier verhalten sich die Pseudo-ArgC-Verdaue ähnlich und weisen deutlich geringere Werte als der konventionelle Trypsin-Verdau auf, allerdings höhere als der ArgC-Verdau.

Bisherige Untersuchungen ergaben, dass Fehlschnitte durch das Vorkommen bestimmter Aminosäuren in der Nachbarregion der Schnittstelle verursacht werden.¹²⁸⁻¹³¹ Es sind häufig Fehlschnitte zu beobachten wenn zwei Schnittstellen direkt nebeneinander liegen. Des Weiteren werden Fehlschnitte oft beobachtet, wenn in der Nähe der Schnittstelle saure Aminosäuren vorkommen. Vermutlich bilden diese intramolekulare Salzbrücken mit Lysin und Argininen, was zu einer reduzierten Basizität/Protonierung dieser Aminosäuren führt und somit die Aktivität des Trypsins an dieser Stelle senkt.

Der einfachste Weg zur Ermittlung des Einflusses einzelner Aminosäure auf das Vorkommen eines Fehlschnittes, ist die Dokumentation, wie oft die Aminosäure in welchem Abstand zu dem Fehlschnitt vorkommt. Denn je häufiger eine Aminosäure in der Nähe eines Fehlschnittes ist, desto wahrscheinlicher ist die Einflussnahme dieser Aminosäure. Diese Auszählung erfolgte für das Trypsin-Protokoll mit Hilfe der *Heatmap* in der Abbildung 4.33 auf der nächsten Seite.

Auf der x-Achse dieser Abbildung ist die Position relativ zu dem Fehlschnitt abgebildet, wobei positive Werte in Richtung des C-Terminus hinweisen und negative Werte in Richtung des N-Terminus des Proteins, aus dem das Peptid stammt. Die y-Achse enthält alle 20 proteinogenen Aminosäuren, so dass eine Matrix mit der Ordnung [20x11] entstand. Je höher die Anzahl der Aminosäure in einer bestimmten Position ist, desto dunkler ist das Feld. Ab einem bestimmten Schwellenwert ändert sich die Schriftfarbe von schwarz zu weiß. Die Werte auf der Position 0 wurden auf null gesetzt, da sonst die extrem hohen Werte für Arginin und Lysin keine Differenzierung der anderen

4. ERGEBNISSE UND DISKUSSION

| | Position | | | | | | | | | | |
|---|----------|-----|-----|-----|-----|---|------|------|------|-----|-----|
| | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 |
| A | 750 | 982 | 637 | 812 | 658 | 0 | 467 | 452 | 534 | 709 | 847 |
| R | 433 | 355 | 344 | 508 | 738 | 0 | 59 | 5 | 29 | 33 | 48 |
| N | 343 | 219 | 364 | 251 | 170 | 0 | 163 | 191 | 300 | 269 | 252 |
| D | 377 | 329 | 539 | 338 | 490 | 0 | 727 | 633 | 752 | 651 | 754 |
| C | 115 | 39 | 79 | 64 | 61 | 0 | 6 | 32 | 24 | 8 | 108 |
| E | 546 | 599 | 678 | 765 | 570 | 0 | 771 | 1022 | 1023 | 857 | 761 |
| Q | 227 | 310 | 209 | 321 | 221 | 0 | 158 | 340 | 280 | 214 | 157 |
| G | 533 | 396 | 452 | 464 | 553 | 0 | 375 | 708 | 472 | 577 | 360 |
| H | 74 | 149 | 96 | 118 | 82 | 0 | 136 | 147 | 159 | 225 | 121 |
| I | 453 | 489 | 314 | 373 | 438 | 0 | 354 | 395 | 367 | 371 | 552 |
| L | 662 | 649 | 853 | 671 | 737 | 0 | 641 | 653 | 670 | 623 | 684 |
| K | 389 | 362 | 451 | 465 | 540 | 0 | 1372 | 553 | 166 | 261 | 216 |
| M | 175 | 103 | 279 | 187 | 199 | 0 | 106 | 206 | 82 | 166 | 194 |
| F | 176 | 302 | 222 | 277 | 258 | 0 | 518 | 200 | 283 | 143 | 212 |
| P | 183 | 295 | 183 | 242 | 108 | 0 | 129 | 373 | 225 | 380 | 326 |
| S | 502 | 313 | 339 | 232 | 236 | 0 | 220 | 265 | 227 | 504 | 299 |
| T | 325 | 248 | 340 | 242 | 252 | 0 | 200 | 181 | 395 | 308 | 438 |
| W | 40 | 69 | 35 | 66 | 56 | 0 | 30 | 34 | 21 | 72 | 53 |
| Y | 142 | 195 | 148 | 156 | 134 | 0 | 153 | 158 | 180 | 116 | 110 |
| V | 575 | 617 | 458 | 468 | 519 | 0 | 435 | 472 | 831 | 533 | 528 |

Abbildung 4.33: *Heatmap* zur Auszählung der Aminosäuren an Positionen -5 bis 5 der Fehlschnittregion für den tryptischen Ansatz

Werte erlauben.

Bei näherer Betrachtung der *Heatmap* fallen bestimmte Aminosäuren durch ihr hohes Aufkommen auf. Diese lassen sich in drei Gruppen gliedern: die sauren Aminosäuren (D und E), die basischen Aminosäuren (R und K) und die lipophilen Aminosäuren (A,L,V und G). Die sauren Aminosäuren scheinen den stärksten Einfluss zu haben, wenn sie C-terminal zur Schnittstelle liegen. Wie bereits erwähnt, war bereits bekannt, dass zwei benachbarte Schnittstellen häufig zu Fehlschnitten führen, dies kann durch das hohe Vorkommen der Arginine und Lysine in unmittelbarer Nachbarschaft zum Fehlschnitt untermauert werden. Erstaunlicherweise hat Lysin von der (+1)-Position und Arginin von der (-1)-Position den größten Einfluss. Bei den lipophilen Aminosäuren ist kein derartiges Muster zu erkennen.

In der Abbildung 4.34 auf der nächsten Seite ist analog die *Heatmap* für das PA-Protokoll zu sehen. Das Auftreten eines Fehlschnittes wird auch hier durch eine saure Umgebung begünstigt und auch hier hat Arginin seinen größten Einfluss von der (-1)-Position. Im Unterschied zum Trypsin-Protokoll lässt sich kein Einfluss der lipophilen Aminosäuren auf das Auf-

4.4. Vergleich mit Trypsin

| | Position | | | | | | | | | | |
|---|----------|-----|-----|-----|-----|---|-----|-----|-----|-----|-----|
| | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 |
| A | 233 | 255 | 183 | 284 | 13 | 0 | 149 | 56 | 238 | 170 | 333 |
| R | 153 | 129 | 72 | 296 | 918 | 0 | 129 | 14 | 10 | 39 | 51 |
| N | 159 | 43 | 183 | 99 | 26 | 0 | 47 | 46 | 69 | 107 | 42 |
| D | 104 | 72 | 213 | 75 | 295 | 0 | 348 | 319 | 123 | 142 | 121 |
| C | 28 | 21 | 3 | 19 | 8 | 0 | 1 | 4 | 15 | 0 | 7 |
| E | 194 | 176 | 235 | 262 | 123 | 0 | 471 | 668 | 408 | 249 | 153 |
| Q | 69 | 41 | 33 | 53 | 6 | 0 | 65 | 70 | 86 | 48 | 79 |
| G | 103 | 59 | 116 | 76 | 55 | 0 | 160 | 81 | 106 | 186 | 124 |
| H | 20 | 50 | 28 | 11 | 18 | 0 | 40 | 16 | 30 | 33 | 17 |
| I | 66 | 113 | 81 | 135 | 46 | 0 | 86 | 53 | 101 | 118 | 85 |
| L | 135 | 167 | 207 | 112 | 198 | 0 | 97 | 64 | 198 | 147 | 289 |
| K | 59 | 89 | 84 | 63 | 34 | 0 | 39 | 44 | 104 | 84 | 97 |
| M | 51 | 24 | 50 | 39 | 77 | 0 | 4 | 52 | 5 | 20 | 53 |
| F | 64 | 140 | 84 | 57 | 34 | 0 | 64 | 49 | 101 | 78 | 47 |
| P | 63 | 59 | 35 | 70 | 2 | 0 | 1 | 177 | 33 | 88 | 76 |
| S | 93 | 124 | 94 | 89 | 24 | 0 | 30 | 71 | 31 | 196 | 79 |
| T | 100 | 61 | 115 | 46 | 23 | 0 | 14 | 67 | 59 | 75 | 125 |
| W | 20 | 3 | 38 | 9 | 3 | 0 | 7 | 16 | 14 | 9 | 9 |
| Y | 30 | 39 | 22 | 42 | 15 | 0 | 51 | 25 | 39 | 16 | 10 |
| V | 212 | 291 | 80 | 119 | 38 | 0 | 153 | 64 | 186 | 151 | 159 |

Abbildung 4.34: Heatmap zur Auszählung der Aminosäuren an Positionen -5 bis 5 der Fehlschnittregion für den ArgC-ähnlichen Ansatz

treten eines Fehlschnittes erkennen. Das propionylierte Lysin hat durch den Verlust seiner Basizität keinen weiteren Einfluss mehr auf das Vorkommen von Fehlschnittstellen.

Die Aussagekraft einer *Heatmap*, die durch bloße Auszählung der Aminosäuren erstellt wurde, wird durch den Nichteinbezug der Häufigkeit einer Aminosäure in dem untersuchten Proteom geschmälert. Aus der Tabelle 4.11 auf der nächsten Seite kann die berechnete prozentuale Häufigkeit der Aminosäuren aller identifizierten Proteine entnommen werden. Der Bias, der durch den Nichteinbezug der Häufigkeiten zustande kommt, ist für die sauren und basischen Aminosäuren nicht schwerwiegend, da ihr Vorkommen mit 5% dem durchschnittlichen Vorkommen aller Aminosäuren entspricht. Dies gilt jedoch nicht für die lipophilen Aminosäuren, die ein überdurchschnittlich hohes Vorkommen im Proteom aufweisen. Andererseits lässt sich der Bias durch das Ergebnis der zweiten *Heatmap* relativieren, wobei die lipophilen Aminosäuren keine auffällig hohen Zahlen aufwiesen und sich somit die hohen Zahlen der ersten *Heatmap* nicht ausschließlich durch das hohe Vorkommen der Aminosäuren in den identifizierten Proteinen erklären lassen.

4. ERGEBNISSE UND DISKUSSION

Eine weitere Art der Ermittlung von Konsensussequenzen wurde mit Einführung der freien Software „iceLogo“ im Jahr 2009 geschaffen.¹³² Die Software „iceLogo“ erfordert zwei Eingaben, die zu untersuchenden Sequenzen und Referenzsequenzen. Die zu untersuchenden Sequenzen müssen einen gewissen Aufbau aufweisen. Der Fehlschnitt muss sich in der Mitte des elf Aminosäuren langen Sequenzstückes befinden, so dass dieses Sequenzstück die fünf Aminosäuren vor und nach dem Fehlschnitt enthält. Die Kreierung dieser Sequenzen erfolgte mit Hilfe eines Python-Scriptes, welches von den fehlgeschnittenen Peptiden entweder Sequenzteile weggeschnitten oder mit Hilfe der zugehörigen Proteinsequenz aufgefüllt hat. Als Referenzsequenzen diente eine Proteindatenbank, die Sequenzen von Escherichia coli Proteinen enthielt. Aus beiden Eingaben wurden die Verteilungen der einzelnen Aminosäuren und deren Standardabweichungen ermittelt. Die beiden Verteilungen wurden von dem Programm miteinander verglichen und anhand eines iceLogos wird visualisiert, wie stark diese beiden Datensets voneinander abweichen. Diese Art der Darstellung, die die Häufigkeit der Aminosäuren mitberücksichtigt, unterliegt keinem Bias, wie es etwa bei den vorher aufgezeigten *Heatmaps* der Fall war.

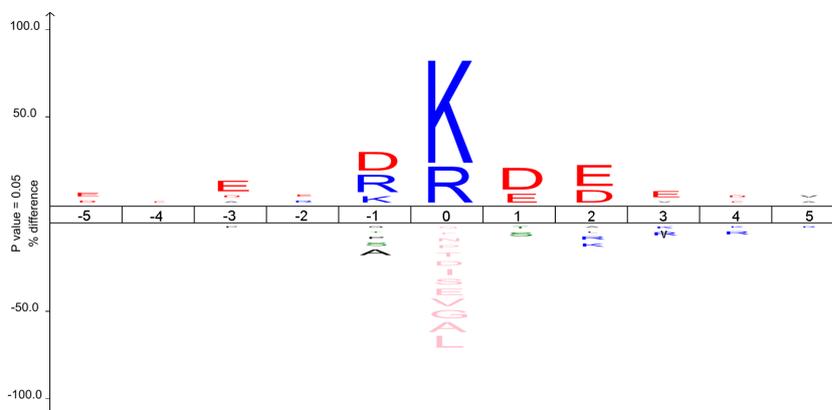


Abbildung 4.35: iceLogo der fehlgeschnittenen Peptide vom Trypsin-Protokoll.

Tabelle 4.11: Prozentuale Häufigkeit der Aminosäuren in dem untersuchten Proteom.

| | | | | | | | | | | |
|----------------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Aminosäure | L | A | G | V | E | R | D | I | S | T |
| Häufigkeit [%] | 10,1 | 9,7 | 7,4 | 7,1 | 6,8 | 6,0 | 5,8 | 5,6 | 5,2 | 5,2 |
| Aminosäure | K | Q | P | N | F | Y | M | H | W | C |
| Häufigkeit [%] | 4,7 | 4,6 | 4,5 | 3,8 | 3,5 | 2,7 | 2,6 | 2,4 | 1,2 | 1,1 |

In den Abbildungen 4.35 und 4.36 sind jeweils die iceLogos für das Trypsin- und das PA-Protokoll zu sehen. In dem für das Trypsin-Protokoll ist erneut ersichtlich, dass Fehlschnitte nach einem Lysin viel häufiger vorkommen als nach einem Arginin, erkennbar durch die Größe der Buchstaben. In Übereinstimmung zu den Ergebnissen aus der *Heatmap* ist auch hier in beiden Abbildungen erkennbar, dass die unmittelbare Nachbarschaft zu anderen Schnittstellen und die Präsenz saurer Aminosäuren die Hauptgründe für das Vorkommen eines Fehlschnittes sind. In dem iceLogo für das PA-Protokoll ist zusätzlich erkennbar, dass die saure Umgebung einen deutlich kleineren Einfluss hat als die Präsenz einer weiteren Schnittstelle. Diese Tendenz ist wiederum für das Trypsin-Protokoll nicht zu erkennen, so dass impliziert werden kann, dass das häufigere Auftreten von Fehlschnitten nach Lysinen mit seiner höheren Empfindlichkeit gegenüber einer sauren Umgebung verbunden ist. Folglich bewirkt die chemische Modifikation der Lysine und der damit einhergehende Verlust ihrer Funktion als Schnittstelle das geringere Auftreten von fehlgeschnittenen Peptiden bei den ArgC-ähnlichen Ansätzen. Die Ursache für die höhere Verdaueffizienz der ArgC-ähnlichen Ansätze liegt in der deutlich höheren Schnitteffizienz von Trypsin nach Argininen als nach Lysinen.

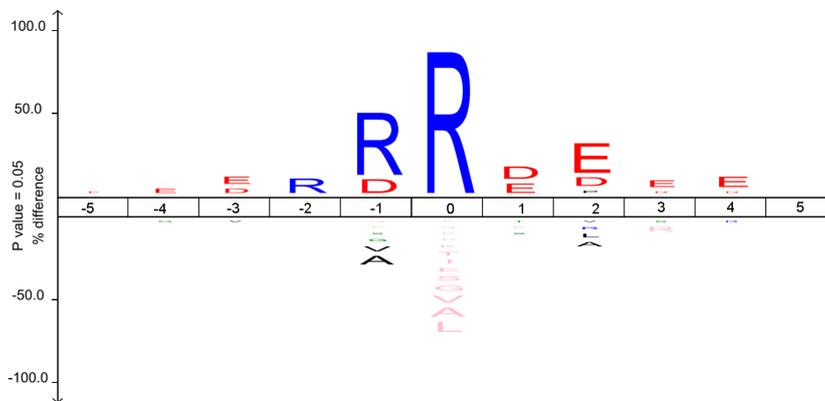


Abbildung 4.36: iceLogo der fehlgeschnittenen Peptide vom PA-Protokoll.

4.4.3 Fragmentierungsverhalten

Im letzten Abschnitt stellte sich heraus, dass sich die Verdaueffizienz in den zwei untersuchten Ansätzen geändert hat, obwohl die selbe Protease verwendet wurde. Auch beim Fragmentierungsverhalten sind Veränderungen zu erwarten, da ArgC-ähnliche Peptide sich in ihrem Aufbau von den trypti-

4. ERGEBNISSE UND DISKUSSION

schen Peptiden unterscheiden. Speziell die Modifikation der Lysin-Gruppen und den fix mit Arginin besetzten C-Terminus der ArgC-ähnlichen Peptide führen zu dieser Erwartung. Als qualitatives Maß für die Güte der Fragmentierung wird die Höhe der *Ionenscores* herangezogen, die von der Suchmaschine ermittelt werden und angeben, mit welcher Sicherheit ein MS/MS-Spektrum einem Peptid zugewiesen werden kann. Je mehr Signale in einem MS/MS-Spektrum vorhanden sind, die den vorausgesagten Fragmentionmassen zugewiesen werden können, und je weniger Signale vorhanden sind, die nicht zuweisbar sind, desto höher fällt der Ionenscore aus. Die Intensitäten der zuweisbaren und nicht zuweisbaren Signale spielen in der Kalkulation eine gewichtige Rolle. In der Abbildung 4.37 sind die Verteilungen der

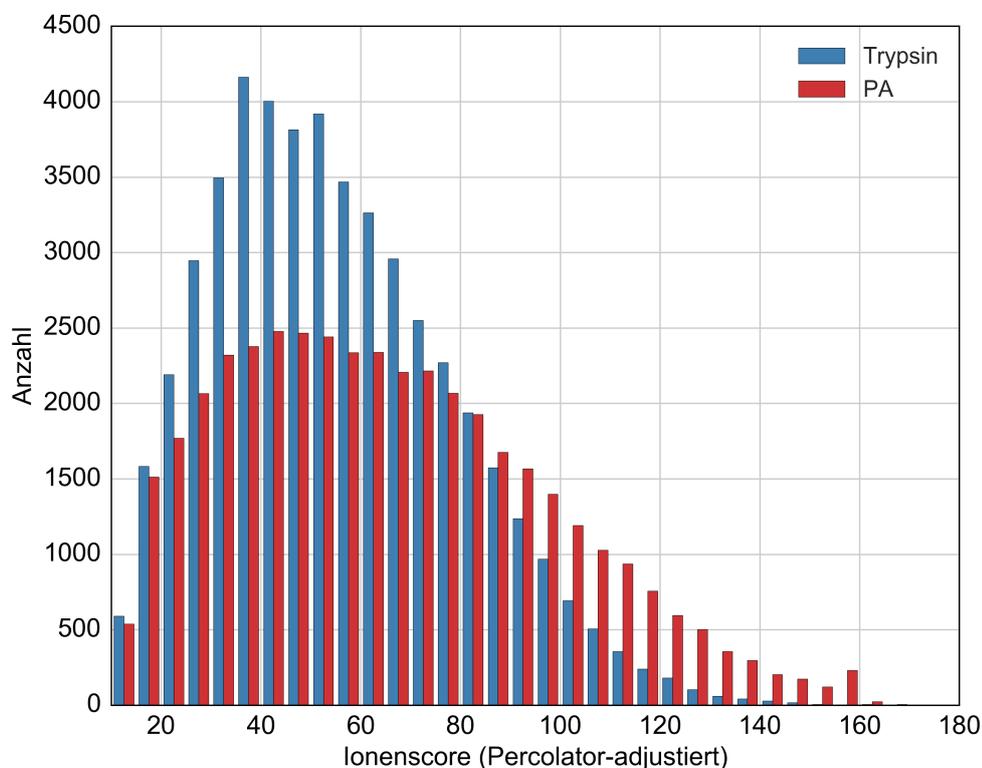


Abbildung 4.37: Verteilung der Percolator-adjustierten Ionenscores für den tryptischen und den ArgC-ähnlichen Ansatz

von der Suchmaschine ermittelten Ionenscores (vom Percolator adjustiert, siehe Anhang C für weitere Erklärungen) für das jeweilige Protokoll zu sehen. Zu erkennen ist, dass das Maximum beider Verteilungen ungefähr die gleiche Lage aufweist und beide Verteilungen zu höheren Massen unsymmetrisch sind, wobei das Maximum bei kleinen Massen beim PA-Protokoll

nicht so stark ausgeprägt ist. Die Verteilung des ArgC-ähnlichen Ansatzes ist somit breiter und weist höhere Werte im hohen Ionenscore-Bereich auf.

Diese Verschiebung der Ionenscores in einem höheren Bereich zeigt deutlich, dass die ArgC-ähnlichen Peptide ein besseres Fragmentierungsverhalten aufweisen. Die Ursachen hierfür sind multifaktoriell. Einer der Ursachen für die Verteilungsverschiebung könnte ein höherer Anteil von Peptiden mit höheren Massen sein, die in der Regel zu höheren Ionenscores führen. Dies kann jedoch durch die Abbildung 4.38, in der die Verteilung der theoretischen Peptidmassen zu sehen ist, nicht eindeutig bestätigt werden. Auch in diesem Histogramm sind beide Verteilungen zu höheren Massen unsymmetrisch und auch hier ist diese Steilheit beim PA-Protokoll nicht so stark ausgeprägt wie für das Trypsin-Protokoll. Trotz einer gewissen Ähnlichkeit der beiden Histogramme, sind die Unterschiede in den Verteilungen der Peptidmasse nicht so stark ausgeprägt, als dass man damit die Unterschiede in den Verteilungen der Ionenscores erklären könnte.

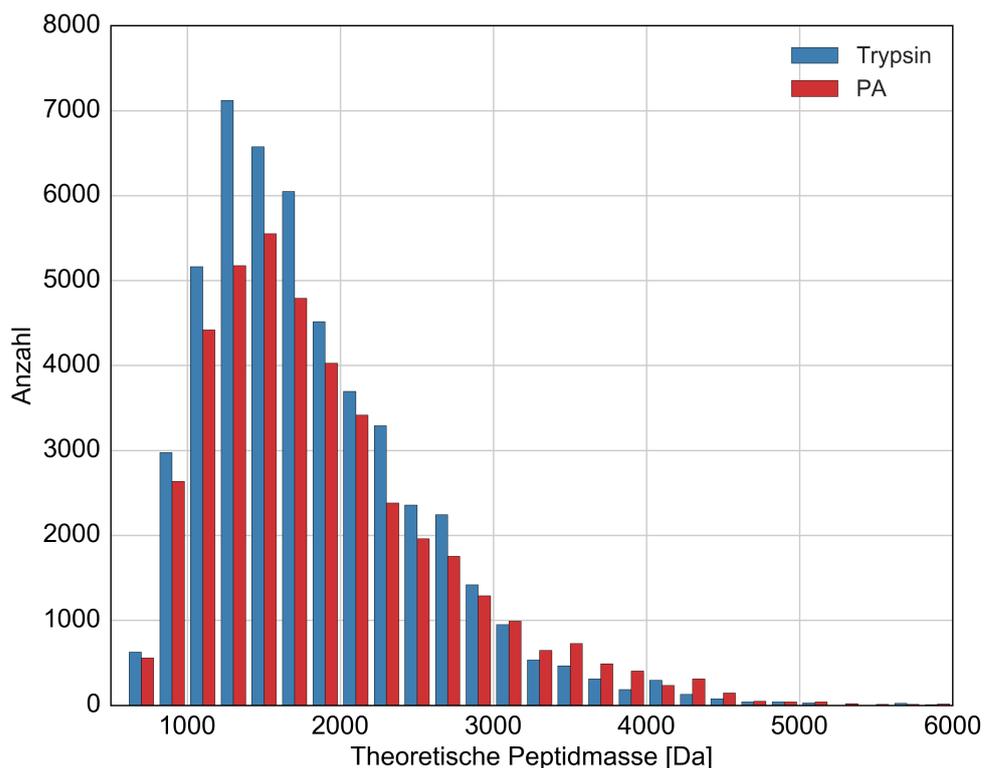


Abbildung 4.38: Verteilung der Peptidmassen für den tryptischen und den ArgC-ähnlichen Ansatz

Zur weiteren Aufklärung der Unterschiede in den Ionenscore-Verteilungen

wurden diese zunächst in Abhängigkeit der Peptidmassen untersucht. Hierzu wurden fünf Massenbereiche definiert und die Peptide beider Protokolle wurden entsprechend ihrer Masse diesen Gruppen zugeordnet. Würde man nun die Verteilungen in den einzelnen Gruppen mit Hilfe von Histogrammen auftragen, wären fünf Abbildungen notwendig. Übersichtlicher sind *Violinplots*, mit denen mehrere Verteilungen in einer Grafik abbildbar sind. In der Abbildung 4.39 ist der *Violinplot* für die Ionenscore-Verteilung in beiden Protokolle, in Abhängigkeit von den definierten Massenbereichen gezeigt. Jede „Violine“ setzt sich zusammen aus zwei *Densityplots*, die jeweils eine normalisierte Verteilung der Ionenscores der einzelnen Protokolle wiedergeben. Aus der Abbildung ist erkennbar, dass sich die Unterschiede in den Verteilungen erst in höheren Massenbereichen ergeben. In den höheren Massenbereichen ist der Anteil an Peptiden in hohen Ionscore-Bereichen beim PA-Protokoll deutlich ausgeprägter. Diese Untersuchung ergab somit, dass die Unterschiede in der Ionenscore-Verteilung massenabhängig sind. Peptide mit hohen Massen zeigen ein besseres Fragmentierungsverhalten auf, wenn sie durch den ArgC-ähnlichen Ansatz generiert werden, als wenn sie tryptischen Ursprungs sind.

Neben dem Ionenscore können noch weitere Kenngrößen zur Beschreibung der Fragmentierung einbezogen werden. Zum einen lässt sich die Güte der Fragmentierung wie in den vorherigen Abschnitten durch die *Intensity Coverage* bestimmen. Zur Ermittlung der *Intensity Coverage* von Fragmentenspektren wurden alle die Signale berücksichtigt, die sich entweder einem y-ion oder einem b-ion zuordnen lassen konnten. Das Verhältnis aus der Summe dieser Ionen durch die Gesamtintensität des Fragmentenspektrums gibt die *Intensity Coverage* wieder. Zusätzlich zu dieser Größe kann der Y-Fragmentüberschuss zur Beschreibung eines Spektrums ermittelt werden:

$$\text{y-Fragmentüberschuss} = \frac{\sum \text{Intensität (y-Ionen - b-Ionen)}}{\sum \text{Intensität (y-Ionen + b-Ionen)}} \quad (4.2)$$

Werte für den y-Fragmentüberschuss liegen im Intervall [-1;1]. Je größer der Wert desto höher der Anteil der y-Ionen an einem Fragmentenspektrum: Liegt der Wert somit bei 1, so liegen nur y-Ionen vor und keine b-Ionen. In der Abbildung 4.40 sind die Verteilungen beider Größen für das jeweilige Protokoll mit jeweils einem Boxplot dargestellt.

Gewisse Modifizierungen führen zum Auftreten von Neutralverlusten. Ein Neutralverlust ist ein ungeladenes Fragment, welches von einem Molekulation abgespalten hat. Ist es bekannt, dass ein Neutralverlust eintritt, kann dies durch die richtige Einstellung der Suchmaschine berücksichtigt werden. Es ist bislang nicht bekannt, dass Propionylierungen zu Neutralverlusten führen. Wären vermehrt Neutralverluste aufgetreten, hätte sich dies durch eine erniedrigte *Intensity Coverage* geäußert. Vergleicht man jedoch die Verteilungen

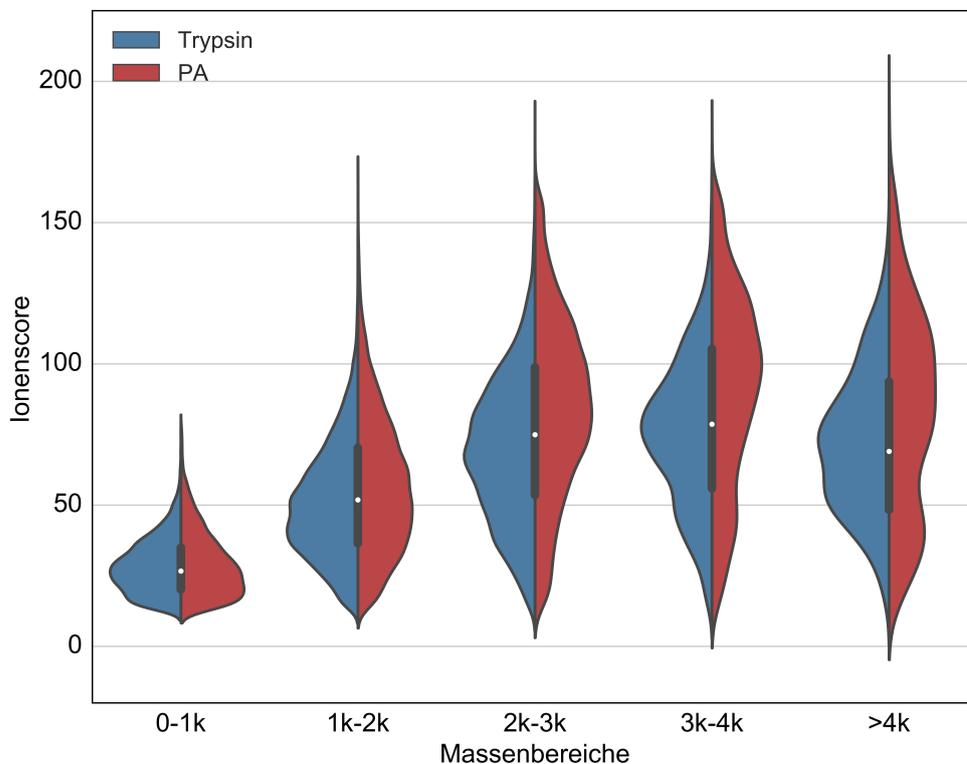


Abbildung 4.39: Violinplot der Ionenscore-Verteilungen in Abhängigkeit der Peptidmassenbereiche

der beiden Protokolle, ist keine Herabsetzung der *Intensity Coverage* zu sehen. Es ergibt sich vielmehr eine Verteilung, deren Werte leicht in höhere Bereiche verschoben ist. Des Weiteren ist eine sichtbare Verschiebung der Verteilung der y-Fragmentüberschüsse für das PA-Protokoll in höhere Bereiche zu sehen.

Die Massenabhängigkeit der Unterschiede in den Ionscore-Verteilungen lieferte den Anlaß, die Verteilungen der *Intensity Coverage* ebenfalls in Abhängigkeit der vorher definierten Massenbereiche aufzutragen (siehe Abbildung 4.41). Auch hier ist zu erkennen, dass die Unterschiede zwischen den Protokollen massenabhängig sind. Die Verteilungen für das Trypsin-Protokolls unterscheiden sich nicht stark voneinander. Dennoch ist eine Tendenz zu erkennen, dass Peptide mit hohen Massen zu tendenziell höheren Werten der *Intensity Coverage* führen. Deutlicher ausgeprägt ist diese Tendenz beim PA-Protokoll. Dies impliziert, dass große ArgC-ähnlich generierte Peptide „sauberer“ fragmentieren. Die Fragmentspektren sind somit leichter auswertbar und führen durch das Fehlen nicht interpretierbarer Signale zu höheren Ionenscores.

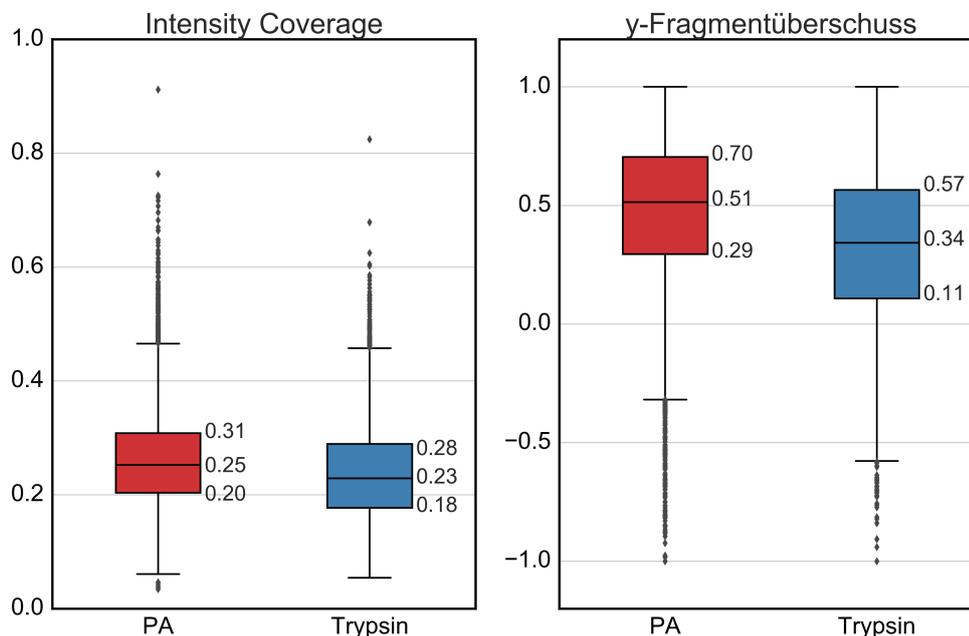


Abbildung 4.40: Boxplots für die *Intensity Coverage* und den *y-Fragmentüberschuss* für das jeweilige Protokoll

Bevor auf den zweiten Parameter, den *y-Fragmentüberschuss*, näher eingegangen wird, ist zunächst folgendes anzumerken. Ein Fragmentüberschuss einer beliebigen Serie führt nicht unmittelbar zu höheren Ionenscores, wenn die genutzte Suchmaschine Datenbank-basiert arbeitet. Erfolgt jedoch die Identifizierung über De-Novo-Sequenzierung erweisen sich vollständige Serien als äußerst nützlich. Dies erleichtert nämlich sowohl die manuelle als auch die maschinelle Interpretation. Auf der rechten Seite der Abbildung 4.40 konnten aus den *Boxplots* abgelesen werden, dass die Verteilung der *y-Fragmentüberschüsse* zu höheren Werten verschoben ist. Dies bedeutet wiederum, dass die Tendenz vollständige *y-Ionenserien* zu detektieren, beim ArgC-ähnlichen Ansatz höher ist und somit deren De-Novo-Interpretation im Vergleich zum tryptischen Ansatz tendenziell erleichtert ist.

Ursachen für die Unterschiede im Fragmentüberschuss und in der *Intensity Coverage* könnten in dem höheren Anteil der fehlgeschnittenen Peptide für das klassische Trypsin-Protokoll liegen, welche bereits im vorangegangenen Abschnitt besprochen wurden. Betrachtet man nämlich den Anteil der fehlgeschnittenen Peptide in Abhängigkeit der oben definierten Massenbereiche wird eine Zunahme der Fehlschnittstellen mit höheren Peptidmassen deutlich. Dieser Trend ist beim tryptischen Ansatz augenscheinlich stärker ausgeprägt. Fehlgeschnittene Peptide machen hier im höheren Massenbereich teilweise einen Anteil von bis zu mehr als 25% aus. Prinzipiell bedeutet das

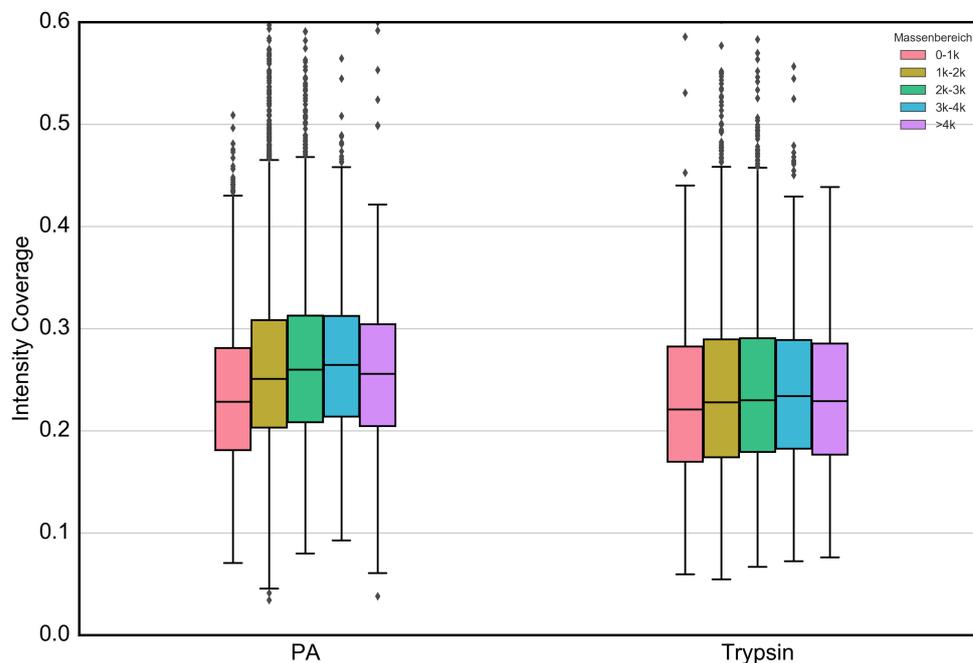


Abbildung 4.41: Nach Massenbereichen kategorisierte Boxplots für die *Intensity Coverage*

Vorkommen eines Fehlschnittes für beide Ansätze die Generierung von Peptiden, die eine basische Aminosäure in ihrer Kette enthalten müssen.

Derartige Peptide können ein abweichendes Fragmentierungsverhalten aufweisen und somit zu schwer interpretierbaren Fragmentionenspektren führen. Der Theorie des mobilen Protonenmodells zufolge liegen die basischen Aminosäuren eines Peptides in der Gasphase protoniert vor. Durch Aktivierung kann das Proton an Peptidbindungsstellen übertragen werden, so dass diese Bindung geschwächt wird und es somit zur Spaltung dieser Bindung kommen kann. Für den vollständigen tryptischen Verdau liegen die basischen Peptide Arginine und Lysin am C-Terminis der generierter Peptide, so dass bei der Fragmentierung leicht interpretierbare Ionenserien entstehen. Dieser Vorgang wird durch das Auftreten von basischen Aminosäuren, die sich innerhalb der Kette befinden, gestört. Das niedrigere Vorkommen von fehlgeschnittenen Peptiden in dem ArgC-ähnlichen Ansatz liefert somit eine Erklärung für das bessere Fragmentierungsverhalten.

Zusammenfassend lässt sich sagen, dass die chemische Modifikation der Lysine zu einem von den üblichen tryptischen Peptiden unterschiedlichem Fragmentierungsverhalten geführt hat. Die Fragmentionenspektren zeigen einen deutlich höheren Anteil an γ -Ionen. Zudem lassen sie sich leichter interpretieren, begründet durch die Verschiebung der Score-Verteilung in

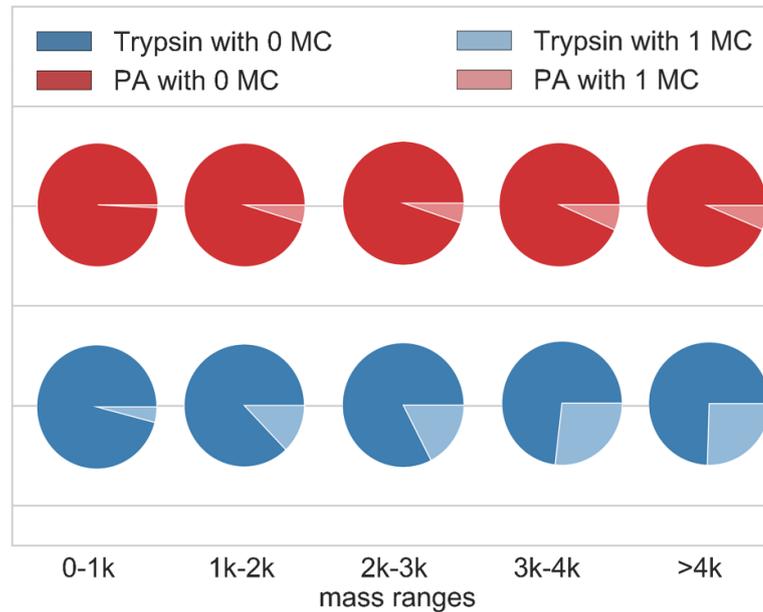


Abbildung 4.42: Anteil fehlgeschnittener Peptide aufgeteilt in Massenbereiche für das jeweilige Protokoll

höhere Bereiche. Dies ließ sich wiederum durch eine „sauberere“ Fragmentierung der Peptide, bedingt durch den geringeren Anteil fehlgeschnittener Peptide, erklären.

4.4.4 Diskussion

Nach Zusammenfassung der Ergebnisse aus dem Vergleich des ArgC-ähnlichen zum tryptischen Verdau lässt sich resümieren, dass der ArgC-ähnliche Verdau nicht nur eine komplementäre Methode zum tryptischen Verdau darstellt. Der ArgC-ähnliche Verdau führte in allen untersuchten Aspekten zu besseren Resultaten. Diese Strategie lieferte einerseits eine höhere Proteom-Abdeckung, angezeigt durch die höhere Zahl identifizierter Peptide. Des Weiteren stellte sich heraus, dass eine höhere Verdaueffizienz vorlag, angezeigt durch den geringeren Anteil von Peptiden mit Fehlschnittstellen. Zusätzlich zeigen die generierten Peptide ein besseres Fragmentierungsverhalten, festgestellt durch die Verschiebung der Verteilung der Ionenscores in höhere Bereiche.

Selbstverständlich muss in diese Betrachtung mit einfließen, dass die Generierung ArgC-ähnlicher Peptide präparativ aufwendiger ist und die identifizierten Proteine eine im Schnitt deutlich geringere Sequenzabdeckung

aufwiesen. Trotzdem kann diese Verdaustrategie anhand der vorliegenden Ergebnisse von einem nur komplementären Ansatz zu einer ernstzunehmenden Alternative zur konventionellen tryptischen Vorgehensweise aufgewertet werden.

4.5 Anwendbarkeit ArgC-ähnlicher Verdau

Im letzten Abschnitt konnte gezeigt werden, dass der ArgC-ähnliche Verdau von Proteinen komplementäre Resultate zum tryptischen Verdau erzielt und in vielen untersuchten Faktoren sogar zu besseren Resultaten geführt hat. Die Protokolle wurden bislang an einem kommerziell erhältlichen *E.coli* Lysat angewendet, welches keiner Aufreinigung bedarf und nach Zugabe eines geeigneten Puffers direkt verwendet werden kann. Im ersten Teil des folgenden Abschnittes wird eine Methode vorgestellt, die es ermöglicht, die Proteine ausgehend von intakten Zellen zu extrahieren und an die magnetischen Beads zu binden. Zudem soll die Anwendbarkeit des ArgC-ähnlichen Verdau auf das humane Proteom untersucht werden. Anschließend wird die ArgC-ähnliche Verdaustrategie auf Ihre Tauglichkeit in Quantifizierungsansätzen untersucht. Im letzten Teil werden Versuche gezeigt, die einem Multienzym-Ansatz zur Erhöhung der Sequence-Coverage eines untersuchten Proteins folgen.

4.5.1 Anwendbarkeit auf das humane Proteom

Wie in der Einleitung erwähnt, war es zunächst notwendig eine Probenvorbereitung zu entwickeln, die von lebendigen Zellen ausgeht. Voruntersuchungen, die hier nicht weiter aufgezeigt werden, ergaben, dass eine chemische Lyse mit anschließender Acetonpräzipitation zu den besten Ergebnissen geführt hat. Als Lysepuffer wurde der kommerziell erhältliche Bugbuster Mastermix[®] verwendet. Die genaue Zusammensetzung dieses Puffers wird vom Hersteller unter Verschluss gehalten. Es ist lediglich bekannt, dass es sich um einen Tensid-haltigen Puffer handelt mit den Zusätzen Benzonase[®] Nuclease und rLysozyme[™]. Die Nuklease spaltet DNA und dient somit zur Viskositätsreduzierung. Lysozym (auch Muramidase) hydrolysiert die Bindung zwischen N-Acetylmuraminsäure- und N-Acetylglucosaminresten und spielt somit nur eine Bedeutung für die Lyse bakterieller Zellen. Nach 15-minütiger Inkubation der Zellen mit dem Lysepuffer erfolgte eine Acetonfällung. Das Präzipitat wurde dann anschließend in einem geeigneten Puffer gelöst und das Verfahren zur Bindung an die Beads erfolgte bereits wie beschrieben (siehe detaillierte Arbeitsanweisungen im 2.2).

Nach erfolgter Proteinbindung an die Beads wurde die Probe reduziert und alkyliert. Für den Verdau wurden drei unterschiedliche Verfahren angewendet. Zum einen zwei Strategien aus dem letzten Abschnitt (Trypsin und ArgC-like mit Propionsäureanhydrid) und zusätzlich wurde ein Verdau mit dem Enzym LysC durchgeführt. LysC ist eine Protease, die Proteine C-Terminal nach Lysinen schneidet. Sie wurde ursprünglich aus dem Bakterium *Lysobacter enzymogenes* gewonnen und wird heutzutage größtenteils rekombinant biotechnologisch hergestellt.^{133,134} Besonders häufig wird LysC in Kombination mit Trypsin benutzt. Wie im vorherigen Abschnitt gezeigt,

treten 70% der Fehlschnitte beim tryptischen Verdau nach Lysin auf. Durch die Kombination beider Proteasen wird eine deutliche Reduktion dieses Anteils beobachtet.¹³⁵

Nach Tsiatsiani *et al.* ist LysC die zweit häufigst genutzte Protease in der Massenspektrometrie-basierten Proteomforschung.³¹ Aufgrund der Proteasespezifität des LysC kommt es zu deutlich weniger Schnitten als es beim Trypsin der Fall ist. Dies verleiht Ansätzen mit LysC eine gewisse Ähnlichkeit zu denen, die ArgC-ähnlich sind. Durch den zusätzlichen Verdauansatz mit LysC sollen im Folgenden zwei Fragestellungen beantwortet werden: Ist der ArgC-ähnliche Verdau anwendbar auf ein humanes Proteom und wie unterscheiden sich seine Resultate von dem des LysC-Verdaues?

Anzahl der Identifikationen

Die Vermessung der Proben erfolgte mit der gleichen Methode und den selben Apparaturen wie die Vermessung des *Escherichia coli*-Proteoms. Die generierten Thermo RAW-Dateien wurden mit MSConvert zu mgf-Dateien umgewandelt und anschließend mit Hilfe von der Suchmaschine Mascot analysiert. Die Mess- und Suchparameter entsprechen denen aus dem Abschnitt 4.3.2. Die Protein- und Peptidlisten zur weiteren Datenanalyse wurden analog in KNIME erstellt. In der Tabelle 4.12 sind die Anzahl der identifizierten Proteine und Peptide mit den dazu gehörigen *False Discovery Rates* (FDR) gelistet.

Tabelle 4.12: Anzahl der Peptid- und Proteinidentifikationen

| Protokoll | Einmalig identifizierte | | False Discovery Rate | |
|----------------|-------------------------|----------|----------------------|----------|
| | Peptide | Proteine | Peptide | Proteine |
| Trypsin | 16379 | 3185 | 0,28 | 1,23 |
| ArgC-like (PA) | 11591 | 3100 | 0,22 | 0,55 |
| LysC | 7409 | 2223 | 0,24 | 0,67 |

Vergleicht man zunächst nur das tryptische Protokoll mit dem ArgC-ähnlichen, so ist das Phänomen aus dem letzten Abschnitt mit der höheren Anzahl identifizierter Proteine trotz geringerer Anzahl an Peptididentifikationen hier nicht zu beobachten. Das Trypsin-Protokoll generiert ungefähr 5000 zusätzlich identifizierte Peptide, wohingegen die Anzahl identifizierter Proteine unter Hinzunahme der FDR in beiden Protokollen gleichwertig ist. Erstellt man analog einen *Scatterplot* wie in Abschnitt 4.4.1 zur Ermittlung der durchschnittlichen Anzahl zugewiesener PSMs pro Protein in Abhängigkeit des verwendeten Protokolls, ist zu erkennen, dass sich die Verhältnisse gedreht haben. Die Anzahl der zugewiesenen PSMs für die höchstabundanten

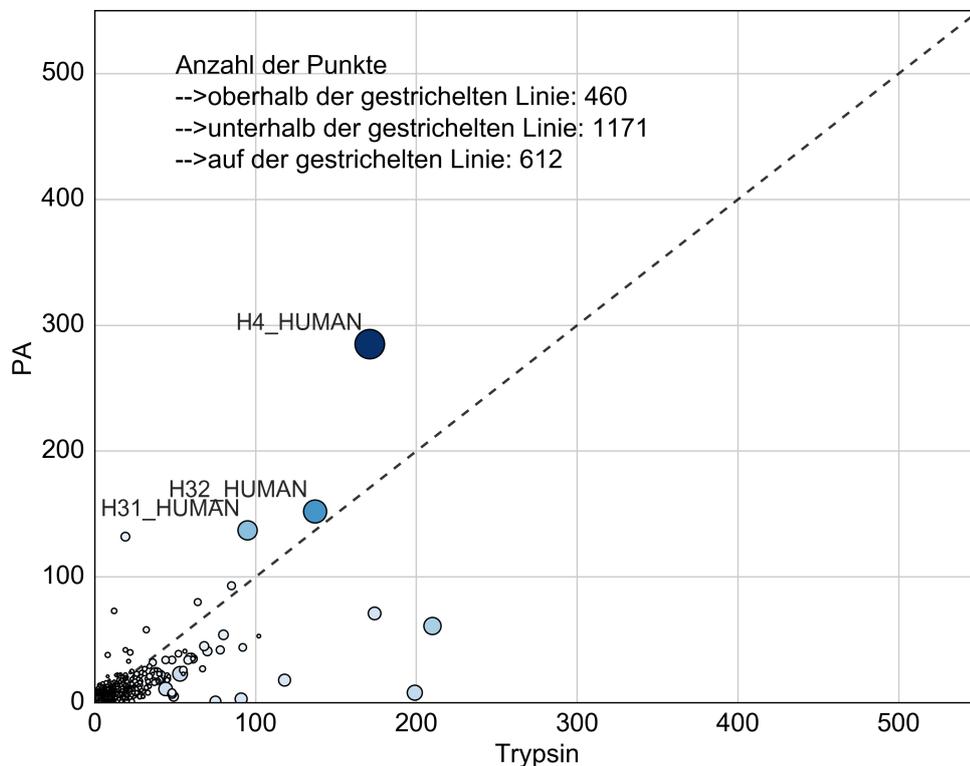


Abbildung 4.43: Scatterplot. x-Achse: Anzahl zugewiesener PSM je Protein für den tryptischen Ansatz; y-Achse: Anzahl zugewiesener PSM je Protein für den ArgC-ähnlichen Ansatz.

Proteine (ermittelt über NSAF, siehe Abschnitt 1.6.2 auf Seite 29) ist beim ArgC-ähnlichen Protokoll in diesem Fall deutlich erhöht. Interessanterweise handelt es sich bei den höchstabundanten Proteinen ausschließlich um Histone, deren bessere Zugänglichkeit mit Hilfe eines ArgC-ähnlichen Verdauens bereits durch die Arbeit von Garcia *et al.* nachgewiesen wurde.⁵² Die bessere Zugänglichkeit eines Proteins für eine Verdaustrategie äußert sich durch eine höhere Anzahl detektierbarer Peptide.

Welches Protokoll für Identifizierungszwecke überlegen ist, entscheidet sich somit durch die Zugänglichkeit der höchstabundanten Proteine. Da in diesem Fall die Histone besser zugänglich sind durch den ArgC-ähnlichen Ansatz und somit durchschnittlich mehr Peptide generiert werden, ist das Trypsin-Protokoll in diesem Fall überlegen. Es ist zusätzlich zu berücksichtigen, dass die Unterschiede in der Proteinidentifikation nicht so stark deutlich werden. Vermutlich ist dies dadurch bedingt, dass die Abundanzunterschiede beim *E. coli* Proteom ausgeprägter waren.

Bei Betrachtung der Ergebnisse der Tabelle 4.12 auf der vorherigen Seite

ist zu erkennen, dass das LysC-Protokoll am schlechtesten abschneidet. Es werden ungefähr 900 Proteine weniger identifiziert und auch die Anzahl identifizierter Peptide ist am geringsten.

Komplementarität der Ergebnisse

Die Komplementarität des LysC- und des ArgC-ähnlichen-Verdaues gegenüber dem tryptischen Verdau werden wie im Abschnitt 4.3.2 auf zwei Ebenen untersucht. Zunächst wird die Komplementarität auf Ebene der Proteinidentifikationen mit Hilfe von Venn-Diagrammen dokumentiert (folgende Abbildung 4.44).

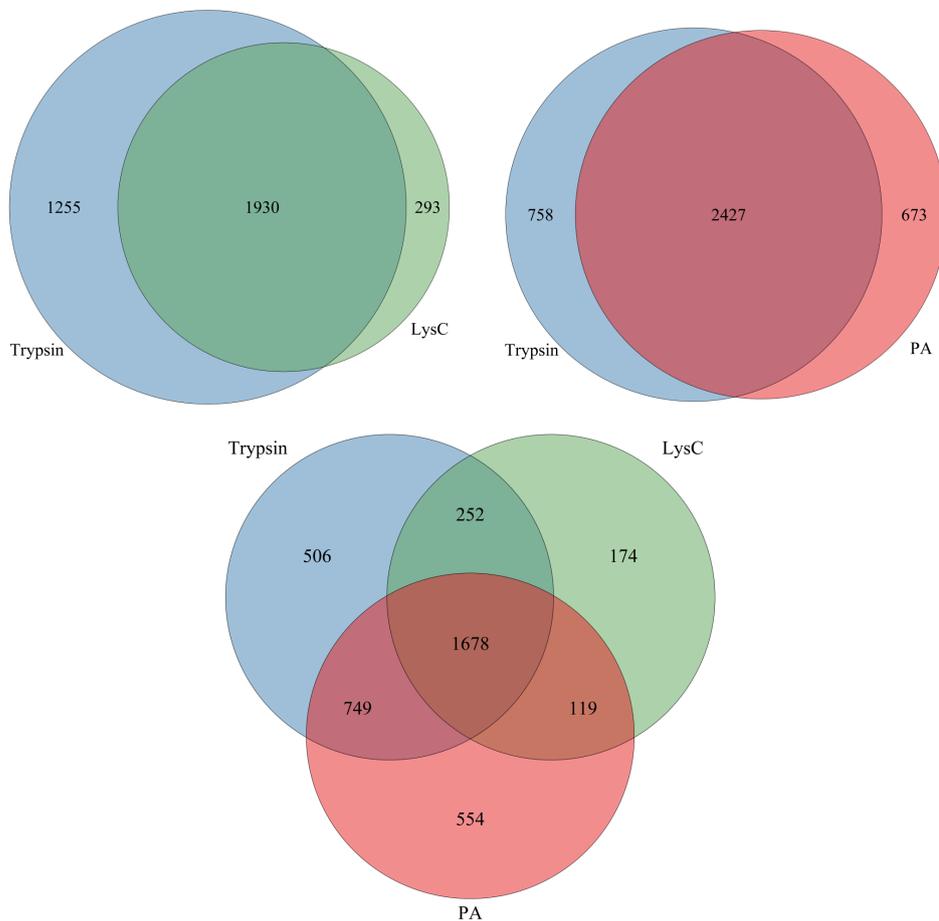


Abbildung 4.44: Venn-Diagramme zur Untersuchung der Komplementarität der verschiedenen Verdauungsansätze auf Ebene der identifizierten Proteine

4. ERGEBNISSE UND DISKUSSION

Im oberen Teil der Grafik werden jeweils zwei Protokolle miteinander verglichen, wobei die Referenzmenge das Resultat des tryptischen Verdauens ist. Beim Vergleich der beiden Venn-Diagramme ist zu erkennen, dass das PA-Protokoll eine höhere Komplementarität aufweist als das LysC-Protokoll. Es können mehr als doppelt so viele zusätzliche Proteine identifiziert werden. Insgesamt ermöglicht das Zusammenfassen der Ergebnisse des LysC- und des Trypsin-Protokolls die Identifikation von 3478 Proteinen. Im Unterschied dazu können nach Zusammenführen der Ergebnisse aus dem ArgC-ähnlichen mit dem tryptischen Verdau 3858 Proteine identifiziert werden. Im unteren Teil der Grafik ist ein Venn-Diagramm zu sehen, welches die Komplementarität aller drei Protokolle aufzeigt. Die Anzahl der exklusiven Identifikationen sind beim Trypsin- und PA-Protokoll ähnlich hoch und bewegen sich in einem Bereich von 500 Proteinen. LysC weist eine deutlich geringere Anzahl exklusiv identifizierter Proteine auf. Alles in allem zeigen der ArgC-ähnliche und der tryptische Verdau die höchste Komplementarität und der Zusatznutzen ist hier am größten.

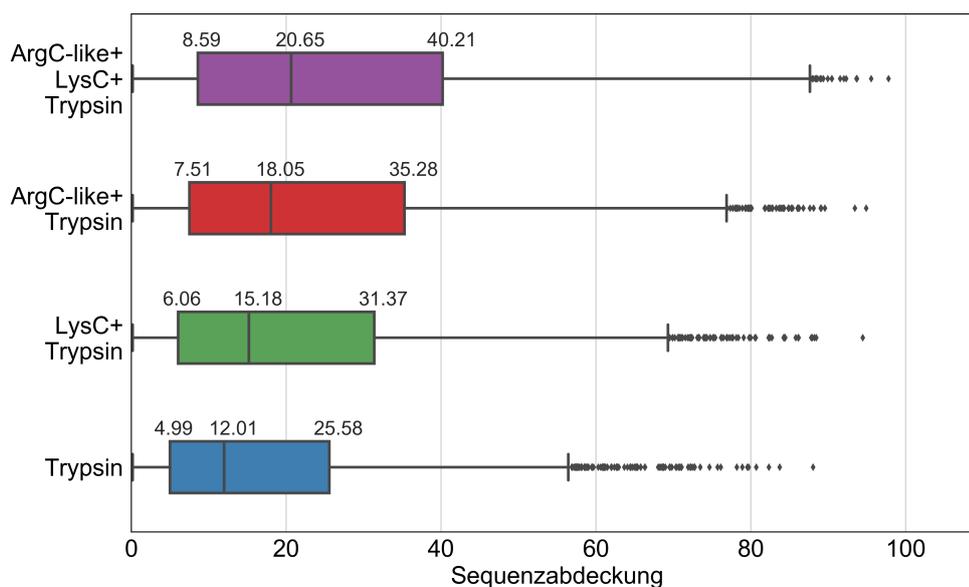


Abbildung 4.45: *Boxplot* zur Visualisierung der Sequenzabdeckungsverteilung und des Zugewinns an Information nach Zusammenführen der Ergebnisse

Die zweite Ebene, auf der sich die Komplementarität der Ergebnisse vergleichen lässt, ist die Sequenzabdeckung der identifizierten Proteine. In der Abbildung 4.45 ist ein *Boxplot* dargestellt, wie er analog in Abschnitt 4.3.2 zu sehen war. Die erste Box von unten zeigt die Verteilung der Sequenzabdeckungen der identifizierten Proteine nach dem tryptischen Verdau. Die zweite und dritte Box zeigen jeweils die Verteilungen nach Hinzunahme der

Sequenzinformationen, die durch den LysC- und den ArgC-like Verdau erhalten wurden. Die letzte Box zeigt wiederum die resultierende Verteilung nach Vereinigung aller Informationen.

Die Verschiebung der Verteilung in höhere Bereiche ist deutlicher ausgeprägt nach Hinzunahme der Informationen, die durch den ArgC-ähnlichen Verdau generiert wurden. Es lässt sich somit zusammenfassen, dass auf beiden Ebenen der Zugewinn zum tryptischen Verdau größer ist bei Verwendung des Arg-ähnlichen Verdau und dass diese Strategie somit bevorzugt eingesetzt werden sollte gegenüber dem LysC-Verdau.

Verdaueffizienz

Der Vergleich der Verdaueffizienz erfolgt analog zum Abschnitt 4.4.2 anhand des Anteils fehlgeschnittener Peptide (MCP, *Missed Cleaved Peptides*). Je höher der Anteil der vollständig geschnittenen Peptide (FCP, *Fully Cleaved Peptides*) ist desto höher ist die Verdaueffizienz. In der Grafik 4.46 sind die Anteile für jede Verdaustrategie mit jeweils einem Kuchendiagramm dargestellt.

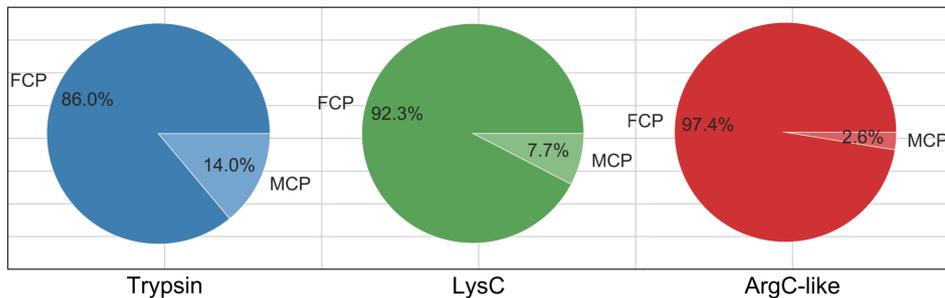


Abbildung 4.46: Anteil fehlgeschnittener Peptida in den verschiedenen Verdau-Ansätzen

Der Anteil fehlgeschnittener Peptide ist beim tryptischen Verdau vergleichbar hoch wie in dem zuvor genannten Abschnitt und stellt in diesem Vergleich mit 14% den höchsten Wert dar. Halb so hohe Werte wurden erhalten beim LysC-Verdau und der niedrigste Wert wurde für den ArgC-ähnliche Verdau ermittelt mit 2,6%. Die höchste Verdaueffizienz weist somit der ArgC-ähnliche Verdau auf, gefolgt vom LysC- und Trypsin-Verdau.

4.5.2 Anwendbarkeit für Quantifizierungsansätze

Neben der Identifikation stellt die Quantifizierung eine weitere wichtige Fragestellung in der Proteomforschung dar. ArgC-ähnliche Verdau-Protokolle

schließen den Einsatzes von magnetischen Beads, welche kritisch sein könnten für Quantifizierungsansätze. Wie bereits im Abschnitt 1.6 diskutiert, können Quantifizierungen in proteomischen Ansätzen absolut und relativ erfolgen. Bei den relativen Verfahren wird das Auftreten unterschiedlicher Proteinabundanzen in unterschiedlichen Proben untersucht. Somit dürfen durch die Probenvorbereitung keine artifiziellen Abundanzunterschiede eingeführt werden. Dies könnte beispielsweise durch eine nicht reproduzierende An- oder Abreicherung von Proteinen an der Beadoberfläche erfolgen.

Die Eignung der Bead-assistierten Probenvorbereitung und des ArgC-ähnlichen Verdauens für Quantifizierungsansätze wurde mit Hilfe eines TMT-Versuches getestet (siehe Abschnitt 1.6.1), der unter den verfügbaren Methoden die höchste Sensitivität zeigt. Verglichen werden die Ergebnisse mit Resultaten eines konventionellen tryptischen Verdauens, der in Lösung erfolgte. Als Probe diente ein E. coli-Proteom, welches über die chemische Lyse mit dem Bugbuster Mastermix[®] und anschließender Acetonfällung gewonnen wurde. Nach dem Verdau erfolgte das TMT-Labeling strikt nach dem Herstellerprotokoll. Die Vermessung der Proben erfolgte über LC-MS/MS unter den in Abschnitt 2.2.9 beschriebenen Bedingungen. Die Identifizierung und Quantifizierung erfolgte mit Hilfe von Mascot. Die Proteinlisten mit den Abundanz-Verhältnissen wurden mit Hilfe Reportfunktion von Mascot erstellt. Die Bedingungen zur Kalkulation der Proteinverhältnisse sind im Anhang C aufgelistet. Die erhaltene CSV-Datei wurde im Anschluss in KNIME zur weiteren Datenanalyse geladen.

Üblicherweise werden für TMT-Experimente Proben genommen, die unterschiedlichen Bedingungen oder Reizen ausgesetzt wurden, die sich in Abundanzunterschieden der Proteine in der Proben bemerkbar machen sollten. Im vorliegenden Versuchsaufbau wurde jedoch die gleiche Probe in verschiedenen Ansätzen mit den unterschiedlichen isobaren Reagenzien umgesetzt. Für beide Ansätze sollten im Idealfall keine Abundanzunterschiede auftreten. Abweichungen vom idealen Verhalten können begründet sein durch das Einbeziehen qualitativ schlechter Daten bzw. von Daten, die außerhalb des linearen Bereichs des Detektors liegen. Da es immer zu Abweichungen vom idealen Verhalten kommt, werden die Resultate des ArgC-ähnlichen mit denen eines konventionell tryptischen Verdauens verglichen.

Die Auswertung erfolgte mit Hilfe von Quantil-Quantil-Diagrammen (*QQ-Plot*). *QQ-Plots* sind Diagramme, in denen die Quantile zweier statistischer Variablen gegeneinander abgetragen werden, um die Verteilung der Ergebnisse zu vergleichen. Eine besondere Form dieser Diagramme wird als *Normal-QQ-Plot* bezeichnet. In diesen Diagrammen sind auf der y-Achse die empirischen Quantile gegen die theoretischen Quantile der Standardnormalverteilung aufgetragen.¹²⁶ Für jeden Verdauansatz wurde ein *Normal-QQ-Plot* erstellt mit den Proteinverhältnissen auf der y-Achse (siehe Abbildung 4.47).

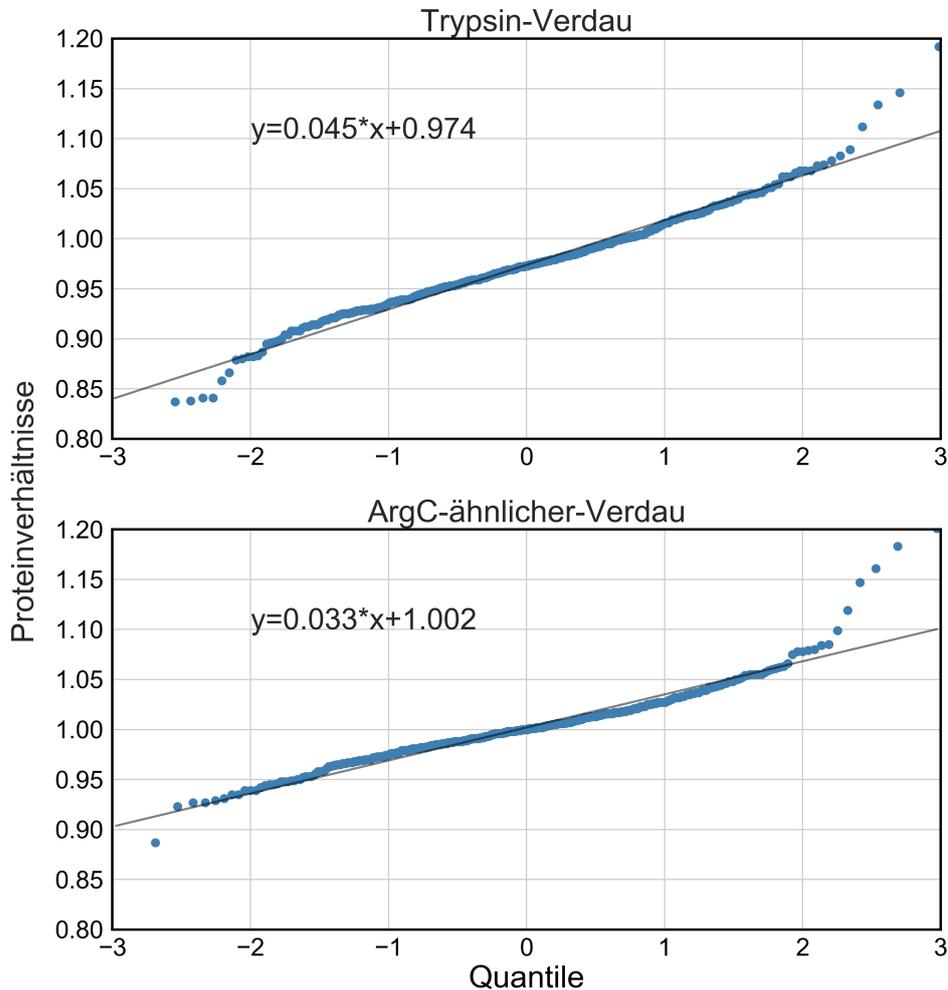


Abbildung 4.47: Normal-QQ-Plot für die Proteinverhältnisse des ArgC-ähnlichen und des tryptischen Ansatzes

Befinden sich die aufgetragenen Punkte zum größten Teil auf einer geraden Linie, sind diese Werte normal verteilt. Dies trifft nach visueller Betrachtung für beide Verteilungen zu. Mit Hilfe eines linearen Regressionsverfahren wurde eine Geradengleichung aufgestellt, deren Steigung ein Maß für die Standardabweichung der Verteilung ist. Je kleiner die Steigung desto schmaler ist die vorliegende Verteilung.

Aus dem Vergleich der beiden Diagramme, ergibt sich, dass der ArgC-ähnliche Verdau für Quantifizierungsansätze geeignet ist und dass die Nutzung der magnetischen Beads keinen messbaren Einfluss auf das Ergebnis hat. Die

aus diesem Versuch erhaltenen Werte sind normal verteilt und weichen vom Mittelwert sogar in einem geringeren Maß ab als der beim konventionellen tryptischen Verdau in Lösung. Der Anteil der Ausreißer ist in beiden Ansätze ähnlich hoch. Die anfänglichen Befürchtungen, dass die Proteine in nicht zu reproduzierender Art und Weise unselektiv an die Beadoberfläche binden, konnten damit entkräftet werden.

4.5.3 Anwendbarkeit für Multi-Enzym-Ansätze

Eine weitere wichtige Fragestellung in der Proteomforschung ist die Charakterisierung der einzelnen Proteine. Die Proteincharakterisierung ist ein sehr weites Feld und kann verschiedene Bereiche umfassen. Je nach Fragestellung liegt der Fokus auf der Sequenzanalyse, der Detektion von Modifikationen oder der Strukturanalyse. In diesem Abschnitt liegt der Fokus auf der Sequenzanalyse und dem Erreichen einer möglichst hohen Sequenzabdeckung mit Hilfe eines proteomischen Bottom-Up-Ansatzes. Eine vollständige Sequenzabdeckung ist dann gegeben, wenn die generierten Peptide eine lückenlose Aufklärung der Aminosäuresequenz des Ursprungproteins und damit auch der chemischen Modifikation zulassen. Es ist allerdings sehr unwahrscheinlich, dass dies mit dem einer einzelnen Protease gelingt, da die ungleichmäßige Verteilung der Schnittstellen häufig zu Peptiden führt, die entweder zu groß oder zu klein sind für eine massenspektrometrische Detektion mit einem einzelnen Gerät.

Eine Lösung dieses Problems bei Untersuchungen einzelner oder weniger Proteine liefern Multi-Enzym-Ansätze, in denen verschiedene Proteasen zum Einsatz kommen und deren Einsatz zusätzlich in der Abfolge optimiert werden kann. Die komplementären Resultate der Proteasen ergeben nach Zusammenführen meist eine sehr hohe Sequenzabdeckung. Proteasen, die häufig eingesetzt werden, sind Trypsin, LysC, LysN, GluC, AspN, Chymotrypsin und ArgC.^{103,104,136} Im Folgenden wird die Tauglichkeit ArgC-ähnlicher Verdau in Multi-Enzym-Ansätzen evaluiert. Weitere Proteasen, die zum Einsatz kommen sind Trypsin und LysC. Bei dem zu charakterisierenden Protein handelt es sich um die Arachidonat-5-Lipoxygenase (5-LO). Diesem Protein wird eine Rolle bei der Pathogenese von Myokardinfarkten und Atherosklerose zugeschrieben. Des Weiteren stellt sie eine potentielle Zielstruktur für die Behandlung von Tumoren dar.¹³⁷ Die Proben werden an der MALDI-Obitrap und dem Voyager vermessen (siehe Abschnitt 2.2.8). Die Rohspektren werden in KNIME-Workflows prozessiert und die Peptididentifikation erfolgt mit Hilfe von Mascot über eine PMF-Suche (siehe Anhang C).

Die Ergebnisse des Multienzym-Ansatzes sind grafisch in den Abbildungen 4.48 und 4.49 dargestellt. Das Protein wird in diesen Abbildungen durch einen Balken repräsentiert. Die ersten Balken von oben bilden ab, an wel-

4.5. Anwendbarkeit ArgC-ähnlicher Verdaue

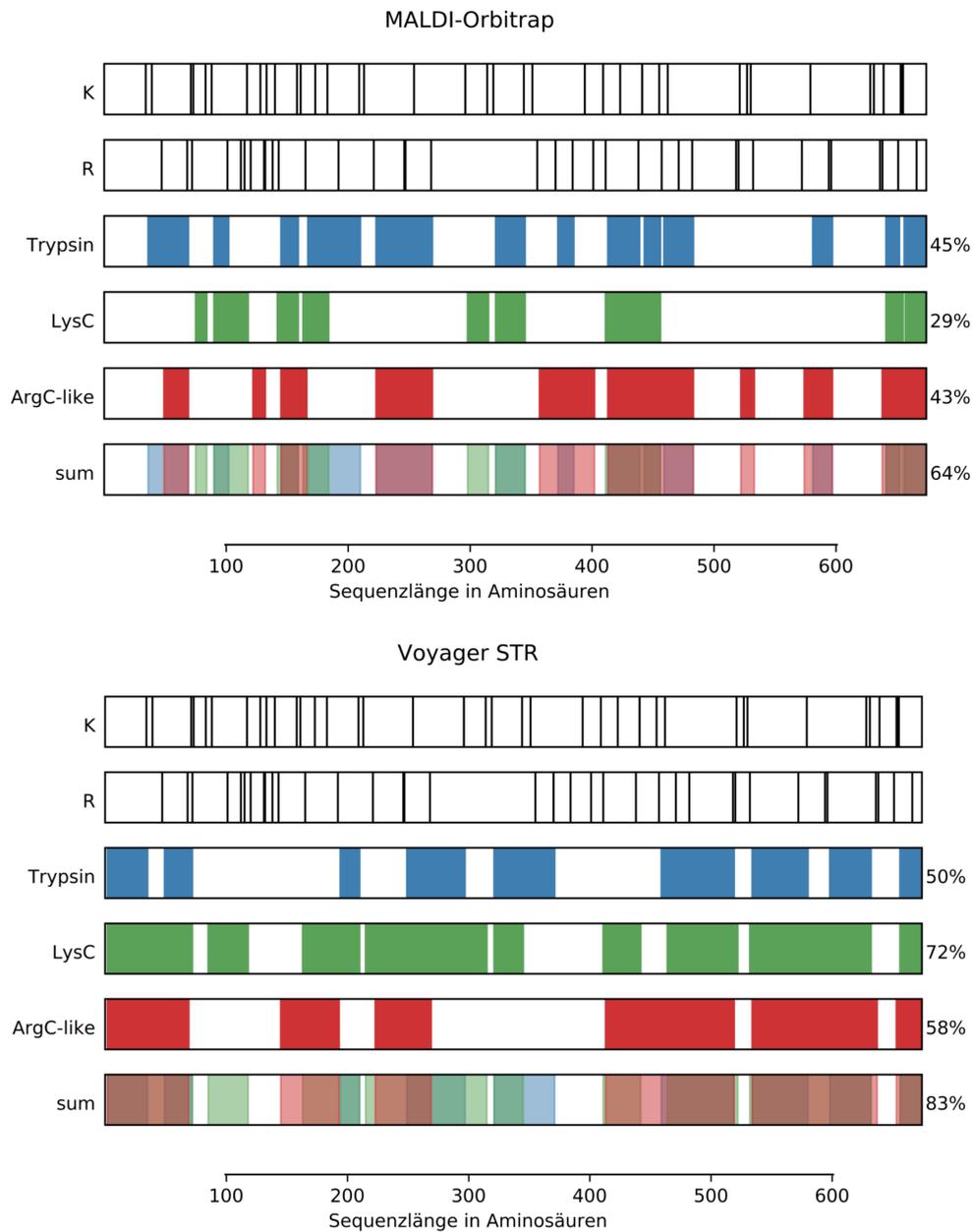


Abbildung 4.48: Übersicht zur erhaltenen Sequenzabdeckung mit verschiedenen Verdauansätzen an verschiedenen Geräten

chen Positionen sich Lysine und Arginine befinden. Die nächsten drei Balken zeigen die Ergebnisse des jeweiligen Verdauansatzes. Die zugehörige prozentuale Sequenzabdeckung befindet sich rechts vom Balken. Der letzte

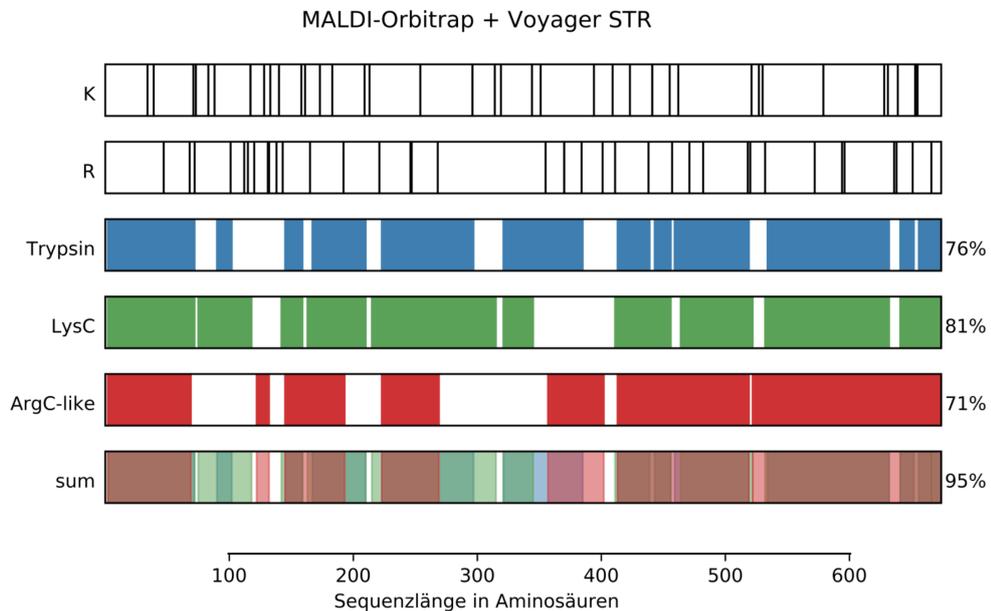


Abbildung 4.49: Übersicht zur erhaltenen Sequenzabdeckung mit verschiedenen Verdauansätzen nach Zusammenführen der Ergebnisse an den verschiedenen Geräten

Balken visualisiert das Ergebnis nach Zusammenführen der Resultate der einzelnen Verdaustrategien.

In der Abbildung 4.48 sind separat die Ergebnisse dargestellt nach Vermessung der generierten Proben an der MALDI-Orbitrap (oben) und dem Voyager STR (unten). Die MALDI-Orbitrap-Messung konnte Peptide erfassen, die sich in einem Massenbereich von 800 - 4000 m/z befanden. Bei Betrachtung der einzelnen Balken ist zu erkennen, dass insbesondere die Sequenzteile mit dem jeweiligen Enzym nicht aufgeklärt werden konnten, die entweder eine zu hohe oder eine zu niedrige Dichte von potentiellen Schnittstellen aufwiesen. Mit Hilfe von LysC konnten speziell die Stellen aufgefüllt werden, die eine hohe Anzahl tryptischer Schnittstellen aufwiesen, deren Anteil an Lysinen jedoch gering war. Analog verhielt es sich mit dem ArgC-ähnlichen Verdau. Sequenzlücken, die aufgrund einer zu geringen Anzahl an tryptischer Schnittstellen entstanden, konnten auch mit Hilfe von LysC- und des ArgC-ähnlichen Verdau nicht aufgeklärt werden. Im Ergebnis konnte durch den Einsatz der verschiedenen Verdaustrategien die Sequenzabdeckung vom tryptischen Verdau mit 45 % auf 64% erhöht werden, wie im letzten Balken zu sehen ist.

Im unteren Teil der Abbildung ist das Ergebnis der Messung der Proben am Voyager STR zu sehen. Die Messung erfolgte innerhalb eines Massenberei-

ches von 2000 - 10'000 m/z. Wie am tryptischen Verdau zu sehen ist, können lediglich die Sequenzbereiche nicht detektiert werden, die einen zu hohen Anteil tryptischer Schnittstellen aufweisen. Die geringen Schnittmöglichkeiten für LysC- und ArgC-like-Verdaue führen in diesen Bereichen zu Peptiden höherer Massen, die detektierbar sind. Aus diesem Grund ist die Erhöhung der Sequenzabdeckung nach Zusammenführen der Ergebnisse der einzelnen Verdaustrategien deutlich ausgeprägt als zuvor. Die Sequenzabdeckung vom tryptischen Verdau konnte in diesem Fall von 50% auf 83% erhöht werden.

Nach Zusammenführen der Ergebnisse aus den Messungen an beiden Geräten, erhält man das Ergebnis, welches in Abbildung 4.49 zu sehen ist. Insgesamt lassen sich somit 95% der Sequenz der 5-LO beschreiben. Die Kombination aus Trypsin-, LysC- und ArgC-ähnlicher Verdaue erwies sich somit als sinnvoller Weg zur Charakterisierung der 5-LO. Des Weiteren ließ sich beobachten, dass die einzelnen Verdaustrategien insbesondere in den Bereichen unbefriedigende Ergebnisse geliefert haben, in denen eine hohe Dichte von tryptischen Schnittstellen vorlag. Die Notwendigkeit an zwei verschiedenen Geräten mit unterschiedlichen Massenbereichen zu messen, entfällt für Geräte mit einer ESI-Quelle, da dort mehrfach geladene Ionen entstehen, die in den detektierbaren Massenbereich fallen.

4.5.4 Diskussion

Nach Entwicklung und Testung des ArgC-ähnlichen Verdaues im letzten Abschnitt, galt es in diesem Abschnitt verschiedene Anwendungsmöglichkeiten herauszuarbeiten. Zu Beginn wurden die Ergebnisse untersucht, nachdem der ArgC-ähnliche Verdau am humanen Proteom angewendet wurde. Die Ergebnisse wurden hierbei mit einem tryptischen und einem LysC-Verdau verglichen. Die Ergebnisse des ArgC-ähnlichen Verdaues wiesen nicht die Überlegenheit gegenüber Trypsin auf, wie sie im letzten Abschnitt vorlagen. Die Anzahl der Proteinidentifikationen war ähnlich hoch, wobei der tryptische Ansatz zu deutlich mehr Peptididentifikationen geführt hat. Des Weiteren konnte gezeigt werden, dass der ArgC-ähnliche Verdau sowohl auf Ebene der identifizierten Proteine als auch auf Ebene der Sequenzabdeckung eine hohe Komplementarität zum tryptischen Verdau aufwies. Im Bereich der Verdaueffizienz konnte die Überlegenheit gegenüber dem tryptischen Verdau erneut gezeigt werden, indiziert durch die geringen Anzahl fehlgeschnittener Peptide. LysC, welches die zweitmeist genutzte Protease ist, lieferte in der Regel in den vergleichenden Untersuchungen die schlechtesten Resultate. In den darauf folgenden Abschnitten konnte die Anwendbarkeit des ArgC-ähnlichen Verdaues für Quantifizierungen und Multienzym-Ansätze gezeigt werden.

Anhang A

Reagenzien und Verbrauchsmaterialien

Tabelle A.1: Reagenzien

| Reagenz | Abkürzung | Hersteller |
|---------------------------------|-----------|---------------------------------------|
| 2-Picolinboran | | Sigma-Aldrich Chemie GmbH, München |
| Aceton | | Merck KGaA, Darmstadt |
| Acetonitril (HPLC grade) | ACN | VWR Chemicals, Darmstadt |
| Ameisensäure | FA | Carl Roth GmbH + Co. KG, Karlsruhe |
| Ammoniumbicarbonat | Ambic | Carl Roth GmbH + Co. KG, Karlsruhe |
| Ammoniumhydroxid -Lösung 25% | | Sigma-Aldrich Chemie GmbH, München |
| ArgC | | Roche AG, Mannheim |
| BugBuster Master Mix | | Merck KGaA, Darmstadt |
| Calciumchlorid-Dihydrat | | Sigma-Aldrich Chemie GmbH, München |
| Diethyl-Pyrocarbonat | DEPC | Sigma-Aldrich Chemie GmbH, München |
| Dithiothreitol | DTT | Carl Roth GmbH + Co. KG, Karlsruhe |
| Formaldehyd 36% | | Sigma-Aldrich Chemie GmbH, München |
| Hydroxylamin 50% (wt/wt) | | Sigma-Aldrich Chemie GmbH, München |

A. REAGENZIEN UND VERBRAUCHSMATERIALIEN

| Reagenz | Abkürzung | Hersteller |
|---|-----------|---|
| Iodacetamid | IAA | Sigma-Aldrich Chemie GmbH, München |
| Methanol | | VWR Chemicals, Darmstadt |
| Propionsäureanhydrid | PA | Sigma-Aldrich Chemie GmbH, München |
| Salzsäure | HCl | Merck KGaA, Darmstadt |
| Sulfo-NHS-Acetate | | Thermo Fisher Scientific GmbH, Dreieich |
| TMTzero/duplex/sixplex | | Thermo Fisher Scientific GmbH, Dreieich |
| Triethylammoniumbicarbonat (1M, pH=8.5) | TEAB | Sigma-Aldrich Chemie GmbH, München |
| Trifluoressigsäure | TFA | Sigma-Aldrich Chemie GmbH, München |
| Tris(hydroxymethyl)-aminomethan | Tris | Carl Roth GmbH + Co. KG, Karlsruhe |
| Trypsin | | Sigma-Aldrich Chemie GmbH, München |
| rLysC | | Promega GmbH, Mannheim |
| α -Cyano-4-hydroxymizsäure | CHCA | Bruker Daltonik GmbH, Bremen |

Tabelle A.2: Verbrauchsmaterialien

| Material | Herrsteller |
|--|---|
| 2.4 μ m Reprosil C18 resin | Dr. Maisch HPLC GmbH, Ammerbuch-Entringen |
| BcMag TM Silica-modifizierten Beads (1 μ m) | Bioclone Inc, San Diego (USA) |
| NHS Mag sepharose beads | GE Healthcare GmbH, Solingen |
| Pierce TM C18 Spin Columns | Thermo Fisher Scientific GmbH, Dreieich |

Anhang B

Software

Tabelle B.1: Programme

| Software | Version | Funktion |
|-----------|----------|--|
| Knime | 3.3.1 | Datenanalyse-Plattform |
| MODDE | 11.0.1 | Programm zum Erstellen und Auswerten von Versuchsplänen (DoE) |
| MSConvert | 3.0.9134 | Konversionstool für massenspektrometrische Rohdateien |
| Mascot | 2.4.2 | Datenbankbasierte Suchmaschine |
| R | 3.2.0 | Freie Programmiersprache für statistische Berechnungen und Grafiken. |
| mMass | 5.5.0 | Programm zur Auswertung einzelner Massenspektren |
| python | 2.7.13 | Programmiersprache |

Mascot Suchparameter

Wie in den Abschnitten 1.5.3 auf Seite 25 und 1.5.4 auf Seite 25 erläutert stellt Mascot verschiedene Methoden zur Protein- und Peptididentifikation zur Verfügung. Die mit MALDI-MS aufgenommenen Spektren wurden mit Hilfe einer PMF-Suche einem Protein zugewiesen und die ESI-MS/MS-Spektren wurden mit Hilfen einer PFF-Suche einem Peptid zugewiesen. Eine erfolgreiche Suche setzt voraus, dass die Suchparameter richtig gewählt wurden. Zu diesen Suchparametern gehört die Auswahl einer Protein-Datenbank, die die Sequenzen der möglichen Proteinkandidaten enthält. Je kleiner der Suchraum ist, umso geringer fallen die falsch positiven Zuweisungen aus. Der Suchraum lässt sich durch die Auswahl der Taxonomie des Organismus, aus dem das Protein stammt sinnvoll verkleinern. Des Weiteren muss die genutzte Protease festgelegt werden und die Anzahl der möglichen Fehlschnitte. Sollten chemische Modifikationen der Aminosäuren bekannt sein, können auch diese entweder als fix oder variabel festgelegt werden.

Instrumentenabhängig werden die Abweichungen der Peptid- und Fragmentmassen festgelegt. Wird eine zu große Abweichung toleriert, könnte die Anzahl der falsch positiven Zuweisungen steigen. Aus diesem Grund steigt die Sensitivität des Suchergebnisses bei Instrumenten mit hohem Auflösungsvermögen. Sie ermöglichen Datenbanksuchen mit sehr kleinen Abweichungen. Ist es gewünscht, dass die Datenbanksuche gegen eine Decoy-Suchmaschine durchgeführt wird und die Ergebnisse im Anschluss percoliert werden, muss auch dies angegeben sein.

Quantifizierungsparameter aus Abschnitt 4.5.2 auf Seite 123

| | |
|-------------------------|-------------------|
| Protein ratio type | weighted |
| Normalisation | average |
| Min. precursor charge | 1 |
| Outlier removal | auto |
| Min. number of peptides | 2 |
| Peptide threshold | at least homology |
| Unique peptides only | no |

Suchparameter aus Abschnitt 4.2 auf Seite 63

Die modifizierten Proteine wurde über eine PMF-Suche identifiziert. Die Datenbank gegen die Suche erfolgte, wurde selbst konstruiert und enthielt nur die Standardproteine, die im Arbeitskreis genutzt werden. Die Auswahl einer Taxonomie war somit auch nicht notwendig.

| Suchparameter | Trypsin | ArgC | PA | NHS | DM | DEPC |
|------------------------|----------------------|----------------------|-------------------------------------|----------------------------------|------------------------------------|--------------------------------|
| Enzyme | Trypsin | ArgC | ArgC | ArgC | ArgC | ArgC |
| Missed cleavage | 0 | 0 | 0 | 0 | 0 | 0 |
| Fixed modifications | Carbamido-methyl (C) | Carbamido-methyl (C) | Carbamido-methyl (C), Propionyl (K) | Carbamido-methyl (C), Acetyl (K) | Carbamido-methyl (C), Dimethyl (K) | Carbamido-methyl (C), DEPC (K) |
| Variable modifications | Oxidation (M) | Oxidation (M) | Oxidation (M), Propionyl (N-term) | Oxidation (M), Acetyl (N-term) | Oxidation (M), Dimethyl (N-term) | Oxidation (M), DEPC (N-term) |
| Peptide tolerance | 5 ppm | 5 ppm | 5 ppm | 5 ppm | 5 ppm | 5 ppm |
| Mass value | mono-isotopic | mono-isotopic | mono-isotopic | mono-isotopic | mono-isotopic | mono-isotopic |

Suchparameter aus Abschnitt 4.3.1 auf Seite 82

| Suchparameter | Trypsin | ArgC | PA | NHS | DM | DEPC |
|------------------------|---------------------|---------------------|------------------------------------|---------------------------------|-----------------------------------|-------------------------------|
| Database | Swissprot | Swissprot | Swissprot | Swissprot | Swissprot | Swissprot |
| Taxonomy | Escherichia coli | Escherichia coli | Escherichia coli | Escherichia coli | Escherichia coli | Escherichia coli |
| Enzyme | Trypsin | ArgC | ArgC | ArgC | ArgC | ArgC |
| Missed cleavage | 1 | 1 | 1 | 1 | 1 | 1 |
| Fixed modifications | Carbamidomethyl (C) | Carbamidomethyl (C) | Carbamidomethyl (C), Propionyl (K) | Carbamidomethyl (C), Acetyl (K) | Carbamidomethyl (C), Dimethyl (K) | Carbamidomethyl (C), DEPC (K) |
| Variable modifications | Oxidation (M) | Oxidation (M) | Oxidation (M), Propionyl (N-term) | Oxidation (M), Acetyl (N-term) | Oxidation (M), Dimethyl (N-term) | Oxidation (M), DEPC (N-term) |
| Peptide tolerance | 5 ppm | 5 ppm | 5 ppm | 5 ppm | 5 ppm | 5 ppm |
| MS/MS tolerance | 0.1 Da | 0.1 Da | 0.1 Da | 0.1 Da | 0.1 Da | 0.1 Da |
| Mass value | mono-isotopic | mono-isotopic | mono-isotopic | mono-isotopic | mono-isotopic | mono-isotopic |
| Decoy | Yes | Yes | Yes | Yes | Yes | Yes |
| Percolator | Yes | Yes | Yes | Yes | Yes | Yes |

Suchparameter aus Abschnitt 4.4 auf Seite 101 (zweiter Suchlauf)

| Suchparameter | Trypsin | ArgC | PA | NHS | DM | DEPC |
|------------------------|----------------------|----------------------|--------------------------------------|-----------------------------------|-------------------------------------|---------------------------------|
| Database | Swissprot | Swissprot | Swissprot | Swissprot | Swissprot | Swissprot |
| Taxonomy | Escherichia coli | Escherichia coli | Escherichia coli | Escherichia coli | Escherichia coli | Escherichia coli |
| Enzyme | Trypsin | ArgC | ArgC | ArgC | ArgC | ArgC |
| Missed cleavage | 8 | 8 | 8 | 8 | 8 | 8 |
| Fixed modifications | Carbamido-methyl (C) | Carbamido-methyl (C) | Carbamido-methyl (C) | Carbamido-methyl (C) | Carbamido-methyl (C) | Carbamido-methyl (C) |
| Variable modifications | Oxidation (M) | Oxidation (M) | Oxidation (M), Propionyl (K, N-term) | Oxidation (M), Acetyl (K, N-term) | Oxidation (M), Dimethyl (K, N-term) | Oxidation (M), DEPC (K, N-term) |
| Peptide tolerance | 5 ppm | 5 ppm | 5 ppm | 5 ppm | 5 ppm | 5 ppm |
| MS/MS tolerance | 0.1 Da | 0.1 Da | 0.1 Da | 0.1 Da | 0.1 Da | 0.1 Da |
| Mass value | mono-isotopic | mono-isotopic | mono-isotopic | mono-isotopic | mono-isotopic | mono-isotopic |
| Decoy | Yes | Yes | Yes | Yes | Yes | Yes |
| Percolator | Yes | Yes | Yes | Yes | Yes | Yes |

Suchparameter aus Abschnitt 4.5.3 auf Seite 126

| Suchparameter | Trypsin | ArgC | PA | NHS | DM | DEPC |
|------------------------|---------------------|---------------------|------------------------------------|---------------------------------|-----------------------------------|-------------------------------|
| Enzyme | Trypsin | ArgC | ArgC | ArgC | ArgC | ArgC |
| Missed cleavage | 2 | 2 | 0 | 2 | 2 | 2 |
| Fixed modifications | Carbamidomethyl (C) | Carbamidomethyl (C) | Carbamidomethyl (C), Propionyl (K) | Carbamidomethyl (C), Acetyl (K) | Carbamidomethyl (C), Dimethyl (K) | Carbamidomethyl (C), DEPC (K) |
| Variable modifications | Oxidation (M) | Oxidation (M) | Oxidation (M), Propionyl (N-term) | Oxidation (M), Acetyl (N-term) | Oxidation (M), Dimethyl (N-term) | Oxidation (M), DEPC (N-term) |
| Peptide tolerance | 1 Da | 1 Da | 1 Da | 1 Da | 1 Da | 1 Da |
| Mass value | average | average | average | average | average | average |

Literaturverzeichnis

- [1] Biomedizinische Massenspektrometrie. <https://campus.uni-muenster.de/hygiene/forschung/forschungsgebiete/biomedizinische-massenspektrometrie/>.
- [2] Collins, F. S.; Green, E. D.; Guttmacher, A. E.; Guyer, M. S.; US National Human Genome Research Institute, A vision for the future of genomics research. *Nature* **2003**, *422*, 835–847.
- [3] Imanishi, T. et al. Integrative annotation of 21,037 human genes validated by full-length cDNA clones. *PLoS biology* **2004**, *2*, e162.
- [4] Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921.
- [5] Jensen, O. N. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Current Opinion in Chemical Biology* **2004**, *8*, 33 – 41.
- [6] Nuhn, P. *Naturstoffchemie. Mikrobielle, pflanzliche und tierische Naturstoffe.*; Hirzel Stuttgart., 1990.
- [7] Fenyö, D.; Qin, J.; Chait, B. T. Protein identification using mass spectrometric information. *ELECTROPHORESIS* **1998**, *19*, 998–1005.
- [8] Tyers, M.; Mann, M. From genomics to proteomics. *Nature* **2003**, *422*, 193–197.
- [9] Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*; W. H. Freeman, 2004.
- [10] Kyte, J.; Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **1982**, *157*, 105–132.

- [11] Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proceedings of the National Academy of Sciences of the United States of America* **1993**, *90*, 5011–5015.
- [12] Mann, M.; Højrup, P.; Roepstorff, P. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biological Mass Spectrometry* **1993**, *22*, 338–345.
- [13] Pappin, D. J.; Højrup, P.; Bleasby, A. J. Rapid identification of proteins by peptide-mass fingerprinting. *Current biology: CB* **1993**, *3*, 327–332.
- [14] James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G. Protein identification by mass profile fingerprinting. *Biochemical and Biophysical Research Communications* **1993**, *195*, 58–64.
- [15] Yates, J. R.; Speicher, S.; Griffin, P. R.; Hunkapiller, T. Peptide Mass Maps: A Highly Informative Approach to Protein Identification. *Analytical Biochemistry* **1993**, *214*, 397–408.
- [16] Taylor, J. A.; Johnson, R. S. Implementation and Uses of Automated de Novo Peptide Sequencing by Tandem Mass Spectrometry. *Analytical Chemistry* **2001**, *73*, 2594–2604.
- [17] Zhang, Z. De Novo Peptide Sequencing Based on a Divide-and-Conquer Algorithm and Peptide Tandem Spectrum Simulation. *Analytical Chemistry* **2004**, *76*, 6374–6383.
- [18] Dancík, V.; Addona, T. A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A. De novo peptide sequencing via tandem mass spectrometry. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* **1999**, *6*, 327–342.
- [19] Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid communications in mass spectrometry: RCM* **2003**, *17*, 2337–2342.
- [20] Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W. Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue. *Proceedings of the National Academy of Sciences* **2002**, *99*, 1774–1779.
- [21] Bogdanov, B.; Smith, R. D. Proteomics by FTICR mass spectrometry: Top down and bottom up. *Mass Spectrometry Reviews* **2005**, *24*, 168–200.

- [22] Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198–207.
- [23] Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chemical reviews* **2013**, *113*, 2343–2394.
- [24] Bodzon-Kulakowska, A.; Bierczynska-Krzysik, A.; Dylag, T.; Drabik, A.; Suder, P.; Noga, M.; Jarzebinska, J.; Silberring, J. Methods for samples preparation in proteomic research. *Journal of Chromatography B* **2007**, *849*, 1–31.
- [25] Jiang, L.; He, L.; Fountoulakis, M. Comparison of protein precipitation methods for sample preparation prior to proteomic analysis. *Journal of Chromatography. A* **2004**, *1023*, 317–320.
- [26] Manza, L. L.; Stamer, S. L.; Ham, A.-J. L.; Codreanu, S. G.; Liebler, D. C. Sample preparation and digestion for proteomic analyses using spin filters. *Proteomics* **2005**, *5*, 1742–1745.
- [27] Wiśniewski, J. R.; Zougman, A.; Mann, M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *Journal of Proteome Research* **2009**, *8*, 5674–5678.
- [28] Hughes, C. S.; Foehr, S.; Garfield, D. A.; Furlong, E. E.; Steinmetz, L. M.; Krijgsveld, J. Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular Systems Biology* **2014**, *10*, 757.
- [29] Gross, E.; Witkop, B. Nonenzymatic Cleavage of Peptide Bonds: The Methionine Residues in Bovine Pancreatic Ribonuclease. *Journal of Biological Chemistry* **1962**, *237*, 1856–1860.
- [30] Li, J.; Shefcheck, K.; Callahan, J.; Fenselau, C. Extension of microwave-accelerated residue-specific acid cleavage to proteins with carbohydrate side chains and disulfide linkages. *International journal of mass spectrometry* **2008**, *278*, 109–113.
- [31] Tsiatsiani, L.; Heck, A. J. R. Proteomics beyond trypsin. *FEBS Journal* **2015**, *282*, 2612–2626.
- [32] Dongré, A. R.; Jones, J. L.; Somogyi, r.; Wysocki, V. H. Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *Journal of the American Chemical Society* **1996**, *118*, 8365–8374.

- [33] Tabb, D. L.; Huang, Y.; Wysocki, V. H.; Yates, J. R. Influence of Basic Residue Content on Fragment Ion Peak Intensities in Low-Energy Collision-Induced Dissociation Spectra of Peptides. *Analytical chemistry* **2004**, *76*, 1243.
- [34] Giansanti, P.; Tsiatsiani, L.; Low, T. Y.; Heck, A. J. R. Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nature Protocols* **2016**, *11*, 993–1006.
- [35] Lucas, F.; Barber, M.; Wolstenholme, W. A.; Geddes, A. J.; Graham, G. N.; Morris, H. R. Mass-spectrometric determination of the amino acid sequences in peptides isolated from the protein silk fibroin of *Bombyx mori*. *The Biochemical Journal* **1969**, *114*, 695–702.
- [36] Ahn, J.; Cao, M.-J.; Yu, Y. Q.; Engen, J. R. Accessing the Reproducibility and Specificity of Pepsin and other Aspartic Proteases. *Biochimica et biophysica acta* **2013**, *1834*, 1222–1229.
- [37] Rietschel, B.; Arrey, T. N.; Meyer, B.; Bornemann, S.; Schuerken, M.; Karas, M.; Poetsch, A. Elastase Digests. *Molecular & Cellular Proteomics : MCP* **2009**, *8*, 1029–1043.
- [38] Morris, H. R.; Batley, K. E.; Harding, N. C.; Bjur, R. A.; Dann, J. G.; King, R. W. Dihydrofolate reductase: low-resolution mass-spectrometric analysis of an elastase digest as a sequencing tool. *The Biochemical Journal* **1974**, *137*, 409–411.
- [39] Rohmer, M.; Baeumlisberger, D.; Karas, M. Advancing the analysis of less specific proteolytic digests via nLC-ESI-QoTOF-MS/MS. *International Journal of Mass Spectrometry* **2013**, *345*, 37–44.
- [40] Boersema, P. J.; Divecha, N.; Heck, A. J. R.; Mohammed, S. Evaluation and Optimization of ZIC-HILIC-RP as an Alternative MudPIT Strategy. *Journal of Proteome Research* **2007**, *6*, 937–946.
- [41] Di Palma, S.; Hennrich, M. L.; Heck, A. J. R.; Mohammed, S. Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *Journal of Proteomics* **2012**, *75*, 3791–3813.
- [42] Boersema, P. J.; Mohammed, S.; Heck, A. J. R. Hydrophilic interaction liquid chromatography (HILIC) in proteomics. *Analytical and Bioanalytical Chemistry* **2008**, *391*, 151–159.
- [43] Wolters, D. A.; Washburn, M. P.; Yates, J. R. An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. *Analytical Chemistry* **2001**, *73*, 5683–5690.

- [44] Rappsilber, J.; Mann, M.; Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols* **2007**, *2*, 1896–1906.
- [45] Rappsilber, J.; Ishihama, Y.; Mann, M. Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. *Analytical Chemistry* **2003**, *75*, 663–670.
- [46] Ishihama, Y.; Rappsilber, J.; Mann, M. Modular stop and go extraction tips with stacked disks for parallel and multidimensional Peptide fractionation in proteomics. *Journal of Proteome Research* **2006**, *5*, 988–994.
- [47] Washburn, M. P.; Wolters, D.; Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology* **2001**, *19*, 242–247.
- [48] Yang, F.; Shen, Y.; Camp, D. G.; Smith, R. D. High pH reversed-phase chromatography with fraction concatenation as an alternative to strong-cation exchange chromatography for two-dimensional proteomic analysis. *Expert Review of Proteomics* **2012**, *9*, 129–134.
- [49] Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. Orthogonality of Separation in Two-Dimensional Liquid Chromatography. *Analytical Chemistry* **2005**, *77*, 6426–6434.
- [50] Strauss, R. L. L. S. *Chemical Reagents for Protein Modification, Third Edition*; CRC Press, 2004.
- [51] Grant, G. A. *Current Protocols in Protein Science*; John Wiley & Sons, Inc., 2001; DOI: 10.1002/cpps.22.
- [52] Garcia, B. A.; Mollah, S.; Ueberheide, B. M.; Busby, S. A.; Muratore, T. L.; Shabanowitz, J.; Hunt, D. F. Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nature Protocols* **2007**, *2*, 933–938.
- [53] Boersema, P. J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J. R. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nature Protocols* **2009**, *4*, 484–494.
- [54] Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K. A.; Hamon, C. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical Chemistry* **2003**, *75*, 1895–1904.

- [55] Gross, J. H. *Massenspektrometrie: Ein Lehrbuch (German Edition)*; Springer Spektrum, 2012.
- [56] Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science (New York, N.Y.)* **1989**, *246*, 64–71.
- [57] Yamashita, M.; Fenn, J. B. Electrospray ion source. Another variation on the free-jet theme. *The Journal of Physical Chemistry* **1984**, *88*, 4451–4459.
- [58] Karas, M.; Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry* **1988**, *60*, 2299–2301.
- [59] Taylor, S. G.; S, F. R. Disintegration of water drops in an electric field. *Proc. R. Soc. Lond. A* **1964**, *280*, 383–397.
- [60] Fenn, J. B. Ion formation from charged droplets: roles of geometry, energy, and time. *Journal of the American Society for Mass Spectrometry* **1993**, *4*, 524–535.
- [61] Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Analytical Chemistry* **2000**, *72*, 1156–1162.
- [62] Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. The Orbitrap: a new mass spectrometer. *Journal of mass spectrometry: JMS* **2005**, *40*, 430–443.
- [63] Kingdon, K. H. A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Physical Review* **1923**, *21*, 408–418.
- [64] Guilhaus, M. Special feature: Tutorial. Principles and instrumentation in time-of-flight mass spectrometry. Physical and instrumental concepts. *Journal of Mass Spectrometry* **1995**, *30*, 1519–1532.
- [65] Bahr, U.; Stahl-Zeng, J.; Gleitsmann, E.; Karas, M. Delayed extraction time-of-flight MALDI mass spectrometry of proteins above 25000 Da. *Journal of Mass Spectrometry* **1997**, *32*, 1111–1116.
- [66] Vestal, M. L.; Juhasz, P.; Martin, S. A. Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* **1995**, *9*, 1044–1050.
- [67] Roepstorff, P. *Proteomics in Functional Genomics*; EXS; Birkhäuser, Basel, 2000; pp 81–97, DOI: 10.1007/978-3-0348-8458-7_6.

- [68] Mass Spectrometry Fundamental LC-MS Orbitrap Mass Analyzers. http://www.chromacademy.com/lms/sco156/Fundamental_LC-MS_Orbitrap_Mass_Analyzers.pdf.
- [69] Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry* **1994**, *5*, 976–989.
- [70] McLafferty, F. W. Tandem mass spectrometry. *Science* **1981**, *214*, 280–287.
- [71] Roepstorff, P.; Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical Mass Spectrometry* **1984**, *11*, 601.
- [72] Johnson, R. S.; Martin, S. A.; Biemann, K. Collision-induced fragmentation of (M + H)⁺ ions of peptides. Side chain specific sequence ions. *International Journal of Mass Spectrometry and Ion Processes* **1988**, *86*, 137–154.
- [73] Brechi, L. A.; Tabb, D. L.; Yates, J. R.; Wysocki, V. H. Cleavage N-terminal to proline: analysis of a database of peptide tandem mass spectra. *Analytical Chemistry* **2003**, *75*, 1963–1971.
- [74] Deutsch, E. mzML: A single, unifying data format for mass spectrometer output. *PROTEOMICS* **2008**, *8*, 2776–2777.
- [75] Holman, J. D.; Tabb, D. L.; Mallick, P. *Current Protocols in Bioinformatics*; John Wiley & Sons, Inc., 2002; DOI: 10.1002/0471250953.bi1324s46.
- [76] Adusumilli, R.; Mallick, P. *Proteomics; Methods in Molecular Biology*; Humana Press, New York, NY, 2017; pp 339–368, DOI: 10.1007/978-1-4939-6747-6_23.
- [77] Yang, C.; He, Z.; Yu, W. Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis. *BMC Bioinformatics* **2009**, *10*, 4.
- [78] Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *ELECTROPHORESIS* **1999**, *20*, 3551–3567.
- [79] Nesvizhskii, A. I. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *Journal of Proteomics* **2010**, *73*, 2092–2123.

- [80] López-Ferrer, D.; Martínez-Bartolomé, S.; Villar, M.; Campillos, M.; Martín-Maroto, F.; Vázquez, J. Statistical Model for Large-Scale Peptide Identification in Databases from Tandem Mass Spectra Using SEQUEST. *Analytical Chemistry* **2004**, *76*, 6853–6860.
- [81] Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. Open Mass Spectrometry Search Algorithm. *Journal of Proteome Research* **2004**, *3*, 958–964.
- [82] Craig, R.; Beavis, R. C. A method for reducing the time required to match protein sequences with tandem mass spectra. *Rapid Communications in Mass Spectrometry* **2003**, *17*, 2310–2316.
- [83] Craig, R.; Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics (Oxford, England)* **2004**, *20*, 1466–1467.
- [84] Tabb, D. L.; Fernando, C. G.; Chambers, M. C. MyriMatch: Highly Accurate Tandem Mass Spectral Peptide Identification by Multivariate Hypergeometric Analysis. *Journal of Proteome Research* **2007**, *6*, 654–661.
- [85] Kim, S.; Gupta, N.; Pevzner, P. A. Spectral Probabilities and Generating Functions of Tandem Mass Spectra: A Strike against Decoy Databases. *Journal of Proteome Research* **2008**, *7*, 3354–3363.
- [86] Kim, S.; Mischerikow, N.; Bandeira, N.; Navarro, J. D.; Wich, L.; Mohammed, S.; Heck, A. J. R.; Pevzner, P. A. The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search. *Molecular & cellular proteomics: MCP* **2010**, *9*, 2840–2852.
- [87] Dorfer, V.; Pichler, P.; Stranzl, T.; Stadlmann, J.; Taus, T.; Winkler, S.; Mechtler, K. MS Amanda, a Universal Identification Algorithm Optimized for High Accuracy Tandem Mass Spectra. *Journal of Proteome Research* **2014**, *13*, 3679–3684.
- [88] Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of Proteome Research* **2011**, *10*, 1794–1805.
- [89] Shevchenko, A.; Jensen, O. N.; Podtelejnikov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Shevchenko, A.; Boucherie, H.; Mann, M. Linking genome and proteome by mass spectrometry: Large-scale identification of yeast proteins from two dimensional gels. *Proceedings of the National Academy of Sciences* **1996**, *93*, 14440–14445.

- [90] Smith, J. C.; Macklin, A. M. Chapter 5: Modern techniques in quantitative proteomics. 2016; <https://www.bioanalysis-zone.com/2016/02/08/chapter-5-modern-techniques-in-quantitative-proteomics/>.
- [91] Beynon, R. J.; Doherty, M. K.; Pratt, J. M.; Gaskell, S. J. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nature Methods* **2005**, *2*, 587–9.
- [92] Pratt, J. M.; Simpson, D. M.; Doherty, M. K.; Rivers, J.; Gaskell, S. J.; Beynon, R. J. Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nature Protocols* **2006**, *1*, 1029–1043.
- [93] Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 6940–6945.
- [94] Wong, J. W. H.; Cagney, G. *Proteome Bioinformatics; Methods in Molecular Biology*TM; Humana Press, 2010; pp 273–283, DOI: 10.1007/978-1-60761-444-9_18.
- [95] Zybailov, B.; Mosley, A.; Sardu, M.; Coleman, M.; Florens, L.; Washburn, M. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *Journal of Proteome Research* **2006**, *5*, 2339–2347.
- [96] Bubis, J. A.; Levitsky, L. I.; Ivanov, M. V.; Tarasova, I. A.; Gorshkov, M. V. Comparative evaluation of label free quantification methods for shotgun proteomics. *Rapid Communications in Mass Spectrometry* **2017**, *31*, 606–612.
- [97] Michalski, A.; Damoc, E.; Hauschild, J.-P.; Lange, O.; Wiegand, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S. Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Molecular & Cellular Proteomics* **2011**, *10*, M111.011015.
- [98] Smith, P. K.; Krohn, R. I.; Hermanson, G. T.; Mallia, A. K.; Gartner, F. H.; Provenzano, M. D.; Fujimoto, E. K.; Goeke, N. M.; Olson, B. J.; Klenk, D. C. Measurement of protein using bicinchoninic acid. *Analytical Biochemistry* **1985**, *150*, 76–85.
- [99] Willard, B. B.; Kinter, M. Effects of the position of internal histidine residues on the collision-induced fragmentation of triply protonated tryptic peptides. *Journal of the American Society for Mass Spectrometry* **2001**, *12*, 1262–1271.

- [100] She, Y.-M.; Rosu-Myles, M.; Walrond, L.; Cyr, T. D. Quantification of protein isoforms in mesenchymal stem cells by reductive dimethylation of lysines in intact proteins. *PROTEOMICS* **2012**, *12*, 369–379.
- [101] Karas, M.; Ehring, H.; Nordhoff, E.; Stahl, B.; Strupat, K.; Hillenkamp, F.; Grehl, M.; Krebs, B. Matrix-assisted laser desorption/ionization mass spectrometry with additives to 2,5-dihydroxybenzoic acid. *Organic Mass Spectrometry* **28**, 1476–1481.
- [102] Käll, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature Methods* **2007**, *4*, 923–925.
- [103] Swaney, D. L.; Wenger, C. D.; Coon, J. J. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *Journal of Proteome Research* **2010**, *9*, 1323–1329.
- [104] Choudhary, G.; Wu, S.-L.; Shieh, P.; Hancock, W. S. Multiple enzymatic digestion for enhanced sequence coverage of proteins in complex proteomic mixtures using capillary LC with ion trap MS/MS. *Journal of Proteome Research* **2003**, *2*, 59–67.
- [105] Guo, X.; Trudgian, D. C.; Lemoff, A.; Yadavalli, S.; Mirzaei, H. Confetti: a multiprotease map of the HeLa proteome for comprehensive proteomics. *Molecular & cellular proteomics: MCP* **2014**, *13*, 1573–1584.
- [106] Schilling, O.; Barré, O.; Huesgen, P. F.; Overall, C. M. Proteome-wide analysis of protein carboxy termini: C terminomics. *Nature Methods* **2010**, *7*, 508–511.
- [107] McDonald, L.; Beynon, R. J. Positional proteomics: preparation of amino-terminal peptides as a strategy for proteome simplification and characterization. *Nature Protocols* **2006**, *1*, 1790–1798.
- [108] Miles, E. W. *Methods in Enzymology*; Enzyme Structure Part E; Academic Press, 1977; Vol. 47; pp 431–442, DOI: 10.1016/0076-6879(77)47043-5.
- [109] Mühlrad, A.; Hegyi, G.; Horányi, M. Studies on the properties of chemically modified actin. 3. Carbethoxylation. *Biochimica Et Biophysica Acta* **1969**, *181*, 184–190.
- [110] Strohalm, M.; Kavan, D.; Novák, P.; Volný, M.; Havlíček, V. mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data. *Analytical Chemistry* **2010**, *82*, 4648–4651.

- [111] Orchard, S.; Taylor, C. F.; Hermjakob, H.; Weimin-Zhu, J.; Julian, R. K.; Apweiler, R. Advances in the development of common interchange standards for proteomic data. *PROTEOMICS* **2004**, *4*, 2363–2365.
- [112] Lin, S. M.; Zhu, L.; Winter, A. Q.; Sasinowski, M.; Kibbe, W. A. What is mzXML good for? *Expert Review of Proteomics* **2005**, *2*, 839–845.
- [113] Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **2008**, *24*, 2534–2536.
- [114] Falkner, J. A.; Falkner, J. W.; Andrews, P. C. ProteomeCommons.org IO Framework: reading and writing multiple proteomics data formats. *Bioinformatics* **2007**, *23*, 262–263.
- [115] Mascot Parser | an object-oriented API to Mascot result files and configuration files. <http://www.matrixscience.com/msparser.html>.
- [116] Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Thiel, K.; Wiswedel, B. KNIME - the Konstanz Information Miner: Version 2.0 and Beyond. *SIGKDD Explor. Newsl.* **2009**, *11*, 26–31.
- [117] Sturm, M.; Bertsch, A.; Gröpl, C.; Hildebrandt, A.; Hussong, R.; Lange, E.; Pfeifer, N.; Schulz-Trieglaff, O.; Zerck, A.; Reinert, K.; Kohlbacher, O. OpenMS – An open-source software framework for mass spectrometry. *BMC Bioinformatics* **2008**, *9*, 163.
- [118] Aiche, S.; Sachsenberg, T.; Kenar, E.; Walzer, M.; Wiswedel, B.; Kristl, T.; Boyles, M.; Duschl, A.; Huber, C. G.; Berthold, M. R.; Reinert, K.; Kohlbacher, O. Workflows for automated downstream data analysis and visualization in large-scale computational mass spectrometry. *PROTEOMICS* **2015**, *15*, 1443–1447.
- [119] Demeure, K.; Quinon, L.; Gabelica, V.; De Pauw, E. Rational Selection of the Optimum MALDI Matrix for Top-Down Proteomics by In-Source Decay. *Analytical Chemistry* **2007**, *79*, 8678–8685.
- [120] Köcher, T.; Engström, k.; Zubarev, R. A. Fragmentation of Peptides in MALDI In-Source Decay Mediated by Hydrogen Radicals. *Analytical Chemistry* **2005**, *77*, 172–177.
- [121] Schnaible, V.; Wefing, S.; Resemann, A.; Suckau, D.; Bücken, A.; Wolf-Kümmeth, S.; Hoffmann, D. Screening for Disulfide Bonds in Proteins by MALDI In-Source Decay and LIFT-TOF/TOF-MS. *Analytical Chemistry* **2002**, *74*, 4980–4988.

- [122] Siebertz, K.; van Bebber, D.; Hochkirchen, T. *Statistische Versuchsplannung - Design of Experiments (DoE)*; Springer, 2010.
- [123] Nebrich, G.; Herrmann, M.; Sagi, D.; Klose, J.; Giavalisco, P. High MS-compatibility of silver nitrate-stained protein spots from 2-DE gels using ZipPlates and AnchorChips for successful protein identification. *ELECTROPHORESIS* **2007**, *28*, 1607–1614.
- [124] MODDE Pro Umetrics. <https://umetrics.com/product/modde-pro>.
- [125] Eriksson, L. *Design of Experiments: Principles and Applications*; Umetrics AB, 2008.
- [126] Fahrmeir, L.; Heumann, C.; Künstler, R.; Pigeot, I.; Tutz, G. *Statistik*; Springer-Lehrbuch; Springer Berlin Heidelberg: Berlin, Heidelberg, 2016; DOI: 10.1007/978-3-662-50372-0.
- [127] Glatter, T.; Ludwig, C.; Ahrné, E.; Aebersold, R.; Heck, A. J. R.; Schmidt, A. Large-Scale Quantitative Assessment of Different In-Solution Protein Digestion Protocols Reveals Superior Cleavage Efficiency of Tandem Lys-C/Trypsin Proteolysis over Trypsin Digestion. *Journal of Proteome Research* **2012**, *11*, 5145–5156.
- [128] Gershon, P. D. Cleaved and Missed Sites for Trypsin, Lys-C, and Lys-N Can Be Predicted with High Confidence on the Basis of Sequence Context. *Journal of Proteome Research* **2014**, *13*, 702–709.
- [129] Siepen, J. A.; Keevil, E.-J.; Knight, D.; Hubbard, S. J. Prediction of Missed Cleavage Sites in Tryptic Peptides Aids Protein Identification in Proteomics. *Journal of Proteome Research* **2007**, *6*, 399–408.
- [130] Thiede, B.; Lamer, S.; Mattow, J.; Siejak, F.; Dimmler, C.; Rudel, T.; Jungblut, P. R. Analysis of missed cleavage sites, tryptophan oxidation and N-terminal pyroglutamylation after in-gel tryptic digestion. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 496–502.
- [131] Keil, B. *Specificity of Proteolysis*; Springer Berlin Heidelberg: Berlin, Heidelberg, 1992; pp 43–228, DOI: 10.1007/978-3-642-48380-6_5.
- [132] Colaert, N.; Helsens, K.; Martens, L.; Vandekerckhove, J.; Gevaert, K. Improved visualization of protein consensus sequences by iceLogo. *Nature Methods* **2009**, *6*, 786–787.
- [133] Jekel, P. A.; Weijer, W. J.; Beintema, J. J. Use of endoproteinase Lys-C from *Lysobacter enzymogenes* in protein sequence analysis. *Analytical Biochemistry* **1983**, *134*, 347–354.

- [134] Shah, H. N., Gharbia, S. E., Eds. *Mass Spectrometry for Microbial Proteomics: Shah/Mass Spectrometry for Microbial Proteomics*; John Wiley & Sons, Ltd: Chichester, UK, 2010; DOI: 10.1002/9780470665497.
- [135] Saveliev, S.; Bratz, M.; Zubarev, R.; Szapacs, M.; Budamgunta, H.; Urh, M. Trypsin/Lys-C protease mix for enhanced protein mass spectrometry analysis. *Nature Methods* **2013**, *10*.
- [136] Wu, F.; Zhao, M.; Xiong, Z.; Xu, P. [Application of optimized multi-enzyme combination and sample pretreatment in proteomics]. *Sheng Wu Gong Cheng Xue Bao = Chinese Journal of Biotechnology* **2016**, *32*, 306–316.
- [137] Roos, J.; Oancea, C.; Heinssmann, M.; Khan, D.; Held, H.; Kahnt, A. S.; Capelo, R.; Buscató, E. I.; Proschak, E.; Puccetti, E.; Steinhilber, D.; Fleming, I.; Maier, T. J.; Ruthardt, M. 5-Lipoxygenase Is a Candidate Target for Therapeutic Management of Stem Cell-like Cells in Acute Myeloid Leukemia. *Cancer Research* **2014**, *74*, 5244–5255.

Abkürzungsverzeichnis

| | |
|--------|---|
| 5-LO | 5-Lipoxygenase |
| BCA | Bicinchoninsäure |
| CE | Capillar Electrophoresis |
| CID | Collision Induced Dissociation |
| DE | Delayed Extraction |
| DEPC | Diethylpyrocarbonat |
| DM | Dimethyl |
| DMSO | Dimethylsulfoxid |
| DNA | Deoxyribonucleic acid |
| DOE | Design of Experiment |
| DTT | Dithiothreitol |
| EI | Elektronenstoss Ionisation |
| ESI | Elektrospray Ionisation |
| FASP | Filter assisted sample preparation |
| FDR | False Discovery Rate |
| FT-ICR | Fourier-Transform Ion Cyclotron Resonance |
| FWHM | Full Width at Half Maximum |
| GRAVY | Grand Averages of Hydrophobicity |
| HCD | Higher-energy collisional dissociation |
| HPLC | High Performance Liquid Chromatography |
| IAA | Iodacetamid |

ABKÜRZUNGSVERZEICHNIS

| | |
|--------|--|
| ISD | Ion Source Decay |
| LC | Liquid Chromatography |
| LOF | Lack of Fit |
| MALDI | Matrix-Assisted Laser Desorption/Ionization |
| MS | Massenspektrometrie |
| MudPIT | Multidimensional Protein Identification |
| NHS | N-Hydroxysuccinimide |
| NSAF | Normalized Spectral abundance Factor |
| PA | Propionsäureanhydrid |
| PFF | Peptidfragmentmassen-Fingerprint |
| PMF | Peptidmassen-Fingerprint |
| PSD | Post Source Decay |
| PSM | Peptide to Spectra Matches |
| PTM | Posttranslationale Modifikation |
| RP | Reversed Phase |
| SAX | Strong Anion Exchange |
| SCX | Strong Cation Exchange |
| SP3 | Single-Pot Solid-Phase-enhanced Sample preparation |
| TCA | Trichloressigsäure |
| TCEP | Tris(2-carboxyethyl)phosphin |
| TEAB | Triethylammoniumbicarbonat |
| TMT | Tandem Mass Tag |
| TOF | Time of Flight |
| UPLC | Ultra Performance Liquid Chromatography |

Eidesstattliche Versicherung

Ich erkläre hiermit an Eides Statt, dass ich die vorgelegte Dissertation über „Etablierung eines ArgC-analogen Verdaues für Bottom-Up-Ansätze in der Proteomforschung“ selbständig angefertigt und mich anderer Hilfsmittel als der in ihr angegebenen nicht bedient habe, insbesondere, dass alle Entlehnungen aus anderen Schriften mit Angabe der betreffenden Schrift gekennzeichnet sind.

Ich versichere, nicht die Hilfe einer kommerziellen Promotionsvermittlung in Anspruch genommen zu haben.

Frankfurt am Main, den 27.10.2018



Vahid Golghalyani

Lebenslauf und Publikationsliste

Angaben zur Person

Name Vahid Golghalyani
Geburtsdatum 13.01.1987
Geburtsort Düsseldorf
Nationalität Deutschland
Familienstand Ledig

Promotionsstudium

Beginn Januar 2013
Titel Etablierung eines ArgC-analogen Verdaues für Bottom-Up-Ansätze in der Proteomforschung
Betreuer Prof. Dr. Michael Karas
Institut Institut für Pharmazeutische Chemie, Goethe-Universität-Frankfurt.

Hochschulbildung

2009–2012 **Pharmazie**, *Goethe-Universität*, Frankfurt am Main, 3. Staatsexamen.
2007–2009 **Pharmazie**, *Universität des Saarlandes*, Saarbrücken, 1. Staatsexamen.
2006–2007 **Mathematik**, *Freie Universität*, Berlin.

Schulbildung

1999–2006 **Gymnasium**, *Luise-Henritte-Oberschule*, Berlin, Abitur.
1993–1997 **Grundschule**, *Rudolf-Hildebrandt-Grundschule*, Berlin.

Publikationen als Erstautor

1. **ArgC-like Digestion: Complementary or Alternative to Tryptic Digestion.** *Journal of proteome research* (2017). Vahid Golghalyani, Moritz Neupärtl, Ilka Wittig, Ute Bahr, Micheal Karas.

■ Weitere Publikationen

1. **Enhanced on-plate digestion of proteins using a MALDI-digestion chamber.** *International Journal of Mass Spectrometry (2016)*. Michael Rühl, Vahid Golghalyani, Günes Barka, Ute Bahr, Michael Karas.
2. **Michael acceptor containing drugs are a novel class of 5-lipoxygenase inhibitor targeting the surface cysteines C416 and C418.** *Biochemical Pharmacology (2016)*. Isabelle V. Maucher, Michael Rühl, Simon B.M. Kreschmer,...,Vahid Golghalyani,...,Jessica Roos, Thorsten J. Maier.
3. **APP Deletion Accounts for Age-Dependent Changes in the Bioenergetic Metabolism and in Hyperphosphorylated CaMKII at Stimulated Hippocampal Presynaptic Active Zones.** *Front. Synaptic Neurosci. (2017)*. Melanie Laßek, Jens Weingarten, Moritz Neupärtl, Tabiwang N. Array, Eva Harde, Benedikt Beckert, Vahid Golghalyani, Jörg Ackermann, Ina Koch, Ulrike C. Müller, Michael Karas, Amparo Acker-Palmer, Walter Volkmandt.

■ Vorträge

1. **ArgC-like Digestion: Complementary or Alternative to Tryptic Digestion.** Vahid Golghalyani, Ute Bahr, Ilka Wittig und Michael Karas. (DGMS 2017 in Kiel)

■ Posterpräsentationen

1. **Quatitative and Specific Derivatization of Arginine Residues by an Improved Protocol using Malondialdehyde.** Vahid Golghalyani, Marion Bäumlisberger, Daniel Flesch und Michael Karas. (DGMS 2014 in Frankfurt)
2. **Comparison of Peptide Separation Techniques using Strong Cation Exchange Materials.** Vahid Golghalyani, Marion Bäumlisberger, Ute Bahr und Michael Karas. (DGMS 2015 in Wuppertal)
3. **Comparative Study on Chemical Derivatization of Lysines to mimic Arg-C digestion.** Vahid Golghalyani, Ute Bahr, Ilka Wittig und Michael Karas. (ASMS 2016 in San Antonio, Texas)