

Methodenentwicklung für die
Proteomanalyse
proteolytischer Prozesse

Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von
Thomas Jakoby-Gregorius
2014

Erster Gutachter:

Prof. Dr. Andreas Tholey

Zweite Gutachterin:

Prof. Dr. Thisbe K. Lindhorst

Tag der mündlichen Prüfung:

31.01.2014

Zum Druck genehmigt:

07.02.2014

gez.

Prof. Dr. Wolfgang J. Duschl, Dekan

Inhaltsverzeichnis

Inhaltsverzeichnis	I
Abkürzungen und Symbole	IV
Zusammenfassung	VIII
Abstract	X
Kapitel I: Theoretischer Teil – Einleitung:	1
1.2 Die Proteomik	1
1.2 Die quantitative Proteomik	4
1.2.1 Quantifizierung auf MS-Ebene.....	5
1.2.2 Quantifizierung auf MS/MS-Ebene.....	7
1.3 Hochleistungs-Flüssigkeitschromatographie für die Auftrennung von Peptid- bzw. Proteingemischen	10
1.3.1 Umkehrphasen- und Ionenpaar-Umkehrphasen-HPLC.....	12
1.3.2 Stationäre Phasen für die Trennung von Peptid- bzw. Proteingemischen in Verbindung mit Mikro-Hochleistungs-Flüssigkeitschromatographie.....	14
1.4 Massenspektrometrie in der Peptid- bzw. Proteinanalytik	15
1.4.1 Die MALDI-Technik.....	17
1.4.2 Der Flugzeit-Massenanalysator.....	21
1.4.3 Die Vielkanalplatte als Detektor.....	23
1.4.4 LC-MALDI MS.....	24
1.4.5 Peptid- bzw. Proteinidentifizierung.....	25
1.5 Analytik von Proteasen und Proteasen-katalysierten Reaktionen	28
1.5.1 Chemische und molekular-biologische Methoden der Proteaseforschung.....	30
1.5.2 Massenspektrometrisch (MS)-basierte Methoden der Proteaseforschung.....	31
Kapitel II: Zielsetzung	36
Kapitel III: Der Q-PICS Ansatz - Methodenentwicklung und MALDI-TOF/TOF MS/MS Parameteroptimierung	37

3.1 Einleitung	37
3.2 Material und Methoden	38
3.2.1 Arbeiten mit einem 10-Proteinstandard	38
3.2.2 Arbeiten mit zytosolischem Hefeproteom.....	44
3.3 Ergebnisse	49
3.3.1 MALDI-TOF/TOF MS/MS Parameteroptimierung mittels Peptidbibliotheken auf Basis des 10-Proteinstandards	52
3.3.2 MALDI-TOF/TOF MS/MS Parameteroptimierung mittels der Peptidbibliothek des zytosolischen Hefeproteoms	60
3.3.3 Validierung des Q-PICS Ansatzes mittels Analyse der Spaltspezifität (<i>cleavage site specificity</i>) der Endoprotease GluC als <u>bekannte</u> Protease	64
3.3.4 Validierung des Q-PICS Ansatzes mittels Analyse der Spaltspezifität (<i>cleavage site specificity</i>) der Endoprotease GluC als <u>unbekannte</u> Protease	66
3.4 Diskussion.....	68
3.4.1 MALDI-TOF/TOF MS/MS Parameteroptimierung mittels der Peptidbibliotheken auf Basis des 10-Proteinstandards	71
3.4.2 MALDI-TOF/TOF MS/MS Parameteroptimierung mittels der Peptidbibliothek des zytosolischen Hefeproteoms	79
3.4.3 Validierung des Q-PICS Ansatzes mittels Analyse der Spaltspezifität (<i>cleavage site specificity</i>) der Endoprotease GluC als bekannte Protease	83
3.5 Zusammenfassung und Ausblick	85
Kapitel IV: Untersuchungen des Einflusses verschiedener Puffer auf die Spaltspezifität(<i>cleavage site specificity</i>)der Endoprotease GluC	88
4.1 Einleitung	88
4.2 Material und Methoden	89
4.3 Ergebnisse	92
4.3.1 Untersuchung des Einflusses des Puffersystems auf die Spaltspezifität der Endoprotease GluC	94
4.3.2 Untersuchung des Einflusses des Puffersystems auf die Spaltspezifität der Endoprotease GluC mittels <i>reversed</i> Markierung	97
4.4 Diskussion.....	99
4.4.1 Untersuchung des Einflusses des Puffersystems auf die Spaltspezifität der Endoprotease GluC	101

4.4.2 Untersuchung des Einflusses des Puffersystems auf die Spaltspezifität der Endoprotease GluC mittels <i>reversed</i> Markierung	104
4.5 Zusammenfassung und Ausblick	107
Kapitel V: Schnittstellen-Identifizierung in ARTC2.2 mittels LC-MALDI-TOF/TOF MS/MS 111	
5.1 Einleitung	111
5.2 Material und Methoden	113
5.3 Ergebnisse	117
5.3.1 Ergebnisse des tryptischen Verdau des Proteins ARTC2.2tace	119
5.3.2 Ergebnisse des Verdau mittels der Endoprotease GluC des Proteins ARTC2.2tace	121
5.4 Diskussion.....	124
5.5 Zusammenfassung und Ausblick	131
Kapitel VI: Optimierte <i>Data Conversion</i> für die quantitative Proteomik, basierend auf isobarer stabiler Isotopenmarkierung und LC-MALDI-TOF/TOF MS/MS	133
6.1 Einleitung	133
6.2 Material und Methoden	135
6.3 Ergebnisse	139
6.4 Diskussion.....	145
6.5 Zusammenfassung und Ausblick	149
Literaturverzeichnis.....	151
Anhang	166
Erklärung.....	194
Danksagung.....	195
Lebenslauf.....	198
Publikationsliste und Beiträge zu Kongressen	199

Abkürzungen und Symbole

Abb.	=	Abbildung
AC	=	Affinitätschromatographie
ACN	=	Acetonitril
bzw.	=	beziehungsweise
ca.	=	circa
CE	=	engl.: <i>capillary electrophoresis</i> ; Kapillarelektrophorese
CEM	=	engl.: <i>channel electron multiplier</i>
CF	=	Chromatofokussierung
CHCA	=	engl.: <i>α-cyano-4-hydroxycinnamic acid</i> ; α-Cyano-4-hydroxycimtsäure
(k)Da	=	(Kilo-)Dalton
CI	=	Chemische Ionisierung
C.I.	=	engl.: <i>Confidential Intervals</i>
CID	=	engl.: <i>collision induced decomposition</i> ; kollisions-induzierte Dissoziation
DHB	=	engl.: <i>dihydroxybenzoic acid</i> ; Dihydroxybenzoesäure
DIGE	=	Differentielle Gelelektrophorese
DOTA	=	1,4,7,10-Tetraazacyclododekan-1,4,7-10-tetraessigsäure
ECD	=	engl.: <i>electron capture dissociation</i> ; Elektronen Einfang Dissoziation
EDC	=	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimid
EI	=	Elektronenstoß Ionisierung
ESI	=	engl.: <i>electrospray ionisation</i>
ETD	=	engl.: <i>electron transfer dissociation</i> ; Elektronen Transfer Dissoziation
FAB	=	engl.: <i>fast atom bombardement</i>
FDR	=	engl.: <i>false discovery rate</i> ; falsch-positiv Rate
FT-ICR	=	Fouriertransformation-Ionenzyklotronresonanz
FRET	=	engl.: <i>fluorescent resonance energy transfer</i>
(2D-)GE	=	(zweidimensionale) Gelelektrophorese
HPLC	=	engl.: <i>high performance liquid chromatography</i> ; Hochleistungsflüssigchromatographie
HEPES	=	2-(4-(2-Hydroxyethyl)-1-piperazinyl)-ethansulfonsäure
IAA	=	engl.: <i>iod acetic acid</i> ; Iodessigsäure
ICAT	=	<i>isotope-coded affinity tag</i>

ICP	=	engl.: <i>inductively coupled plasma</i> , Induktiv gekoppeltes Plasma
ICPL	=	engl.: <i>isotop-coded protein label</i>
I.D.	=	Innendurchmesser
i.d.R.	=	in der Regel
IEC	=	engl.: <i>ion exchange chromatography</i> ; Ionenaustausch-Chromatographie
IMAC	=	engl.: <i>immobilized metal affinity chromatography</i>
IP	=	engl.: <i>ion pair</i> ; Ionenpaar
IPI	=	International Protein Index
IPTL	=	engl.: <i>isobaric peptide termini labeling</i>
IQR	=	engl.: <i>interquartile range</i>
iTRAQ	=	engl.: <i>isobaric tag for relative and absolute quantitation</i>
LC	=	engl.: <i>liquid chromatography</i> ; Flüssigkeitschromatographie
LPtMt	=	engl.: <i>Launch Peaks to Mascot tool</i>
m/m	=	Masse zu Masse
MALDI	=	engl.: <i>Matrix-assisted Laser Desorption/Ionisation</i> ; Matrix-unterstützte Laser Desorption/Ionisierung
MCP	=	engl.: <i>multi channel plate</i> ; Vielkanalplatte
MDHI	=	2-Methoxy-4,5-dihydro-(1H)-imidazol
mgf	=	engl.: <i>Mascot generic format</i>
MIO	=	engl.: <i>monoisotopic only</i>
MMTS	=	Methyl-methanthiosulfonat
MMP	=	Matrixmetalloprotease
MΩ	=	Megaohm
MS	=	engl.: <i>mass spectrometry</i> ; Massenspektrometrie
MS/MS	=	Tandem Massenspektrometrie
Mowse	=	engl.: <i>Molecular weight search</i>
<i>m/z</i>	=	Masse-zu-Ladungs-Verhältnis
MudPit	=	engl.: <i>multi-dimensional protein identification technology</i> ; Multidimensionale Proteinidentifizierungs-Technologie
NHS	=	N-Hydroxysuccinimid
NPC	=	engl.: <i>normal phase chromatography</i> ; Normalphasenchromatographie
OD	=	Optische Dichte
PBS	=	engl.: <i>phosphate buffered saline</i> ; Phosphat-gepufferte Salzlösung
pH	=	engl.: <i>potential of hydrogen</i>

PICS	=	engl.: <i>proteomic identification of cleavage sites</i>
PMF	=	Peptidmassenfingerprint
ppm	=	engl.: <i>parts per million</i>
PSD	=	engl.: <i>post source decay</i> ; metastabiler Zerfall
PS-DVB	=	Polystyrol-Divinylbenzol
PS-SCL	=	engl.: <i>positional scanning substrate combinatorial libraries</i>
RIK	=	Reporterionenkanal
RP	=	engl.: <i>reversed phase</i> ; Umkehrphase
(IP-)RPC	=	engl.: (ion pair) <i>reversed phase chromatography</i> ; (Ionenpaar-) Umkehrphasenchromatographie
RT	=	Raumtemperatur
SA	=	engl.: <i>sinapinic acid</i> ; Sinapinsäure
SA	=	Succinanhydrid
SAS	=	engl.: <i>substrate activity screening</i>
SC	=	engl.: <i>sequence coverage</i> ; Sequenzabdeckung
SCX	=	engl.: <i>strong cation exchange</i> ; starker Kationenaustausch
SDS	=	engl.: <i>sodium dodecyl sulfate</i> ; Natriumdodecylsulfat
SEC	=	engl.: <i>size exclusion chromatography</i> ; Größenausschluss-Chromatographie
SEV	=	Sekundärelektronen Vervielfacher
SILAC	=	engl.: <i>stable isotope labeling by amino acids in cell culture</i>
S/N-	=	engl.: <i>signal to noise</i> ; Signal-zu-Rauschen
SPE	=	engl.: <i>solid phase extraction</i> ; Festphasenextraktion
SVMP	=	engl.: <i>snake venom metalloproteinase</i>
Tab.	=	Tabelle
TAILS	=	engl.: <i>terminal amine isotopic labeling of substrates</i>
TCEP	=	Tris-(2-carboxyethyl)phosphin
TEAB	=	Triethylammoniumbicarbonat
TFA	=	engl.: <i>trifluoroacetic acid</i> ; Trifluoressigsäure
TMT	=	engl.: <i>tandem mass tag</i>
TOF	=	engl.: <i>time of flight</i> ; Flugzeit
TRIS	=	Tris(hydroxymethyl)-aminomethan
u. a.	=	unter anderem
U/min	=	Umdrehungen pro Minute

UV	=	Ultraviolett
v. a.	=	vor allem
vs.	=	versus, gegen
YPD	=	engl.: <i>yeast peptone dextrose</i>
z. B.	=	zum Beispiel

Zusammenfassung

In der vorliegenden Arbeit wurde auf Basis des so genannten *proteomic identification of cleavage sites* (PICS)-Ansatzes eine massenspektrometrische Methode, der Q-PICS Ansatz entwickelt. Dieser Ansatz kann zur Bestimmung der Spaltspezifität (unbekannter) Proteasen und gleichzeitig zur Quantifizierung proteolytischer Prozesse verwendet werden. Während im klassischen PICS-Ansatz biotinylierte Peptide identifiziert werden, werden im Q-PICS Ansatz TMT-markierte Peptide analysiert. Dadurch können bis zu sechs verschiedene TMT-Reporterionenkanäle verwendet werden, die dann zur relativen Quantifizierung genutzt werden. Die Analysen wurden mit einem MALDI-TOF/TOF Massenspektrometer durchgeführt.

Zunächst musste eine Optimierung der LC-MALDI-TOF/TOF MS/MS Methode durchgeführt werden, wobei folgende Parameter untersucht wurden: **i)** Anzahl der akkumulierten Laserschüsse zur Generierung der MS/MS-Spektren, **ii)** Gasdruck in der CID-Kollisionszelle und **iii)** Reihenfolge der *precursor*-Auswahl für die Fragmentierung in der MS/MS-Analyse. Durch die Verwendung von Proteom- bzw. Standardprotein-basierten Peptidbibliotheken wurde die beste Parameterkombination an akkumulierten Laserschüssen pro MS/MS Spektrum, dem Gasdruck in der CID-Kollisionszelle und der Reihenfolge bei der *precursor*-Auswahl für die Fragmentierung ermittelt. Als Auswahlkriterien wurden die Anzahl an identifizierten Spaltsequenzen, die Standardabweichung der Quantifizierungsdaten und die benötigte Messzeit des MALDI-Massenspektrometers herangezogen.

Das quantitative Potential des Q-PICS Ansatzes wurde anschließend mit der optimierten LC-MALDI-TOF/TOF MS/MS Methode am Beispiel der Endoprotease GluC in zwei verschiedenen Puffersystemen (HEPES- und TEAB-Puffer) gezeigt. In diesen Experimenten wurde die These bestätigt, dass GluC eine höhere Affinität für Glu-X gegenüber Asp-X Peptidbindungen unabhängig von dem verwendeten Puffersystem aufweist. Dabei resultiert ein inhibitorischer Effekt von Bicarbonatpuffer auf die GluC Gesamtaktivität in einer scheinbar ausschließlichen Affinität für Glu-X Peptidbindungen in diesem Puffer.

Da die generelle Anwendbarkeit des Q-PICS Ansatzes gezeigt wurde, sollten sich nun die unterschiedlichsten Problemstellungen analysieren lassen. Als Beispiel seien hier die Analyse einer Protease unter Verwendung verschiedener Inhibitoren, oder die Analyse des Einflusses verschiedener Bedingungen des Verdaus wie Temperatur, Inkubationszeiten (oder wie gezeigt: Puffersysteme) auf die Spaltspezifität einer Protease genannt.

Darüber hinaus wurde in einem weiteren Projekt die TACE-Spaltstelle innerhalb des Proteins ARTC2.2 identifiziert. Bei diesem Protein handelt es sich um ein GPI-verankertes Ekto-Enzym, das nach Aktivierung des P2X7 Rezeptors durch ATP oder NAD über eine Metalloprotease (vermutlich TACE bzw. ADAM 17) proteolytisch prozessiert wird. Durch die Durchführung eines In-Gel-Verdaus des Proteins ARTC2.2 mit unterschiedlichen Proteasen in ^{18}O -markiertem Wasser konnten verschiedene Peptide identifiziert werden, die zur Bestimmung der TACE-Spaltstelle nach Phe270 führten. Nur die TACE-Spaltstelle kann unter diesen Bedingungen keine ^{18}O -Markierung am C-Terminus enthalten.

In einem weiteren Projekt konnte gezeigt werden, dass ein *peak picking* Problem besteht, wenn LC-MALDI-TOF/TOF MS/MS Rohdaten auf Basis einer Quantifizierung über isobare Isotopenmarkierung in mgf-Dateien exportiert werden. Dieses Problem besteht darin, dass Signale eines Reporterionenkanals mit geringer Intensität teilweise nicht in die mgf-Datei exportiert werden, wenn sie auf ein Signal eines Reporterionenkanals mit einer hohen Intensität folgen. Insgesamt wurden drei verschiedene isobare *tags* (iTRAQ 4-plex, iTRAQ 8-plex und TMT 6-plex) und drei *data conversion* Programme LPTMt, TS2Mascot und MS DATA Converter untersucht. Das Problem konnte durch die Verwendung eines Perl-Skriptes umgangen werden.

Abstract

This work was focused on the development, optimization and evaluation of a mass spectrometric approach, the Q-PICS method, which simultaneously serves as tool for the determination of unknown cleavage site specificities –based on the so called PICS assay- and the quantification of the proteolytic process. While the classical PICS approach used biotinylated peptides, the application of TMT labeled peptides in the Q-PICS method offered the possibility to compare up to 6 different reporter ion channels, which were analyzed with MALDI TOF/TOF MS and MS/MS.

For the optimization of the LC-MALDI-TOF/TOF MS/MS method the following parameters were tested: **i)** number of accumulated laser shots to generate MS/MS spectra **ii)** CID pressure and **iii)** precursor selection/fragmenting order (“*weakest precursor first*” vs. “*strongest precursor first*”). In the experiments both a standard protein peptide library and a proteome based library were used. The evaluation criteria were represented by the number of identified cleavage sites, the standard deviation of quantification results and the attended measurement time.

Using the optimized LC-MALDI-TOF/TOF MS/MS conditions, the quantitative power of the Q-PICS approach was then demonstrated by the example of endoproteinase GluC in two different buffer systems (HEPES- and TEAB-buffer). The results of the following experiments confirmed the theory assuming that GluC exhibits a higher affinity for Glu-X peptide bonds compared to Asp-X bonds without a relation to the reaction buffer. The thesis additionally states an inhibitory effect of bicarbonate buffer on the GluC activity in general, that finally pretend an exclusively affinity for Glu-X bonds.

In conclusion, the general capability of the Q-PICS approach was shown here, the prerequisite to use it as tool for manifold questions, concerning proteolytic processes. For example the influence of different inhibitors or different reaction conditions like temperature, incubation times could be analyzed on the cleavage site specificity in a quantitative manner.

In a further project the TACE cleavage site inside of the protein ARTC2.2 was identified. ARTC2.2 represents a GPI-anchored ecto-enzyme, which is subjected to proteolytic processing after the activation of receptor P2X7 through ATP or NAD by a metallo protease (proposingly TACE/ADAM 17). Due to the identification of several carboxyterminal unlabeled peptides by the in-gel-digestion of ARTC2.2 in the presence of H₂¹⁸O with different proteases, the TACE cleavage site was located after Phe270. Exclusively the TACE cleavage

site can exhibit unlabeled (^{16}O) C-terminus under ^{18}O conditions (unless tryptic cleavage introduces H_2^{18}O molecules into all other peptides).

Furthermore, a peak picking problem, concerning LC-MALDI-TOF/TOF MS/MS raw data from quantification with isobaric tags, which are exported to mgf-files, was clearly demonstrated at a data set of a 15 protein mixture. Low intensity signals of reporter ion channels were partially not exported to mgf-files whenever they followed high intensity signals of another reporter ion channel in the MS/MS spectrum. Summing up, three different isobaric tags (iTRAQ 4-plex, iTRAQ 8-plex and TMT 6-plex) and three data conversion tools (LPtMt, TS2Mascot and MS Converter) were analyzed. The problem was extensively characterized and finally solved or circumvented by the application of an in-house Perl script.

Kapitel I: Theoretischer Teil – Einleitung:

1.2 Die Proteomik

Das menschliche Genom besteht aus 20.000 bis 25.000 Genen (International Human Genome Sequencing Consortium, 2004). Aus diesen werden vermutlich mehr als 500.000 verschiedene Proteine codiert. Die exakte Anzahl ist jedoch noch unbekannt. Ein Forschungsgebiet der Bioanalytik, die Proteomik (Proteomanalyse), beschäftigt sich mit der Identifizierung bzw. Quantifizierung der Gesamtheit aller Proteine einer Zelle oder eines Gewebes inklusive ihrer Varianten (z. B. durch posttranslationale Modifikationen) unter definierten Bedingungen und zu einem bestimmten Zeitpunkt (Lottspeich, 1999). Diese Gesamtheit der Proteine wird durch den Term Proteom beschrieben (Wilkins *et al.*, 1996). Erst das Zusammenspiel der Proteomik mit verschiedenen anderen Forschungsgebieten wie der Metabolomik, Transkriptomik oder Genomik ermöglicht die vollständige Analyse fundamentaler biologischer Prozesse. Vor allem die Proteomik und die Genomik sind dabei eng miteinander verknüpft. So wurde im Jahr 2007 durch die Arbeitsgruppe Mann die Proteomik als potentiell neue Genomik bezeichnet (Cox and Mann, 2007).

Die Entwicklung schonender massenspektrometrischer (MS) Ionisierungsverfahren, wie die Matrix-unterstützte Laser Desorption/Ionisierung (MALDI, (Karas and Hillenkamp, 1988, Tanaka *et al.*, 1988)) oder die Weiterentwicklung der Elektrospray-Ionisation (ESI, (Meng *et al.*, 1988)) ermöglichten die Analyse biologischer Makromoleküle. Zusätzlich weist das Proteom, im Gegensatz zum eher statischen Genom, einen sehr hohen dynamischen Bereich auf. Dieser überspannt vermutlich bis zu zwölf Größenordnungen (Corthals *et al.*, 2000). Diese hohe Komplexität stellt eine der größten Herausforderungen der Proteomik dar. Nur mit Analysemethoden, welche ein hohes Auflösungsvermögen, eine hohe Peakkapazität und ein niedriges Detektionslimit aufweisen ist sie zu lösen. Um solche Makromoleküle tatsächlich massenspektrometrisch untersuchen zu können, müssen die meist sehr komplexen Peptid- bzw. Proteingemische erst aufgereinigt bzw. aufgetrennt werden. Dadurch soll v. a. Ionensuppression vermieden werden. Diese kann dann auftreten, wenn während der Ionisierung zu viele Probenmoleküle eines hoch abundanten Analyten in die Gasphase übertragen werden. Die Ionisierung niedrig abundanten Analyten kann dann unterdrückt werden. Dies würde letztendlich dazu führen, dass die niedrig abundanten Analyten nicht detektiert werden.

In der Proteomik unterscheidet man im Allgemeinen zwei Strategien, um die Aufreinigung bzw. die Auftrennung der komplexen Proben durchzuführen, die *top-down*- (Kelleher, 2004, Sze *et al.*, 2002) und die *bottom-up*-Methode (Aebersold and Mann, 2003, Washburn *et al.*, 2001), die im speziellen auch als *shotgun*-Methode (Wolters *et al.*, 2001) bezeichnet wird (siehe Abbildung 1.1).

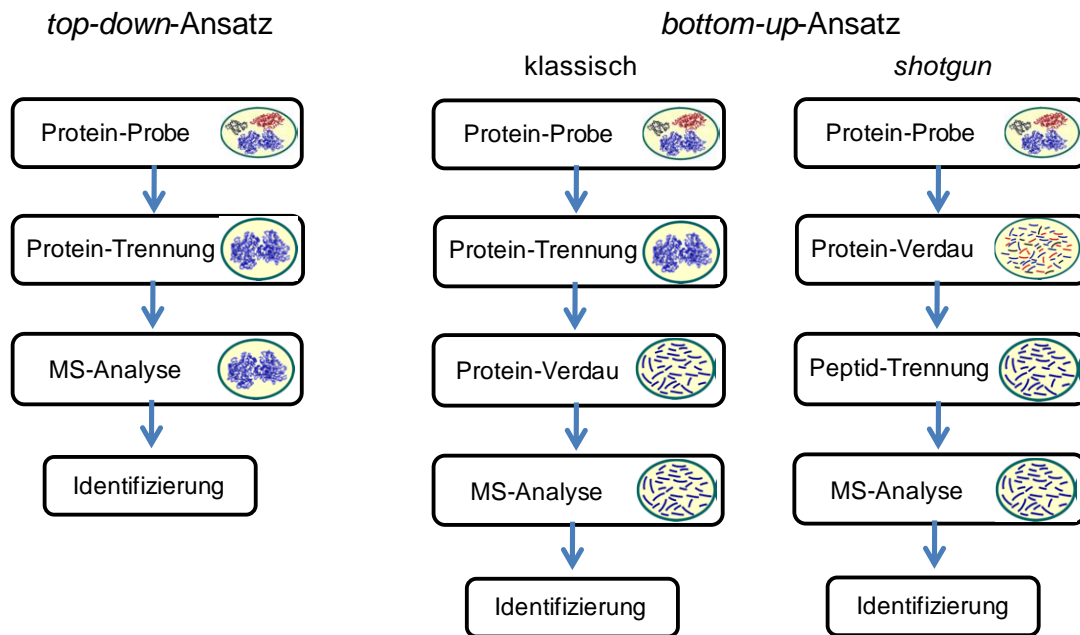


Abb. 1.1: Die beiden generellen Strategien für die Proteomanalyse: *top-down*-Analyse (links) und *bottom-up*-Analyse mit dem Spezialfall der *shotgun*-Methode (rechts).

Der wesentliche Unterschied der Methoden beruht auf den Analyten, die massenspektrometrisch untersucht werden. Während in der *top-down*-Methode die Analyse von intakten Proteinen erfolgt, werden in der *bottom-up*-Methode zuerst die Proteine (enzymatisch) gespalten und dann eine MS-Analyse der resultierenden Peptide durchgeführt.

Die massenspektrometrische Analyse intakter Proteine kann sowohl mit MALDI, als auch mit ESI MS bzw. MS/MS durchgeführt werden. ESI-Instrumente konnten sich jedoch durchsetzen, da FT-ICR- oder Orbitrap-Massenspektrometer eine hohe Massengenauigkeit und Auflösung besitzen (Reid and McLuckey, 2002, Scigelova and Makarov, 2006). Zusätzlich verfügen diese Massenspektrometer häufig über zusätzliche Fragmentierungsmethoden wie die Elektronen Transfer Dissoziation, die oft Anwendung für *top-down*-Experimente findet (Bunger *et al.*, 2008).

In der klassischen *bottom-up*-Methode wird die 2D-GE verwendet, bei der die Auftrennung der Analyte zunächst noch mit den intakten Proteinen durchgeführt wird (Klose, 1975,

O'Farrell, 1975). In dieser Methode wird in der 1. Dimension eine isoelektrische Fokussierung und in der 2. Dimension eine klassische SDS-Gelelektrophorese durchgeführt. Die beiden Trenntechniken basieren auf der Auftrennung nach dem isoelektrischen Punkt bzw. nach der Größe der Proteine und sind somit orthogonal zueinander. Durch diese Orthogonalität wird eine große Auflösung erreicht.

Zur Identifizierung werden die intakten Proteine nach der Auftrennung mit einer Protease (z. B. Trypsin oder Chymotrypsin) zu Spaltpeptiden hydrolysiert. Diese werden anschließend massenspektrometrisch untersucht, wobei man von einer so genannten *Fingerprint*-Analyse (PMF, *peptide massfinger print*, (Pappin *et al.*, 1993)) spricht. Die Spaltpeptide können zusätzlich mittels Tandem Massenspektrometrie (MS/MS) untersucht werden. Dabei werden direkte Sequenzinformationen über die einzelnen Peptide erhalten (Eng *et al.*, 1994). Die Identifizierung erfolgt i. d. R. über eine Datenbanksuche (Aebersold and Mann, 2003).

Einen Spezialfall der *bottom-up*-Methode hingegen stellt die so genannte *shotgun*-Methode dar. Hier wird die Hydrolyse der intakten Proteine zu Spaltpeptiden als erster Schritt vor der Trennung der Analyte durchgeführt. In der sogenannten *multi-dimensional protein identification technology* (MudPit) erfolgt die Auftrennung mittels HPLC über starken Kationenaustausch in der 1. Dimension (SCX), gefolgt von Umkehrphasen-Chromatographie (RPC) in der 2. Dimension (Wolters, *et al.*, 2001). In diesem speziellen Beispiel wird die Elektrospray-Kapillare direkt mit den beiden unterschiedlichen Materialien an stationären Phasen (SCX- und RPC-Material) gefüllt. Man erhält demnach zwei verschiedene Bereiche an stationären Phasen wodurch die 2D Trennung *online* erfolgt. Sie kann jedoch auch über zwei verschiedene Säulen getrennt, *offline*, durchgeführt werden.

Prinzipiell sind sowohl für die *top-down*-, als auch für die *bottom-up*-Methode die unterschiedlichsten Kombinationen an Trenntechniken anwendbar (siehe Tabelle 1.1).

Tab. 1.1: Kombinationsmöglichkeiten verschiedener Modi der Trenntechniken Hochleistungs-Flüssigkeitschromatographie (HPLC), Kapillarelektrophorese (CE) und Gelelektrophorese (GE) zur Auftrennung von Peptid- oder Proteinmischungen.

1. Dimension	2. Dimension
Isoelektrische Fokussierung (Proteine)	SDS Gel Elektrophorese oder Umkehrphasen HPLC (Proteine)
Größenausschluss HPLC (Proteine)	Umkehrphasen HPLC oder Ionenaustausch (Proteine oder Peptide)
Dichtegradientenzentrifugation (Proteine)	Umkehrphasen HPLC (Proteine und Peptide)
Umkehrphasen HPLC (Proteine und Peptide)	Kapillarelektrophorese (Proteine und Peptide)
Affinitätschromatographie (Proteine, Peptide)	Umkehrphasen HPLC (Peptide)
Ionenaustausch HPLC (Proteine und Peptide)	Umkehrphasen HPLC (Proteine und Peptide)

1.2 Die quantitative Proteomik

Neben der Identifizierung von Proteinen spielt auch ihre Quantifizierung eine wichtige Rolle im Feld der Bioanalytik. Prinzipiell kann man bei Quantifizierungen zwischen der absoluten und der relativen bzw. differentiellen Quantifizierung unterscheiden. Obwohl durch letztere keine Aussagen über die absoluten Konzentrationen von Proteinen in einer Zelle oder einem Gewebe getroffen werden, sondern Proteinkonzentrationen verschiedener Zellen oder Geweben miteinander verglichen werden, zählen solche Untersuchungen zu den am häufigsten angewendeten. Traditionell wurde die quantitative Proteomik zu Beginn mittels 2D GE durchgeführt. Während die Identifizierung der Proteine dabei wieder über eine MS-Analyse erfolgt, wurde die quantitative Analyse über spezielle Programme wie z. B. Tycho durchgeführt (Anderson *et al.*, 1981). Dabei wird eine statistische Analyse eines Satzes an reproduzierten Gelen angewendet, in der die Spotintensitäten bzw. -volumina der Kontroll-Gele und der eigentlichen Proben miteinander verglichen werden. Ein großer Nachteil der Methode ist jedoch der große präparative Aufwand, da auf Grund der hohen Gel-zu-Gel-Variabilität mehrere Replikate angefertigt werden müssen (Righetti *et al.*, 2004).

Die Differentielle Gelelektrophorese (DIGE) umgeht dieses Problem, da in dieser Methode die Kontrolle und die zu untersuchende Probe mit zwei unterschiedlichen Cyanin-Farbstoffen (*CyDye*) markiert und gleichzeitig in einem 2D-Gel analysiert werden. Dadurch wird der experimentelle Fehler minimiert. Überlagern sich Proteine aus zwei verschiedenen Proben, so entstehen fluoreszierende Mischfarben, aus denen die Konzentrationen der Proteine beider Proben relativ zueinander bestimmt werden können (Unlu *et al.*, 1997). Dennoch weisen die Gel-basierten Strategien eine Reihe von anwendungstechnischen Grenzen auf. Die Analyse von Membran- oder niedrig abundanten Proteinen erweist sich z. B. als sehr schwierig. Deshalb wurden für die quantitative Proteomik zunehmend MS-basierte Methoden entwickelt, die v. a. auf der Markierung mit Verbindungen, welche stabile Isotope besitzen, beruhen. Es können aber auch markierungsfreie (*label-free*) Quantifizierungen durchgeführt werden (Bantscheff *et al.*, 2007, Ong and Mann, 2005, Zhu *et al.*, 2010).

Die MS-basierte Quantifizierung kann prinzipiell in zwei verschiedene Methoden unterteilt werden:

- i) Quantifizierung auf MS-Ebene
- ii) Quantifizierung auf MS/MS-Ebene

Während bei Strategien, die auf stabiler Isotopenmarkierung beruhen, auf MS-Ebene und bei denjenigen, die auf isobarer stabiler Isotopenmarkierung beruhen, auf MS/MS-Ebene

quantifiziert wird, kann bei der markierungsfreien Strategie die Quantifizierung auf beiden Ebenen erfolgen. Im Zusammenhang der vorliegenden Arbeit wird auf die markierungsfreie Quantifizierung nicht weiter eingegangen.

1.2.1 Quantifizierung auf MS-Ebene

Die Einführung der stabilen Isotope kann auf verschiedene Arten erfolgen. In Experimenten mittels des *protein metabolic labeling* wird beispielsweise die Proteinbiosynthese genutzt und die Einführung erfolgt *in vivo* (siehe Abbildung 1.2). Hierzu werden Organismen wie Hefen (Oda *et al.*, 1999) mit ^{15}N -isotopenmarkierten Substraten, die in Form anorganischer Salze ins Medium gegeben werden, herangezogen. Selbst der Säugetierorganismus *Rattus norvegicus* wurde mit zuvor ^{15}N -isotopenmarkierten Algenzellen gefüttert und für eine *protein metabolic labeling*-Studie analysiert (Wu *et al.*, 2004). Damit die unmarkierten Proteine während des Zellwachstums durch Proteinumsatz ausschließlich durch isotopenmarkierte ersetzt werden, darf nur ^{15}N als Stickstoffquelle zur Verfügung stehen. Dadurch erhält man eine „schwere“ Probe (siehe Abbildung 1.2, Zustand B). Nach Vereinigung, mit der unmarkierten „leichten“ Probe können beide relativ zueinander quantifiziert werden. Die sogenannte *Stable isotope labeling by amino acids in cell culture* (SILAC)-Methode beruht auf dem gleichen Prinzip wie das *protein metabolic labeling*. Hier werden jedoch nicht alle Substrate mit einem stabilen Isotop markiert, sondern isotopenmarkierte essentielle Aminosäuren ebenfalls über biosynthetische Stoffwechselreaktionen in die Proteine eingebaut (Ong *et al.*, 2002).

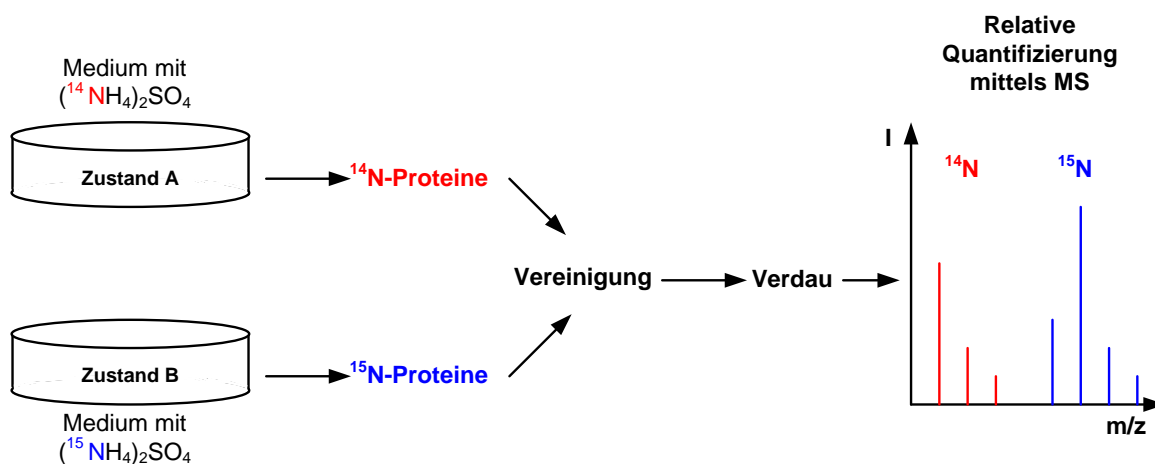


Abb. 1.2: Vorgehensweise in der relativen Quantifizierung mittels *protein metabolic labeling* auf MS-Ebene (Tholey, 2006).

Neben der Einführung der stabilen Isotope *in vivo* gibt es auch zahlreiche Methoden, in denen diese Einführung *in vitro* erfolgt. In der sogenannten *isotope-coded protein label* (ICPL)-Methode erfolgt die Derivatisierung auf Proteinebene (Schmidt *et al.*, 2005). Hierzu werden alle freien Aminogruppen der N-Termini und die ϵ -Aminogruppe der Lysinseitenketten über den Einsatz von *N*-Hydroxysuccinimid (NHS-Chemie) mit den unterschiedlichen ICPL-Reagenzien markiert. Bei diesen Reagenzien handelt es sich um durch verschiedene Isotope substituiertes Nicotinoyl-NHS (siehe Abbildung 1.3). Die Proben werden nach der Derivatisierung vereinigt und können danach z. B. über 2D-GE aufgetrennt werden. Nach enzymatischem Verdau erfolgt die MS-Analyse. Die Identifizierung erfolgt dabei auf MS/MS-Ebene während die Quantifizierung auf MS-Ebene durchgeführt wird. Bis zu vier verschiedene Proben lassen sich auf diese Weise miteinander vergleichen.

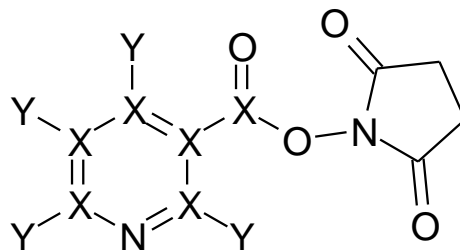


Abb. 1.3: Strukturformel des ICPL-Reagenzes. Insgesamt gibt es vier verschiedene ICPL-Reagenzien, bei denen der Kohlenstoff (X) und Wasserstoff (Y) folgendermaßen substituiert sind: 1) $^{12}\text{C}_6^1\text{H}_4$, 2) $^{12}\text{C}_6^2\text{D}_4$, 3) $^{13}\text{C}_6^1\text{H}_4$ und 4) $^{13}\text{C}_6^2\text{D}_4$.

Eine mit dieser Strategie eng verwandte Methode, stellt die sogenannte *isotope coded affinity tag* (ICAT)-Strategie dar (Gygi *et al.*, 1999). Hier erfolgt die Derivatisierung nicht Amin-, sondern Cystein-spezifisch. Normalerweise wird vor der enzymatischen Hydrolyse intakter Proteine zu Spaltpeptiden die hoch reaktive Thiolgruppe der Cysteinseitenketten durch Alkylierung mittels Iodacetamid oder Iodessigsäure geschützt. Dadurch soll die Rückbildung von Disulfidbrücken der Proteine und damit verbundene Kreuzkupplungen unterbunden werden. Genau diese hohe Reaktivität der Thiolgruppe kann aber auch dazu genutzt werden, unterschiedliche stabile Isotopenmarkierungen wie das ICAT-Reagenz, einzufügen. Auch hier erfolgt die Derivatisierung auf Proteinebene. Nach der Auftrennung wird ein enzymatischer Verdau durchgeführt. Der große Vorteil und zugleich Nachteil des ICAT-Reagenzes besteht darin, dass es ein Biotin *tag* enthält. Über dieses *tag* kann eine Aufreinigung bzw. Anreicherung durch Affinitätschromatographie über Streptavidin-Säulen erfolgen. Auf diese Weise können cysteinhaltige Peptide, die mit dem ICAT-Reagenz markiert wurden,

angereichert werden. Allerdings können auch nur solche zur Generierung quantitativer Datenherangezogen werden. Zudem lassen sich mit dem ICAT-Reagenz nur zwei unterschiedliche Proben miteinander vergleichen.

Eine weitere Methode, mit der auch ein *multiplexing* durchführbar ist, stellt die Quantifizierung von Proteinen mittels *Element-coded Affinity Tag* (ECAT, (Whetstone *et al.*, 2004)) bzw. *Metal-coded Affinity Tag* (MeCAT)-Lanthanid-Markierung dar (Ahrends *et al.*, 2007). Hierzu wird z. B. der makrozyklische Chelat-Komplexbildner 1,4,7,10-Tetraazacyclododekan-1,4,7,10-tetraessigsäure (DOTA) Cystein- oder Amin-spezifisch an Peptide bzw. Proteine angebracht. Dieser kann dann mit den verschiedenen Lanthaniden komplexiert werden. Der große Vorteil dieser Methode besteht v. a. in der Möglichkeit des *multiplexing* und der Tatsache, dass mittels ICP MS auch absolut quantifiziert werden kann. Dies ist prinzipiell jedoch nur dann möglich, wenn eine Basislinientrennung der Analyten innerhalb der chromatographischen Trennung erreicht wird. Das Fragmentierungsverhalten von DOTA-markierten Peptiden wurde sowohl mit ESI (Pieper *et al.*, 2009) als auch mit MALDI MS (Gregorius *et al.*, 2010) untersucht. Dabei wurde bestätigt, dass sich beide MS-Plattformen für die Identifizierung und Quantifizierung DOTA-markierter Peptide eignen.

Auch Methoden, in denen die Quantifizierung über eine Markierung der Carboxylgruppe erfolgt, wurden etabliert. Hier kann auf Peptid-Ebene z. B. eine Derivatisierung durch die Veresterung der Carboxylgruppen mittels d_0/d_3 -Methanol/Salzsäure durchgeführt werden (Goodlett *et al.*, 2001).

Aber auch die Grundlage einer jeden *bottom-up*-Analyse, die enzymatische Hydrolyse der Proteine zu Peptiden, kann zur Einführung stabiler Isotopen genutzt werden. Während der Spaltung der Peptidbindung werden formell ein (oder zwei) Wassermolekül(e) an diese addiert. Findet der Verdau also mittels $H_2^{18}O$ als Lösungsmittel statt, so kann auf diese Weise das schwerere ^{18}O -Sauerstoffisotop in die Carboxylgruppe eingeführt werden. Wird parallel der Verdau der zu vergleichenden Probe in normalem bzw. unmarkiertem $H_2^{16}O$ als Lösungsmittel durchgeführt, so können die beiden Proben ebenfalls relativ zueinander quantifiziert werden (Schnolzer *et al.*, 1996).

1.2.2 Quantifizierung auf MS/MS-Ebene

Wie zuvor bereits kurz beschrieben, handelt es sich bei den Methoden, in denen auf MS/MS-Ebene quantifiziert wird, um so genannte isobare stabile Isotopenmarkierungen. Hier sind v. a. die „*Isobaric Tags for Relative and Absolute Quantification*“ (iTRAQ, (Ross *et al.*,

2004)) und „Tandem Mass Tags“ (TMT, (Thompson *et al.*, 2003)) zu nennen. Diese beiden Reagenzien unterscheiden sich lediglich in ihrer Struktur und der Masse der daraus resultierenden Reporterionen. Während iTRAQ als Amin-spezifisches Reagenz für vier bzw. acht verschiedene Proben zur Verfügung steht, gibt es TMT sowohl Amin- als auch Cystein-spezifisch für bis zu sechs verschiedene Proben. Die generelle Vorgehensweise ist bei beiden Quantifizierungsreagenzien gleich und am Beispiel für das Amin-reaktive TMT in Abbildung 1.4 dargestellt. Beide Quantifizierungsreagenzien bestehen aus einer Reporter-, einer Balance- und einer Amin- (bzw. Cystein-) reaktiven-Gruppe. Die Reporter- und Balance-Gruppe bilden das isobare tag, das über die reaktive Gruppe an die Peptide bzw. Proteine gebunden wird. Die isobaren tags besitzen eine identische Struktur und über die verschiedenen Isotopenmarkierungen in der Reporter- und Balance-Gruppe auch die gleiche Molekülmasse. Deshalb weisen sie auch identische physikochemische Eigenschaften auf. Dies hat den Vorteil, dass sie beispielsweise unter chromatographischen Bedingungen, sowie in der Analyse auf MS-Ebene nicht zu unterscheiden sind. Erst während der MS/MS-Analyse fragmentiert die Reporter-Gruppe in Form der unterschiedlichen Reporterionen an einer bestimmten MS/MS-Spaltstelle. Über die Intensitäten bzw. Flächen der Reporterionen können die unterschiedlichen Proben dann relativ zueinander quantifiziert werden.

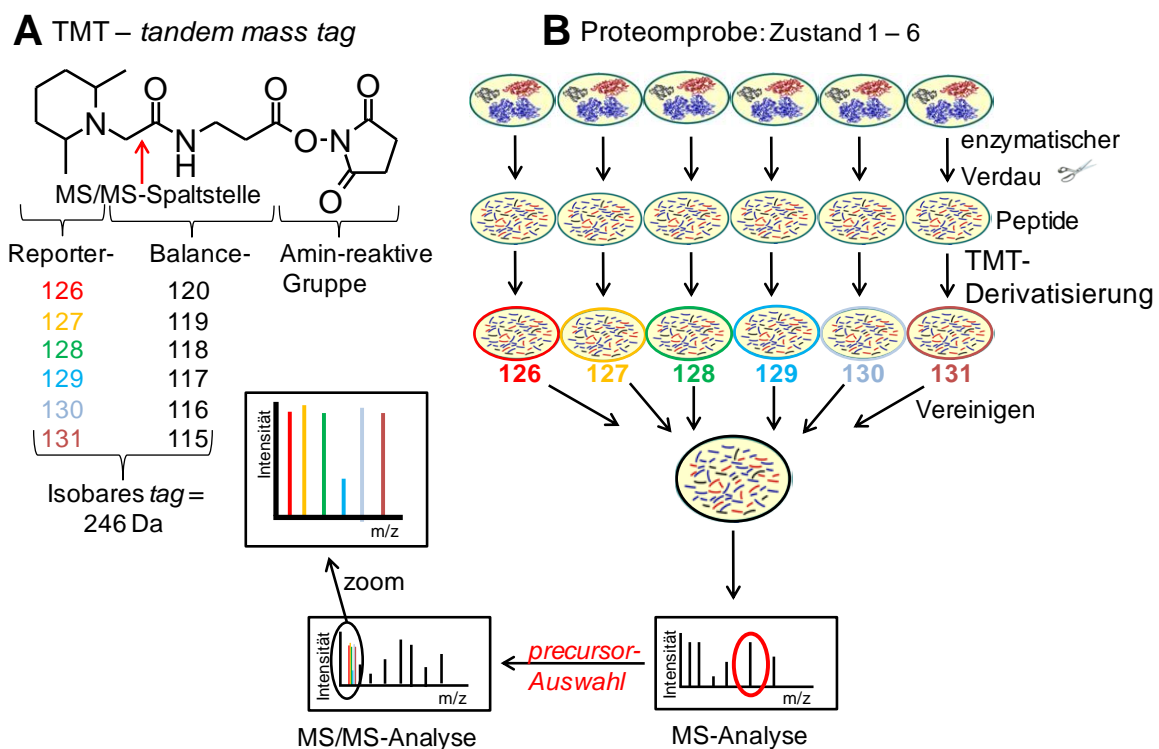


Abb. 1.4: Schematische Darstellung der Amin-spezifischen TMT-Methode. A) Aufbau des Amin-spezifischen TMT-Reagenzes: Es ist in eine Reporter-, eine Balance- und eine Amin-

reaktive-Gruppe unterteilt. Reporter- und Balance-Gruppe ergeben das isobare *tag* mit einer Masse von 246 Da. Dieses wird über die Amin-spezifische Gruppe an die Peptide bzw. Proteine gebunden. Insgesamt stehen sechs verschiedene *tags* zur Verfügung, die sich in der Isotopenmarkierung der Reporter- und Balance-Gruppe unterscheiden. B) Bis zu sechs verschiedene Proteomproben können vor oder nach dem enzymatischen Verdau mit den TMT-Reagenzien markiert werden. Während der massenspektrometrischen Analyse auf MS-Ebene erhält man für alle unterschiedlich markierten Peptide auf Grund der gleichen physikochemischen Eigenschaften nur ein Signal. Erst während der Analyse auf MS/MS-Ebene fragmentieren die isobaren *tags* an der MS/MS-Spaltstelle und setzen somit die unterschiedlichen Reporterionen frei. Eine relative Quantifizierung kann dann über die Intensitäten bzw. Flächen der Reporterionen durchgeführt werden.

Im Vergleich dazu kann es unter der Verwendung von Methoden der Quantifizierung auf MS-Ebene schnell zu einer unerwünschten Erhöhung der Probenkomplexität kommen, da hier jedes unterschiedlich markierte Peptid als ein einzelnes Signal detektiert wird. Dies ist beispielsweise für Experimente mit DOTA-markierten Peptiden der Fall (Gregorius *et al.*, 2013).

Das so genannte *Isobaric Peptide Termini Labeling* (IPTL, (Arntzen *et al.*, 2011, Koehler *et al.*, 2009)) stellt ebenfalls eine MS/MS-basierte Quantifizierungsmethode dar. Mit dieser Methode können allerdings nur zwei verschiedene Zustände miteinander verglichen werden. Um Peptide zu erhalten, die C-terminal mit einem Lysin enden, muss der enzymatische Verdau zweier unterschiedlicher Proteomproben A und B mit der Protease Lys-C durchgeführt werden. Die Lysine der Proteomprobe A werden mit 2-Methoxy-4,5-dihydro-(1*H*)-imidazol (MDHI) und die der Proteomprobe B mit vierfach deuteriertem MDHI-*d*₄ derivatisiert. Isobare Peptide werden erzeugt, indem die N-Termini in Proteomprobe A mit vierfach deuteriertem Succinanhydrid (SA-*d*₄) und die in Proteomprobe B mit unmarkiertem SA derivatisiert werden. Der große Vorteil der Methode besteht nun darin, dass man für jede Probe nicht nur einen Quantifizierungswert, nämlich das Reporterion erhält. Für jedes b- und y-Ion erhält man also ein Signalkonzept, das einen Abstand von 4 Da aufweist. Auf diese Weise kann man über jedes dieser Paare relativ zueinander quantifizieren, und erhält somit eine Vielzahl an quantitativen Daten.

Diese Technik wurde dann zum so genannten triplex-IPTL weiterentwickelt. Durch eine abgeänderte Derivatisierung der Peptide können mit diesem Ansatz drei verschiedene Proben bzw. Zustände miteinander verglichen werden (Koehler *et al.*, 2013).

1.3 Hochleistungs-Flüssigkeitschromatographie für die Auftrennung von Peptid- bzw. Proteingemischen

Neben der Gaschromatographie (GC) stellt die Hochleistungs-Flüssigkeitschromatographie (HPLC) die wohl wichtigste Trenntechnik im Bereich der Bioanalytik dar. Durch die Vielfalt an physikochemischen Eigenschaften wie Größe, Ladung, Polarität oder biologischer Affinität können Peptide bzw. Proteine durch eine Vielzahl unterschiedlichster chromatographischer Methoden aufgetrennt werden (Creighton, 1993). Da in der vorliegenden Arbeit ausschließlich Ionenpaar-Umkehrphasen- (IP-RP)-HPLC verwendet wurde, sollen die übrigen Methoden nur kurz genannt werden und sind in Tabelle 1.2 aufgelistet.

Tab. 1.2: Aufzählung chromatographischer Trennmethode in der Peptid- und Proteinanalyse und deren Trennprinzipien. Mit Ausnahme der SEC findet in allen Methoden eine Wechselwirkung zwischen den Analyten und der jeweiligen stationären Phase statt.

Trennmethode	Abkürzung	Trennprinzip
Affinitätschromatographie	AC	biospezifische Wechselwirkung
Chromatofokussierung	CF	unterschiedliche isoelektrische Punkte
Ionenaustauschchromatographie	IEC	elektrostatische Wechselwirkungen
Metallchelate-Affinitätschromatographie	IMAC	Komplexbildung
Ionenpaar-Umkehrphasen-chromatographie	IP-RPC	elektrostatische und solvophobe Wechselwirkungen
Normalphasenchromatographie	NPC	polare Wechselwirkungen
Umkehrphasenchromatographie	RPC	solvophobe Wechselwirkungen
Größenausschlusschromatographie	SEC	Unterschiede in der Molekülgröße

Vergleicht man die physikochemischen Eigenschaften von kleinen Molekülen, die meistens in der Metabolomik oder der organischen Analytik untersucht werden, mit denen von Peptiden bzw. Proteinen, so unterscheiden sich diese stark in ihrer Größe und den damit verbundenen komplexen Strukturen. Biologische Makromoleküle wie Peptide und Proteine adsorbieren nicht nur an einer, sondern an mehreren Stellen der jeweiligen stationären Phase. Man spricht in diesem Zusammenhang von der so genannten Mehrpunktsadsorption (Geng and Regnier, 1984). Aufgrund der Mehrpunktadsorption erfolgt eine komplette Desorption der Makromoleküle erst dann, wenn die Stärke des Elutionsmittels ausreicht, um alle Adsorptionspunkte zwischen dem Peptid oder dem Protein und der stationären Phase zu lösen. Deshalb spricht man hier vom so genannten „On-Off“-Modell.

Während für kleine Moleküle meist eine isokratische Elution durchgeführt wird, verwendet man für Analyten mit Mehrpunktsadsorption ausschließlich eine Gradientenelution. Dadurch kann während der Peptid- bzw. Proteinauftrennung mittels (IP-) RPC eine Aufkonzentrierung und/oder Entsalzung der Probe bei niedriger Elutionsstärke auf einer so genannten *Trap*- bzw. Aufkonzentriersäule durchgeführt werden. Auf diese Weise lassen sich auch größere Probenvolumina bei vergleichsweise geringerer Bandenverbreiterung injizieren.

Die Bandenverbreiterung und die damit verbundene Effizienz einer chromatographischen Trennung werden durch die Van-Deemter-Kurve beschrieben. Diese setzt sich aus A-, B- und C-Term zusammen und ist in Abbildung 1.5 und Gleichung 1.1 dargestellt. Während der A-Term (Eddy-Diffusion) die Inhomogenität der Teilchengröße bzw. des monolithischen Trennmediums berücksichtigt, beinhaltet der B-Term (Longitudinaldiffusion) die Diffusion des Analyten während der Verweilzeit in der mobilen Phase. Der C-Term (verzögerter Massentransfer) beschreibt den unterschiedlichen Stoffaustausch zwischen mobiler und stationärer Phase.

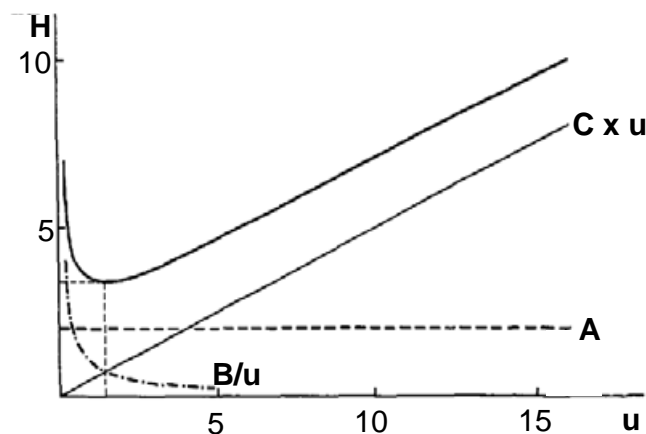


Abb. 1.5: Auftragung der Bodenhöhe H gegen die Flußgeschwindigkeit u einer klassischen Van-Deemter Kurve..(Quelle: Prof. Dr. C.G. Huber, Universität Salzburg; per Abbildung).

$$H = A + \frac{B}{u} + C \times u \quad (\text{Gleichung 1.1})$$

H = Bodenhöhe

A = A-Term

B = B-Term

C = C-Term

u = Flußgeschwindigkeit

1.3.1 Umkehrphasen- und Ionenpaar-Umkehrphasen-HPLC

Wie in Tabelle 1.2 bereits gezeigt, werden Analyten, wie Peptide bzw. Proteine in der RP-HPLC auf Grund ihrer unterschiedlichen solvophoben Eigenschaften aufgetrennt. Dabei spielen zum einen hydrophobe Wechselwirkungen zwischen dem Analyten und der unpolaren stationären Phase und zum anderen solvophobe Wechselwirkungen zwischen Analyt und mobiler Phase eine wichtige Rolle. Während sich die mobile Phase aus einem polaren (Wasser) und einem apolaren (i. d. R. ACN oder Methanol) Lösungsmittel zusammensetzt, handelt es sich bei der stationären Phase um unpolares Material. Dieses besteht meist aus alkyliertem Kieselgel mit Kettenlängen von C4 bis C18. Aber auch organische Polymere wie Polystyrol-Divinylbenzol (PS-DVB) kommen als Säulenmaterial zum Einsatz. Die prinzipiellen Wechselwirkungen zwischen dem Analyten und der mobilen bzw. stationären Phase sind in Abbildung 1.6 dargestellt.

In der Ionenpaar-Umkehrphasen-Chromatographie (IP-RP-HPLC) ist neben der Hydrophobie der Peptide bzw. Proteine auch ihre Ladung ein wichtiger Faktor für die Trennung der Analyten. Durch die Zugabe so genannter Ionenpaar-Reagenzien und deren Adsorption an der stationären Phase wird ein elektrisches Oberflächenpotential erzeugt (Pathy, 1994). Als Ionenpaar-Reagenzien können dabei Amphiphile wie hydrophobe Säuren (z. B. Trifluoressigsäure) oder hydrophobe Basen (z. B. Triethylammoniumacetat) eingesetzt werden.

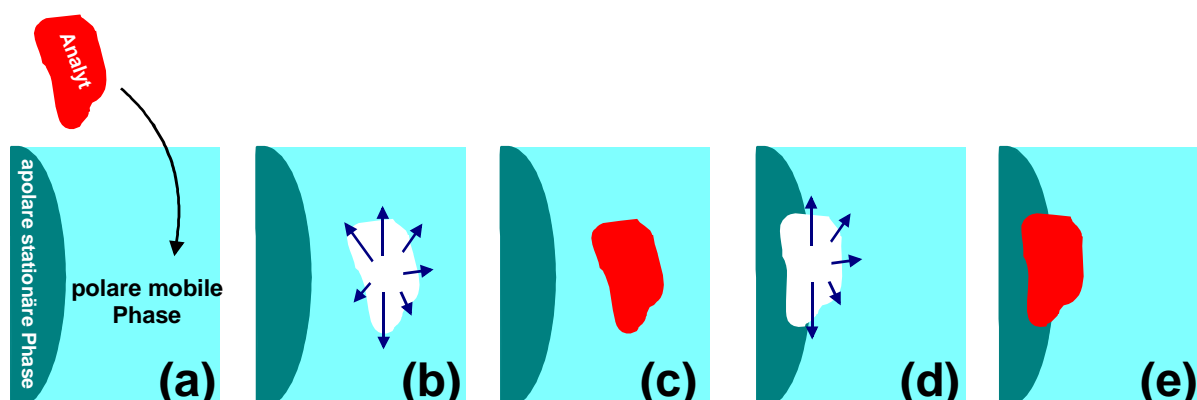


Abb. 1.6: Schematische Darstellung der prinzipiellen Wechselwirkungen zwischen dem Analyten mit der polaren mobilen bzw. apolaren stationären Phase während einer Trennung über RP-HPLC. Um den Analyten in der mobilen Phase (Wasser) zu lösen (a), muss zunächst eine Kavität in der mobilen Phase geschaffen werden (b). Da hierzu erst die Wasserstoffbrücken aufgebrochen werden müssen, handelt es sich dabei um einen energetisch äußerst ungünstigen Zustand (c). Wird die Kavität an der stationären Phase

gebildet (d), ist dies ein energetisch günstigerer Zustand, da zum einen hydrophobe Wechselwirkungen aufgebaut werden (e) und zum anderen weniger Wasserstoffbrücken aufgebrochen werden müssen. (Quelle: Prof. Dr. C.G. Huber, Universität Salzburg; per Abbildung).

In Abbildung 1.7 ist der Mechanismus der IP-RP-HPLC am Beispiel von Trifluoressigsäure als Ionenpaar-Reagenz gezeigt.

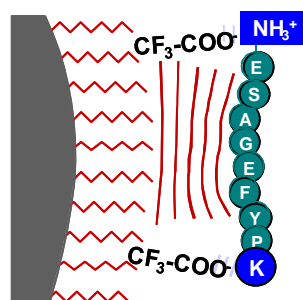


Abb. 1.7: Mechanismus der IP-RP-HPLC mit Trifluoressigsäure als verwendetes Ionenpaar-Reagenz (Quelle: Prof. Dr. C. G. Huber, Universität Salzburg, per Abbildung).

Die Konzentration des Ionenpaar-Reagenzes reicht nicht aus, um die gesamte Oberfläche der stationären Phase zu bedecken. Die Trennung der Analyten findet daher sowohl über hydrophobe bzw. solvophobe Wechselwirkungen (wie in der RPC) als auch über elektrostatische Wechselwirkungen zwischen Analyt und dem, an der stationären Phase adsorbiertem Ionenpaar-Reagenz (wie in der IEC), statt (Chen *et al.*, 1993, Melander *et al.*, 1989).

Durch die denaturierenden Bedingungen und das dadurch verhinderte Auftreten von Isomerenpeaks werden in der RP- bzw. in der IP-RP-HPLC, verglichen mit anderen chromatographischen Trennmethoden, die höchsten Trenneffizienzen und Peakkapazitäten erreicht. Dies wird v. a. durch die größere Moleküloberfläche erreicht, die für Wechselwirkungen mit der stationären Phase zur Verfügung steht (Mehrpunktsadsorption). Verschiedene stationäre und mobile Phasen, Ionenpaar-Reagenzien oder Parameter, wie z. B. Säulentemperatur, ermöglichen die Bearbeitung der unterschiedlichsten Aufgabenstellungen in der Peptid- bzw. Proteinanalytik. Ein weiterer, entscheidender Vorteil z. B. gegenüber der IEC besteht in der Verwendung flüchtiger Komponenten wie Wasser oder ACN als mobile Phase oder Trifluoressigsäure als Ionenpaar-Reagenz. Dadurch entsteht die Möglichkeit der Kopplung mit verschiedenen Massenspektrometern. So können die RP- bzw. die IP-RP-

HPLC sowohl *online* mit ESI MS (Banks, 1996, Whitehouse *et al.*, 1985) als auch *offline* mit MALDI MS (Hsieh *et al.*, 1998) gekoppelt werden.

1.3.2 Stationäre Phasen für die Trennung von Peptid- bzw. Proteingemischen in Verbindung mit Mikro-Hochleistungs-Flüssigkeitschromatographie

Eines der größten Probleme während der Trennung von Peptid- bzw. Proteingemischen stellt der langsame Massentransfer dar (vgl. Van-Deemter-Kurve in Abbildung 1.5). Dieser begründet sich auf den sehr viel niedrigeren Diffusionskoeffizienten der biologischen Makromoleküle im Vergleich zu niedermolekularen organischen Verbindungen. Durch den langsamen Ein- und Austritt in die Poren der stationären Phase über Diffusion entsteht eine Bandenverbreiterung der chromatographisch aufgetrennten Peaks. Um dem Problem des verzögerten Massentransfers entgegenzuwirken, kann man die Partikelgröße der stationären Phase reduzieren. Dadurch werden die Diffusionswege verringert (Abbildung 1.8, (a)). Je kleiner jedoch die Partikelgröße, desto größer ist der Druckabfall über der Säule. Verwendet man eine bimodale Packung, die aus Perfursionsteilchen aufgebaut ist (Abbildung 1.8, (b)), so führt dies zu einer höheren Permeabilität und damit zu einem schnelleren Massentransfer. Kleinere, nicht-poröse Teilchen als stationäre Phase wirken sich ebenfalls positiv auf den verzögerten Massentransfer aus. Hier findet eine Adsorption der Analyten ausschließlich an der Oberflächenschicht der stationären Phase statt. Dies ist allerdings mit einer geringeren Beladungskapazität verglichen mit der poröser Partikel verbunden (Abbildung 1.8, (c)). Einen Spezialfall stellen die so genannten monolithischen Packungen dar (Abbildung 1.8, (d)). Diese bestehen nicht aus sphärischen Teilchen, sondern aus einem zusammenhängenden Partikel, der z. B. durch *in situ* Polymerisation hergestellt werden kann (Hansen and Sievers, 1974, Svec and Frechet, 1992).

Da somit keine Zwischenkornvolumina auftreten, kann das gesamte Laufmittel durch die Porenkanäle der stationären Phase strömen, wodurch die diffusionsabhängige Bandenverbreiterung verringert wird (Viklund *et al.*, 1996). Eine Reihe von Studien konnte zeigen, dass durch die hohe Permeabilität und den schnellen Massentransfer in monolithischen Säulen eine schnelle Trennung von Peptiden bzw. Proteinen mit hohen Flussgeschwindigkeiten möglich ist (Huber *et al.*, 1993, Huber *et al.*, 1997, Premstaller *et al.*, 2001). Da die Van-Deemter-Kurve mit hohen Flussgeschwindigkeiten relativ flach verläuft, werden bei Verwendung monolithischer Säulen sehr gute Trennleistungen erreicht, die nur kurze Analysenzeiten benötigen (Oberacher *et al.*, 2004).

Während sich auf Grund ihrer Robustheit und einfachen Handhabung in der klassischen HPLC überwiegend Säulen von 2,0 bis 4,6 mm Innendurchmesser durchgesetzt haben, werden v. a. im Bereich der Bioanalytik häufig so genannte mikro- oder nanoHPLC-Systeme eingesetzt (Vissers, 1999). Diese ermöglichen auch eine direkte Kopplung der Chromatographie ohne *Split*-System mit der Massenspektrometrie (Whitehouse, *et al.*, 1985). Das *downscaling* der eingesetzten Apparaturen führt zusätzlich zu einer verbesserten Trenneffizienz (Karlsson and Novotny, 1988). Zusätzlich wird weniger Probe benötigt und unter Verwendung von *Split*-freien Pumpen-Systemen wird auch der Verbrauch an mobiler Phase deutlich minimiert. So werden in modernen nanoHPLC-Systemen Säulen mit bis zu 15 μm Innendurchmesser verwendet (Premstaller, *et al.*, 2001, Walcher *et al.*, 2004).

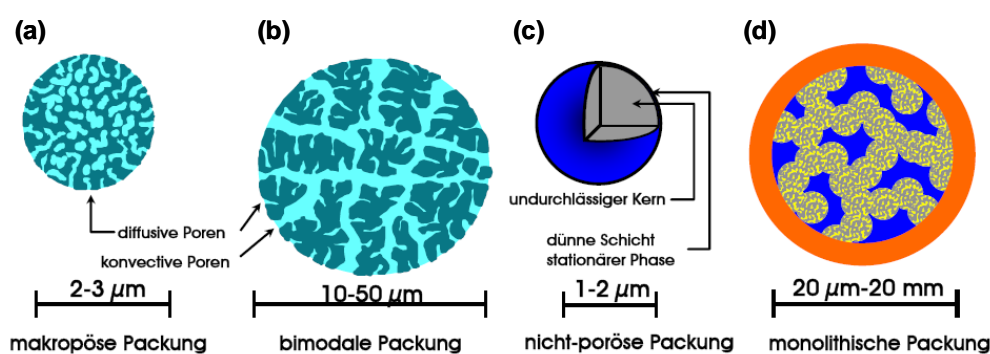


Abb. 1.8: Darstellung verschiedener stationärer Phasen zur chromatographischen Auftrennung von Peptid- bzw. Proteingemischen (Quelle: Prof. Dr. C. G. Huber, Universität Salzburg, per Abbildung).

Zusammenfassend kann man sagen, dass in Kapillarsäulen alle chromatographischen Trennungen, die auch in herkömmlichen HPLC-Säulen Verwendung finden, durchgeführt werden können. Es ist jedoch zu beachten, dass mikro- bzw. nanoHPLC-Systeme generell anfälliger für Störungen sind. So müssen beispielsweise Totvolumina unbedingt minimiert werden, da sich diese in zusätzlichen Bandenverbreiterungseffekten äußern, die die Trenneffizienz stark verringern können (Novotny, 1988).

1.4 Massenspektrometrie in der Peptid- bzw. Proteinanalytik

Die Massenspektrometrie (MS) ist eine physikalische Detektionsmethode. Sie hat nicht nur im Bereich der Peptid- und Proteinanalytik in den letzten Jahren und Jahrzehnten immer größere Bedeutung erlangt und stellt derzeit eine der wichtigsten Methoden der analytischen

Chemie dar. Mit ihrer Hilfe werden Molekülmassen von Ionen im Hochvakuum bestimmt. Hierdurch, sowie durch die Analyse von Fragmentierungsmustern, können auch Informationen über die Struktur der Analyten erhalten werden. Einen zusätzlichen Informationsgehalt stellt die Möglichkeit der Bestimmung der Elementzusammensetzung eines Analyten aus der Isotopenverteilung dar. Es sind Nachweisgrenzen bis in den zeptomol-Bereich möglich (Andren *et al.*, 1994). Wie zuvor bereits kurz beschrieben, war es zunächst nicht bzw. kaum möglich die Massenspektrometrie in der Peptid- und Proteinanalyse einzusetzen. Erst die Entwicklung schonender massenspektrometrischer Ionisierungsverfahren, wie die MALDI (Karas and Hillenkamp, 1988, Tanaka, *et al.*, 1988) oder die Weiterentwicklung der ESI MS, (Meng, *et al.*, 1988) konnten hierbei Abhilfe schaffen. Ab diesem Zeitpunkt konnten auch biologische Makromoleküle ausreichend in der Gasphase ionisiert werden. Die Moleküle blieben dabei überwiegend intakt erhalten. Dies war entscheidend, um die unfragmentierten Analyten untersuchen zu können.

Der generelle Aufbau eines Massenspektrometers ist in Abbildung 1.9 schematisch dargestellt. Es besteht prinzipiell aus drei verschiedenen Bauteilen, die in ein Vakuumsystem eingebunden sind. In der so genannten Ionenquelle werden die Analyten in der Gasphase in Ionen überführt (Ionenerzeugung). Diese Ionen werden dann im Massenanalysator hinsichtlich ihres m/z -Verhältnisses aufgetrennt (Ionentrennung). Anschließend werden die Ionen in einem Detektor aufgezeichnet (Ionennachweis). Beispiele für verschiedene Ionenquellen, Massenanalysatoren und Detektoren eines Massenspektrometers sind in Tabelle 1.3 aufgelistet.

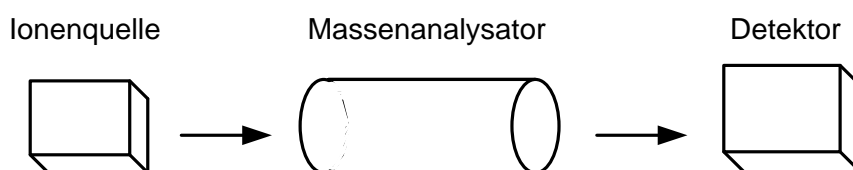


Abb. 1.9: Schematischer Aufbau eines Massenspektrometers.

Da in der vorliegenden Dissertation ausschließlich ein MALDI-TOF/TOF MS verwendet wurde, wird in den folgenden Abschnitten näher auf diese Technik eingegangen.

Tab. 1.3: Aufzählung verschiedener Ionenquellen, Massenanalytoren und Detektoren eines Massenspektrometers.

Ionenquelle	Massenanalysator	Detektor
Elektronenstoß-Ionisation (EI)	Flugzeitanalysator (TOF)	Faraday-Cup
Elektrospray-Ionisation (ESI)	Fouriertransformations-Ionenzyklotronresonanz-Massenspektrometer (FT-ICR)	Konversionsdynode mit Sekundärelektronen-vervielfacher (SEV)
Chemische Ionisation (CI)	Ionenfalle-Massenspektrometer	Szintillationszähler
<i>Fast Atom Bombardement</i> (FAB)	Ionenfalle-Orbitrap-Hybridmassenspektrometer	Vielkanalplatte
<i>Matrix-assisted Laser Desorption/Ionisation</i> (MALDI)	Quadrupol-Massenspektrometer	

1.4.1 Die MALDI-Technik

Seit dem ersten Aufkommen der MALDI MS im Bereich der Peptid- bzw. Proteinanalytik (Karas and Hillenkamp, 1988, Tanaka, *et al.*, 1988) sind ihre fundamentalen Prozesse untersucht und charakterisiert worden (Dreisewerd, 2003, Hillenkamp and Peter-Katalinic, 2007, Jaskolla and Karas, 2011, Karas and Kruger, 2003, Knochenmuss and Zenobi, 2003, Lehmann, 1996).

Im Gegensatz zur ESI MS, zeichnet sich die MALDI MS v. a. durch ihre relativ hohe Toleranz gegenüber Verunreinigungen wie Salzen und Detergenzien aus (Kusmann *et al.*, 1997). Die Abläufe einer massenspektrometrischen Analyse mittels MALDI-Ionisierung lassen sich in fünf verschiedene Prozesse einteilen.

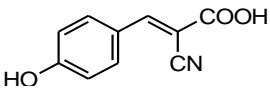
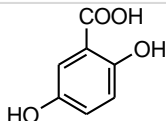
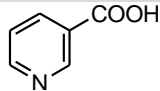
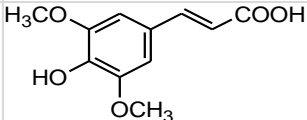
- i) Kokristallisation der Analyten mit einer Matrix
- ii) Laserbeschuss des Kokristalls
- iii) Desorption/Ionisierung des Analyten und der Matrix
- iv) Beschleunigung der Ionen
- v) Massenanalyse

Diese verschiedenen Prozesse werden nun näher erläutert.

i) Kokristallisation der Analyten mit einer Matrix

Als Matrix werden häufig kleine organische Verbindungen gewählt, die in der Lage sind UV-Licht zu absorbieren. Dies ist für den MALDI-Prozess entscheidend. Ihre Hauptaufgabe besteht nicht nur in der Trennung der Analytmoleküle voneinander, sondern auch in der Absorption des Laserlichtes und einer schonenden Übertragung der Energie auf die Probenmoleküle. Hierbei dient sie vermutlich auch als Protonendonator bzw. -akzeptor. Die Matrix wird im hohen molaren Überschuss gegenüber dem Analyten eingesetzt, wobei Verhältnisse von 1.000:1 (Matrix:Analyt, mol/mol) oder höher eingesetzt werden (Stump *et al.*, 2002). Analyt und Matrix werden normalerweise getrennt voneinander in einem geeigneten Lösungsmittel gelöst und dann auf eine Metallplatte, das so genannte *Target*, pipettiert (*dried droplet* Methode). Die Ausbildung des Kokristalls erfolgt durch Verdampfung des Lösungsmittels. Die wichtigsten Matrices, die in der Peptid- bzw. Proteinanalytik verwendet werden, sind in Tabelle 1.4 aufgelistet.

Tab. 1.4: Matrices, die routinemäßig in der Peptid- und Proteinanalytik mittels MALDI MS eingesetzt werden: α -Cyano-4-Hydroxyzimtsäure (Beavis *et al.*, 1992), 2,5-Dihydroxybenzoesäure (Strupat *et al.*, 1991), Nicotinsäure (Karas and Hillenkamp, 1988) und Sinapinsäure (Beavis and Chait, 1989a, Beavis and Chait, 1989b).

Matrix	Wellenlänge [nm]	Analyt	Struktur
α -Cyano-4-Hydroxyzimtsäure (CHCA)	337 und 355	Peptide	
2,5-Dihydroxybenzoesäure (DHB)	266, 337 und 355	Peptide, Proteine, Phosphopeptide	
Nicotinsäure	266	Peptide, Proteine	
Sinapinsäure (SA)	266, 337 und 355	Proteine	

ii) Laserbeschuss des Kokristalls

Nachdem die Ausbildung des Kokristalls aus Analyt und Matrix abgeschlossen ist, kann der Laserbeschuss erfolgen (siehe Abbildung 1.10). Je nach verwendetem Laser werden Frequenzen zwischen 5 und 1000 Hz verwendet, wobei die Beschussdauer meistens im Nanosekunden-Bereich liegt. Der Beschuss kann jedoch auch in kontinuierlicher Bewegung

des Lasers über den Spot (*continous stage motion*) erfolgen. Um qualitativ hochwertige Spektren zu erhalten, werden häufig bis 1.000 oder mehr Einzelspektren akkumuliert. Die am häufigsten eingesetzten Laser operieren im UV-Bereich. Stickstofflaser arbeiten beispielsweise mit einer Wellenlänge von 337 nm und Nd:YAG-Feststoff-Laser mit einer Wellenlänge von 355 nm (Hillenkamp and Peter-Katalinic, 2007). Aber auch Infrarotlaser können zur Desorption/Ionisation eingesetzt werden.

iii) Desorption/Ionisation des Analyten und der Matrix

In der Literatur werden zwei verschiedene Modelle für die Desorption/Ionisierung der Analyten während des MALDI-Prozesses beschrieben. Zum einen wird das Modell der *Lucky Survivor* (Karas *et al.*, 2000, Karas and Kruger, 2003) und zum anderen das Modell der *gas phase protonation* (Knochenmuss and Zenobi, 2003, Knochenmuss, 2006, Zenobi and Knochenmuss, 1998) postuliert. Im ersten Modell wird davon ausgegangen, dass Peptide bzw. Proteine bereits mit positiver Ladung zusammen mit ihren Gegenionen in dem Kokristall eingebaut werden. Durch den Laserbeschuss werden dann so genannte Cluster aus geladenen Analyten und ihren Gegenionen gebildet. Diese Cluster tragen insgesamt keine Ladung, können aber in ungeladene oder positiv geladene Bruchstücke zerfallen, wobei letztere durch stabile Anionen neutralisiert werden können. Aus der Zerstörung des Kokristalls aus Analyten und Matrix können jedoch, eingeleitet durch eine Ladungstrennung, auch positiv geladene Cluster entstehen, die dann durch den Mangel an Gegenionen nicht neutralisiert werden. Diese werden als die *Lucky Survivor* bezeichnet (Karas, *et al.*, 2000). Eine Erweiterung des Modells bietet auch eine Erklärung für die Ausbildung von negativ geladenen Analytionen: Die Deprotonierung der Peptid- bzw. Proteinionen ist endergon, da ihre Carboxylsäuregruppen weniger sauer sind, als die der Gegenionen wie Trifluoracetat oder Chlorid. In dem erweiterten Modell wird die Ladungsneutralisation, die letztendlich zu den *Lucky Survivors* führt, sowie die Deprotonierung der Analyten durch geladene (negative oder positive) Matrixionen durchgeführt (Karas and Kruger, 2003). Da es sich bei negativ geladene Matrixionen i. d. R. um Carboxylate oder delokalisierte Phenolate handelt, die ähnlich sauer sind wie die Carboxylsäuregruppen der Analyten, ist eine Deprotonierung der Analyten zwar prinzipiell möglich, aber nicht sehr ausgeprägt.

Im Gegensatz dazu wird im Modell der *gas phase protonation* zunächst von ungeladenen Analyten in der Gasphase ausgegangen. Durch Zusammenstöße dieser ungeladenen Analytmoleküle mit geladenen (negativ oder positiv) Matrixionen entstehen durch Protonen-Übertragungsreaktionen deprotonierte oder protonierte Analytionen (Knochenmuss and

Zenobi, 2003, Knochenmuss, 2006, Zenobi and Knochenmuss, 1998). Eine aktuelle Studie konnte nun belegen, dass beide Modelle gleichzeitig ablaufen, und dass ihr Verhältnis sowohl von experimentellen Bedingungen als auch von den Eigenschaften der Matrix und den Analyten abhängen (Jaskolla and Karas, 2011).

Darüber hinaus werden in der MALDI MS überwiegend einfach geladene Ionen $[M+H]^+$ detektiert. Unter Abhängigkeit der chemischen Eigenschaften der Analyten und deren Konzentration, kommt es mit zunehmendem Molekulargewicht (z. B. bei Proteinen) neben diesen einfach geladenen Ionen auch zur Ausbildung von Dimeren der Form $[2M+H]^+$ und/oder mehrfach geladenen Monomeren $[M+nH]^{n+}$.

iv) Beschleunigung der Ionen

Die Analytmoleküle liegen nach der Desorption/Ionisation als Ionen in der Gasphase vor und erhalten die gleiche kinetische Energie E_{kin} , indem sie in einem elektrischen Feld beschleunigt werden. Diese Energie ist u. a. von der Masse und der Ladung des jeweiligen Ions abhängig. Dies ist die Grundlage für die Massenanalyse mittels Flugzeitanalysatoren (siehe Abbildung 1.10).

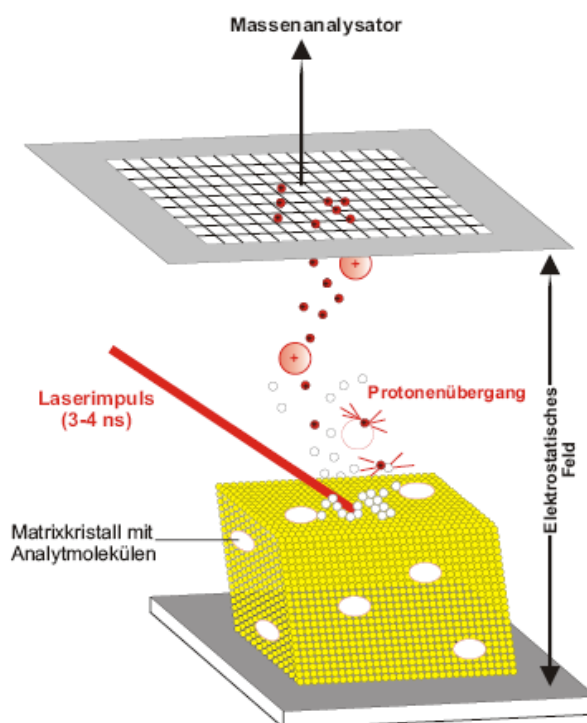


Abb. 1.10: Schematischer Ablauf der Schritte ii), iii) und iv) des MALDI-Prozesses (Lottspeich and Zorbas, 1998).

Beim MALDI-Prozess können sowohl Inhomogenitäten des Kokristalls aus Matrix und Analyten, als auch andere Prozesse eine Energieunschärfe verursachen. Deshalb ist es notwendig, der daraus resultierenden, reduzierten Auflösung entgegenzuwirken. Aus diesem Grund wurden verschiedene Methoden, wie beispielsweise die verzögerte Extraktion (*delayed extraction*) der Ionen entwickelt (Brown and Lennon, 1995, Colby *et al.*, 1994). Diese ermöglichen eine signifikante Verbesserung der Massengenauigkeit und der Auflösung. Hierbei wird das elektrische Feld zur Beschleunigung der Ionen zeitversetzt zum Laserimpuls eingeschaltet. Durch die geeignete Wahl an Verzögerungszeit und Extraktionsfeldstärke lässt sich der Einfluss der Startgeschwindigkeitsverteilung auf die Flugzeitverteilung der Ionen ausgleichen.

v) Massenanalyse

Obwohl MALDI MS am häufigsten mit dem so genannten Flugzeitanalysator (TOF) gekoppelt wird, ist sie prinzipiell auch mit anderen Massenanalysatoren kombinierbar. So gibt es beispielsweise auch ein MALDI MS, das mit einem Ionenfallen-Orbitrap Hybridmassenanalysator ausgestattet ist.

1.4.2 Der Flugzeit-Massenanalysator

Wie zuvor bereits kurz beschrieben, ist der Flugzeit-Massenanalysator in Kombination mit MALDI-Ionisierung der wohl am häufigsten verwendete Massenanalysator. Nachdem die Analytmoleküle nach der Desorption/Ionisation als Ionen in der Gasphase vorliegen und in einem elektrischen Feld beschleunigt wurden, erhalten sie die gleiche kinetische Energie E_{kin} . Analyten unterschiedlicher Masse können nun in einem feldfreien Raum, der so genannten Flugröhre bzw. Driftstrecke, aufgetrennt werden. Grundlage hierfür ist die Abhängigkeit der Geschwindigkeit der Ionen (mit gleicher kinetischer Energie) von ihren m/z -Verhältnissen. Diese Abhängigkeit wird durch Gleichung 1.2 ausgedrückt. Somit können, bei bekannter Beschleunigungsspannung und Flugstrecke durch Messung der Flugzeit, die m/z -Werte der Ionen bestimmt werden. Dies ergibt sich im Folgenden aus den Gleichungen 1.2 bis 1.4. Typische Flugzeiten in der MALDI MS liegen zwischen wenigen μs und einigen 100 μs , wobei die Driftstrecken meist 1 bis 4 m betragen.

$$E_{kin} = \frac{1}{2} \cdot m \cdot v^2 = z \cdot e \cdot U \quad (\text{Gleichung 1.2})$$

E_{kin} = kinetische Energie

m = Masse des Ions

v = Geschwindigkeit des Ions nach der Beschleunigungsstrecke im feldfreien Flugrohr

z = Ladungszahl

e = Elementarladung

U = Beschleunigungsspannung

Die Gesamtflugzeit t ergibt sich aus der Länge der feldfreien Driftstrecke des Flugrohres L und der Geschwindigkeit v :

$$t = \frac{L}{v} \quad (\text{Gleichung 1.3})$$

Durch Kombination der beiden Gleichungen erhält man Gleichung 1.4:

$$\frac{m}{z} = \frac{2 \cdot e \cdot U}{L^2} \cdot t^2 \quad (\text{Gleichung 1.4})$$

Die Ladung der detektierten Ionen erhält man mit genügend hoher Auflösung R des Massenspektrometers aus den Isotopenpeaks, wobei es sich wie zuvor bereits beschrieben bei MALDI MS fast ausschließlich um einfach geladene Ionen handelt.

Die Auflösung R stellt einen der wichtigsten Parameter eines Massenspektrometers dar. Sie bezeichnet die Fähigkeit, Ionen mit geringen Differenzen bezüglich ihres m/z -Verhältnisses noch getrennt voneinander detektieren zu können. Sie wird durch Gleichung 1.5 ausgedrückt.

$$R = \frac{m}{\Delta m} = \frac{m_1}{(m_2 - m_1)} \quad (\text{Gleichung 1.5})$$

R = Auflösung

m_1 = Masse des Ions 1

m_2 = Masse des Ions 2

Verwendet man einen Flugzeit-Massenanalysator, so wird die Auflösung aus einem einzigen Peak bestimmt und Δm als die Halbwertsbreite FWHM (engl. *Full width at half maximum*) des Peaks, d.h. bei 50% der Gesamtpeakhöhe, definiert.

Um einer Peakverbreiterung, die durch die Geschwindigkeitsverteilung der Ionen verursacht wird, und der daraus resultierenden geringeren Auflösung entgegenzuwirken, wurden zwei

verschiedene Techniken entwickelt: Zum einen die bereits zuvor genannte verzögerte Extraktion der Ionen (Brown and Lennon, 1995, Colby, *et al.*, 1994) und zum anderen die Entwicklung des so genannten Reflektors (Mamyrin, 1994). Dieser ist in Abbildung 1.11 dargestellt. Er besteht z. B. aus einer Reihe von Metallringen mit zunehmendem Potential und einer Ladung, die ca. 1,1-mal höher ist, als diejenige der Ionenquelle. Das dadurch aufgebaute Feld ist gleichpolig zu der Ladung der Ionen, wodurch diese abgebremst, in entgegengesetzter Richtung wieder beschleunigt und in Richtung des Detektors weitergeleitet werden. Besitzen Ionen mit gleichem m/z -Verhältnis auf Grund der Energieverteilung unterschiedliche kinetische Energien, so dringen sie unterschiedlich tief in das Reflektorfeld ein (siehe Abbildung 1.11). Dadurch wird die Energieverteilung minimiert und eine verbesserte Auflösung erreicht (Hillenkamp and Peter-Katalinic, 2007).

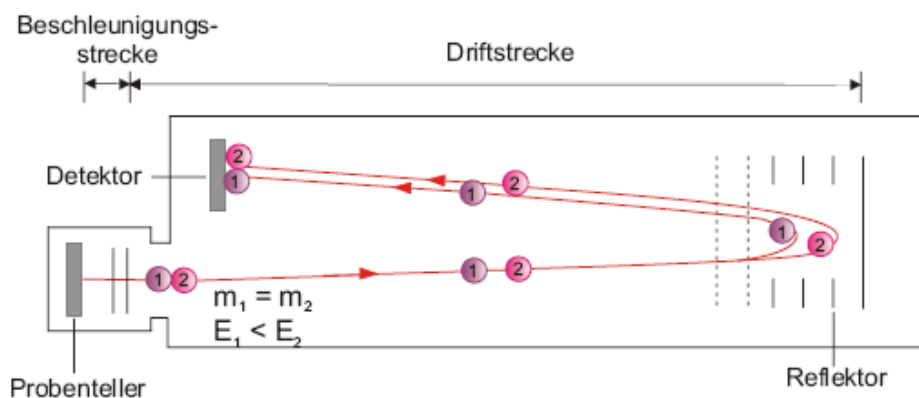


Abb. 1.11: Schematischer Aufbau eines Reflektor-Flugrohrs. Die Ionen mit gleichem m/z -Verhältnis mit der größeren kinetischen Energie (E_2) dringen tiefer in den Reflektor ein, als die Ionen mit der kleineren kinetischen Energie (E_1). Sie legen damit eine längere Flugstrecke zurück (äußere Trajektorie), so dass eine bessere Fokussierung am Detektor erreicht wird (Lottspeich and Zorbas, 1998).

1.4.3 Die Vielkanalplatte als Detektor

Die Vielkanalplatte oder *multi channel plate* (MCP) ist der wohl am häufigsten verwendete Detektor in Verbindung mit Flugzeit-Massenanalysatoren (siehe Abbildung 1.12).

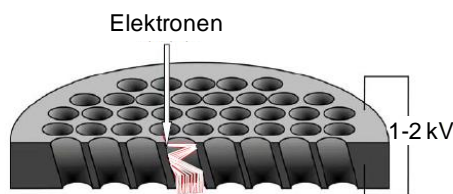


Abb. 1.12: Darstellung einer Vielkanalplatte (Quelle: Prof. DR C. G. Huber, Universität Salzburg, per Abbildung).

Sie besteht aus Millionen *channel electron multiplier* (CEM), deren Innendurchmesser in den μm -Bereich minimiert sind. Durch die hohe Anzahl wird ein ausreichender elektrischer Strom gewährleistet, wobei die angelegte Spannung 1 bis 2 kV beträgt. Auf diese Weise können Verstärkungen des Elektronensignals von bis zu 10^5 erreicht werden (De Hoffmann *et al.*, 1996). Analytione, die auf einen Kanal der MCP treffen, schlagen dort ein Elektron heraus. Dieses fliegt innerhalb des Kanals weiter und setzt dort weitere Elektronen frei. Dadurch wird der Elektronenstrom amplifiziert, bis ein Signal detektiert werden kann.

1.4.4 LC-MALDI MS

Im Gegensatz zu ESI MS ist die Kopplung der MALDI MS mit chromatographischen Trennmethode immer noch eine relativ selten verwendete Methode zur Analyse komplexer Peptid- und Protein-Mischungen. Dennoch wurde eine Reihe von Anwendungen entwickelt (Chen *et al.*, 2005, Ericson *et al.*, 2003, Mirgorodskaya *et al.*, 2005). Diese reichen von Studien, in denen LC-MALDI MS qualitativ für die *bottom-up*-Strategie im Feld der Proteomanalyse (Tholey *et al.*, 2008) eingesetzt wurde bis hin zu quantitativen Analysen (Tiberti *et al.*, 2010). Nachdem die Analyten von der Säule eluieren, werden sie über ein $\mu\text{-T}$ -Stück mit MALDI-Matrix gemischt und dann automatisch auf ein MALDI-Target gespottet (siehe Abbildung 1.13).

Während in der LC-ESI MS die Analyse *online* erfolgt, wird sie in der LC-MALDI MS *offline* durchgeführt. Die Analyten eluieren hier nicht während der chromatographischen Trennung kontinuierlich in das Massenspektrometer. Vielmehr wird ein gespotteter LC-Lauf erst nach seiner Beendigung im Massenspektrometer analysiert. Die einzelnen Spots werden i. d. R. zuerst im MS-Modus gemessen und Vorläuferionen (*precursor*) anschließend in denjenigen Spots fragmentiert, in denen sie in der höchsten Konzentration vorliegen.

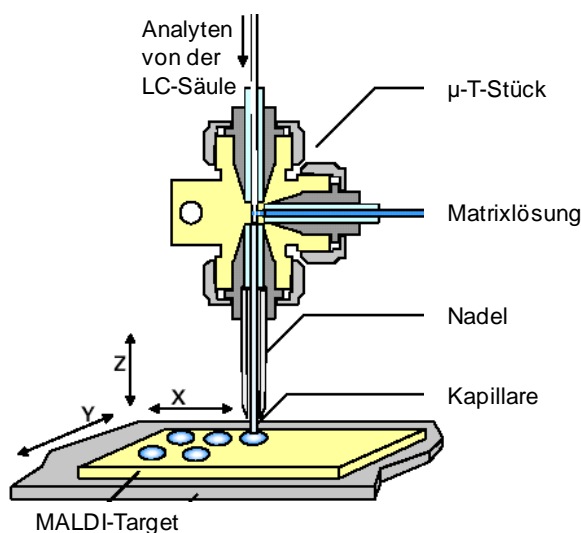


Abb. 1.13: Schematische Darstellung der LC-MALDI-Kopplung über ein μ -T-Stück. Die von der Säule eluierten Analyten werden über das μ -T-Stück mit MALDI-Matrix gemischt und automatisch auf ein MALDI-Target gespottet. Die MALDI MS- und MS/MS-Analyse erfolgt anschließend *offline* (Tholey, 2006).

1.4.5 Peptid- bzw. Proteinidentifizierung

Prinzipiell kann zwischen zwei verschiedenen Vorgehensweisen zur Identifizierung der Peptide bzw. der Proteine unterschieden werden: Zum einen gibt es das so genannte *de novo sequencing* und zum anderen die Datenbankanalyse. *De novo sequencing* wird v. a. dann angewendet, wenn das Proteom eines Organismus noch unbekannt ist und somit keine Datenbank zur Verfügung steht (Nesvizhskii *et al.*, 2007, Seidler *et al.*, 2010). Die MS/MS-Spektren werden in diesem Fall ohne den Vergleich mit einer Datenbank interpretiert.

Im Gegensatz dazu werden die MS/MS-Spektren in einer Datenbanksuche mit theoretischen MS/MS-Spektren verglichen.

Zusätzlich zur Fragmentierung von Analyten mittels der so genannten *post source decay* (PSD) Methode, in der eine metastabile Fragmentierung in der feldfreien Driftstrecke nach der Beschleunigungsstrecke erfolgt (Spengler *et al.*, 1992), sind in MALDI Tandem Massenspektrometern auch MS/MS-Experimente mittels kollisionsinduzierter Dissoziation (CID) möglich (Hayes and Gross, 1990). Hier wird die Fragmentierung durch Stöße des Analytions mit Gasmolekülen bzw. -atomen (z. B. Helium) in einer Stoßkammer eingeleitet. Die Translationsenergie der Ionen wird dabei teilweise in Schwingungsenergie überführt. Dies führt wiederum zu den sequenzspezifischen Fragmentierungen im Peptidrückgrat (siehe Abbildung 1.14).

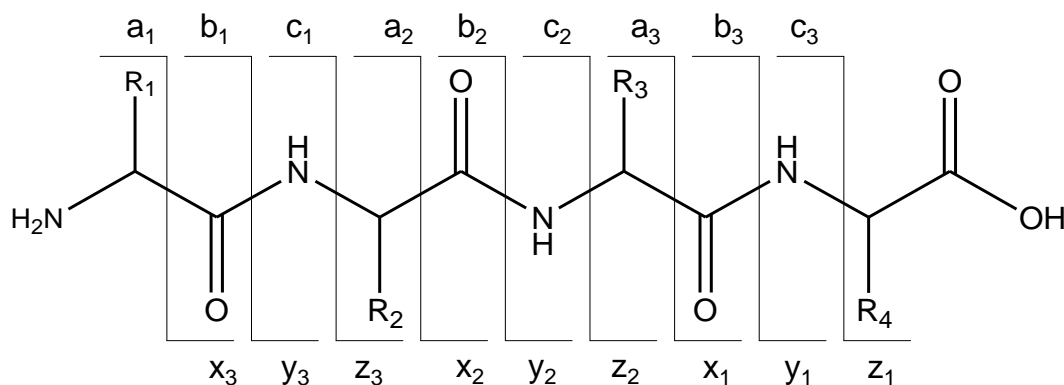


Abb. 1.14: Mögliche Fragmentierungsstellen des Peptidrückgrats unter Verwendung des *post source decay* (PSD) oder der kollisionsinduzierten Dissoziation (CID). Die Nomenklatur der Fragmente erfolgt nach Biemann (Biemann, 1990).

Die N-terminalen Sequenzen a_n , b_n , c_n entstehen, wenn die Ladung am N-terminalen Ende des Peptidions verbleibt. Erfolgt dies am C-terminalen Ende, so entstehen die C-terminalen Sequenzen x_m , y_m und z_m (Biemann, 1990, Roepstorff and Fohlman, 1984). Darüber hinaus kann auch eine Fragmentierung der Seitenketten erfolgen. Werden durch die Fragmentierungen Wasser- oder Ammoniakmoleküle abgespalten, spricht man von sogenannten Neutralverlusten. Genaue Kenntnisse dieser Fragmentierungen sind notwendig, um die MS/MS-Spektren richtig interpretieren zu können. Hauptsächlich entstehen unter den herkömmlich verwendeten Fragmentierungsbedingungen, bei denen in der Stoßkammer Drücke von 5×10^{-7} bis 4×10^{-5} Torr herrschen, die a_n , b_n und y_m Fragmentionen.

Im Jahre 1998 wurde eine weitere Art der Fragmentierung vorgestellt, die so genannte Elektronen Einfang Dissoziation (ECD). Diese Methode beruht darauf, dass freie Elektronen auf ein (mehrfach) positiv geladenes Ion übertragen werden und dieses dadurch fragmentiert (Zubarev *et al.*, 1998). Werden die Elektronen nicht als freie Elektronen übertragen, sondern stammen sie aus Radikal-Anionen von Anthracen oder Azobenzen, so spricht man von der Elektronen Transfer Dissoziation (ETD), die ebenfalls eine sanfte Art der Fragmentierung darstellt (Pitteri *et al.*, 2005, Syka *et al.*, 2004). Im Gegensatz zu CID- und PSD-Fragmentierungen werden bei der ETD-Fragmentierung v. a. c_n und z_m Fragmentionen gebildet.

Auf Grundlage der Fragmentierungsregeln wird eine Datenbank mit simulierten MS/MS-Spektren erstellt, und jedes gemessene Spektrum wird mit den simulierten Spektren der Datenbank verglichen. Dadurch kann die Sequenz der Peptide bestimmt und damit letztendlich die Proteine identifiziert werden. Hierzu wird ein *in silico* Verdau des gesamten Proteoms des zu untersuchenden Organismus durchgeführt, die theoretischen m/z -

Verhältnisse der Peptidfragmente berechnet und somit für jedes mögliche Peptid eines Proteins ein simuliertes MS/MS-Spektrum erstellt. Zusätzlich müssen auch Parameter wie z. B. die verwendete Protease oder chemische Modifikationen berücksichtigt werden.

Beispiele für Datenbanken sind der *International Protein Index* (IPI), *Entrez Protein* oder *UniProt/Swiss-Prot*. Als erste Suchmaschine für Datenbanksuchen wurde SEQUEST entwickelt (Eng, *et al.*, 1994, Yates *et al.*, 1995). Dieser folgten weitere Suchmaschinen wie Mascot (Perkins *et al.*, 1999), Protein Prophet (Keller *et al.*, 2002) oder OMSSA (Geer *et al.*, 2004). In der hier vorliegenden Arbeit wurde ausschließlich Mascot als Suchmaschine verwendet. Diese Software nutzt eine wahrscheinlichkeitsbasierte Integration, den Mowse-Scoring-Algorithmus (*Molecular weight search*, (Perkins, *et al.*, 1999)). Je nach Grad an Übereinstimmung der gemessenen MS/MS-Spektren mit den simulierten MS/MS-Spektren der Datenbank, wird ein bestimmter Wert (*score*) zugewiesen. Dieser drückt die Wahrscheinlichkeit für einen Zufallstreffer in der Peptid/Protein-Identifizierung aus. Je mehr MS/MS-Spektren von Peptiden eines Proteins zugeordnet werden können, desto höher wird der *score* und desto niedriger ist die Wahrscheinlichkeit eines Zufallstreffers. Um die Wahrscheinlichkeit eines Zufallstreffers zu erniedrigen wurde beispielsweise die Abschätzung der so genannten falsch-positiv Rate (FDR) über eine *reversed* Datenbanksuche eingeführt (Elias and Gygi, 2007). Alle Spektren werden zusätzlich auch gegen eine *decoy* Datenbank gesucht, die ebenfalls die Aminosäuresequenzen der eigentlichen Datenbank enthält. Die Anordnung der Aminosäuren erfolgt in der *decoy* Datenbank dann z. B. in umgekehrter Reihenfolge, also vom C- zum N-Terminus. Unter der Annahme, dass eine falsch-positive Zuordnung sowohl für die eigentliche als auch die *decoy* Datenbanksuche gleich wahrscheinlich ist, kann man so die Anzahl an falsch-positiven Zuordnungen abschätzen. Über die Bestimmung eines *cutoff scores* kann die maximale Anzahl an korrekten Zuordnungen, bei minimaler Anzahl an falsch-positiven Zuordnungen, ermittelt werden. Man beachte dabei jedoch, dass nur eine Annahme getroffen werden kann wie viele falsch-positive Zuordnungen erfolgt sind, aber nicht, um welche es sich dabei handelt.

In Abbildung 1.15 ist das hier verwendete Massenspektrometer schematisch abgebildet.

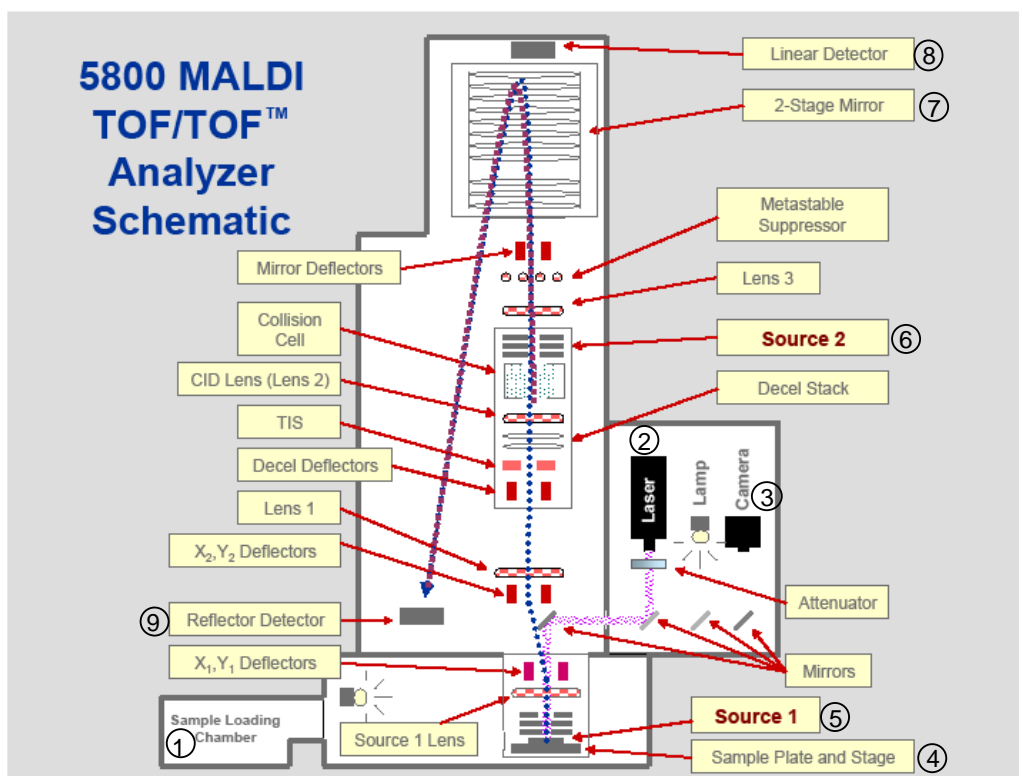


Abb. 1.15: Schematische Darstellung des 5800 MALDI-TOF/TOF Analyser (AB Sciex). Die wichtigsten Bauteile sind mit einer Nummerierung versehen. (1) Ladekammer der *Targets*; (2) Laser; (3) Kamera; (4) Halterung des *Targets*; (5) Ionenquelle 1, (6) Ionenquelle 2/Stoßkammer; (7) Reflektor; (8) linearer Detektor, (9) Reflektor-Detektor. (Quelle: AB SCIEX, Darmstadt, Deutschland)

1.5 Analytik von Proteasen und Proteasen-katalysierten Reaktionen

Als Proteasen werden Enzyme bezeichnet, welche die hydrolytische Spaltung von Peptidbindungen katalysieren. Sie sind auch unter dem Namen Peptidasen oder Proteinasen bekannt. Diese können in Abhängigkeit der Lage der Spaltbindung im Inneren oder am terminalen Ende der Polypeptidkette eines Proteins in so genannte Endopeptidasen und Exopeptidasen eingeteilt werden. Da es Proteasen wie beispielsweise Cathepsin B gibt, die sowohl eine Endo- als auch eine Exopeptidase-Aktivität besitzen (Bond and Barrett, 1980), wurde der Begriff proteolytische Enzyme eingeführt (Barrett and McDonald, 1986). Alle bislang bekannten humanen Proteasen der MEROPS Datenbank (Rawlings *et al.*, 2010, Rawlings *et al.*, 2012) bilden 1,3% des kompletten menschlichen Proteoms (*UniProt*, 31.05.2010). Nach Einteilung der *Enzyme Commission* (EC) Nummerierung werden die Proteasen zur Hauptklasse 3 (Hydrolasen) und Unterklasse 4 (Peptidbindungen als Substrate)

gezählt. Des Weiteren lassen sie sich z. B. nach ihrem Katalysemechanismus oder genauer nach funktionellen Gruppen innerhalb des aktiven Zentrums einteilen (siehe Tabelle 1.5).

Tab. 1.5: Einteilung der Proteasen nach der funktionellen Gruppe im aktiven Zentrum. Diese ist direkt an der Hydrolyse der Peptidbindung beteiligt.

Protease	Funktionelle Gruppe im aktiven Zentrum	Beispiele
Aspartatprotease	Aspartat, aktiviertes Wassermolekül	Renin und Pepsin
Cysteinprotease	Cystein	Papain und Cathepsin B
Serinprotease	Serin	Trypsin und Chymotrypsin
Metalloprotease	sehr häufig Zn^{2+} , aktiviertes Wassermolekül	Thermolysin und Carboxypeptidase A
Threoninprotease	Threonin	katalytische Untereinheiten des Proteasoms

Um die Substratspezifität einer Protease richtig beschreiben zu können, wurde von Schechter und Berger eine Nomenklatur eingeführt, die in Abbildung 1.16 dargestellt ist (Schechter and Berger, 1967). Hier werden die N-terminalen Spaltsequenzen als *nonprime* und die C-terminalen Spaltsequenzen als *prime* (engl.: Strich) bezeichnet. Die zu spaltende Peptidbindung wird N-terminal von den Aminosäuren P1, P2, P3... und P1', P2', P3'... eingegrenzt. Diese Aminosäuren binden dann an die Bindungsstellen S1, S2, S3... bzw. S1', S2', S3'... im aktiven Zentrum der Protease.

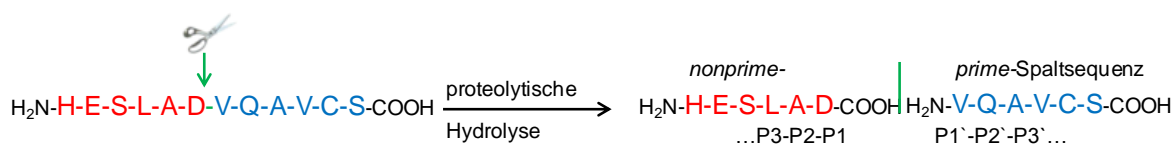


Abb. 1.16: Darstellung der Nomenklatur für die Aminosäuren N- und C-terminal einer proteolytischen Spaltstelle nach Schechter und Berger (1967). Die proteolytisch spaltbare Peptidbindung ist innerhalb der Aminosäuresequenz des Beispielpeptides HESLADVQAVCS grün markiert. Die N-terminale Spaltsequenz HESLAD wird als *nonprime*, die C-terminale Spaltsequenz VQAVCS als *prime* (engl.: Strich) bezeichnet. Die Spaltstelle wird N-terminal von den Aminosäuren P1 (D), P2 (A), P3 (L)... und C-terminal von P1' (V), P2' (Q), P3' (A)... umgeben.

Proteolyse kann als eine irreversible posttranslationale Modifikation betrachtet werden, die *in vivo* eine wichtige Rolle sowohl in Regulationsmechanismen als auch in zahlreichen pathophysiologischen Prozessen spielt (Hojilla *et al.*, 2008, Kessenbrock *et al.*, 2010, Lopez-Otin and Hunter, 2010). Darüber hinaus werden proteolytische Prozesse auch *in vitro* biotechnologisch genutzt. So werden Proteasen beispielsweise in Waschmitteln eingesetzt (Poldermans, 1990) oder finden Verwendung in der Proteomanalyse zur Hydrolyse intakter Proteine zu Spaltpeptiden, die anschließend über massenspektrometrische Methoden identifiziert werden.

Wichtige Ziele innerhalb der Proteasenforschung sind z. B. die folgenden Parameter:

- i) Die Identifizierung von Substraten und aus der Proteolyse resultierender Produkte.
- ii) Die Bestimmung der Substratspezifität, also der Spaltsequenzen, die von einer bestimmten Protease hydrolysiert werden.
- iii) Die Gewinnung quantitativer Informationen bezüglich der Aktivität oder kinetischer Eigenschaften einer bestimmten Protease.
- iv) Die Bindungspartner, die dazu nötig sind eine Protease zu aktivieren/deaktivieren (Signalkaskade)
- v) Die Wirkungsweise von Inhibitoren auf die Aktivität von Proteasen.

1.5.1 Chemische und molekular-biologische Methoden der Proteasenforschung

Da sich die vorliegende Dissertation ausschließlich mit massenspektrometrisch (MS)-basierten Methoden zur Bestimmung proteolytischer Prozesse beschäftigt, wird an dieser Stelle nur ein kleiner Überblick der chemischen und molekular-biologischen Methoden gegeben.

Bei der Bestimmung der Spaltspezifität einer Protease sind v. a. Techniken wie *isokinetic peptide mixture assays* (Ostresh *et al.*, 1994), *Positional Scanning - Substrate Combinatorial Libraries* (PS-SCL) (Rano *et al.*, 1997), und *substrate-activity screening assays* (SAS) (Patterson *et al.*, 2007) zu nennen. Die *fluorescent resonance energy transfer* (FRET) Methode wird zur Analyse der kinetischen Eigenschaften von Proteasen mittels Fluoreszenzmarkierter Substrate verwendet (Matayoshi *et al.*, 1990). Diese Techniken sind in der Literatur ausführlich beschrieben (auf dem Keller and Schilling, 2010, Diamond, 2007, Fonovic and Bogyo, 2008, Lim and Craik, 2009, Poreba and Drag, 2010, van Domselaar *et al.*, 2010). Der Nachteil der meisten Techniken besteht v. a. darin, dass man Vorkenntnisse

über die zu untersuchende Protease und/oder ihre Substrate benötigt. Dadurch wird die Charakterisierung gänzlich unbekannter Proteasen sehr erschwert.

1.5.2 Massenspektrometrisch (MS)-basierte Methoden der Proteaseforschung

MS-basierte Methoden erlauben sowohl die Bestimmung der Aktivität von Proteasen als auch die Identifizierung der Substratspezifität von Proteasen und wurden in vielen Studien etabliert (auf dem Keller and Schilling, 2010, Rogers and Overall, 2013, van den Berg and Tholey, 2012). Sie zeichnen sich durch hohe Sensitivität und hohe Selektivität aus und können generell in zwei verschiedene Klassen eingeteilt werden. So kann der Verdau mit der zu testenden Protease entweder auf Peptid- (Becker-Pauly *et al.*, 2011, Biniossek *et al.*, 2011, Paes Leme *et al.*, 2011, Schilling and Overall, 2008, Schilling *et al.*, 2011) oder Proteinebene (Bredemeyer *et al.*, 2004, Enoksson *et al.*, 2007, Impens *et al.*, 2010, Overall *et al.*, 2004, Prudova *et al.*, 2010, Van Damme *et al.*, 2010) durchgeführt werden.

Zur letzteren Klasse gehört beispielsweise die *Amino-terminal oriented mass spectrometry of substrates* (ATOMS)-Strategie (Doucet and Overall, 2011a, Doucet and Overall, 2011b). Hier werden aufgereinigte Substrat-Proteine entweder mit einer *target*-Protease oder zur Kontrolle nur mit einem Puffer versetzt. Durch die Zugabe der *target*-Protease werden neo-N-Termini gebildet, die in der Kontrollprobe nicht vorhanden sind. Nachdem in beiden Aliquoten eine Reduzierung und Alkylierung der Cysteinseitenketten durchgeführt wurde, erfolgt eine reduktive Dimethylierung der ϵ -Aminogruppen der Lysine, sowie der N-Termini. In der Kontrollprobe wird dies mit herkömmlichem Formaldehyd und in der Probe mit der *target*-Protease mit schwerem Formaldehyd durchgeführt, dadurch ergibt sich ein Unterschied von 6 Da, der in der MS-Analyse leicht festgestellt werden kann. Die beiden Aliquote werden dann vereinigt und tryptisch verdaut. Die durch die *target*-Protease generierten Peptide weisen an ihren (neo-)N-Termini die schwere Dimethylierung auf und sind nur als einzelne Signale im MS-Spektrum enthalten. Peptide, die auch in der Kontrollprobe enthalten sind erscheinen dahingegen als Doublets im MS-Spektrum. Über eine bioinformatische Auswertung können dann die Spaltsequenzen der *target*-Protease innerhalb der Substrat-Proteine identifiziert werden.

Einen weiteren Ansatz stellt die so genannte *combined fractional diagonal chromatography* (COFRADIC)-Methode dar (Gevaert *et al.*, 2003). Sie dient ebenfalls dazu die N-Termini von Substrat-Proteinen zu identifizieren. Zuerst werden die N-Termini und ϵ -Aminogruppen der Lysinseitenketten acetyliert und damit blockiert, dann erfolgt ein tryptischer Verdau. Die

dabei generierten Peptide werden mittels einer RP-HPLC aufgetrennt und fraktioniert und die einzelnen Fraktionen dann mit 2,4,6-Trinitrobenzolsulfonsäure (TNBS) behandelt. Die durch den tryptischen Verdau entstandenen neo-N-Termini werden aryliert und weisen dadurch bei der erneuten RP-HPLC Auftrennung und Fraktionierung einen starken (hydrophoben) Retentionszeitshift auf. Auf diese Weise können die Peptide mit den acetylierten N-Termini von den anderen Peptiden getrennt und anschließend mit (LC)-MS/MS analysiert werden.

Die COFRADIC-Methode wurde später um zwei zusätzliche Schritte erweitert um die Anreicherung der N-terminalen Peptide noch zu verbessern (Staes *et al.*, 2008). Hierzu wurde zum einen eine zusätzliche Auftrennung der Probe über eine SCX-Kartusche, sowie die enzymatische Behandlung der Probe mit einer Glutaminsäure-Aminopeptidase durchgeführt. Dadurch konnte gezeigt werden, dass fast 95% der über COFRADIC sortierten Peptide acetylierte N-Termini besitzen. Außerdem konnten, verglichen mit der unmodifizierten COFRADIC-Methode, die Proteinidentifizierungen erhöht werden.

Ein Beispiel für die Anwendung des COFRADIC-Ansatzes ist die Kombination mit dem so genannten *ribosome profiling*, wodurch die Identifizierungsrate der Proteine um 2,5% erhöht werden konnte. Hierdurch wurden z. B. neue Proteinprodukte oder neue *splice* Varianten gefunden (Menschaert *et al.*, 2013).

Auch bei der *terminal amine isotopic labeling of substrates* (TAILS)-Methode werden die N-Termini und ϵ -Aminogruppen der Lysinseitenketten blockiert, indem sie dimethyliert werden (Kleifeld *et al.*, 2010). Dies kann aber auch über andere Reagenzien erfolgen, wodurch die Messung verschiedener Proben in nur einem Experiment ermöglicht wird (Doucet *et al.*, 2011, Kleifeld *et al.*, 2011). Vor dieser Blockierungsreaktion kann auch eine *target*-Protease zugegeben werden. Eine Unterscheidung ob ein Peptid durch die *target*-Protease generiert wurde, oder ob es bereits in der Probe enthalten war kann dann durch die Verwendung von Reagenzien durchgeführt werden, die unterschiedlich isotopenmarkiert sind. Zu diesem Zweck wird wiederum die Probe einmal mit der *target*-Protease und einmal nur mit Puffer behandelt. Nach Vereinigung der Proben erfolgt dann ein tryptischer Verdau. Dadurch werden freie neo-N-Termini gebildet, die durch ein aminospezifisches Polymer abgefangen werden. Vorteil dieser negativen Aufreinigung ist, dass auf diese Weise in einer LC-MS/MS Analyse auch Peptide identifiziert werden können, deren N-Termini derivatisiert sind, wie z. B. acetylierte oder zyklische N-Termini von Proteinen.

Zusätzlich wurde auch CLIPPER entwickelt. Dabei handelt es sich um ein add-on der *Trans-Proteomic Pipeline* (TPP), das die automatische Auswertung der Daten von TAILS-Experimenten ermöglicht (auf dem Keller and Overall, 2012). Solche Experimente wurden

bereits in zahlreichen Studien durchgeführt. So wurde der TAILS-Ansatz zur Identifizierung natürlicher Substrate der Dipeptidyl-Peptidasen 8 und 9 (Wilson *et al.*, 2013), zur Analyse der Metalloproteasen Meprin α und Meprin β im Zusammenhang der Kollagen-Beschaffenheit der Haut (Broder *et al.*, 2013), sowie für das Profiling von N-Termini für den Import von Proteinen in Plastide innerhalb des Modelorganismus *Thalassiosira pseudonana* (Huesgen *et al.*, 2013) verwendet.

Die Weiterentwicklung und das große Potential des TAILS-Ansatzes zeigt sich auch durch den *iTRAQ-based terminal amine isotopic labeling of substrates* (iTRAQ-TAILS)-Ansatz (Prudova, *et al.*, 2010). Hierbei handelt es sich ebenfalls um eine LC-MS/MS-basierte Strategie, die zur Analyse der Spaltspezifität von Matrixmetalloproteasen (MMP) entwickelt wurde. Zusätzlich zu der Kontrolle können bis zu drei unterschiedliche Proben zeitgleich analysiert werden. Dieser Ansatz wurde dann auch für die Erforschung proteolytischer Prozesse bei Hautentzündungen angewandt (auf dem Keller *et al.*, 2013).

Es gibt auch weitere Methoden auf Proteinebene, die sich die Vorteile der Quantifizierung auf MS/MS-Ebene unter der Verwendung von isobaren stabilen Isotopenmarkierungen, wie iTRAQ (Ross, *et al.*, 2004), zu Nutze machen. So wurde z. B. ein Ansatz für die Identifizierung von Caspase-3 Spaltstellen-Motiven (N-terminale proteolytische Peptide) in einer Mischung aus aktiver und inaktiver Protease (Enoksson, *et al.*, 2007) etabliert. Hierzu wurde eine Mischung aus sieben vermuteten Caspase-3 Substratproteinen verwendet, die in zwei gleiche Mischungen A und B geteilt wurde. Während Mischung A mit inaktiver Caspase-3 versetzt wurde, kam in Mischung B die aktive Variante zum Einsatz. Nach Guanidierung zum Schutz der Lysin- ϵ -Aminogruppen werden die verbleibenden N-Termini mit iTRAQ 114 (Mischung A) bzw. iTRAQ 117 (Mischung B) markiert. Die Proben werden im Verhältnis 1:1 vereinigt und nach tryptischem Verdau anschließend mittels LC-MALDI MS analysiert, wobei folgende Vorgehensweise gewählt wurde:

- i) Analyse auf MS-Ebene zur Ermittlung der möglichen *precursor*
- ii) *Survey scan* auf MS/MS-Ebene mit geringer Laserenergie zur Ermittlung der Spektren, die die Signale des Reporterions 117 enthalten
- iii) Analyse auf MS/MS-Ebene über eine *inclusion list*

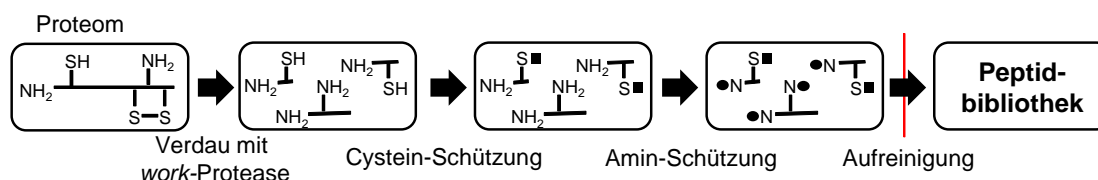
Über die Signale der Reporterionen kann dann eine Aussage getroffen werden, welchen Ursprungs die iTRAQ-markierten Peptide sind. So weisen Peptide, die bereits in beiden Mischungen A und B vorlagen, auch beide Signale der Reporterionen 114 und 117 auf.

Peptide, die durch den Verdau mit Caspase-3 entstanden sind, enthalten auch nur das Signal des Reporterions 117, bzw. zeigen ein hohes Verhältnis von 117/114.

Auf Peptid-Ebene ist der *proteomic identification of cleavage sites* (PICS)-Ansatz zu nennen. Dieser beruht auf der Verwendung von Proteom-basierten Peptidbibliotheken (Schilling and Overall, 2008). Die Vorgehensweise des PICS-Ansatzes ist in Abbildung 1.17 dargestellt.

Der PICS-Ansatz wurde von verschiedenen Arbeitsgruppen aufgegriffen und viele unterschiedliche Proteasen wurden charakterisiert. So wurden PICS-Analysen zur Untersuchung der Spaltspezifität der *Snake Venom Metallo Proteinase* (SVMP) (Paes Leme, *et al.*, 2011), der Astacin Metalloprotease-Familie (Becker-Pauly, *et al.*, 2011), der Cathepsine B, L und S (Biniossek, *et al.*, 2011), sowie in Verbindung mit der Web-basierten Datenanalyse-Pipeline WebPICS dem Faktor Xa (Schilling, *et al.*, 2011), durchgeführt. Darüber hinaus wurden auch Matrix-Metallo-Proteasen von *Arabidopsis thaliana* auf ihre proteolytische Aktivität und Spaltspezifität (Marino *et al.*, 2013) sowie Clostridial-Kollagenasen auf ihre Spaltspezifität bezüglich des Abbaus von Kollagen untersucht (Eckhard *et al.*, 2013).

A Herstellung der Peptidbibliothek



B Protease Assay

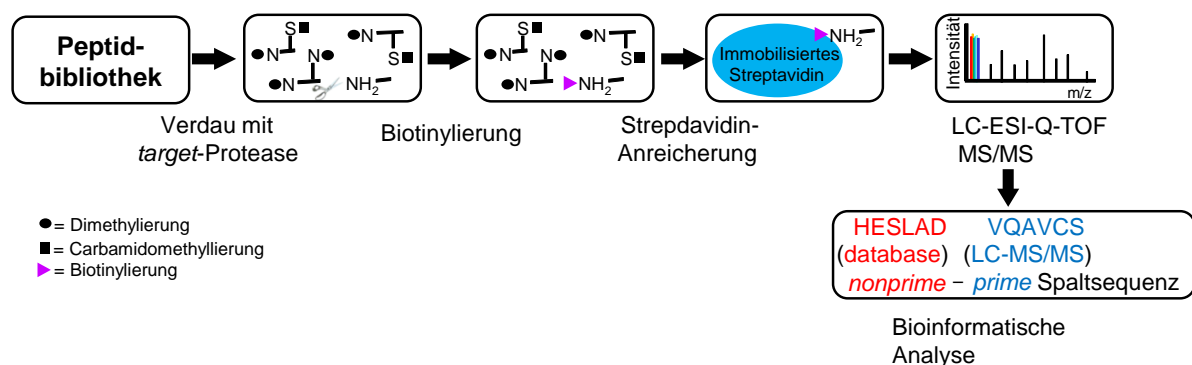


Abb. 1.17: Generelle Vorgehensweise des PICS Ansatzes (Schilling and Overall, 2008).
A) Herstellung der Peptidbibliothek. Nach dem Verdau der intakten Proteine des Basisproteoms durch eine bekannte *work*-Protease (z. B. Trypsin oder Chymotrypsin) erfolgt die Reduzierung der Disulfidbrücken, Alkylierung der freien Cysteinseitenketten und Schutz der primären Amine mittels reduktiver Dimethylierung. Nach intensiver Aufreinigung der Peptide ist die Herstellung der

Peptidbibliothek abgeschlossen. **B)** Protease Assay. Verdau mit der (unbekannten) *target*-Protease. Die durch die *target*-Protease gebildeten neo-N-Termini werden mit Biotin markiert und in einer anschließenden Streptavidin-Anreicherung aufgereinigt. Die Identifizierung der biotinylierten Peptide (*prime* Spaltsequenzen) erfolgt nach der LC-ESI MS/MS Analyse über eine Mascot-Datenbanksuche, während die *nonprime* Spaltsequenzen bioinformatisch mit der PICS Datenanalyse-Pipeline bestimmt werden. (Abbildung adaptiert aus (Schilling and Overall, 2008))

Es ist jedoch zu beachten, dass der PICS-Ansatz für eine qualitative Bestimmung von Spaltspezifitäten unbekannter Proteasen entwickelt wurde, der keine Analyse quantitativer Daten bezüglich proteolytischer Reaktionen erlaubt. Bislang gab es somit keinen Ansatz auf Peptid-Ebene, der eine solche quantitative Analyse ermöglicht.

Neben den hier gezeigten Methoden zur Analyse N-terminaler Peptide bzw. Proteine gibt es bislang auch zwei Methoden, die sich mit der Analyse von C-Termini von Proteinen beschäftigen. Dabei handelt es sich zum einen um den so genannten C-TAILS-Ansatz (Schilling *et al.*, 2010). Auch hier werden zunächst die N-Termini und ϵ -Aminogruppen der Lysinseitenketten dimethyliert und damit blockiert. Nachdem die Carboxylgruppen durch die Derivatisierung mit EDC, NHS und Ethanolamin ebenfalls blockiert wurden erfolgt ein tryptischer Verdau. Die bei diesem Verdau gebildeten neo-N-Termini werden dimethyliert und die neo-C-Termini durch ein Polyallylamin-Polymer abgefangen. Ähnlich wie in dem TAILS-Ansatz für natürlich modifizierte N-Termini von Proteinen können im C-TAILS-Ansatz die natürlich modifizierten C-Termini von Proteinen identifiziert werden.

Bei der zweiten Methode handelt es sich um eine Modifizierung des COFRADIC-Ansatzes (Kleifeld, *et al.*, 2011), bei dem ebenfalls eine SCX-Trennung durchgeführt wird. Die freien α -Aminogruppen der C-terminalen Peptide werden hier aber nicht mit TNBS, sondern mit NHS-Butyrat modifiziert. Dadurch können Sie dann von den N-terminalen Peptiden durch RP-HPLC abgetrennt werden (Van Damme, *et al.*, 2010).

Kapitel II: Zielsetzung

Ziel der vorliegenden Arbeit war es massenspektrometrische Methoden für die Proteomanalyse proteolytischer Prozesse innerhalb des Sonderforschungsbereich 877 „*Proteolysis as a Regulatory Event in Pathophysiology*“ der Christian-Albrechts-Universität zu Kiel zu entwickeln. Dieser Sonderforschungsbereich hat es sich zur Aufgabe gemacht proteolytische Prozesse zu untersuchen, die in der Zellregulation und in Signalwegen eine wichtige Rolle spielen. Damit sollen grundlegende Kenntnisse erhalten und ausgebaut werden, wie diese Reaktionen mit der humanen Pathophysiologie zusammenhängen. Um diese Zusammenhänge aufzuklären sollen die Details der proteolytischen Prozesse, die ein Bestandteil von Signalprozessen sind bis hin zur Molekularebene analysiert werden, um die Regulationsmechanismen der proteolytischen Prozesse in Signalkaskaden zu definieren.

Der Fokus der vorliegenden Arbeit lag v. a. auf der quantitativen Analyse von N-terminalen (*nonprime*-) Proteolyseprodukten. Mit deren Hilfe sollte ein Ansatz zur Bestimmung der Spaltspezifität (unbekannter) Proteasen, die gleichzeitig die Quantifizierung proteolytischer Prozesse ermöglicht, erstellt und etabliert werden. Zu diesem Zwecke wurde eine LC-MALDI-TOF/TOF MS/MS Methode entwickelt, die zunächst hinsichtlich verschiedener Parameter optimiert werden musste. Anhand einer biologischen Problemstellung sollte dann diese Methode auf ihr Potential untersucht werden.

Hierzu mussten auch Problemstellungen bearbeitet werden, die unabhängig von der Methodenentwicklung für den Protease-Assay zu sehen, aber allgemeingültig für quantitative MALDI MS/MS Methoden mittels isobaren *tags* sind. So wurde eine Methode zur Verbesserung des so genannten *peak-pickings*, einem Teil der MS-Rohdaten Bearbeitung eingeführt.

Darüber hinaus sollten bereits bestehende Methoden innerhalb des Sonderforschungsbereichs durch massenspektrometrische Methoden der Proteomanalyse ergänzt und ausgebaut werden. Damit sollten spezifische Problemstellungen anderer Arbeitsgruppen des Sonderforschungsbereichs bearbeitet und gelöst werden. In diesem Zusammenhang wurde dann auch eine Analysenmethode zur Identifizierung eines Protein-C-Terminus entwickelt.

Kapitel III: Der Q-PICS Ansatz - Methodenentwicklung und MALDI-TOF/TOF MS/MS Parameteroptimierung

3.1 Einleitung

Wie in Kapitel 1.2 bereits beschrieben, gibt es verschiedene Strategien um Proteine bzw. Peptide mittels MS-basierter Methoden relativ oder absolut zu quantifizieren. Insbesondere sei hier wieder auf die Methoden mittels isobarer Isotopenmarkierung zur MS/MS-basierten Quantifizierung, (Bantscheff, *et al.*, 2007) wie die iTRAQ- (Ross, *et al.*, 2004) oder die TMT-Technologie (Thompson, *et al.*, 2003), hingewiesen.

Die meisten Studien, die eine Quantifizierung über solche isobaren *tags* verwenden, wurden mittels ESI MS/MS Analysen durchgeführt und die Optimierung vieler verschiedener Parameter zur Generierung robuster quantitativer Daten wurden hierfür beschrieben (Bantscheff *et al.*, 2008, Gan *et al.*, 2007, Griffin *et al.*, 2007, Kocher *et al.*, 2009, Ow *et al.*, 2009, Pichler *et al.*, 2011).

Bantscheff *et al.* geben beispielsweise eine praktische Anleitung für die MS/MS-Analyse iTRAQ-markierter Peptide. Dabei gehen sie insbesondere auf Parameter ein, auf die geachtet werden muss, wenn die so genannte *Pulsed Q Dissociation* (PQD) mit einem LTQ Orbitrap Massenspektrometer (Thermo Fisher Scientific) verwendet wird (2008). Diese Methode ist normalerweise bekannt dafür, dass die Peptidionen nur eine geringe Fragmentierungseffizienz aufweisen. Um diesem Problem entgegenzuwirken sind in der Studie Optimierungen der Parameter für die angewandte Kollisionsenergie, die *activation Q*, die Verzögerungszeit, die Isolationsbreite der *precursor*, die Anzahl an *microscans* sowie die Anzahl an *getrappten* Ionen beschrieben. Dadurch wurde eine akkurate und sensitive Methode (bis zu 100 amol) auf Basis der PQD-Fragmentierung für iTRAQ-markierte Peptide entwickelt. Zusätzlich wurde sie mit einer *higher energy collision-induced dissociation* (HCD) Methode auf dem gleichen Massenspektrometer und mit einer kollisionsinduzierten Dissoziation (CID) Methode auf einem Q-TOF Ultima (Waters) verglichen.

Die PQD-Fragmentierung wurde auch für eine lineare Ionenfalle (LTQ, Thermo Fisher Scientific) verwendet (Griffin, *et al.*, 2007). In dieser Studie wurde der Einfluss der Parameter wie der relativen Kollisionsenergie, der Aufnahme von *microscans* oder der *repeat count* auf die quantitativen Daten, analysiert. Dabei wurden ähnliche Ergebnisse wie unter der Verwendung eines Quadrupol-TOF Massenspektrometers (Qstar, AB SCIEX) erhalten.

Darüber hinaus wurden auch Innovationen bezüglich der Hardware der Massenspektrometer für quantitative proteomische Analysen auf Basis isobarer *tags* beschrieben: Pichler *et al.* verwendeten ein axiales Extraktionsfeld in einer Orbitrap HCD-Zelle, das normalerweise dazu verwendet wird, Fragmentationen in die *C-trap* von Orbitrap XL ETD oder Orbitrap Velos Instrumenten zu leiten (2011). Durch die Verwendung dieses axialen Extraktionsfeldes wurde eine hohe Anzahl an HCD MS/MS-Spektren mit Reporterionen erhalten. Diese wiesen nur eine geringe Varianz auf und führten somit zu einer verbesserten analytischen Präzision.

Im Gegensatz dazu sind nur wenige Studien beschrieben, die überhaupt eine Kombination der MALDI MS/MS Analyse mit isobaren *tags* verwenden (Baeumlisberger *et al.*, 2010, Dayon *et al.*, 2008, Dayon *et al.*, 2010b, Enoksson, *et al.*, 2007, Hu *et al.*, 2006, Kuzyk *et al.*, 2009, Shirran and Botting, 2010, Tiberti, *et al.*, 2010). Diese reichen zwar von der verbesserten Identifizierung schlecht detektierbarer Peptide (Baeumlisberger, *et al.*, 2010), bis hin zur Identifizierung von Biomarkern (Tiberti, *et al.*, 2010). Es scheint jedoch bislang keine Studien zu geben, die sich ausschließlich mit der Parameteroptimierung für die MALDI-TOF/TOF MS/MS Analyse isobarer *tags* beschäftigen.

Wie in Kapitel 3.4 beschrieben wird, weist die Kombination der LC-MALDI-TOF/TOF MS/MS Analyse mit der Quantifizierung isobarer *tags* für den in dieser Arbeit beschriebenen Q-PICS Ansatz zur simultanen Bestimmung der Spaltspezifität (*cleavage site specificity*) einer unbekannt Protease und der Quantifizierung von proteolytischen Prozessen große Vorteile auf. Der Ansatz beruht auf dem in Kapitel 1.5 beschriebenen PICS Ansatz (Schilling and Overall, 2008), beinhaltet jedoch über die Verwendung des isobare *tags* TMT, zusätzlich zu der qualitativen auch eine quantitative Ebene. Um sowohl eine robuste Quantifizierung als auch eine robuste Identifizierung der TMT-markierten Peptide gewährleisten zu können, mussten in einem ersten Teil jedoch zunächst systematisch einige LC-MALDI-TOF/TOF MS/MS Parameter optimiert werden. Im Folgenden wird diese Parameteroptimierung beschrieben.

3.2 Material und Methoden

3.2.1 Arbeiten mit einem 10-Proteinstandard

Die Proteine des 10-Proteinstandards (Tabelle 3.1), Endoprotease GluC, Acetonitril (E Chromasolv), α -Cyano-4-hydroxy-zimtsäure (CHCA, $\geq 98\%$), Harnstoff ($\geq 98\%$) und Glu¹-Fibrinopeptid B wurden von Sigma-Aldrich (Steinheim, Deutschland) bezogen. Entionisiertes

Wasser (18.2 MΩ*cm) wurde mit einem arium®611VF System (Sartorius, Göttingen, Deutschland) hergestellt und für alle Experimente verwendet. Chymotrypsin, 2-(4-(2-Hydroxyethyl)-1-piperazinyl)-ethansulfonsäure (HEPES-Puffer), Formaldehyd (37 Gew%, ACS Reagenz), Iodacetamid (Ultra), Isopropanol (Chromasolv), Triethylammoniumbicarbonat (1M, puriss. P.a.), Tris(2-carboxyethyl)phosphin-Hydrochlorid (TCEP, ≥ 98%) und Trifluoressigsäure (ca. 98%) wurden von Fluka und *Sequencing grade modified* Trypsin von Promega (Madison, WI, USA) erworben. Das zur MS-Kalibrierung verwendete 4700 Proteomics Analyzer Standard Kit wurde von AB SCIEX (Darmstadt, Deutschland), das *TMTsixplexTM Label Reagent Set* von Thermo Fisher Scientific (Bremen, Deutschland) und Natriumcyanoborhydrid (*ALD coupling solution*) von Sterogene, (Carlsbad, CA, USA) bezogen.

Tab. 3.1: 10 Proteine des 10-Proteinstandards.

Protein	Organismus/Herkunft	Accession Nummer	Lot-Nr	Reinheit
Cytochrom c	Pferd/Herz	gi 255683507	18H7001	≥ 95 %
BSA	Rind/Serum	gi 30794280	10K7600	≥ 97 %
β-Lactoglobulin B	Rind/Milch	gi 87196497	11K7032	≥ 90 %
Carboanhydrase	Rind/Erythrocyten	gi 78369302	111K1102	keine Angabe
Catalase*	Rind/Leber	gi 78369302	63H7090	keine Angabe
Lysozym	Huhn/Eiweiß	gi 45384212	53H7145	≥ 90 %
Myoglobin	Pferd/Skelettmuskulatur	gi 255683511	80K7052	95-100 %
Ribonuklease A	Rind/Pancreas	gi 68299789	93H7245	keine Angabe
Transferrin	Mensch	gi 4557871	36755	≥ 98 %
α-Lactalbumin	Rind/Milch	gi 27805979	109F8185	≥ 85 %
*Angabe der Masse des Monomers, liegt in Natur als Tetramer vor				

Nasschemische Methoden - Probenvorbereitung

Die verwendeten Proteine des 10-Proteinstandards sind Tabelle 3.1 zu entnehmen, die Konzentration der Stocklösung ist in Tabelle 3.2 dargestellt.

Tab. 3.2: Konzentration der eingesetzten Stocklösung des 10-Proteinstandards.

Protein	M (g/mol)	Konzentration [mg/ml]	Konzentration [pmol/μl]
Cytochrom c (Pferd)	12360	0,32	26,0
BSA (Rind)	66534	0,41	6,1
β-Lactoglobulin B (Rind)	18276	0,39	21,2
Carboanhydrase (Rind)	29022	0,34	11,8
Catalase (Rind)*	57585	0,38	6,6
Lysozym (Huhn)	14306	0,39	27,0
Myoglobin (Pferd)	16951	0,38	22,5
Ribonuklease A (Rind)	13682	0,37	27,1
Transferrin (Mensch)	79593	0,33	4,1
α-Lactalbumin (Rind)	14178	0,40	28,4
*Angabe der Masse des Monomers, liegt in Natur als Tetramer vor			

Herstellung der Peptidbibliothek

Die Generierung der tryptischen und chymotryptischen Peptidbibliotheken wurde in einer modifizierten Version nach dem PICS Ansatz durchgeführt (Schilling and Overall, 2008). Es wurden je 270 μL (entspricht ca. 1 mg Gesamtprotein) des 10-Proteinstandards eingesetzt und mit entionisiertem Wasser auf je 800 μl aufgefüllt. Zur Denaturierung der intakten Proteine wurden 480 μL 8 M Harnstoff-Lösung (gelöst in 1 M Triethylammoniumbicarbonat-Puffer, TEAB) hinzugefügt und für 30 min bei RT inkubiert. Die Reduktion der Disulfidbrücken erfolgte jeweils durch Zugabe von 128 μL Reduzierungsreagenz (50 mM Tris-(2-carboxyethyl)phosphin, TCEP) und 1 h Inkubationszeit im Thermomixer bei 60°C und 350 U/min. Um eine Rückreaktion der Cysteine zu Cystin zu verhindern, wurde eine Alkylierung durch Zugabe von 64 μL Sulfhydryl-Blockierungsreagenz (200 mM Iodacetamid, IAA) durchgeführt. Die Inkubationszeit betrug 30 min bei RT im Dunkeln. Eine Umpufferung der Reaktionsgemische auf 200 mM HEPES pH 7,4 mit einem Endvolumen von 1 mL erfolgte mit Amicon Ultra 4 3K Centrifugal Filtern (Millipore, Schwalbach, Deutschland) mit einem *cutoff* von 3000 Da nach den Herstellerangaben. Um die Proteine in Peptide zu spalten, wurde ein tryptischer bzw. chymotryptischer Verdau durch Zugabe der jeweiligen (*work-*) Protease im Verhältnis Protease:Protein1:50 (m/m) durchgeführt. Die Inkubation erfolgte über Nacht bei 37°C und 300 U/min im Thermomixer.

Zum Schutz der N-Termini sowie der Lysin- ϵ -Aminogruppen wurden die Reaktionsgemische mit je 10 μL 1 M Formaldehyd- bzw. Cyanoborhydrid-Lösung versetzt und für 16 h bei 22°C und 300 U/min im Thermomixer inkubiert. Um eine vollständige Reaktion der primären Amine zu gewährleisten, wurde die reduktive Dimethylierung noch einmal wiederholt, bevor eine Aufreinigung mit 60 mg Oasis SPE-Kartuschen (Waters, Milford, MA USA) nach Herstellerangaben durchgeführt wurde. Die Peptide wurden dann in je 1 mL 200 mM HEPES-Puffer bei pH 7,4 gelöst, wobei die Konzentration 1 mg/mL betrug.

Verdau mittels zu testender Protease und TMT-Derivatisierung

Die tryptische und chymotryptische Peptidbibliothek wurden mit GluC als zu testende Protease (*target*-Protease) weiterbehandelt. Hierzu wurden jeweils sechs interne technische Replikate zu je 30 μg der Peptidbibliotheken mit GluC im Verhältnis Protease:Protein1:50 (m/m) verdaut. Die Inkubation erfolgte über Nacht bei 37°C und 300 U/min im Thermomixer.

Die in dem Verdau entstandenen neo-N-Termini wurden mittels dem *TMTsixplexTM Label Reagent Set* derivatisiert. Hierzu wurden je 41 μL Acetonitril (ACN) zu einem der sechs

TMT-Reagenzien (126-131) gegeben und die Lösung 5 min lang bei RT inkubiert. Um das vollständige Lösen der TMT-Reagenzien zu gewährleisten wurden die Eppendorfgläser während dieser Zeit immer wieder mit einem Vortexer behandelt. Dann wurden je 13,5 μL der einzelnen TMT-Reagenzien zu einer mit GluC verdauten Probe gegeben und die Reaktionsgemische für 2 h bei RT inkubiert. Überschüssiges TMT-Reagenz wurde durch Zugabe von 20 μL 100 mM Ammoniumbicarbonat-Lösung abgefangen, wobei die Inkubationszeit 15 min betrug. Die TMT-markierten Proben wurden jeweils 1:1:1:1:1 vereinigt.

Trennung mittels IP-RP-HPLC bei pH 2,1

Die Proben wurden 1:175 (v/v) mit 0,1%iger (v/v) wässriger Trifluoressigsäure-Lösung auf eine Konzentration von 250 fmol/ μL verdünnt. Primärziel war es zunächst, die optimalen massenspektrometrischen Messbedingungen für die TMT-markierten Peptide zu ermitteln. Hierzu wurden zwei verschiedene Konzentrationen getestet. Es wurden jeweils 1 μL (entspricht 250 fmol verdautes Gesamtprotein) bzw. 10 μL (entspricht 2,5 pmol verdautes Gesamtprotein) mit einem U3000 nanoHPLC-System (Dionex, Idstein, Deutschland) über eine 50 μL Probenschleife injiziert und auf einer Acclaim PepMap100 C18 *Trap*-Säule (300 μm I.D. x 10 mm, 5 μm , Dionex, Idstein, Deutschland) aufkonzentriert. Die *Trap*-Säule wurde 6 min mit 0,1%iger (v/v) wässriger Trifluoressigsäure-Lösung mit 3% (v/v) ACN mit einer Flussrate von 30 $\mu\text{L}/\text{min}$ mit Peptiden beladen. Die Analyten wurden anschließend über eine Acclaim PepMap100 C18 Trennsäule (75 μm I.D. x 150 mm, Dionex, Idstein, Deutschland) separiert. Die Trennung wurde unter Verwendung eines Gradienten aus Eluent A (0,05% (v/v) wässrige Trifluoressigsäure) und Eluent B (0,04% (v/v) wässrige Trifluoressigsäure in 80% (v/v) Acetonitril) durchgeführt. Dieser ist Tabelle 3.3 zu entnehmen. Die Flussrate betrug 0,3 $\mu\text{L}/\text{min}$ bei einer Säulentemperatur von 30°C. Die Detektion erfolgte mittels eines UV-Detektors mit Z-förmiger Detektorzelle mit einem Detektorvolumen von 3 nL und einer Wellenlänge von 214 nm. Alle LC-Auftrennungen wurden in Form von vier technischen Replikaten durchgeführt.

Nach Elution von der Säule wurden die Peptide ab der 25. Minute 1:3 (v/v) mit MALDI-Matrix (CHCA) über ein T-Stück mittels Spritzenpumpe gemischt und in einem Zeitintervall von 15 s auf ein MALDI-*Target* (Opti-TOFTM LC MALDI Insert, AB SCIEX, Darmstadt, Deutschland) gespottet. Das Auftragen erfolgte mit einem Probot Mikrofraktionssammler (Dionex, Idstein, Deutschland) über einen Zeitraum von 78 min. Die Matrix-Stockkonzentration betrug 3 mg/mL (in 70% (v/v) Acetonitril/0,1% (v/v) wässriger

Trifluoressigsäure) und enthielt ca. 5 nM Glu¹-Fibrinopeptid B als internen Kalibranten für die anschließende Analyse auf MS-Ebene.

Tab. 3.3: Zeitlicher Verlauf des Gradienten für die Trennung der derivatisierten tryptischen und chymotryptischen Peptidbibliotheken auf Basis des 10-Proteinstandards mittels IP-RP-HPLC.

Zeit [min]	Eluent A [%]	Eluent B [%]
0,0	95	5
6,0	95	5
91,0	40	60
101,0	5	95
106,0	5	95
107,0	95	5
125,0	95	5

Massenspektrometrische Analyse mittels MALDI-TOF/TOF MS/MS

Zur massenspektrometrischen Analyse der mittels LC aufgetrennten Proben und Datenprozessierung (GPS-Software, Version 3.6, AB SCIEX, Darmstadt, Germany) wurde ein MALDI-TOF/TOF 5800 Massenspektrometer (AB SCIEX, Darmstadt, Deutschland) verwendet. Die Einstellungen für die Analysen auf MS-Ebene wurden konstant gehalten und im *positive ion mode* in einem Massenbereich von 700 – 4000 m/z und einer Akkumulation von 1000 Laserschüssen durchgeführt. Die MS-Kalibrierung erfolgte über sechs Standardpeptide (4700 Proteomics Analyzer Mass Standard Kit, AB SCIEX, Darmstadt, Deutschland), welche den erforderlichen Messbereich abdecken, in Form einer *default calibration*. Zusätzlich wurde während der LC-MALDI-Analysen mit Glu¹-Fibrinopeptid B als internem Standard kalibriert.

Alle MS/MS-Spektren wurden im *positive ion mode* mit 1 kV Kollisionsenergie mittels CID generiert. Die Auswahl der Vorläuferionen (*precursor*) erfolgte über die Software des Massenspektrometers unter Verwendung einer *inclusion list*, auf der 50 TMT-markierte Peptide enthalten waren, die durch vorangegangene Pilotexperimente identifiziert wurden (siehe Anhang Kapitel 3 Tabelle 3.I für die tryptische und Tabelle 3.II für die chymotryptische Peptidbibliothek).

Die Anzahl an akkumulierten Laserschüssen für die Analyse auf MS/MS-Ebene war ein Parameter, der für die Qualität eines Spektrums und die Identifizierung TMT-markierter Peptide optimiert werden sollte. Aus diesem Grund wurden MS/MS-Spektren unter

Akkumulation von 3000 und 5000 Laserschüssen generiert. Um die Fragmentierungseffizienz unter verschiedenen CID-Drücken zu untersuchen, wurde der Druck in der CID-Kollisionszelle auf 1×10^{-6} Torr (CIDniedrig) und 4×10^{-6} Torr (CIDhoch) eingestellt. Die Reihenfolge der analysierten Parametereinstellungen ist in Tabelle 3.4 dargestellt. Die MS/MS-Kalibrierung wurde mit Glu¹-Fibrinopeptid B aus dem 4700 *Proteomics Analyzer Mass Standard Kit* (AB SCIEX, Darmstadt, Deutschland) in Form einer *default calibration* durchgeführt.

Tab. 3.4: Parametereinstellungen der MS/MS-Analysen der auf dem 10-Proteinstandard basierenden Peptidbibliotheken.

		3000 laser shots, CID 1x10-6 Torr	5000 laser shots, CID 1x10-6 Torr	3000 laser shots, CID 4x10-6 Torr	5000 laser shots, CID 4x10-6 Torr
LC run1	MSMS run 1			X	
	MSMS run 2				X
	MSMS run 3	X			
	MSMS run 4		X		
LC run2	MSMS run 1				X
	MSMS run 2	X			
	MSMS run 3		X		
	MSMS run 4			X	
LC run3	MSMS run 1	X			
	MSMS run 2		X		
	MSMS run 3			X	
	MSMS run 4				X
LC run4	MSMS run 1		X		
	MSMS run 2			X	
	MSMS run 3				X
	MSMS run 4	X			

Protein- bzw. Peptididentifizierung mittels Datenbankanalyse

Die Rohdaten wurden unter Verwendung der TS2Mascot Software (Matrix Sciences, London, England) in *mascot generic format* (mgf)-Dateien umgewandelt. Dabei wurden folgende Einstellungen verwendet: Massenbereich von 60 Da bis zu der Masse des *precursors* minus 35 Da; minimales S/N-Verhältnis von 10, wobei nur die monoisotopischen Signale exportiert wurden. Die Datenbanksuche erfolgte in einer selbsterstellten Datenbank, die die Aminosäuresequenzen der zehn Proteine enthielt. Semi Trypsin und semi GluC wurde als Enzym eingestellt, wobei keine C- oder N-terminale Orientierung berücksichtigt wurde. Die maximal zulässige Anzahl an ausgelassenen Schnittstellen (*missed cleavages*) betrug 4. Die Massentoleranz für die *precursor* wurde auf ± 50 ppm und die der MS/MS-Fragmentationen auf

$\pm 0,5$ Da festgesetzt. Als feste Modifikationen wurde Dimethylierung (K) Carbamidomethylierung (C) sowie Markierung TMT (N-Term), als variable Modifikation Oxidation (M) eingestellt. Der Vertrauensbereich für die Proteinidentifizierung wurde auf $\geq 95\%$ ($p < 0,05$) und der *identity threshold* der Peptide auf einen *ion score* von 13 festgesetzt. Die Auswertung der quantitativen Daten wurde ebenfalls mit der Mascot Software durchgeführt. Die TMT-Faktoren für die Isotopenkorrektur wurden nach Angaben von Thermo Fisher Scientific (Bremen, Deutschland) für die Charge LJ150159A verwendet. Alle Flächen der einzelnen TMT-Reporterionen wurden durch die Flächenwerte des Kanals 126 geteilt. Die daraus resultierenden TMT-Flächenverhältnisse der Reporterionen wurden zusammen mit den Ergebnissen der Mascot Datenbanksuche im csv-Format exportiert, um statistische Berechnungen durchzuführen. Der Erwartungswert der TMT-Flächenverhältnisse der sechs internen technischen Replikate betrug 1,0. Deshalb wurden negative TMT-Flächenverhältnisse, die aus der Quantifizierung der Mascot Software stammen, entfernt und mathematische Ausreißer der TMT-Flächenverhältnisse mittels der freizugänglichen *outlier* Software (http://www.sediment.uni-goettingen.de/staff/dunkl/software/out_liier.html) über einen *interquartile range* (IQR)-Test identifiziert und von der weiteren Analyse ausgeschlossen. Die geometrischen Mittelwerte, sowie die zugehörige Standardabweichung der TMT-Flächenverhältnisse wurden, wie zuvor von Tan *et al.* beschrieben, errechnet (Tan *et al.*, 2008).

3.2.2 Arbeiten mit zytosolischem Hefeproteom

Entionisiertes Wasser ($18.2 \text{ M}\Omega \cdot \text{cm}$) wurde mit einem arium®611VF System (Sartorius, Göttingen, Deutschland) hergestellt und für alle Experimente verwendet. Die Endoprotease GluC, Acetonitril (E Chromasolv), 2-(4-(2-Hydroxyethyl)-1-piperazinyl)-ethansulfonsäure (HEPES-Puffer), Glu¹-Fibrinopeptid B, α -Cyano-4-hydroxy-zimtsäure (CHCA, $\geq 99\%$), Ammoniumsulfat ($\geq 99\%$), Ethylendiamintetraacetat EDTA Dinatriumsalz Dihydrat, Formaldehyd (37 Gew%, ACS Reagenz), Ammoniumhydroxid (ACS-Reagenz), Methanol ($\geq 99.5\%$), L-Arginin-Monohydrochlorid, L-Aspartat, Uracil ($\geq 99\%$), L-Valin ($\geq 98\%$) Glucose-Monohydrat ($\geq 98\%$) und Harnstoff ($\geq 98\%$) wurden von Sigma-Aldrich (Steinheim, Deutschland) bezogen. Isopropanol (Chromasolv), Triethylammoniumbicarbonat (1M, puriss. P.a.), Ammoniumbicarbonat ($\geq 99.5\%$), Iodacetamid (Ultra), Trifluoressigsäure (ca. 98%), Tris(2-carboxyethyl)phosphin-Hydrochlorid (TCEP, $\geq 98\%$), L-Histidin ($\geq 99.5\%$), L-Methionin ($\geq 99\%$), L-Lysin ($\geq 98\%$), L-Leucin ($\geq 99\%$), L-Tryptophan ($\geq 99.5\%$), L-

Isoleucin ($\geq 95\%$) und L-Phenylalanin ($\geq 99\%$) wurden von Fluka und *Sequencing grade modified* Trypsin von Promega (Madison, WI, USA) erworben. Das zur MS-Kalibrierung verwendete 4700 Proteomics Analyzer Standard Kit wurde von AB SCIEX (Darmstadt, Deutschland) bezogen. *Yeast nitrogen base* (ohne Aminosäuren und Ammoniumsulfat) wurde bei Difco/Becton Dickinson (Sparks, MD, USA) und der Protease-Inhibitor-Cocktail Complete Mini bei Roche Diagnostics GmbH (Penzberg, Deutschland) erworben. Das *TMTsixplexTM Label Reagent Set* wurde von Thermo Fisher Scientific (Bremen, Deutschland) und Natriumcyanoborhydrid (*ALD coupling solution*) von Sterogene, (Carlsbad, CA, USA) bezogen.

Verwendete Organismen

Saccharomyces cerevisiae YHUM257 (Stammsammlung der Molekular- und Zellbiologie, Universität des Saarlandes)

Verwendetes Medium

In Tabelle 3.5 ist das zur Kultivierung von *S. cerevisiae* verwendete synthetische Vollmedium aufgelistet.

Tab. 3.5: Zusammensetzung des synthetischen Vollmediums.

synthetisches Vollmedium			
1,7	g/L	YNB (0,17 %); sterilfiltrierte Lagerung bei 4°C	
5,0	g/L	(NH ₄) ₂ SO ₄	
20	g/L	Glukose	
Aminosäuren/ Basen			
0,2	g/L	Adenin, Uracil, Tryptophan, Histidin, Arginin, Methionin	
0,3	g/L	Tyrosin, Lysin, Isoleucin	
0,5	g/L	Phenylalanin	
1,0	g/L	Leucin/ Aspartat	
2,0	g/L	Threonin	
1,5	g/L	Valin	

Bestimmung der Zellkonzentration bzw. Zellzahl

Die Zellkonzentration wurde über photometrische Bestimmung der optischen Dichte bei 600 nm in 1,5 mL Polystyrol-Einwegküvetten (Spectrophotometer Novaspec®II, GE Healthcare, München, Deutschland) ermittelt. Nach Yin *et al.* entspricht einer optischen Dichte von 1,0 eine Zellzahl von $1,66 \pm 0,05 \times 10^7$ Zellen/mL (Yin *et al.*, 2007). Die Zellzahl wurde mit Hilfe mikroskopischer Zählung der *S. cerevisiae* Zellen in einer Neubauer-Zählkammer bestimmt.

Kultivierung von *S. cerevisiae*

Die gesamte Kultivierung von *S. cerevisiae* wurde von Barbara Gregorius, (Institut für Experimentelle Medizin, Christian-Albrechts Universität zu Kiel) durchgeführt und erfolgte bei 30°C und 230 U/min in einem Schüttelinkubator (Multitron HAT, Bottmingen, Schweiz) in Schikanekolben. Als Vorkultur wurden 5 ml synthetisches Vollmedium mit 50 µL einer Kühlkultur (4°C) von *S. cerevisiae* YHUM257 in YPD-Medium angeimpft. Von dieser Übernachtskultur wurden 2 mL verwendet, um eine 200 mL Hauptkultur in synthetischem Vollmedium heranzuziehen. Bei einer optischen Dichte von 2 wurde die Hauptkultur in vier Falkons aliquotiert und pelletiert (8500 U/min, 4°C, Biofuge stratos, Rotor #3046, Kendro, Osterode, Deutschland). Die Zellpellets wurden anschließend zweimal, zur Entfernung aller Stickstoffbestandteile des Mediums, mit 2%iger Glukose gewaschen und in 150 ml des synthetischen Vollmediums resuspendiert. Die Anzucht erfolgte bis zu einer OD von 3.

Probenvorbereitung

Die Zellen wurden pelletiert, (8500 U/min, 5 min, 4°C, Biofuge stratos, Rotor #3046, Kendro, Osterode, Deutschland) mit kaltem Wasser gewaschen und anschließend in 10 mM Tris-HCl, pH 7,5 resuspendiert. Sie wurden dann in 2 mL Reaktionsgefäße aliquotiert und abzentrifugiert (10000 U/min, 5 min 4°C, Biofuge fresco, Rotor #3325, Kendro, Osterode, Deutschland). Nach Abnahme der Überstände wurde das Feuchtgewicht der Zellpellets bestimmt. Die Zellpellets wurden 1:3 (m/v) in Lysepuffer (50 mM Tris, 5 mM EDTA, mit Proteaseinhibitor, pH 7,5) resuspendiert (ca. 200 mg Feuchtgewicht auf 600 µL Lysepuffer), und 1:1 (m/m) mit Glasperlen (0,5 mm Durchmesser) versetzt. Die Zellen wurden dann 15 min in der Schwingmühle (Modell MM301, Retsch, Haan, Deutschland) mit Maximalgeschwindigkeit (30 Hz) homogenisiert. Nach dem Absinken der Glasperlen wurden die Zelltrümmer abzentrifugiert (13000 U/min, 4°C, 10 min, Biofuge fresco, Rotor #3325, Kendro, Osterode, Deutschland) und der Überstand mit dem zytosolischen Proteom abgenommen. Das verbleibende Pellet mit Glasperlen wurde 1:1,5 (m/v) in Lysepuffer resuspendiert und die Extraktionsprozedur noch zweimal wiederholt. Die vereinigten Überstände wurden erneut zentrifugiert (2 h, 19000 U/min, 4°C, Biofuge stratos, Rotor: #3331, Kendro, Osterode, Deutschland) und die Proteinkonzentration mit BCA-Assay bestimmt. Sie betrug ca. 2,3 µg/µL.

Herstellung der Peptidbibliothek

Die Generierung der tryptischen Peptidbibliothek auf Basis des zytosolischen Hefeproteoms wurde, wie bereits zuvor beschrieben, durchgeführt. Hierzu wurde die Proteinkonzentration des zytosolischen Hefeproteoms mit entionisiertem Wasser auf 2 mg/mL eingestellt und dann 1 mL für die weiteren Schritte verwendet. Aus diesem Grund unterscheiden sich lediglich die volumenbezogenen Mengen an zugegebenen Reagenzien, die Molaritäten bleiben erhalten. Hierdurch ergeben sich 600 µL 8 M Harnstoff-Lösung (gelöst in 1 M TEAB), 160 µL Reduzierungsreagenz (50 mM TCEP) sowie 80 µL Sulfhydryl-Blockierungsreagenz (200 mM Iodacetamid). Die Umpufferung der Reaktionsgemische auf 200 mM HEPES pH 7,4 erfolgte auf ein Endvolumen von 1 mL. Der tryptische Verdau wurde im Verhältnis Protease:Protein 1:1000 (m/m) durchgeführt; auf einen chymotryptischen Verdau wurde an dieser Stelle verzichtet. Die Blockierung der N-Termini und der Lysin-ε-Aminogruppen, sowie die anschließende Aufreinigung der Peptidbibliothek erfolgten ohne Modifizierungen, wobei die Konzentration 2 mg/mL in 200 mM HEPES-Puffer bei pH 7,4 betrug.

Verdau mittels zu testender Protease und TMT-Derivatisierung

Auch in der auf dem zytosolischen Hefeproteom basierenden Peptidbibliothek wurde der zweite Verdau mit GluC als *target*-Protease durchgeführt. Hierzu wurden jeweils sechs Aliquote zu je 30 µg der Peptidbibliothek im Verhältnis Protease:Protein 1:50 (m/m) verdaut. Die Inkubation erfolgte über Nacht bei 37°C und 300 U/min im Thermomixer. Die Derivatisierung mittels TMT Kit erfolgte entsprechend den Herstellerangaben und nach der zuvor beschriebenen Vorgehensweise. Anschließend wurden die sechs internen technischen Replikate im Verhältnis 1:1:1:1:1:1 vereinigt.

Trennung mittels IP-RP-HPLC bei pH 2,1

Die Proben der tryptischen Peptidbibliothek auf Basis des zytosolischen Hefeproteoms wurden über das gleiche HPLC-System wie die Peptidbibliotheken auf Basis des 10-Proteinstandards aufgetrennt. Da es sich dabei jedoch um eine Probe mit höherer Komplexität handelt, wurden folgende Modifikationen erstellt: Die Trennung der Peptide wurde mit einer Acclaim PepMap100 C18 Trennsäule (75 µm I.D. x 250 mm, Dionex, Idstein, Deutschland) mittels des in Tabelle 3.6 dargestellten Gradienten in Form von drei technischen Replikaten durchgeführt.

Die Peptide wurden ab der 30. Minute 1:3 (v/v) mit MALDI-Matrix gemischt und in Intervallen von 15 s auf ein MALDI-Target gespottet, wobei das Spotten über einen Zeitraum von 120 min erfolgte.

Tab. 3.6: Zeitlicher Verlauf des Gradienten für die Trennung der Proben der tryptischen Peptidbibliothek auf Basis des zytosolischen Hefeproteoms mittels IP-RP-HPLC.

Zeit [min]	Eluent A [%]	Eluent B [%]
0,0	95	5
6,0	95	5
156,0	50	50
166,0	5	95
171,0	5	95
172,0	95	5
200,0	95	5

Massenspektrometrische Analyse mittels MALDI-TOF/TOF MS/MS

Die massenspektrometrischen Analysen auf MS-Ebene wurden unter den gleichen Bedingungen wie in den Experimenten des 10-Proteinstandards durchgeführt.

Alle MS/MS-Spektren wurden im *positive ion mode* mit 1 kV Kollisionsenergie mittels CID bei dem bereits optimierten Druck von 4×10^{-6} Torr (CIDhoch) generiert. Die Auswahl der *precursor* erfolgte nicht wie zuvor unter Verwendung einer *inclusion list*, sondern über die Software des Massenspektrometers nach folgenden Parametern: Minimales S/N-Verhältnis 20; *precursor* Massentoleranz: ± 200 ppm; das Signal musste nur in einem Spot detektiert werden, um als *precursor* akzeptiert zu werden; maximal 25 *precursor* pro Spot. Die MS/MS-Spektren wurden weiterhin unter Akkumulation von 3000 und 5000 Laserschüssen generiert. Ein weiterer Parameter, der mit der auf zytosolischem Hefeproteom basierenden Peptidbibliothek optimiert werden sollte, war die Reihenfolge der *precursor* Auswahl. Hierzu wurden sowohl Analysen, in denen zuerst der *precursor* mit der geringsten Intensität in der MS-Messung eines Spots fragmentiert wird ("*weakest precursor first*") durchgeführt als auch solche, in denen zuerst derjenige mit der höchsten Intensität für eine MS/MS-Analyse ausgewählt wird ("*strongest precursor first*").

Protein- bzw. Peptididentifizierung mittels Datenbankanalyse

Die Rohdaten wurden, unter Beibehaltung der Einstellungen: Massenbereich von 60 Da bis zu der Masse des *precursors* minus 35 Da, minimales S/N-Verhältnis von 10, unter Verwendung der TS2Mascot Software in mgf-Dateien umgewandelt.

Die Mascot Datenbanksuche wurde mit der Mascot 2.2.04 Software gegen die *Swiss-Prot* FASTA-Datenbank (15. Februar 2011, 7256 Sequenzen) mittels des Taxonomie-Filters *Saccharomyces cerevisiae* durchgeführt. Die zuvor verwendeten Parametereinstellungen wurden bis auf die folgenden beibehalten: Die Massentoleranz für die *precursor* wurde auf ± 25 ppm und die der MS/MS-Fragmentionen auf $\pm 0,3$ Da festgesetzt. Der *identity threshold* der Peptide wurde auf einen *ion score* von 38 eingestellt, wodurch eine falsch-positiv Rate (FDR) $< 4\%$ erhalten wurde. Die Bestimmung der FDR erfolgte über eine Datenbanksuche gegen eine *reversed* Datenbank, in der die gleichen Proteine mit umgekehrter Aminosäuresequenz enthalten waren. Die Auswertung der quantitativen Daten erfolgte, wie zuvor beschrieben, außer dass die zuvor verwendete *outlier* Software auf Grund ihrer Limitation auf 300 Werte nicht mehr eingesetzt werden konnte. IQR-Tests zur Entfernung mathematischer Ausreißer wurden deshalb mit Hilfe eines Perl-Skriptes (Bart HJ van den Berg, Institut für Experimentelle Medizin, Christian-Albrechts Universität zu Kiel, siehe Anhang Kapitel 3.I Perl-Skript TMT_SIXREP_IQR_Outlier v 2.0) durchgeführt. Auch hier wurden die geometrischen Mittelwerte, sowie die zugehörige Standardabweichung der TMT-Flächenverhältnisse nach Tan *et al.* errechnet (2008).

Um die Spaltspezifität der Endoprotease GluC zu ermitteln, wurde eine leicht modifizierte Version der PICS Datenanalyse-Pipeline verwendet (Schilling and Overall, 2008). Hierzu wurden die Excel-Tabellen dahingehend verändert, dass mehr als 1000 Peptide pro Analyse betrachtet werden konnten. *Heat-Maps* wurden mittels Microsoft Office Excel 2007 mit dem *Add-Inn* gradient-contour (<http://j-walk.com/ss/excel/files/gradcontour.htm>) erstellt.

3.3 Ergebnisse

Zwei wesentliche Punkte der quantitativen Proteomanalyse sind die Robustheit einer Methode sowie die Abschätzung der systematischen Fehler, die der Methode zu Grunde liegen. Zusätzlich spielen aber auch die Fehler während der Probenvorbereitung eine wichtige Rolle. Deshalb müssen sowohl geeignete Messbedingungen (z. B. für die Generierung der MS/MS-Spektren), als auch die notwendige Anzahl an technischen bzw. internen technischen Replikaten für eine ausreichende statistische Analyse berücksichtigt werden. Durch die

Interpretation der, in diesem Kapitel beschriebenen Daten, sollte deshalb v. a. eine optimierte MALDI-TOF/TOF MS/MS-Methode für die Identifizierung und stabile Quantifizierung von TMT-markierten Peptiden entwickelt werden. Zusätzlich dienten die Experimente der Validierung des Q-PICS Ansatzes zur simultanen Bestimmung der Spaltspezifität (*cleavage site specificity*) einer unbekannt Protease und der Quantifizierung von proteolytischen Prozessen. Die Spaltspezifität einer Protease richtet sich nach den Aminosäuren, die N- oder C-terminal der Spaltstelle liegen (siehe Abbildung 1.16).

Die generelle Vorgehensweise (*workflow*) des Q-PICS Ansatzes ist in Abbildung 3.1 gezeigt. Als Basis dienen die intakten Proteine eines kompletten Proteoms. Es können auch Subproteome oder weniger komplexe Mischungen verwendet werden. Für die in dieser Arbeit durchgeführten Experimente wurden eine tryptische und eine chymotryptische Peptidbibliothek auf Basis eines 10-Proteinstandards sowie eine tryptische Peptidbibliothek auf Basis des zytosolischen Hefeproteoms der Bäckerhefe *Saccharomyces cerevisiae* verwendet. Nach Denaturierung der intakten Proteine wurden zunächst die Disulfidbrücken reduziert und die freien Cysteinseitenketten alkyliert. Peptide wurden durch den Verdau mit einer (bekannten) *work*-Protease generiert und die primären Amine der N-Termini und Lysinseitenketten mittels reduktiver Dimethylierung geschützt. Die Herstellung der Peptidbibliothek wurde dann durch eine Aufreinigung der derivatisierten Peptide abgeschlossen (Abbildung 3.1, A). Im eigentlichen Protease Assay wurde die Peptidbibliothek dann mit der (unbekannten) *target*-Protease verdaut. Für die Experimente in diesem Kapitel wurde GluC als (in diesem Falle bekannte) *target*-Protease verwendet.

Hierzu wurden jeweils sechs Aliquote der drei verschiedenen Peptidbibliotheken mit GluC verdaut. Insgesamt wurden demnach sechs interne technische Replikate verwendet. Die in dem Verdau entstandenen neo-N-Termini wurden mit einem der sechs unterschiedlichen TMT-Reagenzien markiert und im Verhältnis 1:1:1:1:1:1 vereinigt und die vereinigten Proben jeweils mittels LC-MALDI-TOF/TOF MS/MS analysiert. Während die TMT-markierten Peptide (*prime* Spaltsequenz) über eine Mascot Datenbanksuche identifiziert wurden, werden die *nonprime* Spaltsequenzen mittels einer leicht modifizierten Datenanalyse-Pipeline des PICS-Ansatzes bestimmt (Abbildung 3.1, B) (Schilling and Overall, 2008).

Aus den beschriebenen Gründen war es notwendig zunächst eine optimierte LC-MALDI-TOF/TOF MS/MS Methode für die qualitative und quantitative Analyse TMT-markierter Peptide zu entwickeln. Hierzu wurden systematisch die Parameter **i**) Anzahl der akkumulierten Laserschüsse zur Generierung der MS/MS-Spektren (3000 Laserschüsse vs. 5000 Laserschüsse), **ii**) Gasdruck in der CID-Kollisionszelle (CIDniedrig = 1×10^{-6} Torr vs.

CIDhoch = 4×10^{-6} Torr) und **iii**) Reihenfolge der *precursor*-Auswahl für die Fragmentierung in der MS/MS-Analyse (“*weakest precursor first*” vs. “*strongest precursor first*”) untersucht. Für die Optimierung der Parameter **i**) und **ii**) wurden eine tryptische und eine chymotryptische Peptidbibliothek auf Basis eines 10-Proteinstandards verwendet. Im Anschluss daran wurde der Parameter **iii**) in Abhängigkeit des Parameters **i**) mit einer tryptischen Peptidbibliothek auf Basis eines zytosolischen Hefeproteoms analysiert. Parameter **ii**) wurde in den Experimenten auf Basis des 10-Proteinstandards bereits optimiert und wurde in den Experimenten mit dem zytosolischen Hefeproteoms konstant gehalten.

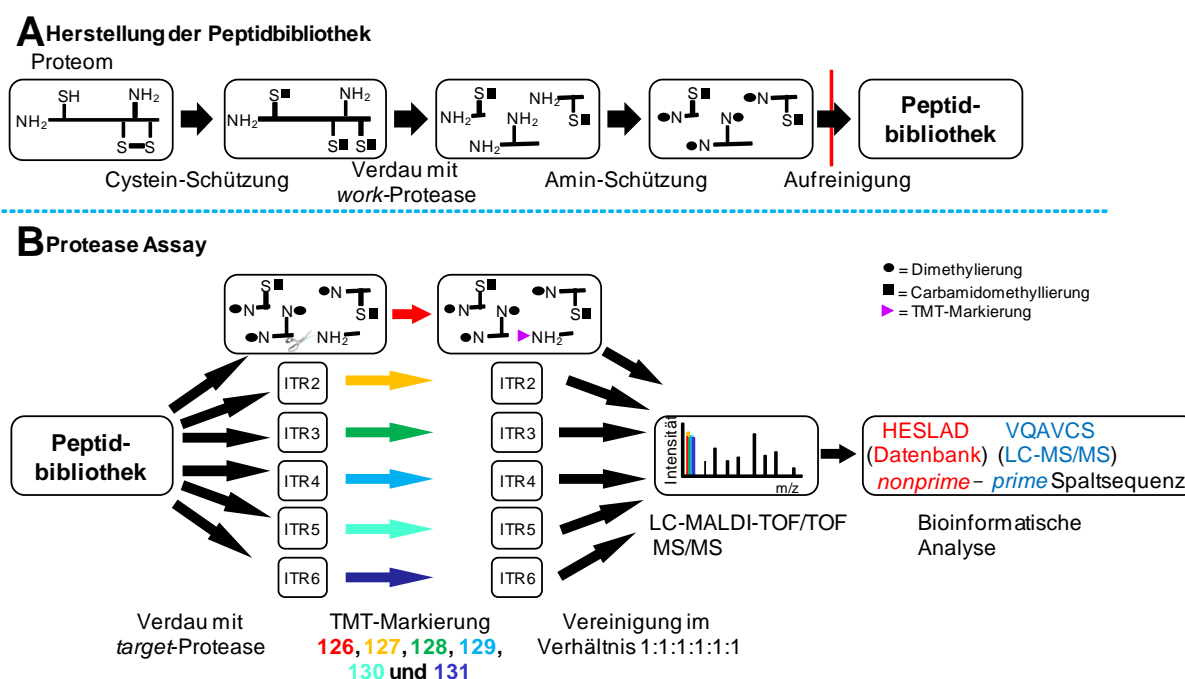


Abb. 3.1: Generelle Vorgehensweise des Q-PICS Ansatzes nach einem modifizierten PICS Ansatz (Schilling and Overall, 2008). **A**) Herstellung der Peptidbibliothek: Nach Denaturierung der intakten Proteine, Reduzierung der Disulfidbrücken und Alkylierung der freien Cysteinseitenketten erfolgt ein Verdau des Basisproteoms mittels einer bekannten *work*-Protease (z. B. Trypsin oder Chymotrypsin). Die primären Amine werden mittels reduktiver Dimethylierung geschützt. Nach Aufreinigung der Peptide ist die Herstellung der Peptidbibliothek abgeschlossen. **B**) Protease Assay: Verdau mit der (unbekannten) *target*-Protease. Bis zu sechs unterschiedliche Proben können mit dem TMT Kit gleichzeitig analysiert werden. Hierzu werden die durch die *target*-Protease gebildeten neo-N-Termini mit den unterschiedlichen TMT-Reagenzien markiert und im Verhältnis 1:1:1:1:1:1 vereinigt. Die Identifizierung der TMT-markierten Peptide (der *prime* Spaltsequenzen) erfolgt nach der LC-MALDI-TOF/TOF MS/MS Analyse über eine Mascot-Datenbanksuche. Die *nonprime* Spaltsequenzen werden bioinformatisch über eine leicht modifizierte

PICS Datenanalyse-Pipeline bestimmt (Schilling and Overall, 2008). Die quantitative Analyse erfolgt über die Flächen der einzelnen Reporterionensignale, die zueinander ins Verhältnis gesetzt werden (ITR = internes technisches Replikat).

3.3.1 MALDI-TOF/TOF MS/MS Parameteroptimierung mittels Peptidbibliotheken auf Basis des 10-Proteinstandards

Die Proteine und verwendeten Konzentrationen des 10-Proteinstandards sind in Tabelle 3.1 und Tabelle 3.2 aufgelistet. Die Herstellung der tryptischen und chymotryptischen Peptidbibliotheken erfolgte nach der in Abbildung 3.1 A dargestellten Vorgehensweise. GluC wurde als *target*-Protease verwendet und die dabei entstandenen freien N-Termini mittels TMT derivatisiert. Insgesamt wurden jeweils 250 fmol und 2,5 pmol an verdautem Gesamtprotein für die Auftrennung mittels nanoHPLC verwendet (siehe Abbildung 3.1 B). In jeweils vier technischen Replikaten wurden dann die insgesamt vier verschiedenen Kombinationen der Parameter **i**) Anzahl der akkumulierten Laserschüsse und **ii**) Gasdruck in der CID-Kollisionszelle und ihr Einfluss auf die Identifizierung und Quantifizierung der TMT-markierten Peptide der *inclusion list* untersucht.

A) Tryptische Peptidbibliothek des 10-Proteinstandards

Von den 50 Peptiden der *inclusion list* wurden sowohl bei der Verwendung von 250 fmol als auch von 2,5 pmol an verdauter Gesamtproteinmenge in der LC-Auftrennung, 47 TMT-markierte Peptide identifiziert (siehe Anhang Kapitel 3 Tabelle 3.III).

Um den Einfluss der beiden Parameter **i**) und **ii**) auf die Identifizierung der TMT-markierten Peptide beurteilen zu können, wurde der Mascot *ion score*, ein Maß für die Wahrscheinlichkeit einer korrekten Identifizierung eines Peptides, verwendet. In Abbildung 3.2 ist der durchschnittliche *ion score* der identifizierten Peptide aus den jeweils vier technischen Replikaten der verschiedenen Parameterkombinationen (CIDniedrig, 3000 Laserschuss; CIDniedrig, 5000 Laserschuss; CIDhoch, 3000 Laserschuss und CIDhoch, 5000 Laserschuss) gezeigt.

Der durchschnittliche Mascot *ion score* schwankt über alle Peptide gemittelt <5% für die auf der LC aufgetrennten Mengen an verdautem Gesamtprotein von 250 fmol bzw. <6% für 2,5 pmol. Der Unterschied des *ion scores* zwischen den beiden Mengen an verdautem Gesamtprotein beträgt <3%, wobei die höhere Menge an eingesetzten Ausgangspeptiden auch die höheren *ion scores* zur Folge hat.

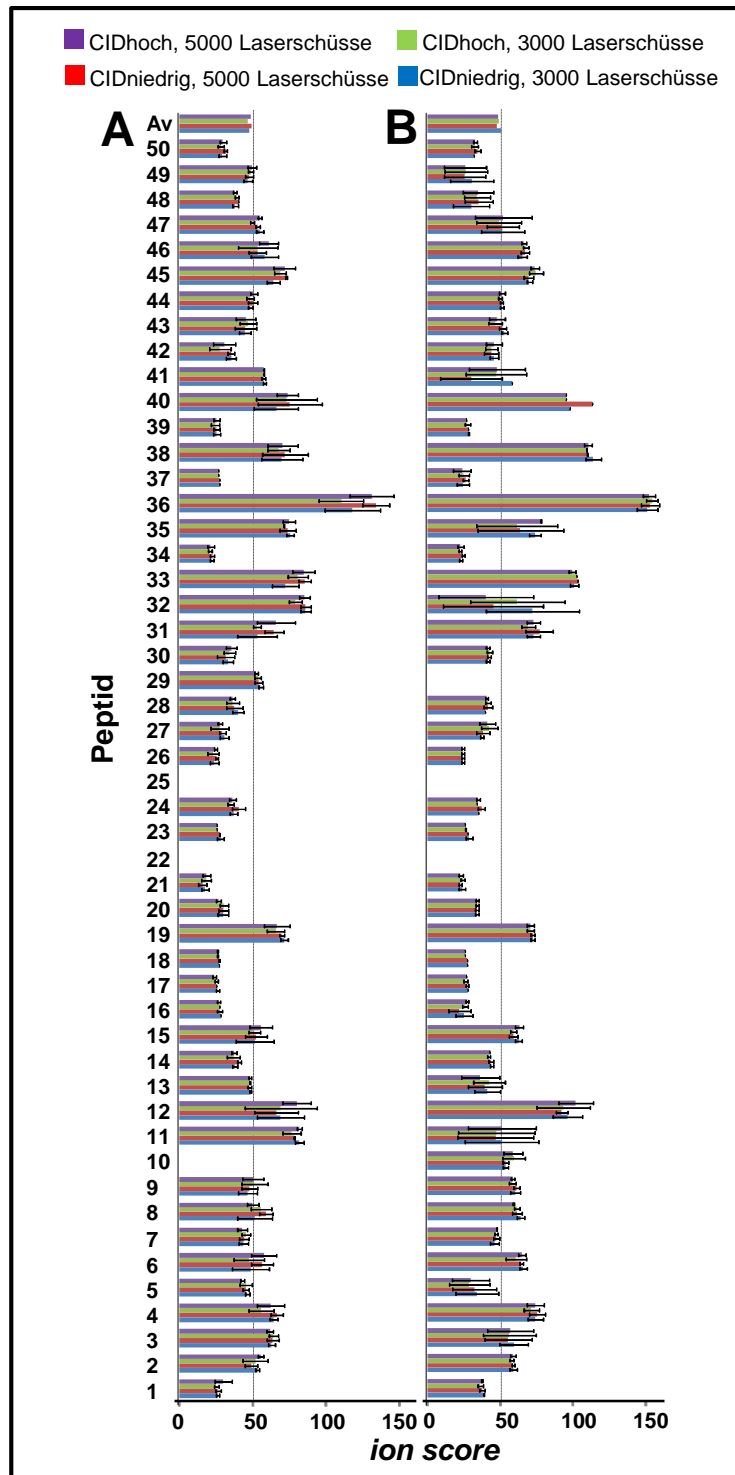


Abb. 3.2: Auftragung des Mascot *ion scores* der TMT-markierten Peptide der tryptischen Peptidbibliothek des 10-Proteinstandards. Die verschiedenen Farben beziehen sich auf die verschiedenen Kombinationen der Parameter **i**) Anzahl der akkumulierten Laserschüsse zur Generierung der MS/MS-Spektren (3000 Laserschüsse vs. 5000 Laserschüsse), **ii**) Gasdruck in der Kollisionszelle (CIDniedrig = 1×10^{-6} Torr vs. CIDhoch = 4×10^{-6}). Der durchschnittliche *ion score* der identifizierten Peptide errechnet sich aus den jeweils vier technischen Replikaten. A) Verwendung von

250 fmol und B) Verwendung von 2,5 pmol an verdauter Gesamtproteinmenge in der LC-Auftrennung. $A_v = average$ (durchschnittlicher *ion score* über alle Peptide einer Parameterkombination gemittelt).

Der Einfluss der unterschiedlichen Kombinationen der Parameter **i**) Anzahl der akkumulierten Laserschüsse und **ii**) Gasdruck in der Kollisionszelle wurde nicht nur bezüglich der Identifizierung sondern auch bezüglich der Quantifizierung der TMT-markierten Peptide betrachtet. Hierzu wurden alle TMT-Reporterionenflächen mit der Mascot Software durch die Flächenwerte des Kanals 126 geteilt. Dadurch wurden insgesamt fünf verschiedene TMT-Flächenverhältnisse (127/126 bis 131/126) errechnet. Da es sich um insgesamt sechs interne technische Replikate handelte, die im Verhältnis 1:1:1:1:1 kombiniert wurden, betrug der Erwartungswert der TMT-Flächenverhältnisse 1,0. Negative Werte der TMT-Flächenverhältnisse wurden entfernt. Diese werden durch die Mascot Software berechnet, wenn beispielsweise ein unkorrektes Isotopenverhältnis im MS/MS-Spektrum erhalten wird, das nicht dem natürlichen Erwartungswert entspricht. Insgesamt wurden in den Experimenten mit der tryptischen Peptidbibliothek des 10-Proteinstandards unter Verwendung von 250 fmol an verdauter Gesamtproteinmenge maximal 3,6% und für 2,5 pmol an verdauter Gesamtproteinmenge 1,3% solch negativer TMT-Flächenverhältnisse beobachtet.

Zusätzlich zu den negativen TMT-Flächenverhältnissen wurden auch mathematische Ausreißer über einen IQR-Test der freizugänglichen *outlier* Software identifiziert und von der weiteren Analyse ausgeschlossen. Die durchschnittlichen geometrischen Mittelwerte und Standardabweichungen über alle fünf TMT-Flächenverhältnisse sind in Tabelle 3.7 und Abbildung 3.3 dargestellt.

Für alle Parameterkombinationen (CIDniedrig, 3000 Laserschüsse; CIDniedrig, 5000 Laserschüsse; CIDhoch, 3000 Laserschüsse und CIDhoch, 5000 Laserschüsse) der beiden unterschiedlichen Mengen an verdautem Gesamtprotein wurde für den durchschnittlichen geometrischen Mittelwert eine Abweichung <3% vom Erwartungswert 1,0 erhalten. Der durchschnittliche geometrische Mittelwert der TMT-Flächenverhältnisse über alle vier Parameterkombinationen errechnete sich auf 0,984, unabhängig von der verwendeten Menge an verdautem Gesamtprotein. Unter Verwendung von 250 fmol an verdauter Gesamtproteinmenge beträgt die durchschnittliche Standardabweichung über alle vier Parameterkombinationen gemittelt <14%, wohingegen sie bei der höheren verdauten Gesamtproteinmenge von 2,5 pmol einen Wert <9% erreicht.

Vergleicht man nun den Einfluss des Parameters **i**) die Anzahl an akkumulierten Laserschüssen miteinander, so erhält man bei der Menge an verdautem Gesamtprotein von 250 fmol bei beiden unterschiedlichen Drücken (CIDniedrig und CIDhoch) unter der Akkumulation von 5000 Laserschüssen eine geringere durchschnittliche Standardabweichung, als unter der Akkumulation von 3000 Laserschüssen (15% vs. 18% und 11% vs. 13%). Dieser Effekt ist bei der Verwendung der höheren verdauten Gesamtproteinmenge weniger ausgeprägt. Hier ist die durchschnittliche Standardabweichung für die unterschiedlichen Drücke (CIDniedrig und CIDhoch) unter Akkumulation von 5000 Laserschüssen im Vergleich zu 3000 Laserschüssen höher (10% vs. 9% und 9% vs. 8%).

Tab. 3.7: Durchschnittliche geometrische Mittelwerte und Standardabweichungen (SD) über alle fünf TMT-Flächenverhältnisse (127/126 bis 131/126) der MALDI-TOF/TOF MS/MS Parameteroptimierung mittels der tryptischen Peptidbibliothek des 10-Proteinstandards. Alle Werte sind über die Summe von vier technischen Replikaten für jede Parametereinstellung gemittelt.

	250 fmol Gesamt-Peptidmenge		2,5 pmol Gesamt-Peptidmenge	
Parameter-einstellung	durchschnittlicher geometrischer Mittelwert	durchschnittliche SD (%)	durchschnittlicher geometrischer Mittelwert	durchschnittliche SD (%)
CIDniedrig, 3000 Laserschüsse	0,975	0,176 (18%)	0,978	0,093 (9%)
CIDniedrig, 5000 Laserschüsse	0,988	0,144 (15%)	0,980	0,094 (10%)
CIDhoch, 3000 Laserschüsse	0,986	0,125 (13%)	0,988	0,075 (8%)
CIDhoch, 5000 Laserschüsse	0,986	0,108 (11%)	0,992	0,089 (9%)

Für beide Mengen an verdautem Gesamtprotein erreicht die Standardabweichung der TMT-Flächenverhältnisse in Bezug auf Parameter **ii**) den Druck in der CID-Kollisionszelle, unter CIDhoch einen geringeren Wert, als unter CIDniedrig unabhängig von der verwendeten Schusszahl. So erhält man für die durchschnittliche Standardabweichung unter Akkumulation von 5000 Laserschüssen unter dem höheren CID Druck eine ähnliche Verbesserung wie unter Akkumulation von 3000 Laserschüssen (13% vs. 18% und 11% vs. 15%). Dieser Trend ist bei

der höheren verdauten Gesamtproteinmenge weniger stark ausgeprägt (8% vs. 9% und 9% vs. 10%).

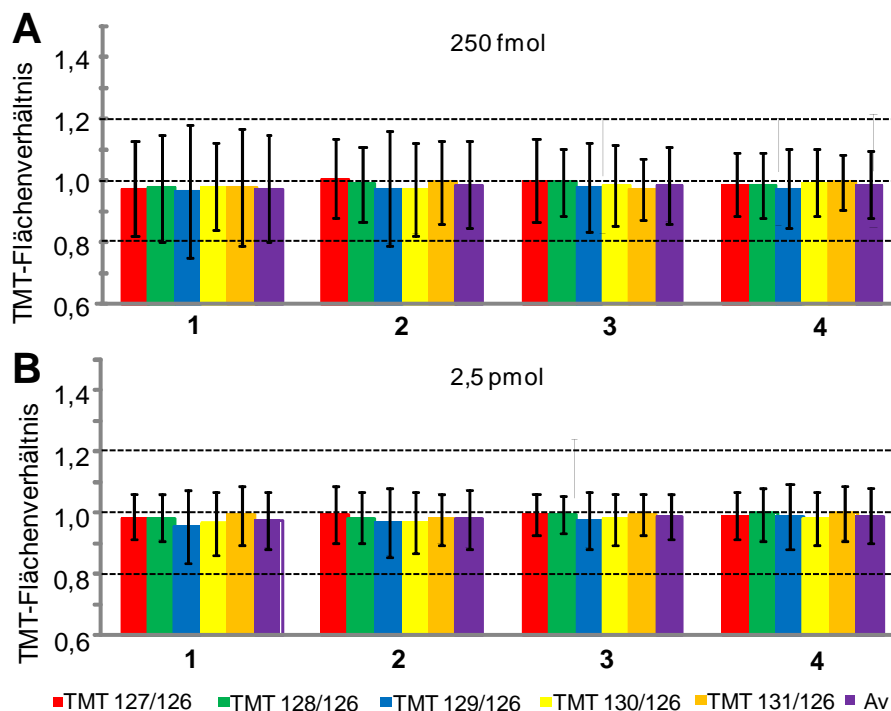


Abb. 3.3: Geometrische Mittelwerte der TMT-Flächenverhältnisse (127/126 bis 131/126) und Standardabweichungen der MALDI-TOF/TOF MS/MS Parameteroptimierung mittels der tryptischen Peptidbibliothek des 10-Proteinstandards. Alle Werte sind über die Summe von vier technischen Replikaten für jede Parametereinstellung (1 bis 4) gemittelt. 1) CIDniedrig, 3000 Laserschüsse; 2) CIDniedrig, 5000 Laserschüsse; 3) CIDhoch, 3000 Laserschüsse und 4) CIDhoch, 5000 Laserschüsse). A) Verwendung von 250 fmol und B) Verwendung von 2,5 pmol an verdauter Gesamtproteinmenge für die LC-Auftrennung. Av = *average* (durchschnittlicher geometrischer Mittelwert über alle fünf TMT-Flächenverhältnisse).

B) Chymotryptische Peptidbibliothek des 10-Proteinstandards

In den Experimenten mit der chymotryptischen Peptidbibliothek des 10-Proteinstandards wurden von den 50 TMT-markierten Peptiden der *inclusion list* 49 TMT-markierte Peptide unter Verwendung von 250 fmol an verdauter Gesamtproteinmenge identifiziert. Unter Verwendung von 2,5 pmol wurden alle 50 TMT-markierten Peptide identifiziert (siehe Anhang Kapitel 3 Tabelle 3.IV).

Um wie zuvor bei der tryptischen Peptidbibliothek den Einfluss der beiden Parameter **i)** Nummer an akkumulierten Laserschüssen und **ii)** Druck in der CID-Kollisionszelle auf die Identifizierung der TMT-markierten Peptide beurteilen zu können, ist in Abbildung 3.4 der

durchschnittliche Mascot *ion score* der identifizierten, TMT-markierten Peptide der verschiedenen Parameterkombinationen gezeigt.

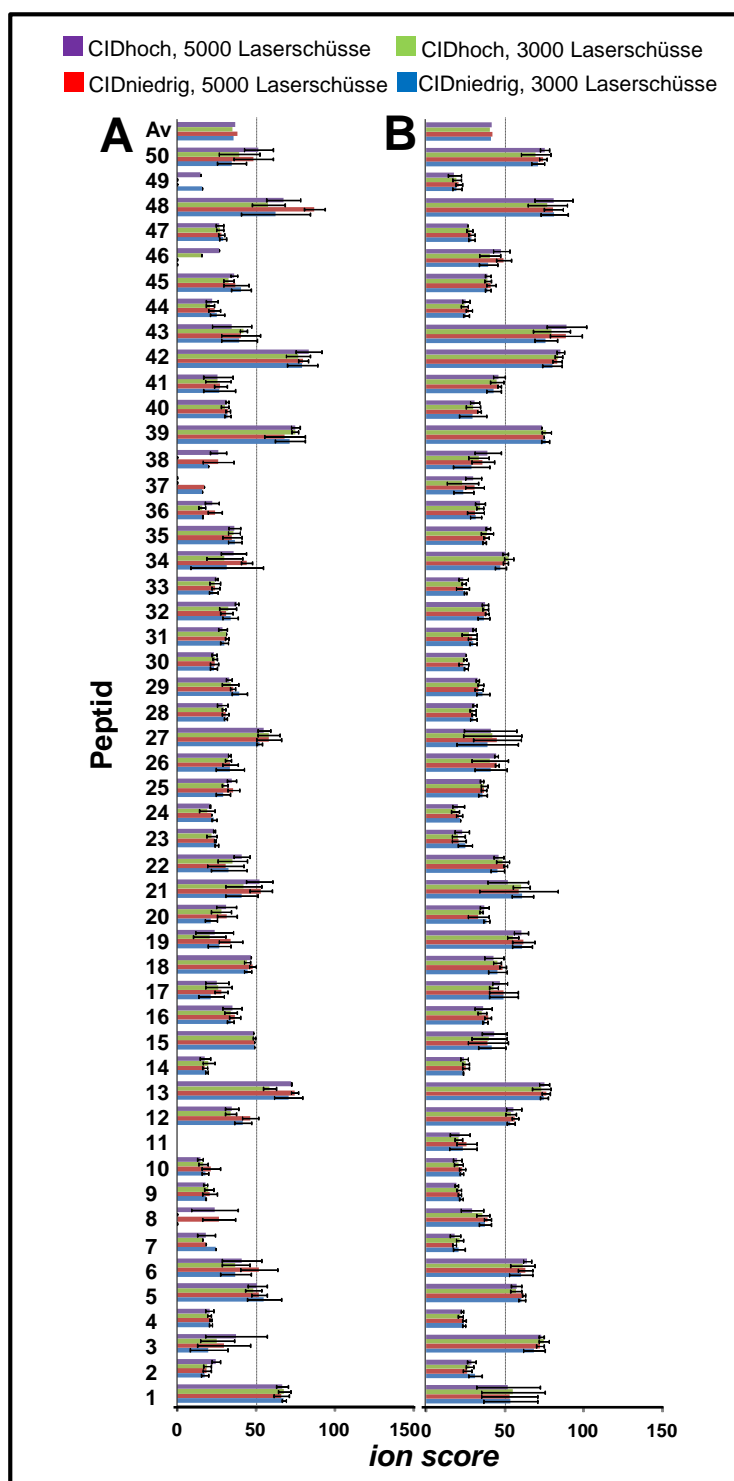


Abb. 3.4: Auftragung des Mascot *ion scores* der TMT-markierten Peptide der chymotryptischen Peptidbibliothek des 10-Proteinstandards. Die verschiedenen Farben beziehen sich auf die verschiedenen Kombinationen der Parameter **i**) Anzahl der akkumulierten Laserschüsse zur Generierung der MS/MS-Spektren (3000 Laserschüsse vs. 5000 Laserschüsse), **ii**) Gasdruck in der Kollisionszelle (CIDniedrig = 1×10^{-6} Torr vs.

CIDhoch = 4×10^{-6}). Der durchschnittliche *ion score* der identifizierten Peptide errechnet sich aus den jeweils vier technischen Replikaten. A) Verwendung von 250 fmol und B) Verwendung von 2,5 pmol an verdauter Gesamtproteinmenge für die LC-Auftrennung. $Av = average$ (durchschnittlicher *ion score* über alle Peptide einer Parameterkombination gemittelt).

Keiner der beiden Parameter hat einen starken Einfluss auf den Mascot *ion score*. Der durchschnittliche Mascot *ion score* schwankt <4% unter Verwendung von 2,5 pmol an verdauter Gesamtproteinmenge. Bei der geringeren Menge an verdautem Gesamtprotein von 250 fmol weist der durchschnittliche *ion score* mit <8% die größten Abweichungen auf.

Auch in den Experimenten zur Parameteroptimierung mit der chymotryptischen Peptidbibliothek des 10-Proteinstandards wurde der Einfluss der unterschiedlichen Parameterkombinationen zusätzlich zur Identifizierung auch in Bezug auf die Quantifizierung untersucht. Wie zuvor wurden hier alle TMT-Reporterionenflächen mit der Mascot Software durch die Flächenwerte des Kanals 126 geteilt. Dadurch wurden erneut insgesamt fünf verschiedene TMT-Flächenverhältnisse (127/126 bis 131/126) errechnet. Die Anzahl an negativen TMT-Flächenverhältnissen unter Verwendung von 250 fmol an verdauter Gesamtproteinmenge erreichten einen Maximalwert von 9,6%. Unter Verwendung von 2,5 pmol an verdauter Gesamtproteinmenge wurde ein Wert von 1,6% an negativen TMT-Flächenverhältnissen beobachtet. Auch in diesen Experimenten wurden zusätzlich zu den negativen TMT-Flächenverhältnissen mathematische Ausreißer über den IQR-Test identifiziert und von der weiteren Analyse ausgeschlossen. Die geometrischen Mittelwerte und Standardabweichungen über alle fünf TMT-Flächenverhältnisse sind in Tabelle 3.8 und Abbildung 3.5 dargestellt.

Der geometrische Mittelwert der einzelnen Parameterkombinationen weist unter Verwendung von 250 fmol an verdauter Gesamtproteinmenge eine Abweichung <5% vom Erwartungswert 1,0 auf, wohingegen er für die größere Menge von 2,5 pmol, wie in den Experimenten mit der tryptischen Peptidbibliothek einen Wert <3% erreicht. Der durchschnittliche geometrische Mittelwert der TMT-Flächenverhältnisse über alle vier Parameterkombinationen gemittelt, errechnete sich auf 0,975 mit einer Standardabweichung <22% unter Verwendung von 250 fmol an verdauter Gesamtproteinmenge. Unter Verwendung von 2,5 pmol errechnete er sich auf 0,984 mit einer Standardabweichung <15%.

Betrachtet man hier Parameter **i**) die Anzahl an akkumulierten Laserschüssen, so wird für beide Mengen an verdautem Gesamtprotein mit beiden unterschiedlichen Drücken

(CIDniedrig und CIDhoch) unter der Akkumulation von 5000 Laserschüssen eine geringere Standardabweichung als unter der Akkumulation von 3000 Laserschüssen erreicht (24% vs. 25% bzw. 17% vs. 20% für 250 fmol und 15% vs. 16% bzw. 13% vs. 14% für 2,5 pmol an verwendeter Menge an verdautem Gesamtprotein).

Tab. 3.8: Durchschnittliche geometrische Mittelwerte und Standardabweichungen (SD) über alle fünf TMT-Flächenverhältnisse (127/126 bis 131/126) der MALDI-TOF/TOF MS/MS Parameteroptimierung mittels der chymotryptischen Peptidbibliothek des 10-Proteinstandards. Alle Werte sind über die Summe von vier technischen Replikaten für jede Parametereinstellung gemittelt.

Parameter-einstellung	250 fmol Gesamt-Peptidmenge		2,5 pmol Gesamt-Peptidmenge	
	durchschnittlicher geometrischer Mittelwert	durchschnittliche SD (%)	durchschnittlicher geometrischer Mittelwert	durchschnittliche SD (%)
CIDniedrig, 3000 Laserschüsse	0,953	0,238 (25%)	0,971	0,156 (16%)
CIDniedrig, 5000 Laserschüsse	0,979	0,233 (24%)	0,981	0,144 (15%)
CIDhoch, 3000 Laserschüsse	0,979	0,198 (20%)	0,995	0,141 (14%)
CIDhoch, 5000 Laserschüsse	0,990	0,172 (17%)	0,990	0,130 (13%)

Bezüglich Parameter **ii)** dem Druck in der CID-Kollisionszelle wurden unter dem höheren CID Druck wieder geringere Standardabweichungen erhalten als unter dem niedrigeren CID Druck. Wie zuvor in den Experimenten mit der tryptischen Peptidbibliothek ist dieser Trend für die höhere Menge an verdautem Gesamtprotein von 2,5 pmol weniger stark ausgeprägt, als unter Verwendung von 250 fmol. Dies ist wiederum unabhängig davon, ob 5000 oder 3000 Laserschüsse akkumuliert werden (17% vs. 24% bzw. 20% vs. 25% für 250 fmol und 13% vs. 15% bzw. 14% vs. 16% für 2,5 pmol an verwendeter Menge an verdautem Gesamtprotein).

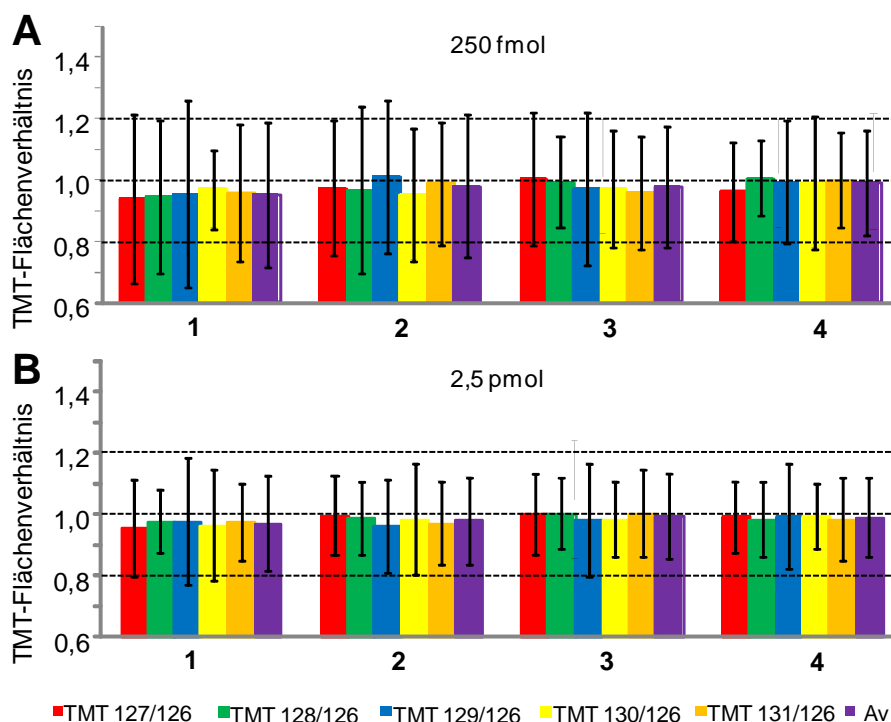


Abb. 3.5: Geometrische Mittelwerte der TMT-Flächenverhältnisse (127/126 bis 131/126) und Standardabweichungen der MALDI-TOF/TOF MS/MS Parameteroptimierung mittels der chymotryptischen Peptidbibliothek des 10-Proteinstandards. Alle Werte sind über die Summe von vier technischen Replikaten für jede Parametereinstellung (1 bis 4) gemittelt. 1) CIDniedrig, 3000 Laserschüsse; 2) CIDniedrig, 5000 Laserschüsse; 3) CIDhoch, 3000 Laserschüsse und 4) CIDhoch, 5000 Laserschüsse). A) Verwendung von 250 fmol und B) Verwendung von 2,5 pmol an verdauter Gesamtproteinmenge für die LC-Auftrennung. Av = *average* (durchschnittlicher geometrischer Mittelwert über alle fünf TMT-Flächenverhältnisse).

3.3.2 MALDI-TOF/TOF MS/MS Parameteroptimierung mittels der Peptidbibliothek des zytosolischen Hefeproteoms

Neben den Parametern **i**) Anzahl der akkumulierten Laserschüsse zur Generierung der MS/MS-Spektren (3000 Laserschüsse vs. 5000 Laserschüsse) und **ii**) Gasdruck in der CID-Kollisionszelle (CIDniedrig = 1×10^{-6} Torr vs. CIDhoch = 4×10^{-6} Torr) wurde auch Parameter **iii**) Reihenfolge der *precursor* Auswahl für die Fragmentierung in der MS/MS-Analyse (“*weakest precursor first*” vs. “*strongest precursor first*”) untersucht. Hierzu wurde eine tryptische Peptidbibliothek auf Basis des zytosolischen Proteoms der Bäckerhefe *S. cerevisiae* verwendet und nach der gleichen Vorgehensweise, wie in den Experimenten mit dem 10-Proteinstandard, hergestellt (siehe Abbildung 3.1 A). Die tryptische Peptidbibliothek wurde

ebenfalls mit GluC als *target*-Protease verdaut und die daraus resultierenden freien N-Termini mit TMT markiert. Wie in den Experimenten mit den Peptidbibliotheken des 10-Proteinstandards wurden jeweils sechs interne technische Replikate im Verhältnis 1:1:1:1:1:1 vereinigt und jeweils 2,5 µg an Gesamt-Peptidmenge mittels nanoHPLC aufgetrennt. Während Parameter **ii**) der Gasdruck in der CID-Kollisionszelle in diesen Experimenten konstant auf CIDhoch eingestellt wurde, wurden die verschiedenen Parameterkombinationen von **i**) Anzahl der akkumulierten Laserschüsse und **iii**) Reihenfolge der *precursor*-Auswahl für die Fragmentierung in der MS/MS-Analyse untersucht. Insgesamt wurde der Einfluss von vier verschiedenen Parameterkombinationen 1) 3000 Laserschüsse, “*weakest precursor first*”, 2) 5000 Laserschüsse, “*weakest precursor first*”; 3) 3000 Laserschüsse, “*strongest precursor first*” und 4) 5000 Laserschüsse, “*strongest precursor first*” auf die Identifizierung und Quantifizierung von TMT-markierten Peptiden analysiert. In Tabelle 3.9 sind die qualitativen Ergebnisse dieser Untersuchungen aufgelistet.

Tab. 3.9: Qualitative Ergebnisse der MALDI-TOF/TOF MS/MS Parameteroptimierung mittels dem zytosolischen Hefeproteom. Alle Werte sind über die Summe von drei technischen Replikaten für jede Parametereinstellung gemittelt.

Parametereinstellung	# <i>precursor</i>	durchschnittlicher <i>ion score</i>	# identifizierte Peptide	# identifizierte Spaltstellen	# identifizierte Spaltstellen (in mindestens 2 aus 3 TR)
3000 Laserschüsse, <i>weakest precursor first</i>	26.672	58	2.393	886	631
5000 Laserschüsse, <i>weakest precursor first</i>	26.450	56	2.109	828	550
3000 Laserschüsse, <i>strongest precursor first</i>	26.872	61	2.814	962	737
5000 Laserschüsse, <i>strongest precursor first</i>	24.395	60	2.261	819	601

Die Anzahl der *precursor*, die für eine MS/MS-Analyse ausgewählt wurden, unterscheidet sich <2% für die verschiedenen Parametereinstellungen. Lediglich für die Parametereinstellung 5000 Laserschüsse, CIDhoch wurden mit 24.395 Vorläuferionen 9% weniger ausgewählt. Dies ist bei den Ergebnissen zu berücksichtigen. Der durchschnittliche Mascot *ion score* ist, wie für die Messungen mit den Peptidbibliotheken des 10-Proteinstandards, nur wenig von den verschiedenen Parametereinstellungen beeinflusst und schwankt <6%. Um einen Vergleich der verschiedenen Parametereinstellungen

durchzuführen, wurde die Anzahl der identifizierten TMT-markierten Peptide verglichen. Dabei wurde für alle Parametereinstellungen ein *ion score cutoff* für den *identity threshold* von 38 gewählt. Damit wurde eine falsch-positiv Rate <4% erreicht. Die Peptidzuordnungen resultieren letztendlich in der Identifizierung von Spaltstellensequenzen, die dann über eine leicht modifizierte Version der PICS Datenanalyse-Pipeline ermittelt wurden (Schilling and Overall, 2008).

Vergleicht man den Einfluss von Parameter **i**) die Anzahl der akkumulierten Laserschüsse mit der Anzahl an identifizierten TMT-markierten Peptiden, so wurden nach Akkumulation von 3000 Laserschüssen mehr Identifizierungen erreicht als nach der Akkumulation von 5000 Laserschüssen, unabhängig von Parameter **iii**) der Reihenfolge der *precursor*-Auswahl (2.393 vs. 2.109 TMT-markierte Peptide und 2.814 vs. 2.261 TMT-markierte Peptide). Bezüglich dieses Parameters wurden mehr TMT-markierte Peptide unter Verwendung der “*strongest precursor first*” Methode identifiziert, als unter Verwendung der “*weakest precursor first*” Methode, unabhängig von der Anzahl an akkumulierten Laserschüssen (2.393 vs. 2.814 TMT-markierte Peptide und 2.109 vs. 2.261 TMT-markierte Peptide). Die meisten TMT-markierten Peptide (2.814) wurden unter der Verwendung der Parameterkombination 3000 Laserschüsse, “*strongest precursor first*” identifiziert. Vergleicht man dies mit der Parameterkombination 5000 Laserschüsse, “*weakest precursor first*”, mit der die wenigsten TMT-markierten Peptide (2.109) identifiziert wurden, so sind es 25% weniger.

In den Experimenten mit der Peptidbibliothek des zytosolischen Hefeproteoms wurde der Einfluss der verschiedenen Parameterkombinationen aus **i**) der Anzahl der akkumulierten Laserschüsse und **iii**) Reihenfolge der *precursor*-Auswahl für die Fragmentierung auch in Bezug auf die Quantifizierung TMT-markierter Peptide untersucht. Um eine möglichst große Anzahl an Quantifizierungswerten pro identifizierter Spaltstellensequenz zu berücksichtigen, wurden für die weitere Analyse nur solche TMT-markierte Peptide als identifiziert betrachtet, die in mindestens zwei von drei technischen Replikaten gefunden wurden.

Es wurden wie zuvor die fünf verschiedenen TMT-Flächenverhältnisse (127/126 bis 131/126) mittels der Mascot Software bestimmt. Insgesamt wurden maximal 1,0% an negativen Werten für TMT-Flächenverhältnisse errechnet und von der weiteren Analyse ausgeschlossen.

Da auch in diesen Experimenten sechs interne technische Replikate zum Einsatz kamen, betrug der Erwartungswert für die TMT-Flächenverhältnisse erneut 1,0. Mathematische Ausreißer wurden über einen IQR-Test mit dem selbsterstellten Perl-Skript TMT_SIXREP_IQR_Outlier v 2.0 entfernt und danach die durchschnittlichen geometrischen

Mittelwerte und Standardabweichungen über alle fünf TMT-Flächenverhältnisse berechnet (siehe Tabelle 3.10 und Abbildung 3.6).

Tab. 3.10: Durchschnittliche geometrische Mittelwerte und Standardabweichungen (SD) über alle fünf TMT-Flächenverhältnisse (127/126 bis 131/126) der MALDI-TOF/TOF MS/MS Parameteroptimierung mittels der tryptischen Peptidbibliothek des zytosolischen Hefeproteoms. Alle Werte sind über die Summe von drei technischen Replikaten für jede Parametereinstellung gemittelt.

Parametereinstellung	durchschnittlicher geometrischer Mittelwert	durchschnittliche SD (%)
3000 Laserschüsse, <i>weakest precursor first</i>	0,975	0,186 (19%)
5000 Laserschüsse, <i>weakest precursor first</i>	0,975	0,187 (19%)
3000 Laserschüsse, <i>strongest precursor first</i>	0,979	0,124 (13%)
5000 Laserschüsse, <i>strongest precursor first</i>	0,985	0,120 (12%)

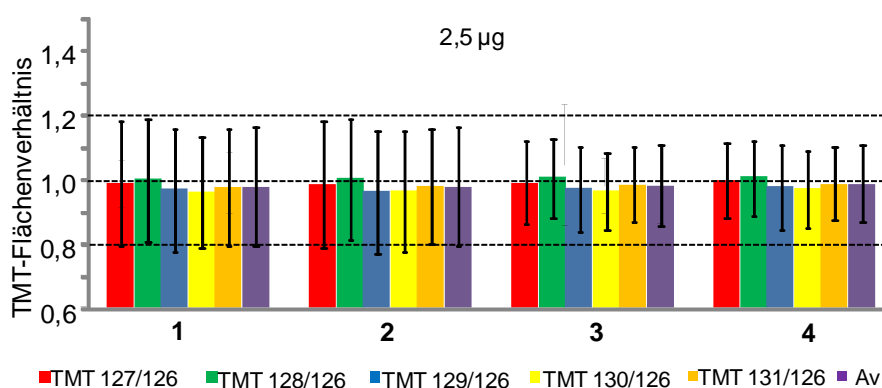


Abb. 3.6: Geometrische Mittelwerte der TMT-Flächenverhältnisse (127/126 bis 131/126) und Standardabweichungen der MALDI-TOF/TOF MS/MS Parameteroptimierung mittels der tryptischen Peptidbibliothek des zytosolischen Hefeproteoms. Alle Werte sind über die Summe von drei technischen Replikaten für jede Parametereinstellung (1 bis 4) gemittelt. 1) 3000 Laserschüsse, “*weakest precursor first*”, 2) 5000 Laserschüsse, “*weakest precursor first*”, 3) 3000 Laserschüsse, “*strongest precursor first*” und 4) 5000 Laserschüsse, “*strongest precursor first*”, Av = *average* (durchschnittlicher geometrischer Mittelwert über alle fünf TMT-Flächenverhältnisse).

Der durchschnittliche geometrische Mittelwert der TMT-Flächenverhältnisse über alle vier Parameterkombinationen 1) 3000 Laserschüsse, “*weakest precursor first*”, 2) 5000 Laserschüsse, “*weakest precursor first*”, 3) 3000 Laserschüsse, “*strongest precursor first*”; und 4) 5000 Laserschüsse, “*strongest precursor first*” gemittelt, errechnete sich auf 0,984 mit einer Standardabweichung <15%. Für alle Parameterkombinationen wurde für den durchschnittlichen geometrischen Mittelwert eine Abweichung <3% vom Erwartungswert 1,0 erhalten.

Der Einfluss des Parameters **i**) der Anzahl an akkumulierten Laserschüssen ist vernachlässigbar, unabhängig von den beiden Einstellungen des Parameters **iii**) der Reihenfolge der *precursor*-Auswahl “*weakest precursor first*” oder “*strongest precursor first*” (19% vs. 19% und 13% vs. 12%). Die durchschnittliche Standardabweichung der TMT-Flächenverhältnisse in Bezug auf diesen Parameter erreichte mittels der “*weakest precursor first*” Einstellung höhere Werte verglichen mit der “*strongest precursor first*” Einstellung, unabhängig von der Anzahl an akkumulierten Laserschüssen (19% vs. 13% und 19% vs. 12%).

Betrachtet man schließlich auch noch die benötigte Messzeit für die MS/MS-Analysen der jeweils drei technischen Replikate, so zeigt es sich, dass man unter der Akkumulation von 3000 Laserschüssen 28% weniger Messzeit benötigt als unter der Akkumulation von 5000 Laserschüssen (durchschnittlich 32,5 h vs. durchschnittlich 45 h).

3.3.3 Validierung des Q-PICS Ansatzes mittels Analyse der Spaltspezifität (*cleavage site specificity*) der Endoprotease GluC als bekannte Protease

Neben der Methodenentwicklung für den Q-PICS Ansatz, in dem v. a. geeignete Parametereinstellungen für die Identifikation und Quantifizierung TMT-markierter Peptide mittels LC-MALDI-TOF/TOF MS/MS bestimmt werden sollten, spielte auch die korrekte Identifizierung von Spaltsequenzen eine sehr wichtige Rolle. Dies stellt mit der simultanen Gewinnung quantitativer Daten die eigentliche Aufgabe des Q-PICS Ansatzes dar. Zur Validierung des Ansatzes wurde v. a. der Datensatz des zytosolischen Hefeproteoms verwendet, der unter der optimalen Parametereinstellung 3000 akkumulierte Laserschüsse, CID_{hoch} = 4×10^{-6} Torr und Reihenfolge der *precursor*-Auswahl “*strongest precursor first*” erstellt wurde. Für die Datenbanksuche wurde zunächst semi Trypsin und semi GluC als verwendetes Enzym gewählt, wobei keine C- oder N-terminale Orientierung berücksichtigt wurde. Dabei wurden 2.814 TMT-markierte Peptide, also Spaltsequenzen identifiziert.

Hieraus resultierten schließlich 962 *unique* Spaltsequenzen. Insgesamt wurden daraus 737 *unique* Spaltsequenzen, die in mindestens zwei der drei technischen Replikate identifiziert wurden. Wie bereits beschrieben, wurde die tryptische Peptidbibliothek des zytosolischen Hefeproteoms mit GluC als *target*-Protease verdaut. Dabei handelt es sich um eine Endoprotease, die Peptidbindungen speziell nach Glutamat oder weniger ausgeprägt nach Aspartat in Position P1 hydrolysiert (Sorensen *et al.*, 1991). Um die Spaltspezifität der Protease GluC zu analysieren, wurde eine leicht modifizierte Version der PICS Datenanalyse-Pipeline verwendet (Schilling and Overall, 2008). Mit dieser konnten mehr als 1000 Peptide und damit Spaltsequenzen untersucht werden. In Abbildung 3.7 ist eine *Heat-Map*, die die Häufigkeit der einzelnen Aminosäuren in Position P6 bis P6' zeigt, dargestellt. Darin wird die Spaltspezifität der Protease GluC eindeutig charakterisiert.

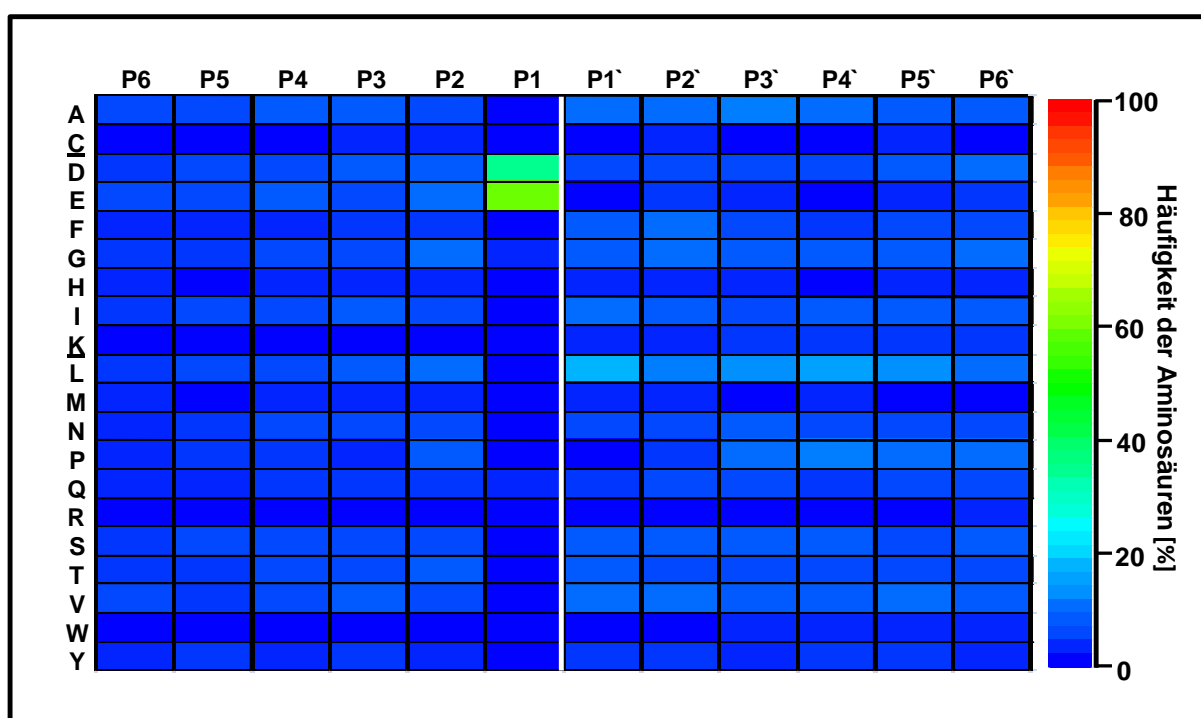


Abb. 3.7: *Heat-Map* zur Darstellung der GluC Spaltstellen Spezifität unter Verwendung des Experimentes der Peptidbibliothek des zytosolischen Hefeproteoms mit den optimalen Parametereinstellungen (3000 akkumulierte Laserschüsse, CIDhoch = 4×10^{-6} Torr und Reihenfolge der *precursor*-Auswahl "*strongest precursor first*"). Für die Datenbanksuche wurde semi Trypsin und semi GluC als Enzymeinstellung gewählt. Die Peptidbindung zwischen Position P1 und P1' ist durch einen weißen Strich abgetrennt. C= carbamidomethyliertes Cystein, K= dimethyliertes Lysin.

Insgesamt wurden 59,1% der identifizierten Spaltsequenzen nach einem C-terminalen Glutamat und 34,9% nach einem C-terminalen Aspartat in Position P1 generiert. Damit wird

ein Wert von insgesamt 94,0% erreicht, der der Spaltspezifität von GluC entspricht. Spaltsequenzen, in denen andere Aminosäuren in Position P1 enthalten waren, waren nur zu 4,4% beobachtbar. Die übrigen 1,6% verteilten sich auf sogenannte nicht eindeutige (*engl.: ambiguous*) Spaltsequenzen. Damit werden C-terminale *prime* Spaltsequenzen bezeichnet, die in verschiedenen Proteinen, mit verschiedenen *nonprime* Spaltsequenzen identifiziert werden. Hier ist eine eindeutige Zuordnung der *nonprime* Spaltsequenz nicht möglich und die Spaltsequenzen werden durch die PICS Datenanalyse-Pipeline ausgeschlossen (Schilling and Overall, 2008).

Zusätzlich wurden die drei anderen Datensätze der MALDI-TOF/TOF MS/MS Parameteroptimierung mit der Peptidbibliothek des zytosolischen Hefeproteoms bezüglich der GluC Spaltspezifität untersucht. Insgesamt wurden dabei 59,1% bis 61,2% der identifizierten Spaltsequenzen nach einem C-terminalen Glutamat und 34,0% bis 34,9% nach einem C-terminalen Aspartat in Position P1 generiert. Der Gesamtwert für die bekannte GluC Spaltspezifität erreicht somit 94,0% bis 95,3%.

3.3.4 Validierung des Q-PICS Ansatzes mittels Analyse der Spaltspezifität (*cleavage site specificity*) der Endoprotease GluC als unbekannte Protease

Da das Hauptziel der vorliegenden Arbeit darin bestand, einen Ansatz zur simultanen Bestimmung der Spaltspezifität einer unbekannten Protease und der Quantifizierung von proteolytischen Prozessen zu entwickeln, wurde die Mascot Datenbanksuche zusätzlich unter der Einstellung ohne verwendetes Enzym durchgeführt. Auch hier wurde der Datensatz des zytosolischen Hefeproteoms verwendet, der unter Einsatz der optimalen Parametereinstellungen generiert wurde. Um auch hier eine falsch-positiv Rate <4% zu erreichen, musste der *ion score cutoff* für den *identity threshold* von 38 auf 41 erhöht werden. Dadurch erhöhte sich auch der durchschnittliche Mascot *ion score* von 61 auf 64. Von den ursprünglich 2.814 TMT-markierten Peptiden, die unter der Einstellung semi Trypsin und semi GluC als verwendetes Enzym für die Datenbanksuche identifiziert wurden, wurden unter der Einstellung kein verwendetes Enzym 2.584 TMT-markierte Peptide identifiziert. Dies stellt eine Abnahme zugeordneter Fragmentspektren um lediglich 8% dar. Bezüglich der unterschiedlichen *unique* Spaltsequenzen, die in mindestens zwei der drei technischen Replikate identifiziert wurden, wurden von den ursprünglich 737 *unique* Spaltsequenzen noch 664 *unique* Spaltsequenzen identifiziert, was einer Abnahme von 10% entspricht. Betrachtet

man die Ergebnisse bezüglich der Quantifizierung (siehe Abbildung 3.8), so wird deutlich, dass hier kaum ein Unterschied zu der vorherigen Datenbanksuche festzustellen ist.

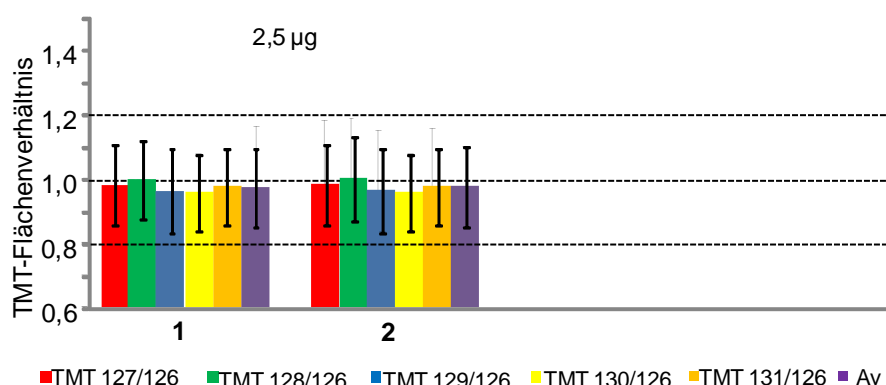


Abb. 3.8: Geometrische Mittelwerte der TMT-Flächenverhältnisse (127/126 bis 131/126) und Standardabweichungen der MALDI-TOF/TOF MS/MS Parameteroptimierung mittels der tryptischen Peptidbibliothek des zytosolischen Hefeproteoms. Alle Werte sind über die Summe von drei technischen Replikaten für die beiden unterschiedlichen Enzym-Einstellungen (1 und 2) der Datenbankanalyse gemittelt. 1) semi Trypsin und semi GluC 2) kein Enzym. Av = *average* (durchschnittlicher geometrischer Mittelwert über alle fünf TMT-Flächenverhältnisse).

So errechnet sich der durchschnittliche geometrische Mittelwert über alle fünf TMT-Flächenverhältnisse nun auf 0,981 mit einer durchschnittlichen Standardabweichung <13% im Vergleich zu 0,979 mit einer durchschnittlichen Standardabweichung <13%.

Auch hier wurde die Spaltspezifität der Protease GluC unter Verwendung der leicht modifizierten Version der PICS Datenanalyse-Pipeline untersucht und eine *Heat-Map* (Abbildung 3.9) erstellt. Insgesamt wurden 61,0% (vorher 59,1%) der identifizierten Spaltsequenzen nach einem C-terminalen Glutamat und 34,3% (vorher 34,9%) nach einem C-terminalen Aspartat in Position P1 generiert. Damit wurde ein Wert von 95,3% (vorher 94,0%) erreicht, der der Spaltspezifität von GluC entspricht. Spaltstellen, in denen andere Aminosäuren in Position P1 enthalten waren, waren zu 3,5% (vorher 4,4%) beobachtbar. Die übrigen 1,2% (vorher 1,6%) verteilen sich auf *ambiguous* Spaltsequenzen.

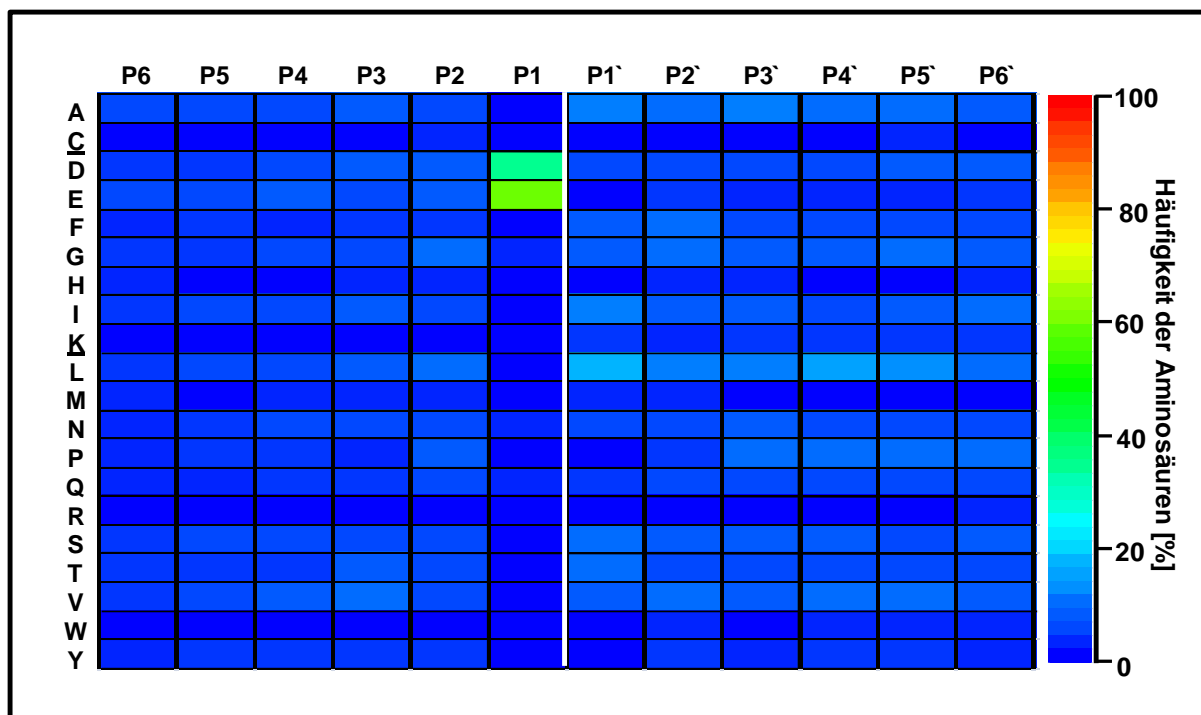


Abb. 3.9: Heat-Map zur Darstellung der GluC Spaltstellen Spezifität unter Verwendung des Experimentes der Peptidbibliothek des zytosolischen Hefeproteoms mit den optimalen Parametereinstellungen (3000 akkumulierte Laserschüsse, CIDhoch = 4×10^{-6} Torr und Reihenfolge der *precursor*-Auswahl "strongest precursor first"). Für die Datenbanksuche wurde kein Enzym als Enzymeinstellung gewählt. Die Peptidbindung zwischen Position P1 und P1' ist durch einen weißen Strich abgetrennt. C = carbamidomethyliertes Cystein, K = dimethyliertes Lysin.

3.4 Diskussion

Wie zuvor bereits beschrieben basiert die generelle Vorgehensweise des Q-PICS Ansatzes (siehe Abbildung 3.1) zur Bestimmung der Spaltspezifität von unbekanntem Proteasen, auf dem von Schilling *et al.* (2008) entwickelten qualitativen PICS Ansatz (siehe Abbildung 1.18). Dieser wurde für den Q-PICS Ansatz jedoch um den quantitativen Aspekt erweitert. Dieser quantitative Aspekt ermöglicht es bis zu sechs internen technischen Replikaten gleichzeitig zu analysieren oder die gleichzeitige Identifizierung von Spaltsequenzen in verschiedenen biologischen Zuständen. Beide Ansätze verwenden als Grundlage Proteombasierte Peptidbibliotheken, die durch den Verdau eines Vollproteoms mit einer (bekanntem) *work*-Protease wie z. B. Trypsin generiert werden. Die Wahl dieser *work*-Protease ist insofern entscheidend, als dass sie keine Spaltspezifität besitzen darf, die der der eigentlichen *target*-Protease entspricht. Hierbei müssen Überlappungen also vermieden werden. So wurden in

dieser Arbeit als *work*-Proteasen Trypsin und Chymotrypsin verwendet, die keinerlei Überlappungen mit der *target*-Protease GluC bezüglich ihrer Spaltstellenspezifität aufweisen. Während im klassischen PICS Ansatz das Proteom humaner Zelllinien verwendet wurde, kommen in dem in dieser Arbeit entwickelten Q-PICS Ansatz Peptidbibliotheken auf Basis eines 10-Proteinstandards bzw. des zytosolischen Proteoms der Bäckerhefe *S. cerevisiae* zum Einsatz. Dabei handelt es sich um weniger komplexe Peptidbibliotheken.

Auch die Probenvorbereitung zur Herstellung der Peptidbibliotheken ist im Q-PICS Ansatz leicht modifiziert. Während beim PICS-Ansatz der Verdau der intakten Proteine mit der *work*-Protease direkt nach der Hitzedenaturierung erfolgt, wird im Q-PICS Ansatz durch die Zugabe von 8 M Harnstoff denaturiert, anschließend die Disulfidbrücken reduziert und die freien Cysteinseitenketten alkyliert. Diese beiden Schritte erfolgen beim PICS Ansatz erst nach dem Verdau. Der Vorteil der Vorgehensweise des Q-PICS Ansatzes, in dem der Verdau mittels *work*-Protease nach der Reduktion der Disulfidbrücken und Alkylierung der Cysteinseitenketten erfolgt, besteht darin, dass die Reagenzien der Reduktion und Alkylierung vor dem Verdau über Amicon Ultra 4 3K Zentrifugenfilter mit einem *cutoff* von 3000 Da entfernt werden können. Darüber hinaus wird durch die Denaturierung der Proteine mittels Harnstoff die Zugänglichkeit der Proteasen zum Ort der Hydrolyse der Peptidbindungen verbessert, wodurch eine größere Anzahl an Spaltpeptiden resultiert (Riviere L. R. and Tempst, 1995).

Durch die Umpufferung der Proben auf 200 mM HEPES pH 7,4 wird außerdem der Proteaseinhibitor entfernt, der dem Lysepuffer beigegeben wurde. Innerhalb des klassischen PICS Ansatz wird hingegen kein Proteaseinhibitor verwendet. Die Blockierung der durch den Verdau mittels *work*-Protease gebildeten *neo*-N-Termini und der Lysinseitenketten mittels reduktiver Dimethylierung, wird innerhalb beider Ansätze gleichermaßen durchgeführt. Für den klassischen PICS-Ansatz müssen sowohl die überschüssigen Reagenzien für die Reduktion und Alkylierung der Cysteinseitenketten als auch diejenigen der reduktiven Dimethylierung über eine umfassende Aufreinigung entfernt werden. Hierzu wird zunächst eine präparative Größenausschluss-Chromatographie mit einem *cutoff* von 700 Da, gefolgt von einer Kationenaustausch- und einer Umkehrphasen-Chromatographie durchgeführt. Für den Q-PICS Ansatz müssen zu diesem Zeitpunkt der Peptidbibliothekherstellung nur die Reagenzien der reduktiven Dimethylierung über eine Aufreinigung mit 60 mg Oasis SPE-Kartuschen entfernt werden. Die überschüssigen Reagenzien zur Derivatisierung der Cysteinseitenketten wurden bereits zuvor durch die Umpufferung abgetrennt.

Der eigentliche Protease Assay wird zunächst in beiden Ansätzen gleichermaßen durchgeführt. So erfolgt als erstes der Verdau der (verschiedenen) Peptidbibliotheken mittels der zu analysierenden *target*-Protease. Beim PICS Ansatz werden die dabei entstandenen neo-N-Termini mit Biotin markiert und in einer anschließenden Streptavidin-Anreicherung aufgereinigt bzw. angereichert. Die Identifizierung der biotinylierten Peptide erfolgt dann über eine LC-ESI MS/MS Analyse mit anschließender Mascot-Datenbanksuche, während die *nonprime* Spaltsequenzen bioinformatisch durch die PICS Datenanalyse-Pipeline bestimmt werden (Schilling and Overall, 2008).

Im Gegensatz dazu werden im Q-PICS Ansatz die neo-N-Termini mit TMT statt mit Biotin markiert. Wie in Kapitel 1.2 bereits erläutert handelt es sich dabei um ein isobares *tag*, das aus einer Reporter-, einer Balance- und einer Amin-reaktiven-Gruppe besteht (siehe Abbildung 1.4). Auf diese Weise wurden, unter Ausnutzung der sechs verschiedenen TMT Reporterionenkanäle (126, 127, 128, 129, 130 und 131) insgesamt jeweils sechs interne technische Replikate verwendet. Da keine weitere Aufreinigung oder Anreicherung des vereinigten Probengemisches durchgeführt wurde, wurden anschließend sowohl die TMT-markierten N-terminalen *nonprime* Spaltsequenzen als auch die C-terminalen *prime* Spaltsequenzen mittels LC-MS/MS analysiert. Dadurch wird die Komplexität der Probe, verglichen mit dem klassischen PICS Ansatz erhöht. In diesem werden im Idealfall nur die angereicherten Biotin-markierten N-Termini mittels LC-MS/MS analysiert. Um der höheren Komplexität entgegenzuwirken, wurde für den Q-PICS Ansatz das im Vergleich zum humanen Proteom weniger komplexe zytosolische Hefeproteom verwendet.

Zusätzlich wurde an Stelle einer LC-ESI eine LC-MALDI-TOF/TOF MS/MS Analyse durchgeführt. Ein großer Vorteil der LC-MALDI-TOF/TOF MS/MS Analyse für den Q-PICS Ansatz, ist die Ausbildung überwiegend einfach geladener Ionen (Karas, *et al.*, 2000). In ESI-basierten Analysen können während der Messung auf MS-Ebene Peptide als verschieden geladene Spezies vorliegen (dies gilt auch für die Fragmentionen auf MS/MS-Ebene). Dies führt zum einen zu einem Verdünnungseffekt der einzelnen Analyten und zum anderen werden zusammen mit den verschieden geladenen Spezies co-eluierender Peptide die MS Spektren durch die zunehmende Anzahl an Signalen immer komplexer.

Prinzipiell wurden sowohl die TMT-markierten N-terminalen *nonprime* Spaltsequenzen als auch die C-terminalen *prime* Spaltsequenzen mittels LC-MALDI-TOF/TOF MS/MS analysiert. Wie im klassischen PICS Ansatz wurden dann die C-terminalen Spaltsequenzen über eine Mascot Datenbanksuche identifiziert und die N-terminalen Spaltsequenzen, die in

Abbildung 3.1, B gezeigt sind, mittels einer leicht modifizierten Datenanalyse-Pipeline des PICS-Ansatzes bestimmt (Schilling and Overall, 2008).

Die Vor- und Nachteile der Quantifizierung mittels isobarer *tags* wurden in ESI-basierten MS/MS-Analysen, wie zuvor beschrieben, für die unterschiedlichsten Instrumententypen bereits untersucht. Hierbei scheint sich v. a. das LTQ-Orbitrap Massenspektrometer durchzusetzen (Bantscheff, *et al.*, 2008, Dayon *et al.*, 2010a, Griffin, *et al.*, 2007, Kocher, *et al.*, 2009, Pichler, *et al.*, 2011, Zhang *et al.*, 2009).

Dennoch wurde für den Q-PICS Ansatz LC-MALDI-TOF/TOF MS/MS eingesetzt, da diese Instrumentierung die beschriebenen Vorteile, wie einfach geladene Ionen, besitzt.

3.4.1 MALDI-TOF/TOF MS/MS Parameteroptimierung mittels der Peptidbibliotheken auf Basis des 10-Proteinstandards

Um den Einfluss unterschiedlicher Probenkonzentrationen auf die Parameteroptimierung untersuchen zu können, wurden von zwei verschiedenen Peptidbibliotheken 250 fmol bzw. 2,5 pmol an verdauter Gesamtproteinmenge auf dem nanoHPLC-System aufgetrennt. Die *precursor*-Auswahl für die MS/MS-Analyse erfolgte nicht messdatenabhängig, sondern über den Einsatz einer sogenannten *inclusion list*. Eine messdatenabhängige *precursor*-Auswahl würde hingegen durch Einstellungen bestimmter Parameter durch die Software des Massenspektrometers durchgeführt werden. Hierbei werden nach der Analyse der Spots auf MS-Ebene, abhängig von der eingestellten Zahl an maximalen MS/MS-Messungen pro Spot, alle *precursor* ausgewählt, die oberhalb eines zuvor eingestellten S/N-Verhältnisses liegen. Die Software errechnet dann für die ausgewählten *precursor* in welchem Spot sie die höchste Intensität aufweisen. In diesen Spots erfolgt dann auch die MS/MS-Analyse. Im Gegensatz dazu werden unter Verwendung einer *inclusion list* die *m/z*-Verhältnisse bereits bekannter Analyten auf diese Liste eingetragen. Die Software wählt nach der Analyse auf MS-Ebene nur solche *precursor* aus, deren *m/z*-Verhältnisse in der *inclusion list* eingetragen sind. Auch hier erfolgt die MS/MS-Analyse in denjenigen Spots, in denen die ausgewählten *precursor* mit der höchsten Intensität vorliegen. Die Verwendung dieser *inclusion lists* ermöglicht den Einfluss der Parameter **i)** Anzahl der akkumulierten Laserschüsse zur Generierung der MS/MS-Spektren und **ii)** Gasdruck in der CID-Kollisionszelle unter sonst möglichst konstanten Bedingungen zu untersuchen.

Insgesamt wurden vier technische Replikate analysiert. Die vier unterschiedlichen Parameterkombinationen wurden dann in alternierender Reihenfolge mit den einzelnen technischen Replikaten durchgeführt (siehe Tabelle 3.4). Dies entspricht einer Gesamtanzahl

von 16 LC-MALDI-TOF/TOF MS/MS-Analysen für jede Probe. Wie in Kapitel 1.4 bereits beschrieben, kann es zur so genannten Ionensuppression kommen, wenn zu viele Probenmoleküle in der Gasphase vorliegen. Die Qualität eines MS/MS-Spektrums hängt also auch von der Spotzusammensetzung ab. Die *precursor*-Auswahl erfolgt nach der MS-Analyse für die vier unterschiedlichen Parameterkombinationen des gleichen technischen Replikats gleichermaßen. Dadurch werden die einzelnen Peptide immer in dem gleichen Spot analysiert. Eine Spot-zu-Spot-Variabilität entfällt somit innerhalb eines technischen Replikats. Damit ist die Spotzusammensetzung innerhalb eines technischen Replikats für die jeweiligen Peptide gleich, so dass auftretende Ionensuppressionseffekte ebenfalls gleich bleiben.

Maximal wurden 25 MS/MS-Messungen pro Spot durchgeführt. Abhängig von der Komplexität der Probe wird dieses Maximum auf den unterschiedlichen Spots erreicht. Je mehr MS/MS-Messungen in einem Spot erfolgen, desto mehr Laserschüsse werden auf diesen abgegeben. Mit zunehmender Anzahl erfolgter Laserschüsse wird erhöht sich der Spotabtrag. Die Qualität der MS/MS-Spektren hängt auch von der Konzentration der Analyten ab. Es spielt also eine Rolle, wie viele Laserschüsse bereits in einem Spot generiert wurden, also wie groß der bereits erfolgte Spotabtrag ist, bevor ein bestimmter *precursor* in diesem Spot fragmentiert wird. Durch Verwendung der *inclusion list* werden nur wenige *precursor* pro Spot analysiert, so dass der Spotabtrag nur eine geringe Rolle spielt. Um dies wirklich sicherzustellen, wurde zusätzlich die zuvor bereits erwähnte, alternierende Reihenfolge der vier unterschiedlichen Parameterkombinationen mit den einzelnen technischen Replikaten durchgeführt.

Zusammenfassend kann also davon ausgegangen werden, dass in den Experimenten mit den Peptidbibliotheken auf Basis des 10-Proteinstandards durch die Verwendung der *inclusion lists* die Spot-zu-Spot-Variabilität, die Spotzusammensetzung und der Spotabtrag vernachlässigbar sind.

A) Tryptische Peptidbibliothek des 10-Proteinstandards

Sowohl bei der Verwendung von 250 fmol als auch von 2,5 pmol an verdauter Gesamtproteinmenge für die LC-Auftrennung wurden 47 TMT-markierte Peptide identifiziert. Damit wurden nicht alle 50 Peptide der *inclusion list* zugeordnet. In den zuvor durchgeführten Pilotexperimenten wurden ebenfalls vier technische Replikate analysiert. Da auch solche TMT-markierten Peptide auf der *inclusion list* hinzugefügt wurden, die nur in einem technischen Replikat der Pilotexperimente identifiziert wurden, ist es nicht verwunderlich, dass nicht alle 50 TMT-markierten Peptide identifiziert wurden.

Betrachtet man die Ergebnisse der tryptischen Peptidbibliothek des 10-Proteinstandards bezüglich der Identifizierung der TMT-markierten Peptide, so wurde sich hier auf den Mascot *ion score* bezogen. Dieser stellt das Maß der Wahrscheinlichkeit dar, ob es sich bei einer Peptidzuordnung um eine korrekte oder zufällige Identifizierung handelt. Dabei wurde angenommen: je höher der *ion score*, also die Wahrscheinlichkeit einer korrekten Zuordnung, desto höher ist auch die Qualität des zugrunde liegenden MS/MS-Spektrums. Es zeigte sich jedoch, dass weder die Anzahl an akkumulierten Laserschüssen, noch der Gasdruck in der Kollisionszelle einen starken Einfluss auf den Mascot *ion score* haben. Vergleicht man die Werte für die vier unterschiedlichen Parameterkombinationen und die beiden verschiedenen Mengen analysierter Probe, so ist kein Trend erkennbar, unter welchen Parameterbedingungen die besten Identifizierungen erhalten wurden.

In Abbildung 3.10 sind die MS/MS-Spektren des Peptides ACFAVEGPK aus dem Protein BSA dargestellt, die unter den vier verschiedenen Parameterkombinationen generiert wurden.

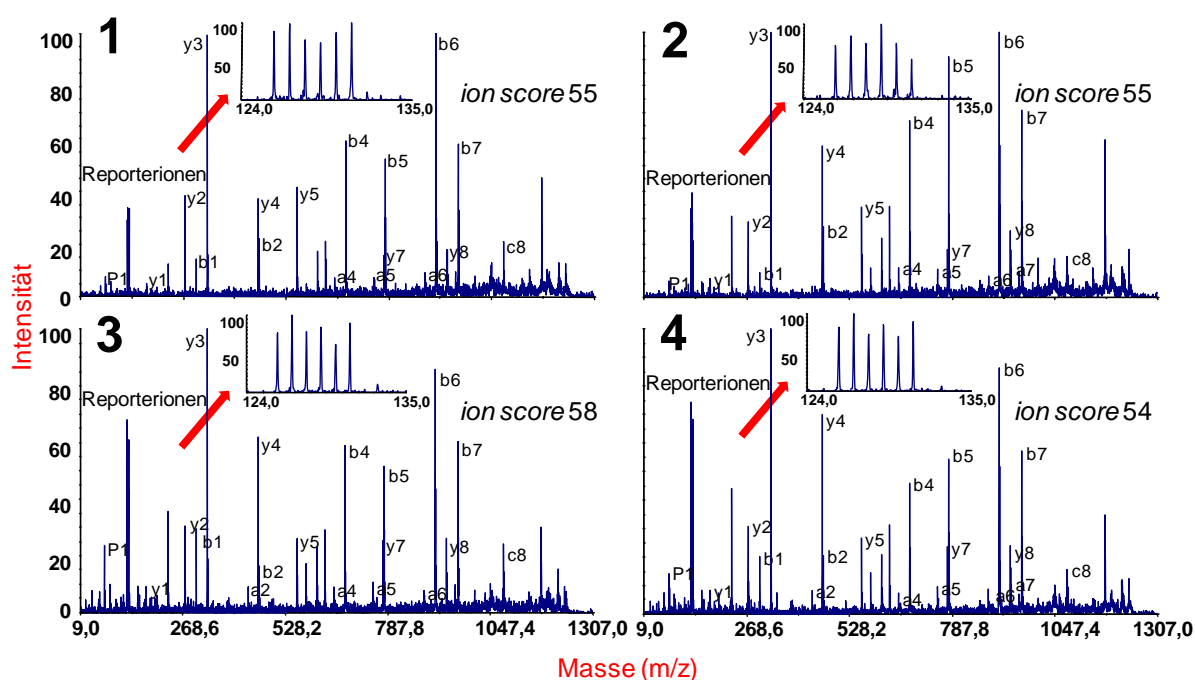


Abb. 3.10: MS/MS-Spektren des Peptides ACFAVEGPK ($m/z = 1235,65$) aus dem Protein BSA der tryptischen Peptidbibliothek auf Basis des 10-Proteinstandards. Die Spektren wurden mit einem MALDI-TOF/TOF Massenspektrometer im *positive ion mode* generiert. Insgesamt wurden vier verschiedene Parameterkombinationen analysiert, um den Einfluss der Parameter **i**) Anzahl der akkumulierten Laserschüsse zur Generierung der MS/MS-Spektren **ii**) Gasdruck in der CID-Kollisionszelle auf die Identifizierung und Quantifizierung TMT-markierter Peptide zu untersuchen: 1) 3000 Laserschüsse, CIDniedrig; 2) 5000 Laserschüsse, CIDniedrig, 3) 3000 Laserschüsse, CIDhoch und 4)

5000 Laserschüsse, CIDhoch. Der m/z -Bereich, der die Signale der Reporterionen enthält, ist jeweils in Vergrößerung dargestellt und mit einem roten Pfeil markiert. Zusätzlich sind die Mascot *ion scores* abgebildet.

Der Unterschied des Mascot *ion scores* beträgt für dieses Peptid bis zu 7% und liegt damit etwas höher als die Schwankung des durchschnittlichen Mascot *ion scores*. Dieser lag über alle Peptide gemittelt bei einem Wert $<5\%$ für die mittels LC getrennte Menge an verdautem Gesamtprotein von 250 fmol bzw. 6% für 2,5 pmol. Betrachtet man sich die b- und y-Ionenserien, so sieht man, dass weder unter Verwendung unterschiedlicher Schusszahlen, noch verschiedener Gasdrücke in der CID-Kollisionszelle stark unterschiedliche Signalmuster beobachtbar sind. Es kommt also zu keiner unterschiedlichen Ausbildung an Immoniumionen, Seitenfragmenten oder b- bzw. y-Ionen.

Um den Einfluss der verschiedenen MALDI-TOF/TOF MS/MS Parameterkombinationen auf die Quantifizierung TMT-markierter Peptide beurteilen zu können, wurden alle TMT-Reporterionenflächen mittels der Mascot Software durch die Flächenwerte des Kanals 126 geteilt, wodurch insgesamt fünf verschiedene TMT-Flächenverhältnisse (127/126 bis 131/126) errechnet wurden.

Die Anzahl an negativen Werten der TMT-Flächenverhältnisse, die für die weitere Datenanalyse entfernt wurden, ist für beide verwendete Mengen an verdautem Gesamtprotein vernachlässigbar. Solche negativen Werte der TMT-Flächenverhältnisse entstehen durch die Berechnung der Isotopenkorrektur durch die Mascot Software. Sind Reporterionensignale entweder nicht detektierbar oder von solch geringer Intensität, dass sie während der Erstellung der mgf-Datei nicht erfasst werden, kann eine Isotopenkorrektur nicht mehr durchgeführt werden, da einzelne Isotope fehlen. Somit kann das Isotopenmuster nicht mehr korrekt berechnet werden und man erhält negative TMT-Flächenverhältnisse. Diese ergeben biologisch keinerlei Sinn.

Die Abweichung des geometrischen Mittelwertes vom Erwartungswert 1,0 beträgt $<3\%$ für alle Parameterkombinationen der beiden unterschiedlichen Mengen an verdautem Gesamtprotein (siehe Tabelle 3.7). Dies sowie die Tatsache, dass auch die durchschnittlichen Standardabweichungen $<20\%$ betragen, steht in guter Übereinkunft mit anderen iTRAQ- (Gluckmann *et al.*, 2007, Ogata *et al.*, 2007, Ross, *et al.*, 2004) oder TMT-Quantifizierungsstudien (Dayon, *et al.*, 2008, Tiberti, *et al.*, 2010).

Die durchschnittliche Standardabweichung, gemittelt über alle vier verschiedenen Parameterkombinationen und die fünf TMT-Flächenverhältnisse der einzelnen Reporterionenkanäle verändert sich in Abhängigkeit der verwendeten Menge an verdautem

Gesamtprotein. Hu *et al.* konnten in einer Studie unter Verwendung eines 4700 MALDI TOF/TOF Massenspektrometers (AB SCIEX, Darmstadt, Deutschland) für das iTRAQ-Reagenz zeigen, dass die analytische Varianz eine Funktion der Signalintensität der Reporterionensignale darstellt (Hu, *et al.*, 2006). Hieraus wurden verschiedene Fehlermodelle für den Umgang mit quantitativen Daten isobarer *tags* erstellt (Karp *et al.*, 2010, Zhang *et al.*, 2010). Die Varianz der gemessenen Intensitäten der Reporterionen (und damit verbunden der Flächen der Reporterionen) sinkt, wenn diese Intensitäten (Flächen) ansteigen. Da eine höhere Konzentration eines Peptides prinzipiell auch zu einem höheren Signal führen sollte, stellt dies eine Erklärung für die geringeren durchschnittlichen Standardabweichungen unter Verwendung der höheren Menge an verdautem Gesamtprotein dar.

Dadurch sind die unterschiedlichen Einflüsse der Parameter **i**) der Anzahl an akkumulierten Laserschüssen und **ii**) dem Druck in der CID-Kollisionszelle unter Verwendung der höheren Menge an verdautem Gesamtprotein weniger stark ausgeprägt, da die Quantifizierung in diesem Fall insgesamt viel robuster zu sein scheint. Zusammenfassend kann man sagen, dass durch den Einsatz des höheren Gasdrucks in der Kollisionskammer, eine geringere durchschnittliche Standardabweichung erreicht wird, als durch den niedrigeren Gasdruck. Vergleicht man die Intensitäten der Reporterionen relativ zu den Signalen der b- und y-Fragmentationen in den MS/MS Spektren in Abbildung 3.10, so erkennt man, dass diese für den höheren Gasdruck im Vergleich zum niedrigeren Gasdruck angestiegen sind. Hier spielt demnach die geringere Varianz der insgesamt höheren gemessenen Intensitäten der Reporterionen eine Rolle (Hu, *et al.*, 2006). Dieser Effekt ist unabhängig von Parameter **ii**) der Anzahl an akkumulierten Laserschüssen pro MS/MS-Spektrum. Der Einfluss dieses Parameters auf die Qualität der Quantifizierung TMT-markierter Peptide ist nicht eindeutig feststellbar. Unter Verwendung von 250 fmol an verdautem Gesamtprotein werden unter der Akkumulation von 5000 Laserschüssen geringere durchschnittliche Standardabweichungen gemessen, als unter der Akkumulation von 3000 Laserschüssen. Unter Verwendung von 2,5 pmol an verdautem Gesamtprotein ist dieser Effekt umgekehrt, wenn auch weniger stark ausgeprägt. Eine eindeutige Aussage, ob die Akkumulation von 5000 Laserschüssen zu geringeren durchschnittlichen Standardabweichungen führt als die Akkumulation von 3000 Laserschüssen, konnte anhand dieser Daten nicht getroffen werden.

B) Chymotryptische Peptidbibliothek des 10-Proteinstandards

Um eine möglichst große Vielfalt an unterschiedlichen Peptiden der Peptidbibliotheken zu erhalten, wurde neben der tryptisch verdauten, auch eine chymotryptisch verdaute

Peptidbibliothek hergestellt. Man erhält z. B. mit der Protease Trypsin Peptide, die die basischen Aminosäuren Lysin oder Arginin am C-Terminus tragen. Damit sind basische Aminosäuren innerhalb der Peptidsequenzen dieser Peptidbibliothek unterrepräsentiert. Durch chymotryptischen Verdau hingegen (spaltet nach großen hydrophoben Resten wie Tryptophan, Tyrosin, Leucin oder Phenylalanin in Position P1) erhält man sowohl basische als auch saure Aminosäuren innerhalb der Sequenzen, wobei weniger hydrophobe Aminosäuren enthalten sind. Für MMP2, einer wichtigen Protease in der Tumormetastasierung und Angiogenese (Overall and Kleifeld, 2006), wurde eine höhere Abdeckung an Spaltstellenidentifizierungen durch die Verwendung von Trypsin, Chymotrypsin oder GluC als *work*-Protease zur Herstellung der Peptidbibliotheken erreicht (Schilling and Overall, 2008).

Auch für die Experimente der chymotryptischen Peptidbibliothek wurden 50 TMT-markierte Peptide auf eine *inclusion list* eingetragen. Aus diesen TMT-markierten Peptiden wurden 44 *unique* (einzigartige) Spaltsequenzen der Endoprotease GluC identifiziert. Unter Verwendung der tryptischen Peptidbibliothek wurden aus den 50 TMT-markierten Peptiden 47 *unique* Spaltsequenzen identifiziert, wobei insgesamt 12 dieser *unique* Spaltsequenzen in beiden Peptidbibliotheken gefunden wurden (siehe Abbildung 3.11). Die Verwendung unterschiedlicher Peptidbibliotheken führt also auch hier zu einer höheren Abdeckung an Spaltstellenidentifizierungen.

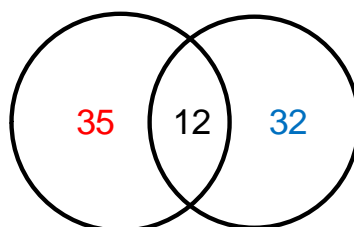


Abb. 3.11: Venn-Diagramm der Anzahl an identifizierten *unique* Spaltsequenzen der Endoprotease GluC in der tryptischen (rot) bzw. der chymotryptischen (blau) Peptidbibliothek auf Basis des 10-Proteinstandards.

In den Experimenten mit der chymotryptischen Peptidbibliothek des 10-Proteinstandards wurden geringfügig mehr TMT-markierte Peptide identifiziert, als in den Experimenten der tryptischen Peptidbibliothek (49 unter der Verwendung von 250 fmol und 50 unter Verwendung von 2,5 pmol an verdauter Gesamtproteinmenge)

Picotti *et al.* konnten in einer Studie (2007) zeigen, dass die meisten Algorithmen von Datenbanksuchen zur Berechnung der so genannten *scores* dazu tendieren längere Peptidsequenzen mit typisch tryptischen Enden zu favorisieren. Die Anzahl an Peptiden mit untypischen tryptischen Enden, die möglicherweise zu geringe *scores* erhalten, ist in der

chymotryptischen Peptidbibliothek höher, als in der tryptischen Peptidbibliothek. Dies stellt eine Erklärung dar, warum die durchschnittlichen Mascot *ion scores* der Experimente mit der tryptischen niedriger waren (siehe Abbildung 3.2), als die der chymotryptischen Peptidbibliothek (siehe Abbildung 3.4).

Die Ergebnisse der tryptischen Peptidbibliothek bezüglich der Identifizierung TMT-markierter Peptide können durch diejenigen der chymotryptischen Peptidbibliothek bestätigt werden. Vergleicht man die Werte für beide Mengen an verdautem Gesamtprotein, so ist auch hier kein Trend erkennbar, welche der vier Kombinationen der Parameter **i**) der Anzahl an akkumulierten Laserschüssen und **ii**) dem Druck in der CID-Kollisionszelle, zu den besten Identifizierungen führt. In Abbildung 3.12 sind die MS/MS-Spektren des Peptides ADIAGHGQE aus dem Protein Myoglobin dargestellt, die unter den vier verschiedenen Parameterkombinationen generiert wurden.

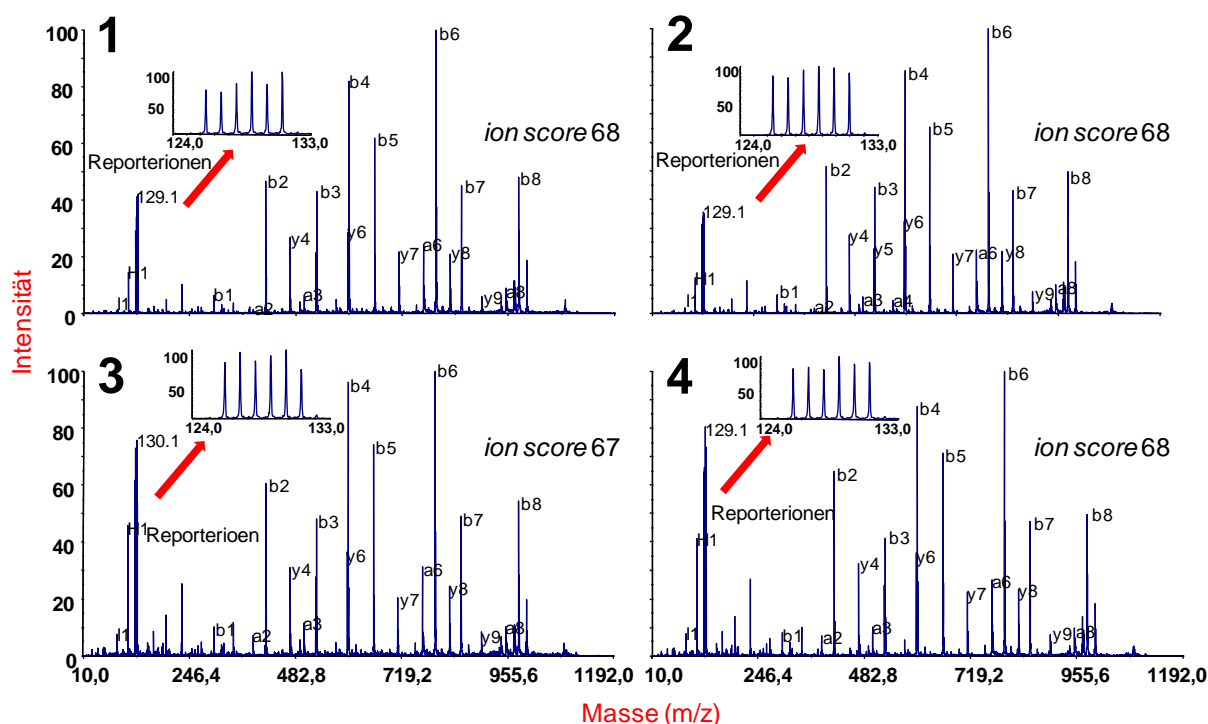


Abb. 3.12: MS/MS-Spektren des Peptides ADIAGHGQE ($m/z = 1126,55$) aus dem Protein Myoglobin der chymotryptischen Peptidbibliothek auf Basis des 10-Proteinstandards. Die Spektren wurden mit einem MALDI-TOF/TOF Massenspektrometer im *positive ion mode* generiert. Insgesamt wurden vier verschiedene Parameterkombinationen analysiert, um den Einfluss der Parameter **i**) Anzahl der akkumulierten Laserschüsse zur Generierung der MS/MS-Spektren **ii**) Gasdruck in der CID-Kollisionszelle auf die Identifizierung und Quantifizierung TMT-markierter Peptide zu untersuchen: 1) 3000 Laserschüsse, CIDniedrig; 2) 5000 Laserschüsse, CIDniedrig, 3) 3000 Laserschüsse, CIDhoch und 4) 5000 Laserschüsse, CIDhoch. Der m/z -Bereich, der die Signale der

Reporterionen enthält, ist jeweils in Vergrößerung dargestellt und mit einem roten Pfeil markiert. Zusätzlich sind die Mascot *ion scores* abgebildet.

Bei diesem Peptid ist praktisch kein Unterschied des Mascot *ion scores* zu beobachten. Betrachtet man wie zuvor die b- und y-Ionenserien, so sieht man, dass weder unter Verwendung der unterschiedlichen Anzahl an Laserschüssen, noch der verschiedenen Gasdrücke in der CID-Kollisionszelle, abweichende Zuordnungen beobachtbar sind. Lediglich die Intensitäten der Signale unterscheiden sich und führen somit zu minimalen Varianzen.

Während die Anzahl an negativen Werten der TMT-Flächenverhältnisse für 2,5 pmol an verdaulichem Gesamtprotein wie in den Experimenten der tryptischen Peptidbibliothek vernachlässigbar ist, so ist sie für 250 fmol an verdaulichem Gesamtprotein fast dreimal so hoch, wie in diesen Experimenten. Solche negative Werte der TMT-Flächenverhältnisse sind prinzipiell auch ein Indiz für qualitativ niedrige MS/MS-Spektren, da sie einen Hinweis für niedrige Intensitäten und/oder ein schlechtes S/N-Verhältnis liefern könnten.

Die Abweichung der durchschnittlichen geometrischen Mittelwerte der TMT-Flächenverhältnisse vom Erwartungswert 1,0 stimmt mit den Ergebnissen der tryptischen Peptidbibliothek sehr gut überein (siehe Tabelle 3.7). Die durchschnittlichen Standardabweichungen der TMT-Flächenverhältnisse der verschiedenen Parameterkombinationen für die chymotryptische Peptidbibliothek sind hingegen höher (siehe Tabelle 3.8). So wurden für 250 fmol an verdauter Gesamtproteinmenge 6% - 9% höhere Werte für die durchschnittlichen Standardabweichungen beobachtet, die z. T. >20% lagen. Für 2,5 pmol an verdauter Gesamtproteinmenge lagen die Werte 5% - 7% über den Werten der tryptischen Peptidbibliothek. Prinzipiell stimmen aber die Werte der durchschnittlichen Standardabweichung der TMT-Flächenverhältnisse auch unter Verwendung der chymotryptischen Peptidbibliothek gut mit den anderen iTRAQ- (Gluckmann, *et al.*, 2007, Ogata, *et al.*, 2007, Ross, *et al.*, 2004) oder TMT-Quantifizierungsstudien (Dayon, *et al.*, 2008, Tiberti, *et al.*, 2010) überein.

In den Experimenten der chymotryptischen Peptidbibliothek ist ebenfalls die Abhängigkeit der durchschnittlichen Standardabweichung der TMT-Flächenverhältnisse von der verwendeten Menge an verdaulichem Gesamtprotein zu beobachten. Auch hier sind, in guter Übereinstimmung mit der Studie von Hu *et al.* (2006), die Standardabweichungen für die höhere Menge an verdaulichem Gesamtprotein niedriger als für die niedrigere Menge. Die Einflüsse der Parameter **i**) der Anzahl an akkumulierten Laserschüssen und **ii**) dem Druck in

der CID-Kollisionszelle sind ebenfalls unter Verwendung der höheren Menge an verdautem Gesamtprotein weniger stark ausgeprägt.

Die Ergebnisse der tryptischen Peptidbibliothek, die zeigten, dass durch den Einsatz des höheren Gasdrucks in der Kollisionskammer eine geringere durchschnittliche Standardabweichung der TMT-Flächenverhältnisse erreicht wird, als durch den niedrigeren Gasdruck, werden bestätigt. Vergleicht man die Intensitäten der Reporterionen relativ zu den Signalen der b- und y-Fragmentationen in den MS/MS-Spektren in Abbildung 3.12, so erkennt man auch für das Peptid ADIAGHGQE, dass diese für den höheren Gasdruck im Vergleich zum niedrigeren Gasdruck angestiegen sind. Dieser Effekt ist, wie zuvor, unabhängig von Parameter **ii**) der Anzahl an akkumulierten Laserschüssen zu beobachten.

Der Einfluss dieses Parameters auf die Qualität der Quantifizierung TMT-markierter Peptide ist unter Verwendung der chymotryptischen Peptidbibliothek eindeutiger, als unter Verwendung der tryptischen Peptidbibliothek. Für beide Mengen an verdautem Gesamtprotein werden nach der Akkumulation von 5000 Laserschüssen geringere durchschnittliche Standardabweichungen erreicht, als nach der Akkumulation von 3000 Laserschüssen. Insgesamt betrachtet scheint, die Akkumulation von 5000 Laserschüssen prinzipiell zu geringeren durchschnittlichen Standardabweichungen zu führen, als die Akkumulation von 3000 Laserschüssen, auch wenn Ausnahmen existieren.

3.4.2 MALDI-TOF/TOF MS/MS Parameteroptimierung mittels der Peptidbibliothek des zytosolischen Hefeproteoms

Die Experimente mit den beiden Peptidbibliotheken auf Basis des 10-Proteinstandards zeigten deutlich, dass die vier verschiedenen Kombinationen der Parameter **i**) Anzahl der akkumulierten Laserschüsse zur Generierung der MS/MS-Spektren und **ii**) Gasdruck in der CID-Kollisionszelle, so gut wie keinen Einfluss auf die Identifizierung TMT-markierter Peptide hat. Bezüglich der Quantifizierung wurde jedoch deutlich, dass die Verwendung des höheren Gasdrucks in der CID-Kollisionskammer zu einer Verringerung der durchschnittlichen Standardabweichung der TMT-Flächenverhältnisse führt. Parameter **ii**) wurde deshalb in den Experimenten mit der Peptidbibliothek auf Basis des zytosolischen Hefeproteoms der Bäckerhefe *S. cerevisiae* konstant auf CIDhoch eingestellt. Im Gegensatz dazu sollte der Einfluss der Anzahl der akkumulierten Laserschüsse weiterhin untersucht werden.

Zusätzlich wurde nun auch Parameter **iii**) die Reihenfolge der *precursor* Auswahl für die Fragmentierung in der MS/MS-Analyse (“*weakest precursor first*” vs. “*strongest precursor first*”) untersucht. Insgesamt wurden also vier verschiedene Parameterkombinationen (1) 3000 Laserschüsse, “*weakest precursor first*”; 2) 5000 Laserschüsse, “*weakest precursor first*”; 3) 3000 Laserschüsse, “*strongest precursor first*” und 4) 5000 Laserschüsse, “*strongest precursor first*”) analysiert.

Sowohl bei der Verwendung einer *inclusion list* als auch bei der datenabhängigen *precursor*-Auswahl, gibt es zwei verschiedene Vorgehensweisen, “*weakest precursor first*” und “*strongest precursor first*”. Benutzt man die “*weakest precursor first*” Bedingung, so wird von den in einem Spot ausgewählten *precursor* zuerst derjenige einer MS/MS-Analyse unterzogen, der die geringste Intensität in dem Spektrum auf MS-Ebene dieses Spots aufweist. Die übrigen *precursor* dieses Spots werden dann in der Reihenfolge nach zunehmender Intensität fragmentiert. Unter der “*strongest precursor first*” Bedingung, wird von den in einem Spot ausgewählten *precursor* zuerst derjenige auf MS/MS-Ebene fragmentiert, der die höchste Intensität in dem Spektrum auf MS-Ebene dieses Spots aufweist. Die übrigen *precursor* dieses Spots werden dann in der Reihenfolge nach abnehmender Intensität in der MS/MS-Analyse untersucht. Wie zuvor bereits beschrieben erfolgt mit jeder MS- oder MS/MS-Messung in einem Spot ein Spotabtrag, der in Abhängigkeit der Anzahl an generierten Laserschüssen in diesem Spot zunimmt.

Verwendet man die “*weakest precursor first*” Bedingung wird davon ausgegangen, dass von den niedrig abundanten *precursor* MS/MS-Spektren mit guter Qualität generiert werden können, bevor die höher abundanten *precursor* analysiert werden, da zu Beginn der MS/MS-Messungen in einem Spot erst ein geringer Spotabtrag erfolgt ist. Die höher abundanten *precursor* sollten auch mit zunehmendem Spotabtrag noch qualitativ gute MS/MS-Spektren liefern und damit ebenfalls identifiziert werden können. Der Spotabtrag ist jedoch auch davon abhängig, wie viele Laserschüsse für jede Messung eines MS/MS-Spektrums akkumuliert werden. Prinzipiell würde man also erwarten, dass man mehr Peptide mit der “*weakest precursor first*” Bedingung identifizieren kann, weil theoretisch mehr niedrig abundante *precursor* zusätzlich zugeordnet werden können.

Es besteht jedoch auch die Gefahr, dass der Spotabtrag nach Messung der niedrig abundanten *precursor* in einem Spot bereits so weit fortgeschritten ist, dass für die höher abundanten *precursor* keine qualitativ hochwertigen MS/MS-Spektren mehr generiert werden können und diese dadurch nicht zugeordnet werden. Aus diesem Grund wurde auch die “*strongest precursor first*” Bedingung untersucht.

In den Experimenten der beiden Peptidbibliotheken auf Basis des 10-Proteinstandards spielte die Reihenfolge, in der die *precursor* auf MS/MS-Ebene analysiert werden, praktisch keine Rolle, da durch die Verwendung der *inclusion lists* nur wenige *precursor* pro Spot gemessen wurden und somit der Spotabtrag vernachlässigbar war. Durch die höhere Komplexität der tryptischen Peptidbibliothek auf Basis des zytosolischen Hefeproteoms und der Kombination mit einer datenabhängigen *precursor*-Auswahl wurden die Effekte wie Spot-zu-Spot-Variabilität, die Spotzusammensetzung und der Spotabtrag in diesen Experimenten nun berücksichtigt. Die zusätzliche Analyse der chymotryptischen Peptidbibliothek auf Basis des 10-Proteinstandards hatte zwar eine höhere Abdeckung an identifizierten TMT-markierten Peptiden und damit GluC Spaltsequenzen zur Folge, führte jedoch auch zu einem geringeren durchschnittlichen Mascot *ion score* und höheren Werten der durchschnittlichen Standardabweichung der TMT-Flächenverhältnisse. Aus diesem Grund wurde in den Experimenten mit der Peptidbibliothek auf Basis des zytosolischen Hefeproteoms auf eine chymotryptische Peptidbibliothek verzichtet.

Auch in den Experimenten mit der tryptischen Peptidbibliothek auf Basis des zytosolischen Hefeproteoms haben die unterschiedlichen Parameterkombinationen praktisch keinen Einfluss auf den durchschnittlichen Mascot *ion score* (siehe Tabelle 3.9). In den Experimenten mit der tryptischen und chymotryptischen Peptidbibliothek auf Basis des 10-Proteinstandards war dieser 17% bzw. 34% niedriger. Eine Erklärung dafür ist die Einstellung des *ion score cutoff* für den *identity threshold*. Dieser wurde in diesen Experimenten nicht auf 38, sondern auf 13 festgesetzt. Damit wurden hier auch Spektren geringerer Qualität zur Auswertung hinzugezogen.

Dennoch haben die Parameter **i**) die Anzahl der akkumulierten Laserschüsse und **iii**) die Reihenfolge der *precursor*-Auswahl für die Fragmentierung einen Einfluss auf die Anzahl an Identifizierungen TMT-markierter Peptide. So wurden unter der Akkumulation von 3000 Laserschüssen mehr TMT-markierte Peptide identifiziert, als unter der Akkumulation von 5000 Laserschüssen, unabhängig von Parameter **iii**). Interessanterweise zeigten die gemessenen Daten den umgekehrten Effekt als den prinzipiell erwarteten. So wurden unter Verwendung der "*strongest precursor first*" mehr TMT-markierte Peptide identifiziert als unter Verwendung der "*weakest precursor first*" Methode, unabhängig von Parameter **i**). Dies ist wohl auf den höheren Spotabtrag zurückzuführen, der, wie zuvor bereits beschrieben, proportional zur, in einem Spot generierten Anzahl an Laserschüssen ist. Damit könnte der Fall eingetreten sein, dass unter Verwendung der "*weakest precursor first*" Methode der Spotabtrag nach Messung der niedrig abundanten *precursor* bereits so fortgeschritten war,

dass für die höher abundanten *precursor* keine qualitativ hochwertigen MS/MS-Spektren mehr generiert wurden. Diese wurden somit auch nicht mehr zugeordnet. Bezüglich der Identifizierung TMT-markierter Peptide stellt also die Parameterkombination 3000 akkumulierte Laserschüsse und “*strongest precursor first*” die beste Methode dar.

Wie in den Experimenten mit den beiden Peptidbibliotheken auf Basis des 10-Proteinstandards, wurden auch in den Experimenten mit der tryptischen Peptidbibliothek auf Basis des zytosolischen Hefeproteoms, die fünf verschiedenen TMT-Flächenverhältnisse (127/126 bis 131/126) mittels der Mascot Software bestimmt.

Vergleicht man die Anzahl an negativen Werten für die TMT-Flächenverhältnisse, so stellt man fest, dass in diesen Experimenten weniger negative Werte erhalten wurden als in den Experimenten mit den beiden Peptidbibliotheken auf Basis des 10-Proteinstandards. Dies könnte ebenfalls mit dem höheren *ion score cutoff* für den *identity threshold* von 38 im Vergleich zu 13 zusammenhängen. Prinzipiell kann man sagen, dass je höher der Mascot *ion score* ist, desto höher ist auch die Qualität des zugrunde liegenden MS/MS-Spektrums. Es scheint also ein direkter Zusammenhang zwischen der Qualität der MS/MS-Spektren und der Anzahl an negativen Werten für die TMT-Flächenverhältnisse zu bestehen.

Die Abweichung der durchschnittlichen geometrischen Mittelwerte der TMT-Flächenverhältnisse vom Erwartungswert 1,0 stimmt mit den Ergebnissen der tryptischen und chymotryptischen Peptidbibliothek auf Basis des 10-Proteinstandards gut überein (siehe Tabelle 3.7, Tabelle 3.8 und Tabelle 3.10). Vergleicht man den Einfluss der Parameter **i**) der Anzahl an akkumulierten Laserschüssen und **iii**) der Reihenfolge der *precursor*-Auswahl für die Fragmentierung auf die Quantifizierung TMT-markierter Peptide, so ist ein minimaler Trend zu erkennen. Dieser besagt, dass mit der höheren Anzahl an akkumulierten Laserschüssen niedrigere durchschnittliche Standardabweichungen der TMT-Flächenverhältnisse erhalten werden. Dieser Trend ist jedoch so wenig ausgeprägt, dass er praktisch vernachlässigbar ist.

Im Gegensatz dazu wird jedoch deutlich, dass die “*strongest precursor first*” Einstellung zu geringeren durchschnittlichen Standardabweichungen der TMT-Flächenverhältnisse führt. Bezüglich der Quantifizierung TMT-markierter Peptide kann also auch nach den Experimenten mit der tryptischen Peptidbibliothek auf Basis des zytosolischen Hefeproteoms noch keine eindeutige Aussage getroffen werden, ob die Akkumulation von 3000 oder 5000 Laserschüssen zu besseren Ergebnissen führt.

Ein weiterer Parameter, der v. a. für LC-MALDI MS/MS-Analysen eine wichtige Rolle spielt, ist der Zeitfaktor. So werden für solche Analysen immer wesentlich längere Messzeiten

benötigt als für vergleichbare LC-ESI MS/MS-Analysen. Für die zweidimensionale LC-MS/MS-Analyse des zytosolischen Proteoms des *Corynebacterium glutamicum* wurde beispielsweise 75% weniger Messzeit (124h vs. 492h) für die LC-ESI MS/MS-Analyse verglichen zur LC-MALDI MS/MS-Analyse benötigt (Lasaosa *et al.*, 2009). Die Messzeit einer LC-MALDI MS/MS-Analyse hängt stark von der Anzahl an akkumulierten Laserschüssen pro MS/MS-Spektrum ab. Betrachtet man die Ergebnisse der Experimente der tryptischen Peptidbibliothek auf Basis des zytosolischen Hefeproteoms, so zeigt es sich, dass man unter der Akkumulation von 3000 Laserschüssen 28% weniger Messzeit benötigt als unter der Akkumulation von 5000 Laserschüssen.

Zusammenfassend kann man also sagen, dass die Optimierung der Parameter **i)** Anzahl der akkumulierten Laserschüsse zur Generierung der MS/MS-Spektren (3000 Laserschüsse vs. 5000 Laserschüsse), **ii)** Gasdruck in der CID-Kollisionszelle (CIDniedrig = 1×10^{-6} Torr vs. CIDhoch = 4×10^{-6} Torr) und **iii)** Reihenfolge der *precursor*-Auswahl für die Fragmentierung in der MS/MS-Analyse (“*weakest precursor first*” vs. “*strongest precursor first*”) bezüglich der Identifizierung und Quantifizierung TMT-markierter Peptide zu folgendem Ergebnis geführt hat:

Die beste Kombination der untersuchten Parameter wurde erreicht bei einer Akkumulation von 3000 Laserschüssen unter dem Gasdruck in der CID-Kollisionszelle von CIDhoch = 4×10^{-6} Torr bei der *precursor*-Auswahl für die Fragmentierung “*strongest precursor first*”.

Diese Parameterkombination resultierte in den Experimenten mit der tryptischen Peptidbibliothek auf Basis des zytosolischen Hefeproteoms in der höchsten Anzahl an Identifizierungen (2.814 TMT-markierte Peptide) mit einer geringen durchschnittlichen Standardabweichung (13%) und geringerer Messzeit (ca. 33h).

3.4.3 Validierung des Q-PICS Ansatzes mittels Analyse der Spaltspezifität (*cleavage site specificity*) der Endoprotease GluC als bekannte Protease

Zur Validierung des Q-PICS Ansatzes wurde die Endoprotease GluC (*Staphylococcus aureus* Protease V8) als *target*-Protease verwendet, die Peptidbindungen speziell nach Glutamat oder weniger ausgeprägt nach Aspartat in Position P1 hydrolysiert (Sorensen, *et al.*, 1991). Die Auswertung des Datensatzes der tryptischen Peptidbibliothek auf Basis des zytosolischen Hefeproteoms unter Verwendung der optimierten Parametereinstellungen führte zur Identifizierung von 962 *unique* Spaltsequenzen. Da im Gegensatz zum klassischen PICS

Ansatz im Q-PICS Ansatz auch die quantitative Ebene analysiert werden kann, wurden jedoch nur solche Spaltstellen als identifiziert angesehen, die in mindestens zwei der drei technischen Replikate identifiziert wurden. Dadurch sollte pro Peptid eine möglichst große Anzahl an TMT-Flächenverhältnissen erhalten werden, um eine robuste statistische Analyse zu gewährleisten. Insgesamt wurden in diesen Experimenten sechs interne technische Replikate verwendet, so dass maximal 18 TMT-Flächenverhältnisse pro Peptid ausgewertet werden können.

Diese Bedingungen resultierten letztendlich in der Zuordnung von 737 *unique* Spaltsequenzen aus insgesamt 2,814 identifizierten TMT-markierten Peptiden. Diese stellen, verglichen mit den 453 Spaltsequenzen, die durch Schilling *et al.* in einer tryptischen Peptidbibliothek auf Basis des Proteoms von HEK-Zellen und ebenfalls GluC als *target*-Protease identifiziert wurden (2008), einen sehr guten Wert dar. In den von Schilling *et al.* durchgeführten Experimenten stimmten 94% - 96% der identifizierten Spaltstellen mit der vorhergesagten Spaltspezifität von GluC überein. Die Experimente der vorliegenden Arbeit erzielten ebenfalls eine Übereinstimmung von 94% (siehe Abbildung 3.7). Die Ergebnisse beider Studien stimmen insgesamt also sehr gut überein. Lediglich der Anteil an Spaltsequenzen nach einem C-terminalen Glutamat und nach einem C-terminalen Aspartat in Position P1 scheint in den beiden Studien leicht unterschiedlich zu sein. Dies ist wohl auf die unterschiedlich verwendeten Peptidbibliotheken zurückzuführen. Prinzipiell wird in beiden Studien jedoch die Spaltspezifität der Protease GluC, Peptidbindungen speziell nach Glutamat oder weniger ausgeprägt nach Aspartat in Position P1 zu hydrolysieren, sehr gut wiedergegeben.

Die Auswertung der drei übrigen Datensätze der MALDI-TOF/TOF MS/MS Parameteroptimierung mit der Peptidbibliothek des zytosolischen Hefeproteoms bestätigen diese Ergebnisse.

Auch die Mascot Datenbanksuche, die mit der Einstellung ohne verwendetes Enzym durchgeführt wurde, zeigt die vorhergesagte Spaltspezifität von GluC (siehe Abbildung 3.9). Unter dieser Einstellung wird bei der Mascot Datenbanksuche keine Spaltspezifität einer Protease vorgegeben. Stattdessen können nun alle Aminosäuren am C- oder N-terminalen Ende der Peptide enthalten sein. Durch den Verzicht auf Spaltspezifität werden die Möglichkeiten für Peptidzuordnungen weniger stark eingeschränkt. Je mehr variable Parameter für eine Mascot Datenbanksuche verwendet werden, desto höher muss i. d. R. auch der Mascot *ion score* sein, ab dem eine Zuordnung als korrekt zu bezeichnen ist (*Confidential Intervall* (C.I.) von 95%). Insgesamt wurden in dieser Datenbanksuche 664 *unique* Spaltsequenzen identifiziert. Dies stellt zwar eine Abnahme von 10% im Vergleich zur

Datenbanksuche unter Verwendung von semi Trypsin und semi GluC dar, ist aber immer noch sehr gut vergleichbar mit den 453 Spaltsequenzen, die durch Schilling *et al.* identifiziert wurden (2008). Bezüglich der Quantifizierung der TMT-markierten Peptide hat die veränderte Mascot Datenbanksuche praktisch keine Auswirkung und es werden fast identische Werte für den durchschnittlichen geometrischen Mittelwert über alle fünf TMT-Flächenverhältnisse und die durchschnittliche Standardabweichung erhalten (siehe Abbildung 3.8). Auch die Unterschiede der prozentualen Anteile der identifizierten Spaltsequenzen nach einem C-terminalen Glutamat bzw. nach einem C-terminalen Aspartat in Position P1 sind vernachlässigbar, so dass die Spaltspezifität von GluC sehr gut wiedergegeben wird.

Insgesamt kann die Anwendbarkeit des PICS-Ansatzes auf die Identifizierung hunderter *prime* und *nonprime* Spaltsequenzen und die daraus resultierende Bestimmung der Spaltspezifität von Proteasen also auch für den Q-PICS Ansatz bestätigt werden. Dieser bietet jedoch zusätzlich die Möglichkeit der Quantifizierung der einzelnen Spaltsequenzen, die in Form TMT-markierter Peptide identifiziert werden.

3.5 Zusammenfassung und Ausblick

Das Ziel der in diesem Kapitel beschriebenen Experimente, eine optimierte LC-MALDI-TOF/TOF MS/MS Methode für die Identifizierung und Quantifizierung TMT-markierter Peptide für den Q-PICS Ansatz zu entwickeln wurde erreicht. Insgesamt wurden drei verschiedene Parameter **i)** Anzahl der akkumulierten Laserschüsse zur Generierung der MS/MS-Spektren (3000 Laserschüsse vs. 5000 Laserschüsse), **ii)** Gasdruck in der CID-Kollisionszelle (CIDniedrig = 1×10^{-6} Torr vs. CIDhoch = 4×10^{-6} Torr) und **iii)** Reihenfolge der *precursor*-Auswahl für die Fragmentierung in der MS/MS-Analyse (“*weakest precursor first*” vs. “*strongest precursor first*”) untersucht.

Der Einfluss der Parameter **i)** der Anzahl der akkumulierten Laserschüsse und **ii)** dem Gasdruck in der CID-Kollisionszelle auf die Identifizierung TMT-markierter Peptide waren in den Experimenten auf Basis des 10-Proteinstandards vernachlässigbar. In Kombination mit Parameter **iii)** der Reihenfolge der *precursor*-Auswahl wurde für Parameter **i)** anhand der Experimente auf Basis des zytosolischen Hefeproteoms der Trend festgestellt, dass unter der geringeren Akkumulation an Laserschüssen mehr Identifizierungen erreicht wurden.

Bezüglich der Quantifizierung TMT-markierter Peptide war in den ersten Experimenten bereits der positive Effekt des Parameters **ii)** dem Gasdruck in der CID-Kollisionszelle auf die

durchschnittliche Standardabweichung des geometrischen Mittelwertes der TMT-Flächenverhältnisse gezeigt und durch die Verwendung der beiden verschiedenen Peptidbibliotheken auf Basis des 10-Proteinstandards bestätigt worden. Auch der Einsatz unterschiedlicher Probenmengen führte hierbei zu dem gleichen Ergebnis. Nach Auswertung der Experimente auf Basis des zytosolischen Hefeproteoms war auch der Trend beobachtbar, dass der Einfluss des Parameters **i)** der Anzahl an akkumulierten Laserschüssen nicht eindeutig festzustellen war, dass jedoch Parameter **iii)** die Reihenfolge der *precursor*-Auswahl einen starken Einfluss auf die durchschnittliche Standardabweichung des geometrischen Mittelwertes der TMT-Flächenverhältnisse ausübt. Insgesamt wurde die LC-MALDI-TOF/TOF MS/MS Methode unter Akkumulation von 3000 Laserschüssen bei einem Gasdruck in der CID-Kollisionszelle von CIDhoch und der “*strongest precursor first*” Reihenfolge für die *precursor*-Auswahl als die optimale Methode bestimmt.

Prinzipiell könnte man in Folgeexperimenten den Einfluss weiterer Parameter auf die Identifizierung und Quantifizierung TMT-markierter Peptide untersuchen. So spielen viele weitere Parameter wie beispielsweise die Kollisionsenergie eine wichtige Rolle. Diese könnte von 1 kV auf 2 kV erhöht werden. Auch die Akkumulation von 2000 oder 4000 Laserschüssen könnte untersucht werden. Ein ebenfalls wichtiger Parameter stellt die maximale Anzahl an *precursors*, die pro Spot fragmentiert werden dar. Dieser könnte beispielsweise von 25 auf 15 erniedrigt werden. Dadurch würde ein unterschiedlicher Spotabtrag zu den hier gezeigten Experimenten auftreten. Dies könnte auch einen Einfluss auf die Wahl der Reihenfolge der *precursor* Auswahl haben. Vergleicht man die Ergebnisse jedoch mit anderen proteomischen Studien, die ebenfalls auf der Quantifizierung mittels isobaren *tags* basieren, so sieht man, dass diese bezüglich der durchschnittlichen Standardabweichung bereits sehr gut vergleichbar sind und somit erst einmal keiner weiteren Optimierung bedürfen.

Vergleicht man die Anzahl der Identifizierungen des Q-PICS Ansatzes mit denjenigen des PICS Ansatzes, so scheint bezüglich der Identifizierung ebenfalls keine zusätzliche Optimierung mehr notwendig zu sein. Auch hier sind jedoch prinzipiell weitere Möglichkeiten vorhanden, die jedoch ausführlicher in Kapitel 4.5 behandelt werden.

Die generelle Anwendbarkeit des Q-PICS Ansatzes unter Verwendung sechs interner technischer Replikate in einer Analyse auf die qualitative Untersuchung der Spaltspezifität einer unbekannt Protease konnte in diesem Kapitel belegt werden. In einem weiteren Experiment, in dem der Einfluss verschiedener Puffersysteme auf die Spaltspezifität der

Endoprotease GluC untersucht wurde, sollte nun auch die Anwendbarkeit der quantitativen Ebene unter Beweis gestellt werden.

Das generelle Potential des Q-PICS Ansatzes für quantitative Analysen soll deshalb im folgenden Kapitel beschrieben werden.

Kapitel IV: Untersuchungen des Einflusses verschiedener Puffer auf die Spaltspezifität(*cleavage site specificity*)der Endoprotease GluC

4.1 Einleitung

In Kapitel 1.5 wurden verschiedene Methoden zur Untersuchung proteolytischer Prozesse bereits vorgestellt. Die Experimente zur Optimierung der LC-MALDI-TOF/TOF MS/MS Methode für die Identifizierung und Quantifizierung TMT-markierter Peptide haben die generelle Anwendbarkeit des Q-PICS Ansatzes auf die qualitative Analyse der Spaltspezifität unbekannter Proteasen gezeigt. Im folgenden Teil sollte nun die Anwendbarkeit dieses Ansatzes auf quantitativer Ebene belegt werden.

Die Endoprotease GluC spaltet Peptidbindungen C-terminal zu sauren Aminosäuren, wie Glutamat oder Aspartat in Position P1 (Drapeau *et al.*, 1972). Dies wurde auch durch die Ergebnisse der tryptischen Peptidbibliothek auf Basis des zytosolischen Hefeproteoms aus Kapitel 3 bestätigt. Zusätzlich wurde eine höhere Affinität der Endoprotease GluC für Glu-X Peptidbindungen in Bicarbonatpuffern im Vergleich zu anderen Puffern beschrieben, so dass in diesem Puffer Asp-X Bindungen praktisch nicht hydrolysiert werden (Houmard and Drapeau, 1972).

In einer späteren Studie wurde diese These dahingehend korrigiert, dass die höhere Affinität der Endoprotease GluC für Glu-X gegenüber Asp-X Peptidbindungen unabhängig von dem verwendeten Puffersystem ist (Sorensen, *et al.*, 1991). Die in dieser Publikation präsentierten Daten zeigten, dass Glu-X Peptidbindungen im Vergleich zu Asp-X Peptidbindungen 3000 – 5000-mal so schnell gespalten werden. In dem in den Experimenten der tryptischen Peptidbibliothek auf Basis des zytosolischen Hefeproteoms verwendeten HEPES-Puffer (pH 7,8) erfolgt diese Hydrolyse 3100-mal so schnell. Sorensen *et al.* stellten jedoch einen Einfluss des für den Verdau verwendeten Puffersystems auf die Gesamtaktivität der Endoprotease GluC fest. So erfolgt die Spaltung von Peptidbindungen durch GluC zehnmal so schnell in Phosphat- im Vergleich zu Bicarbonatpuffer. Der inhibitorische Effekt von Bicarbonatpuffer auf die GluC Gesamtaktivität resultiert dennoch in einer scheinbar ausschließlichen Affinität für Glu-X Peptidbindungen. Asp-X Bindungen werden weniger oft gespalten und die daraus entstehenden weniger abundanten Peptide liegen unterhalb des Detektionslimits.

Um dieses Phänomen nun genauer untersuchen zu können, sollte das quantitative Potential des Q-PICS Ansatzes verwendet werden. So wurden drei interne technische Replikate einer tryptischen Peptidbibliothek auf Basis des bereits zuvor verwendeten zytosolischen Hefeproteoms mit GluC in TEAB-Puffer und drei interne technische Replikate in HEPES-Puffer verdaut und anschließend vereinigt. Nach der LC-MALDI-TOF/TOF MS/MS Analyse mit der zuvor optimierten Methode sollten die beiden Puffersysteme dann qualitativ und quantitativ miteinander verglichen werden. Die genaue Vorgehensweise und die Ergebnisse des Experimentes, werden im Folgenden genauer beschrieben.

4.2 Material und Methoden

Es wurden die gleichen Chemikalien, Lösungsmittel und Organismen wie in Kapitel 3 verwendet.

Herstellung der Peptidbibliothek

Es wurde eine weitere tryptische Peptidbibliothek auf Basis des zytosolischen Hefeproteoms nach der in Kapitel 3 beschriebenen Vorgehensweise generiert. Um die Peptidbibliothek mit verschiedenen Puffern versetzen zu können, wurde sie nach der Aufreinigung mit Oasis SPE-Kartuschen mit einer Konzentration von 2 mg/ml in entionisiertem Wasser gelöst, zu je 50 µl aliquotiert und bei -20°C gelagert.

Verdau mittels zu testender Protease und TMT-Derivatisierung

Sechs Aliquote der in entionisiertem Wasser gelösten tryptischen Peptidbibliothek auf Basis des zytosolischen Hefeproteoms wurden für die Experimente des Puffereinflusses auf die Endoprotease GluC verwendet. Drei der sechs Aliquote wurden in 100 mM TEAB- (pH 7,8), die anderen drei in 50 mM HEPES-Puffer (pH 7,8) inkubiert. Mit allen sechs Aliquoten wurde dann ein Verdau mit GluC im Verhältnis Protease:Protein 1:50 (m/m) über Nacht bei 37°C und 300 U/min im Thermomixer durchgeführt. Die Reaktion wurde durch Temperaturerhöhung auf 99°C für 15 min und 300 U/min im Thermomixer gestoppt. Um die gleichen Pufferbedingungen während der TMT-Derivatisierung zu gewährleisten, wurde die gleiche molare Menge an HEPES- zu den Aliquoten mit TEAB-Puffer gegeben und umgekehrt. Die TMT-Derivatisierung wurde, wie in Kapitel 3.1 beschrieben, durchgeführt. Die Proben des Verdau in TEAB-Puffer wurden mit den Reagenzien 126, 127 und 128 und

die Proben des Verdau in HEPES-Puffer mit den Reagenzien 129, 130 und 131 markiert. Zur Validierung der Daten wurde ein *reversed* Markierungsexperiment über einen Labeltausch durchgeführt. Die drei internen technischen Replikate mit dem Verdau in TEAB-Puffer wurden mit den Reagenzien 129, 130 und 131 und die drei internen technischen Replikate mit dem Verdau in HEPES-Puffer mit den Reagenzien 126, 127 und 128 markiert (siehe Abbildung 4.1). Die weitere Probenvorbereitung wurde, wie in Kapitel 3.1 beschrieben, durchgeführt.

Trennung mittels IP-RP-HPLC bei pH 2,1

Die Proben der tryptischen Peptidbibliothek auf Basis des zytosolischen Hefeproteoms zur Untersuchung des Einflusses verschiedener Puffersysteme auf die Spaltspezifität der Endoprotease GluC wurden über das gleiche HPLC-System und unter den gleichen Bedingungen, wie die Probe in den Experimenten zur Methodenoptimierung in Kapitel 3.1, aufgetrennt.

Massenspektrometrische Analyse mittels MALDI-TOF/TOF MS/MS

Die massenspektrometrische Analyse auf MS-Ebene wurde, wie in den zuvor erfolgten Experimenten (Kapitel 3.1), durchgeführt.

Die Analyse auf MS/MS-Ebene wurde mit der in Kapitel 3 optimierten Methode durchgeführt. Die Parametereinstellungen wurden hierzu wie folgt gewählt: Minimales S/N-Verhältnis 20; *precursor* Massentoleranz: ± 200 ppm; das Signal musste nur in einem Spot detektiert werden, um als *precursor* akzeptiert zu werden; maximal 25 *precursor* pro Spot. Die MS/MS-Spektren wurden mit 1 kV mittels CID, mit einer Akkumulation von 3000 Laserschüssen und mit einem Druck von 4×10^{-6} Torr in der CID-Kollisionszelle (CIDhoch) generiert. Die Reihenfolge der *precursor*-Auswählerfolgte in Form der "*strongest precursor first*" Einstellung. Die MS/MS-Kalibrierung wurde mit Glu¹-Fibrinopeptid B aus dem 4700 Proteomics Analyzer Mass Standard Kit (AB SCIEX, Darmstadt, Deutschland) durchgeführt.

Protein- bzw. Peptididentifizierung mittels Datenbankanalyse

Um dem Problem eines fehlerhaften *peak pickings* durch die TS2Mascot Software (Matrix Science, London, England) entgegenzuwirken (siehe Kapitel 6), wurde die Erstellung der mgf-Dateien leicht modifiziert: Die Parameter Massenbereich von 65 Da bis zu der Masse des *precursors* minus 35 Da und minimales S/N-Verhältnis von 10 wurden beibehalten. Es wurden jedoch zwei verschiedene mgf-Dateien generiert. Hierzu wurden einmal die

monoisotopischen Signale (*monoisotopic only* (MIO) *enabled*) und einmal alle Signale (MIO *disabled*) exportiert. Zusätzlich wurde von Bart HJ van den Berg (Institut für Experimentelle Medizin, Christian-Albrechts Universität zu Kiel) ein Perl-Skript entwickelt, das fehlende Signale (Flächen) der Reporterionen aus den mgf-Dateien der TS2Mascot Software unter Verwendung der MIO *enabled* Option durch die Flächen in den mgf-Dateien unter Verwendung der MIO *disabled* Option ergänzt (siehe Anhang Kapitel 6. Perl-Skript RICmerge).

Die Mascot Datenbanksuche wurde mit der Mascot 2.2.04 Software unter Beibehaltung der zuvor gewählten Parameter aus Kapitel 3.1 durchgeführt. Lediglich der *identity threshold* der Peptide wurde auf einen *ion score* von 39 eingestellt. Dadurch wurde erneut eine falsch-positiv Rate (FDR) <4% erreicht. Die quantitativen Daten wurden, zusammen mit den Ergebnissen der Mascot Datenbanksuche, als csv-file Format exportiert und die folgenden statistischen Berechnungen angewandt.

Im ersten Schritt wurde der Grenzwert der TMT-Flächenverhältnisse berechnet, ab dem eine Spaltstelle (*cleavage site*) entweder zu TEAB- zu HEPES- oder zu beiden Puffern klassifiziert wird. Hierzu wurden die beiden Puffer zunächst getrennt betrachtet. Für TEAB-Puffer wurden die Flächen der Reporterionenkanäle 126 (127/126 und 128/126) und 127 (128/127) als Teiler verwendet (siehe Abbildung 4.1). Im Gegensatz dazu wurden für HEPES-Puffer die Flächen der Kanäle 129 (130/129 und 131/129) und 130 (131/130) als Teiler verwendet. Der Erwartungswert der TMT-Flächenverhältnisse der einzelnen Puffer betrug 1,0 und negative TMT-Flächenverhältnisse wurden entfernt. IQR-Tests zur Entfernung mathematischer Ausreißer wurden mit dem in Kapitel 3 vorgestellten Perl-Skript durchgeführt und der durchschnittliche geometrische Mittelwert aller TMT-Flächenverhältnisse der beiden Puffer, sowie die zugehörige Standardabweichung (SD) nach Tan *et al.* errechnet (2008). Der Grenzwert der TMT-Flächenverhältnisse, ab dem eine Spaltstelle zu TEAB-Puffer klassifiziert wurde, errechnete sich über die Formel $1 + 2 \cdot SD$ (Tan, *et al.*, 2008). Der Grenzwert, ab dem eine Spaltstelle zu HEPES-Puffer klassifiziert wird, über den reziproken Wert. Die Spaltsequenzen mit einem TMT-Flächenverhältnis $\ll 1$ sollten überwiegend in TEAB- und diejenigen mit einem Wert $\gg 1$ in HEPES-Puffer entstanden sein. Spaltstellen, deren TMT-Flächenverhältnisse zwischen den beiden Grenzwerten lagen wurden, zu beiden Puffern zugeordnet.

Im zweiten Schritt wurden die TMT-Flächenverhältnisse der Spaltsequenzen in beiden Puffern miteinander verglichen. Die Flächen der einzelnen TMT-Reporterionen der internen technischen Replikate der beiden Puffer wurden in Bezug zueinander gesetzt. Hierzu wurden

die Flächen der TMT-Reporterionenkanäle für TEAB-Puffer (126, 127 und 128) als Teiler für die Flächen der TMT-Reporterionenkanäle für HEPES-Puffer (129, 130 und 131) verwendet (siehe Abbildung 4.1). Der durchschnittliche geometrische Mittelwert und die zugehörige Standardabweichung wurde unter Berücksichtigung der drei technischen Replikate für jede Spaltsequenz mit einem weiteren Perl-Skript (Bart HJ van den Berg, Institut für Experimentelle Medizin, Christian-Albrechts Universität zu Kiel, siehe Anhang Kapitel 4.1 Perl-Skript BufferComparison version 1.8.1) berechnet, wobei durch das Skript parallel IQR-Tests zur Entfernung mathematischer Ausreißer durchgeführt wurden.

Die einzelnen Spaltsequenzen wurden dann über die TMT-Flächenverhältnisse nach den im ersten Schritt errechneten Grenzwerten klassifiziert. Die gleiche statistische Vorgehensweise wurde für das *reversed* Markierungsexperiment (siehe Abbildung 4.1) verwendet. Die Spaltspezifität der Endoprotease GluC wurde mittels der in Kapitel 3 verwendeten Version der PICS Datenanalyse-Pipeline ermittelt und es wurden *Heat-Maps* generiert.

4.3 Ergebnisse

Um die Anwendbarkeit des Q-PICS Ansatzes auf einer *large-scale* Ebene zu zeigen, wurde die GluC Spaltspezifität in zwei verschiedenen Puffersystemen analysiert. Die Vorgehensweise für dieses Experiment ist in Abbildung 4.1 dargestellt.

Als Grundlage für den Verdau mit GluC als *target*-Protease wurde die gleiche tryptische Peptidbibliothek auf Basis des zytosolischen Hefeproteoms verwendet, die bereits in den Experimenten zur LC-MALDI-TOF/TOF MS/MS Methodenoptimierung zum Einsatz kam. Diese wurde lediglich einmal in TEAB- (pH 7,8) und einmal in HEPES-Puffer (pH 7,8) gelöst. Pro Puffersystem wurden drei interne technische Replikate verwendet, die nach dem Verdau mit GluC mit den unterschiedlichen TMT-Reagenzien markiert (126, 127 und 128 für TEAB- bzw. 129, 130 und 131 für HEPES-Puffer) und im Verhältnis 1:1:1:1:1:1 vereinigt wurden. Während die TMT-markierten Peptide (*prime* Spaltsequenz) nach der LC-MALDI-TOF/TOF MS/MS-Analyse wieder über eine Mascot Datenbanksuche identifiziert wurden, wurden die *nonprime* Spaltsequenzen mittels der leicht modifizierten Datenanalyse-Pipeline des PICS-Ansatzes bestimmt (Schilling and Overall, 2008). Um eine statistisch robuste Quantifizierung durchzuführen, wurden verschiedene TMT-Flächenverhältnisse der Reporterionen berechnet. Diese sind Abbildung 4.1, A-C zu entnehmen.

Zusätzlich wurde die Strategie durch ein *reversed*-Markierungsexperiment validiert. Hierzu wurden die drei Aliquote für TEAB-Puffer statt mit den TMT-Reagenzien 126, 127 und 128 mit 129, 130 und 131 und die drei Aliquote für HEPES-Puffer statt mit 129, 130 und 131 mit 126, 127 und 128 markiert. Die Berechnung für die TMT-Flächenverhältnisse in diesem *reversed* Markierungsexperiment ist in Abbildung 4.1, D dargestellt.

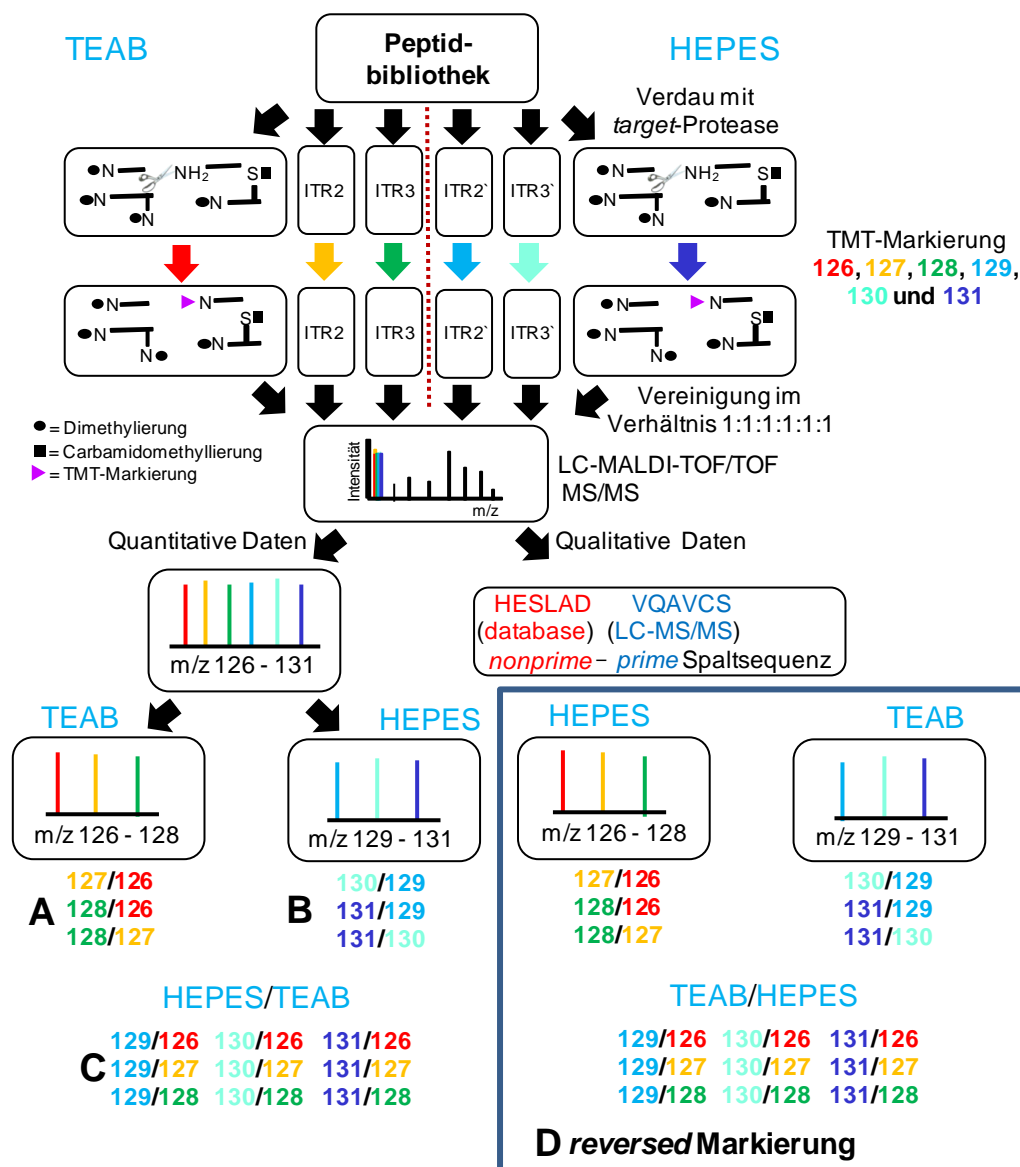


Abb. 4.1: Vorgehensweise des Q-PICS Ansatzes zur Untersuchung des Einflusses von HEPES- (pH 7,8) und TEAB-Puffer (pH 7,8) auf die Spaltspezifität der Endoprotease GluC. Für den Verdau mittels GluC als *target*-Protease wurde eine tryptische Peptidbibliothek auf Basis des zytosolischen Hefeproteoms verwendet. Diese wurde jeweils in den beiden unterschiedlichen Puffersystemen gelöst. Nach dem Verdau wurden je drei Aliquote für TEAB und drei Aliquote für HEPES mit den TMT-Reagenzien 126, 127 und 128 bzw. 129, 130 und 131 markiert und im Verhältnis 1:1:1:1:1:1 vereinigt. Die Identifizierung der TMT-markierten Peptide (der *prime*

Spaltsequenzen) erfolgte nach der LC-MALDI-TOF/TOF MS/MS Analyse über eine Mascot-Datenbanksuche. Die *nonprime* Spaltsequenzen wurden bioinformatisch über eine leicht modifizierte PICS Datenanalyse-Pipeline bestimmt (Schilling and Overall, 2008). Die quantitative Analyse und Klassifizierung der Spaltsequenzen zu den beiden Puffersystemen erfolgte über die Flächen der einzelnen Reporterionensignale, die miteinander ins Verhältnis gesetzt wurden. **A)** Schema der Berechnung der TMT-Flächenverhältnisse zur Bestimmung eines *cutoff*-Wertes, ab dem Peptide zu TEAB- oder **B)** HEPES-Puffer klassifiziert wurden. **C)** Für den Vergleich der beiden Puffer wurde jedes TMT-Reagenz des TEAB-Puffers als Teiler für die TMT-Reagenzien des HEPES-Puffers verwendet. **D)** Die Validierung der Strategie erfolgte über ein *reversed*-Markierungsexperiment. (ITR = internes technisches Replikat).

4.3.1 Untersuchung des Einflusses des Puffersystems auf die Spaltspezifität der Endoprotease GluC

Um eine falsch-positiv Rate $<4\%$ zu erreichen, wurde der Mascot *ion score cutoff* für den *identity threshold* der TMT-markierten Peptide auf 39 gesetzt. Die Ergebnisse bezüglich der Identifizierung TMT-markierter Peptide bzw. Spaltsequenzen sind in Tabelle 4.1 dargestellt. Insgesamt wurden 1.773 TMT-markierte Peptide identifiziert, deren durchschnittlicher Mascot *ion score* 60 betrug. Über die leicht modifizierte PICS Datenanalyse-Pipeline (Schilling and Overall, 2008) wurden daraus 557 *unique* Spaltsequenzen ermittelt, die in mindestens zwei der drei technischen Replikate identifiziert wurden.

Insgesamt wurden 4,7% an negativen Werten für TEAB- bzw. 3,0% an negativen Werten für HEPES-Puffer an TMT-Flächenverhältnissen errechnet und von der weiteren Analyse ausgeschlossen.

In diesem Experiment kamen jeweils drei interne technische Replikate pro Probe zum Einsatz. Der Erwartungswert für die TMT-Flächenverhältnisse innerhalb dieser Replikate betrug 1,0. Mathematische Ausreißer wurden über den IQR-Test mit einem *in-house* Perl-Skript identifiziert und von der weiteren Analyse ausgeschlossen. Danach wurden die geometrischen Mittelwerte und Standardabweichungen über jeweils drei TMT-Flächenverhältnisse für jeden der beiden Puffer errechnet (siehe Tabelle 4.2 und Abbildung 4.1).

Tab. 4.1: Ergebnisse der Q-PICS Analyse zur Untersuchung des Einflusses verschiedener Puffersysteme auf die Spaltspezifität der Endoprotease GluC. Alle Werte sind über die Summe von drei technischen Replikaten für jedes der beiden TMT-Markierungsexperimente gemittelt.

Experiment	# precursor	durchschnittlicher ion score	# identifizierte Peptide	# identifizierte Spaltstellen (in mindestens 2 aus 3 TR)
normale Markierung	28.448	60	1.773	557
reversed Markierung	28.601	61	1.552	485

Tab. 4.2: Durchschnittliche geometrische Mittelwerte und Standardabweichungen (SD) über je drei TMT-Flächenverhältnisse für den Verdau mit der Endoprotease GluC in TEAB- bzw. HEPES-Puffer (siehe Abbildung 4.a). Zusätzlich sind die Grenzwerte abgebildet, ab denen eine *unique* Spaltsequenz zu einem der beiden Puffer klassifiziert wurde. TMT-Flächenverhältnisse die zwischen diesen beiden Grenzwerten liegen, wurden zu beiden Puffersystemen klassifiziert. Alle Werte sind über die Summe von drei technischen Replikaten für jedes der beiden Markierungsexperimente gemittelt.

Experiment	durchschnittlicher geometrischer Mittelwert	durchschnittliche SD (%)	Grenzwert für TEAB-Puffer	Grenzwert für HEPES-Puffer
normale Markierung	0,956	0,208 (22%)	< 0,706	> 1,417
reversed Markierung	0,935	0,244 (26%)	> 1,488	< 0,672

Der durchschnittliche geometrische Mittelwert über die sechs TMT-Reporterionenkanäle (siehe Abbildung 4.1, Teil A und B) wies für dieses Experiment eine Abweichung <5% vom Erwartungswert 1,0 auf und errechnete sich auf 0,956 mit einer Standardabweichung <22%. Aus diesem Wert wurde der Grenzwert, der TMT-Flächenverhältnisse gebildet, ab dem eine *unique* Spaltstelle zu TEAB- oder HEPES-Puffer klassifiziert wurde. Dieser Grenzwert errechnete sich über die Formel $1 + 2 \cdot SD$, (Tan, *et al.*, 2008) für TEAB- und über den reziproken Wert für HEPES-Puffer. *Unique* Spaltsequenzen, deren TMT-Flächenverhältnisse zwischen den beiden Grenzwerten lagen, wurden zu beiden Puffern gezählt.

In Abbildung 4.2, Teil A sind beispielhaft drei *unique* Spaltsequenzen und ihre TMT-Flächenverhältnisse aufgelistet. Wie zuvor bereits beschrieben, wurde hierzu jeder der drei TMT-Reporterionenkanäle des TEAB-Puffers als Teiler für die TMT-Reporterionenkanäle des HEPES-Puffers verwendet.

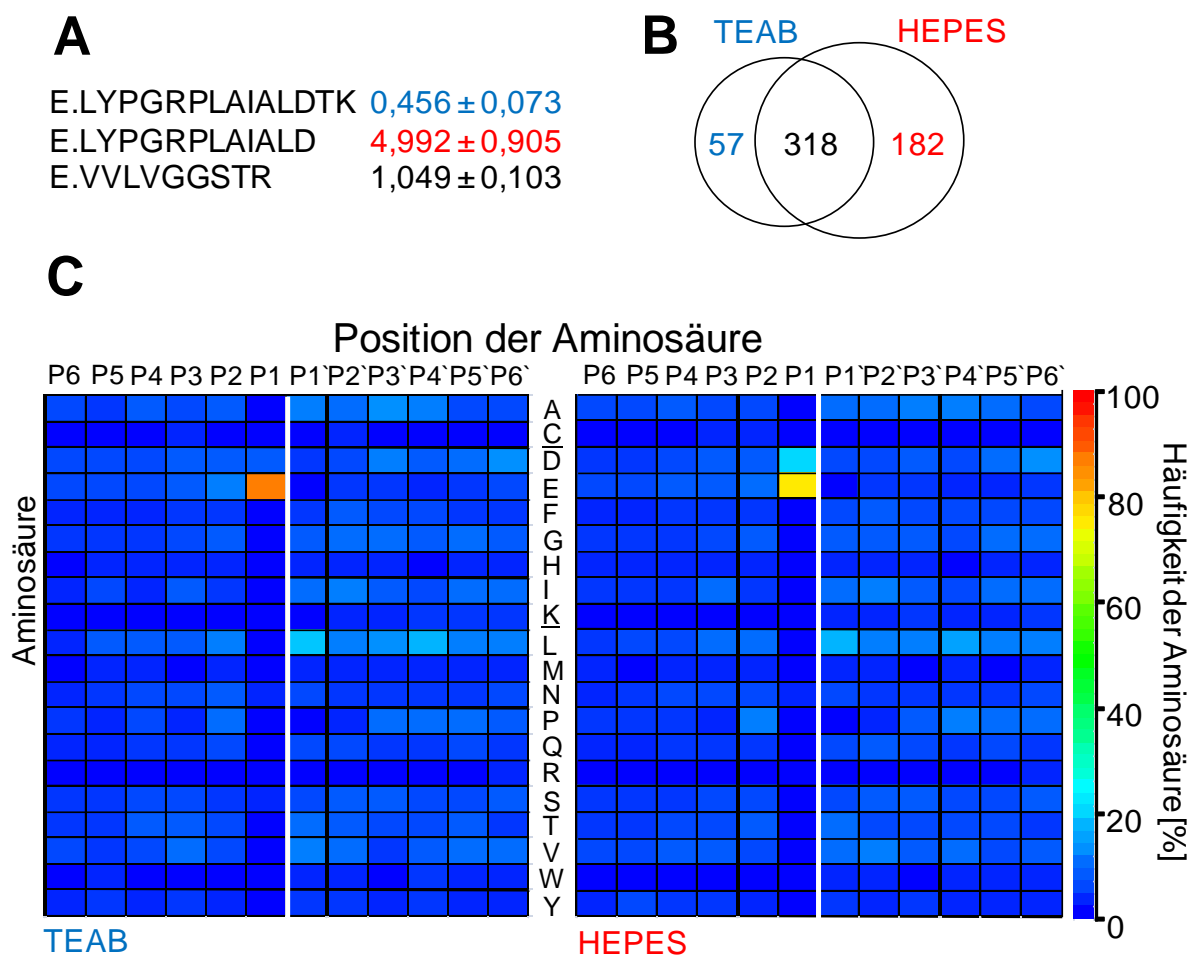


Abb. 4.2: Übersicht der Ergebnisse des Q-PICS Ansatzes zur Untersuchung des Einflusses von TEAB- (pH 7,8) und HEPES-Puffer (pH 7,8) auf die Spaltspezifität der Endoprotease GluC. **A)** Beispiele für identifizierte *unique* Spaltstellen und ihre Klassifizierung zu TEAB-, HEPES- oder beiden Puffern über ihre TMT Flächenverhältnisse. **B)** Venn-Diagramm der identifizierten *unique* Spaltsequenzen und ihre Klassifizierung und **C)** *Heat-Maps* zur Darstellung der GluC Spaltspezifität unter Verwendung der beiden Puffersysteme TEAB- (pH 7,8) und HEPES-Puffer (pH 7,8). Für die Datenbanksuche wurde semi Trypsin und semi GluC als Enzymeinstellung gewählt. Die Peptidbindung zwischen Position P1 und P1' ist durch einen weißen Strich abgetrennt. C= carbamidomethyliertes Cystein, K= dimethyliertes Lysin.

Die Verteilung der identifizierten *unique* Spaltsequenzen und ihre Klassifizierung, ob sie in TEAB-, HEPES oder beiden Puffern generiert wurden, ist in Abbildung 4.2, Teil B gezeigt. Insgesamt wurden 375 *unique* Spaltsequenzen so klassifiziert, dass sie in TEAB-Puffer entstanden sind. Von diesen Spaltsequenzen wurden 57 überwiegend diesem Puffersystem zugeordnet. Insgesamt 500 *unique* Spaltsequenzen wurden einer Spaltung in HEPES-Puffer zugeordnet, wobei 182 überwiegend für dieses Puffersystem klassifiziert wurden. Der Großteil an identifizierten *unique* Spaltstellen (318) wurde jedoch so klassifiziert, dass der Verdau mit der Endoprotease GluC in beiden Puffersystemen stattgefunden hat.

Die Häufigkeit der einzelnen Aminosäuren in Position P6 bis P6' ist in Abbildung 4.2, Teil C in Form zweier *Heat-Maps*, für beide Puffersysteme gezeigt. Unter Verwendung des TEAB-Puffers wurden insgesamt 87,9% der identifizierten Spaltsequenzen nach einem C-terminalen Glutamat und 6,5% nach einem C-terminalen Aspartat in Position P1 generiert. Zusätzlich wurden 3,7% Spaltsequenzen, in denen andere Aminosäuren als Glutamat oder Aspartat in Position P1 enthalten waren, identifiziert. Außerdem wurden 1,9% der identifizierten Peptide als *ambiguous* Peptide zugeordnet. Die Anzahl an korrekten Zuordnungen, die der vorhergesagten Spaltspezifität der Endoprotease Glu C entsprechen, beträgt damit 94,4%.

Im Gegensatz dazu wurden für HEPES-Puffer 75,5% der Spaltsequenzen nach einem C-terminalen Glutamat und 19,8% nach einem C-terminalen Aspartat in Position P1 gebildet. Für dieses Puffersystem wurden 2,6% der Spaltstellen mit anderen Aminosäuren als Glutamat oder Aspartat in Position P1 und 2,1% *ambiguous* Peptide identifiziert. Die Anzahl an korrekten Zuordnungen, die der vorhergesagten Spaltspezifität der Endoprotease Glu C entsprechen, beträgt damit in HEPES-Puffer 95,3%.

4.3.2 Untersuchung des Einflusses des Puffersystems auf die Spaltspezifität der Endoprotease GluC mittels *reversed* Markierung

Zur Validierung des Q-PICS Ansatzes und der in den Experimenten zur Untersuchung des Einflusses von TEAB- (pH 7,8) und HEPES-Puffer (pH 7,8) auf die Spaltspezifität der Endoprotease GluC gemessenen Daten, wurde ein *reversed* Markierungsexperiment durchgeführt. Die Berechnung für die TMT-Flächenverhältnisse ist in Abbildung 4.1, Teil D dargestellt.

Insgesamt wurden 1.552 TMT-markierte Peptide mit einem durchschnittlichen Mascot *ion score* von 61 identifiziert. Diese resultierten in 485 *unique* Spaltsequenzen, die in mindestens zwei der drei technischen Replikaten identifiziert wurden (siehe Tabelle 4.1).

Außerdem wurden 4,4% negative Werte für TEAB-, bzw. 5,1% negative Werte für HEPES-Puffer an TMT-Flächenverhältnissen von der weiteren Analyse ausgeschlossen.

Der durchschnittliche geometrische Mittelwert über die sechs TMT-Reporterionenkanäle (siehe Abbildung 4.1, Teil D) betrug 0,935 mit einer Standardabweichung <26%. Damit wies er eine Abweichung vom Erwartungswert 1,0 <7% auf. Der Grenzwert der TMT-Flächenverhältnisse, ab dem eine *unique* Spaltstelle zu TEAB- oder HEPES-Puffer klassifiziert wurde, wurde über die gleiche Formel $1 + 2 \cdot SD$ (Tan, *et al.*, 2008), errechnet (siehe Tabelle 4.2).

In Abbildung 4.3, Teil A sind die drei gleichen *unique* Spaltsequenzen und ihre TMT-Flächenverhältnisse, wie in Abbildung 4.2, Teil A, gezeigt.

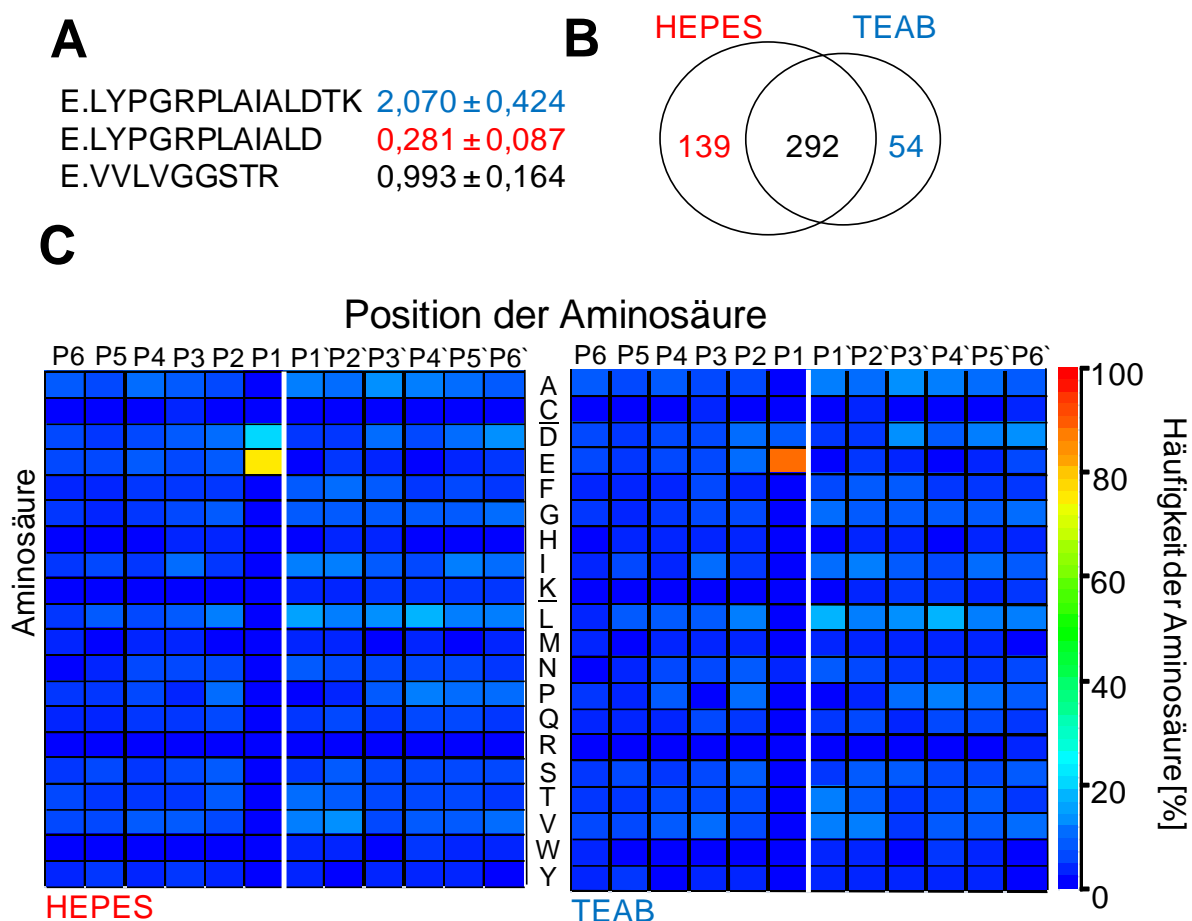


Abb. 4.3: Übersicht der Ergebnisse des Q-PICS Ansatzes zur Untersuchung des Einflusses von TEAB- (pH 7,8) und HEPES-Puffer (pH 7,8) auf die Spaltspezifität der Endoprotease GluC mittels *reversed* Markierung. **A)** Beispiele für identifizierte *unique* Spaltstellen und ihre Klassifizierung zu TEAB-, HEPES- oder beiden Puffern über ihre TMT Flächenverhältnisse. **B)** Venn-Diagramm der identifizierten *unique* Spaltsequenzen und ihre Klassifizierung und **C)** *Heat-Maps* zur Darstellung der GluC Spaltspezifität unter

Verwendung der beiden Puffersysteme TEAB- (pH 7,8) und HEPES-Puffer (pH 7,8). Für die Datenbanksuche wurde semi Trypsin und semi GluC als Enzymeinstellung gewählt. Die Peptidbindung zwischen Position P1 und P1´ ist durch einen weißen Strich abgetrennt. C = carbarmidomethyliertes Cystein, K = dimethyliertes Lysin.

In dem *reversed* Markierungsexperiment wurden insgesamt 346 *unique* Spaltsequenzen so klassifiziert, dass sie in TEAB-Puffer generiert wurden. Es wurden 54 dieser Spaltsequenzen überwiegend diesem Puffersystem zugeordnet. Für HEPES-Puffer wurden insgesamt 431 *unique* Spaltsequenzen zugeordnet, wobei 139 überwiegend für dieses Puffersystem klassifiziert wurden. Wiederum wurde der Großteil von 292 identifizierten *unique* Spaltstellen so klassifiziert, dass sie beiden Puffersystemen zugeordnet wurden (siehe Abbildung 4.3, Teil B).

Die *Heat-Maps* in Abbildung 4.3, Teil C stellen die Häufigkeit der einzelnen Aminosäuren in Position P6 bis P6` für beide Puffersysteme in dem *reversed* Markierungsexperiment dar. Insgesamt 88,4% der identifizierten Spaltsequenzen wurden unter Verwendung des TEAB-Puffers nach einem C-terminalen Glutamat und 6,8% nach einem C-terminalen Aspartat in Position P1 generiert. Die Anzahl an Spaltsequenzen, in denen andere Aminosäuren als Glutamat oder Aspartat in Position P1 enthalten war, betrug 3,1%, die der *ambiguous* Peptide 1,7%. Insgesamt 95,2% der Spaltsequenzen entsprachen damit der vorhergesagten Spaltspezifität der Endoprotease Glu C.

Im Gegensatz dazu wurden für HEPES-Puffer 75,8% der Spaltsequenzen nach einem C-terminalen Glutamat und 20,6% nach einem C-terminalen Aspartat in Position P1 generiert. In dem *reversed* Markierungsexperiment wurden für HEPES-Puffer 1,4% der Spaltsequenzen mit anderen Aminosäuren als Glutamat oder Aspartat in Position P1 und 2,2% *ambiguous* Peptide identifiziert. Hier errechnet sich die Anzahl an korrekten Zuordnungen, die der vorhergesagten Spaltspezifität der Endoprotease Glu C entsprechen, zu 96,4%.

4.4 Diskussion

Die generelle Anwendbarkeit des klassischen PICS Ansatzes (Schilling and Overall, 2008) zur Charakterisierung der Spaltspezifität unterschiedlicher Proteasen konnte bereits in vielen Studien gezeigt werden (Becker-Pauly, *et al.*, 2011, Biniossek, *et al.*, 2011, Paes Leme, *et al.*, 2011, Schilling and Overall, 2008, Schilling, *et al.*, 2011). Wie in Kapitel 1.5 bereits beschrieben, handelt es sich bei diesem Ansatz jedoch um eine rein qualitative Methode.

Diese ermöglicht zwar die Identifizierung von Spaltsequenzen und damit die Charakterisierung der Spaltspezifität einer Protease, enthält jedoch keinerlei Information bezüglich einer quantitativen Ebene. Nachdem in Kapitel 3 die Anwendbarkeit des Q-PICS Ansatzes auf qualitativer Ebene bereits nachgewiesen wurde, sollte dies nun auch auf quantitativer Ebene erfolgen. Prinzipiell konnte dies bereits mit den Experimenten der tryptischen Peptidbibliothek auf Basis des zytosolischen Hefeproteoms gezeigt werden. Hier konnten durch die sechs verschiedenen TMT-Reporterionenkanäle insgesamt sechs interne technische Replikate in einer Messreihe analysiert werden. Dies ist beim klassischen PICS Ansatz nicht der Fall, da hier immer nur eine Probe untersucht werden kann.

Die Vorgehensweise für das quantitative Experiment zur Untersuchung der Spaltspezifität der Endoprotease GluC in unterschiedlichen Puffersystemen ist in Abbildung 4.1 dargestellt. Da in den Experimenten der tryptischen Peptidbibliothek auf Basis des zytosolischen Hefeproteoms zur Optimierung verschiedener LC-MALDI-TOF/TOF MS/MS Parameter bereits GluC verwendet wurde und hunderte von Spaltsequenzen identifiziert wurden, wurde die gleiche Peptidbibliothek auch für das quantitative Experiment benutzt. Ein Unterschied bestand jedoch darin, dass die Peptidbibliothek einmal in TEAB- und einmal in HEPES-Puffer gelöst wurde. Für jedes Puffersystem standen je drei TMT-Reagenzien zur Verfügung (126, 127 und 128 für TEAB- bzw. 129, 130 und 131 für HEPES-Puffer), so dass für die beiden verschiedenen biologischen Zustände (dem Verdau in den beiden Puffersystemen) je drei interne technische Replikate verwendet wurden. Die Zuordnung der identifizierten Spaltsequenzen zu einem oder beiden Puffersystemen erfolgte letztendlich über die TMT-Flächenverhältnisse. Damit sollten die Spaltsequenzen, die in beiden Puffern gleichermaßen generiert wurden, ein Verhältnis nahe 1,0 aufweisen. Die Spaltsequenzen mit einem TMT-Flächenverhältnis $\ll 1$ sollten überwiegend in TEAB- und diejenigen mit einem Wert $\gg 1$ in HEPES-Puffer entstanden sein. Ein genauer Grenzwert, ab wann ein quantitativer Unterschied zwischen verschiedenen biologischen Zuständen feststellbar ist, kann über die Formel $1 + 2 \cdot SD$ (Tan, *et al.*, 2008) für TEAB- und über den reziproken Wert für HEPES-Puffer berechnet werden. Hierzu muss der durchschnittliche geometrische Mittelwert der TMT-Flächenverhältnisse und die zugehörige Standardabweichung (SD) für beide Puffersysteme berechnet werden (siehe Abbildung 4.1, A und B).

Eine Besonderheit des Q-PICS Ansatzes ist die Tatsache, dass die Quantifizierung ausschließlich auf Peptid-Ebene durchgeführt wird. In klassischen proteomischen Studien, die auf der Quantifizierung mittels isobaren *tags* basieren, wird dies prinzipiell in einem ersten Schritt auch auf Peptid-Ebene durchgeführt. Die quantitativen Daten der Peptide werden dann

aber auf das Protein, dem sie durch den Verdau entstammen, bezogen und zusammengefasst. Somit ergibt sich für die einzelnen Proteine oft eine Vielzahl an quantitativen Werten. Diese Anzahl an Messwerten kann dann durch die Verwendung technischer Replikate (also LC-MS/MS Analysen) zusätzlich erhöht werden. In den meisten Fällen wird ein Reporterionenkanal des isobaren *tags* für jeden biologischen Zustand verwendet, um eine größtmögliche Anzahl an verschiedenen biologischen Zuständen in einem Experiment miteinander vergleichen zu können (Ross, *et al.*, 2004, Shirran and Botting, 2010). Es gibt jedoch Ausnahmen, die sich die Verwendung interner technischer Replikate zu Nutze machen (Dayon, *et al.*, 2008, Tiberti, *et al.*, 2010).

Auch isobare *tags* unterliegen methodischen Einschränkungen wie beispielsweise der Markierungseffizienz. Zusätzlich gibt es in den Experimenten auch biologische Variationen oder methodische Fehlerquellen, die durch die Verwendung interner technischer Replikate untersucht werden können.

Auch der in dieser Arbeit entwickelte Q-PICS Ansatz basiert auf der Verwendung von internen technischen Replikaten. Dadurch wird dem zuvor beschriebenen Problem der Quantifizierung auf Peptid-Ebene und der damit einhergehenden geringen Anzahl an quantitativen Messwerten pro Spaltsequenz entgegengewirkt. Die Durchführung der LC-MALDI-TOF/TOF MS/MS-Analysen in Form dreier technischer Replikate in Kombination mit drei internen technischen Replikaten und der ausgewählten Bildung der TMT-Flächenverhältnisse (siehe Abbildung 4.1, A und B) ermöglicht die maximale Anzahl von bis zu 27 Quantifizierungswerten pro identifizierter Spaltsequenz. Diese Zahl kann noch erhöht werden, sofern eine Spaltsequenz mehr als einmal in den technischen Replikaten identifiziert wird. Durch die zusätzliche Einschränkung, dass nur diejenigen Spaltsequenzen als identifiziert betrachtet wurden, die in mindestens zwei der drei technischen Replikaten gefunden wurden, wird eine höhere Anzahl an TMT-Flächenverhältnissen gewährleistet, die für eine robuste Quantifizierung ausreichen sollten.

4.4.1 Untersuchung des Einflusses des Puffersystems auf die Spaltspezifität der Endoprotease GluC

Die durchschnittliche *precursor*-Anzahl, die für die MS/MS-Analyse ausgewählt wurden, lag in den Experimenten der tryptischen Peptidbibliothek auf Basis des zytosolischen Hefeproteoms zur Optimierung verschiedener LC-MALDI-TOF/TOF MS/MS Parameter um ca. 9% niedriger, als in den Experimenten zur Untersuchung des Einflusses verschiedener

Puffersysteme auf die Spaltspezifität der Endoprotease GluC (siehe Tabelle 3.13 und Tabelle 4.1). Dennoch wurden in diesem Experiment 37% weniger TMT-markierte Peptide identifiziert, auch wenn der durchschnittliche Mascot *ion score* praktisch identisch war.

In beiden Experimenten wurde die gleiche Menge an Gesamtpeptid für jedes technische Replikate auf der nanoHPLC aufgetrennt und unter identischen Bedingungen mittels MALDI-TOF/TOF MS/MS analysiert. In dem GluC Verdau in zwei verschiedenen Puffersystemen sollten insgesamt mehr unterschiedliche Peptide gebildet worden sein, als unter der Verwendung nur eines Puffersystems. Dadurch erhält man nach Vereinigung der jeweils drei internen technischen Replikate pro Puffersystem, eine komplexere Probe, als in den Experimenten mit nur einem Puffersystem und sechs internen technischen Replikaten. Da in beiden Experimenten jedoch die gleiche Probenmenge pro Analyse verwendet wurde, sinkt damit die Konzentration pro Einzelpeptid in den Experimenten mit zwei verwendeten Puffersystemen. Es ist also nicht erstaunlich, dass damit auch die Anzahl an identifizierten TMT-markierten Peptiden in diesem Experiment sinkt. Dennoch wurden in diesem Experiment insgesamt 557 *unique* Spaltsequenzen ermittelt, die in mindestens zwei der drei technischen Replikate identifiziert wurden. Dies stellt im Vergleich zu den Ergebnissen des ursprünglichen PICS Ansatzes mit 453 identifizierten Spaltsequenzen (Schilling and Overall, 2008) immer noch ein sehr gutes Ergebnis dar.

Hätte die Anzahl an Identifizierungen nicht ausgereicht, um die Spaltspezifität einer Protease hinreichend beschreiben zu können, so könnte die Anzahl der identifizierten Spaltsequenzen z. B. durch die Verwendung anderer Peptidbibliotheken, die nicht mit Trypsin als *work*-Protease verdaut wurden, erhöht werden. Hier sei noch einmal darauf hingewiesen, dass diese *work*-Proteasen sich in der Spaltspezifität nicht mit der *target*-Protease überschneiden dürfen. Darüber hinaus könnte die geringere Probenmenge pro einzelner Spaltsequenz und die höhere Probenkomplexität auch durch Verwendung einer höheren Gesamtprobenmenge in Kombination mit einer 2D-HPLC Auftrennung kompensiert werden. Die Fraktionierung der Proben in der ersten Dimension über eine starke Kationenaustauschchromatographie (SCX-LC), gefolgt von einer nano (Ionenpaar-) Umkehrphasenchromatographie (IP-RP-LC) in der zweiten Dimension wurde für Peptide, die mit isobaren *tags* markiert wurden, bereits beschrieben (Dayon, *et al.*, 2008, Ross, *et al.*, 2004, Sinclair *et al.*, 2011).

Betrachtet man nun die Ergebnisse bezüglich der Quantifizierung, so zeigt es sich, dass im Vergleich zu den Experimenten der tryptischen Peptidbibliothek auf Basis des zytosolischen Hefeproteoms zur Optimierung verschiedener LC-MALDI-TOF/TOF MS/MS Parameter die Abweichung des durchschnittlichen geometrischen Mittelwertes vom Erwartungswert 1,0

gleichermaßen ausgeprägt ist (<5%). Die durchschnittliche Standardabweichung für die Experimente mit den beiden unterschiedlichen Puffersystemen liegt mit 22% (im Vergleich zu 13%) jedoch höher. In Abbildung 4.4, Teil 1 ist beispielhaft das MS/MS-Spektrum des Peptides LYPGRPLAIALD des Proteins Pyruvatkinase 1 dargestellt.

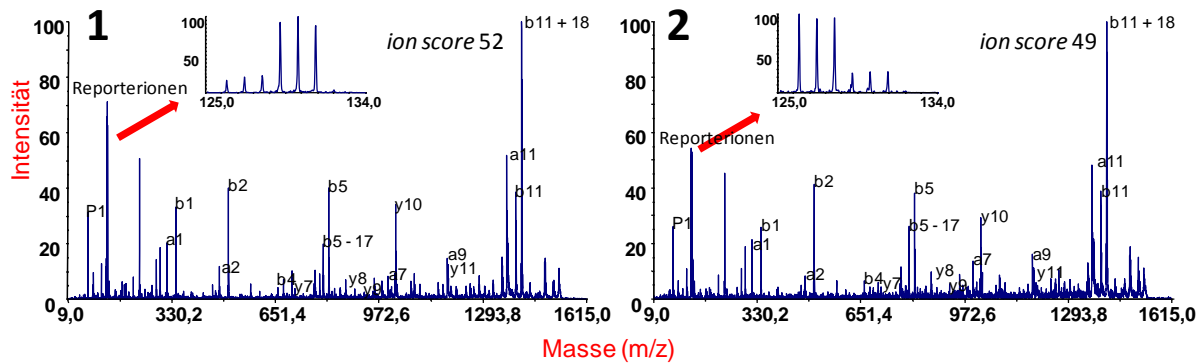


Abb. 4.4: MS/MS-Spektren des Peptides LYPGRPLAIALD ($m/z = 1527,91$) aus dem Protein Pyruvatkinase 1 der Experimente zur Untersuchung des Einflusses unterschiedlicher Puffersysteme auf die Spaltspezifität der Endoprotease GluC mit **1)** normaler und **2)** *reversed* Markierung. Die Spektren wurden im *positive ion mode* generiert. Der m/z -Bereich, der die Signale der Reporterionen enthält, ist jeweils in Vergrößerung dargestellt und mit einem roten Pfeil markiert. Zusätzlich sind auch die Mascot *ion scores* abgebildet.

Dieses Peptid wurde mit einem durchschnittlichen geometrischen Mittelwert der TMT-Flächenverhältnisse von $4,992 \pm 0,905$ zu HEPES-Puffer zugeordnet (siehe Abbildung 4.2). Man sieht in der Vergrößerung des m/z -Bereichs, der die Signale der Reporterionen enthält, dass die Intensitäten/Flächen der Reporterionen 126, 127 und 128 niedriger, als diejenigen der Reporterionen 129, 130 und 131 sind. Zur Berechnung des durchschnittlichen geometrischen Mittelwertes der TMT-Flächenverhältnisse und der durchschnittlichen Standardabweichung zur Grenzwertbestimmung (siehe Tabelle 4.2) wurden also auch diese Signale geringerer Intensität mit eingerechnet. Für die beiden Puffer ergibt sich immer noch der Erwartungswert 1,0. In den Experimenten mit nur einem verwendeten Puffersystem, errechnet sich dieser Wert zwar auch über alle sechs verschiedenen TMT-Reporterionenkanäle, hier weisen sie allerdings auch die gleiche Intensität auf. Wie in Kapitel 3 bereits beschrieben, steht die analytische Varianz in Abhängigkeit von der Intensität der Reporterionensignale (Hu, *et al.*, 2006). Es ist also nicht überraschend, dass die Experimente mit zwei verwendeten Puffersystemen eine höhere durchschnittliche Standardabweichung aufweisen, da hier auch Signale mit geringeren Intensitäten in die Berechnung eingehen.

In Abbildung 4.2, Teil A sind beispielhaft drei TMT-markierte Peptide bzw. *unique* Spaltsequenzen, deren durchschnittliche geometrischen Mittelwerte der TMT-Flächenverhältnisse und ihre Klassifizierung zu TEAB-, HEPES- oder beiden Puffersystemen, gezeigt. Die Spaltstelle der Spaltsequenz LYPGRPLAIALDTK (dargestellt als “E.”) wurde über ihren durchschnittlichen geometrischen Mittelwert ($0,456 \pm 0,073$) zu TEAB-Puffer klassifiziert. Man beachte, dass in diesem Peptid die potentielle Spaltstelle nach einem Aspartat (Asp12) in diesem Puffersystem nicht hydrolysiert wurde. Die um zwei Aminosäuren verkürzte Spaltsequenz LYPGRPLAIALD, bei der diese Hydrolyse stattgefunden hatte, wurde hingegen HEPES-Puffer zugeordnet ($4,992 \pm 0,905$). Diese Zuordnung korreliert sehr gut mit der vorhergesagten GluC Spaltspezifität, da beide Spaltsequenzen N-terminal nach einem Glutamat in Position P1 gespalten wurden und in HEPES-Puffer zusätzlich eine Spaltstelle nach einem Aspartat hydrolysiert wurde. Die *unique* Spaltstelle VVLVGGSTR wurde nach einem Glutamat in Position P1 gebildet und wurde somit beiden Puffersystemen zugeordnet ($1,049 \pm 0,103$).

Wie zuvor beschrieben, wurde durch Sorensen *et al.* (1991) die These aufgestellt, dass GluC eine höhere Affinität für Glu-X gegenüber Asp-X Peptidbindungen aufweist. Dies ist von dem verwendeten Puffersystem unabhängig. Der inhibitorische Effekt von Bicarbonatpuffer auf die GluC Gesamtaktivität könnte dennoch in einer scheinbar ausschließlichen Affinität für Glu-X Peptidbindungen resultieren. Diese These wird durch die Ergebnisse der Q-PICS Experimente zur Untersuchung des Einflusses von TEAB- im Vergleich zu HEPES-Puffer auf die GluC Spaltspezifität bestärkt. Zum einen wurden mehr *unique* Spaltsequenzen in HEPES-, als in TEAB-Puffer identifiziert (500 vs. 375) und zum anderen wurden die meisten Spaltsequenzen, die nach einem Aspartat in Position P1 generiert wurden, zu HEPES-Puffer klassifiziert. Auch die beiden *Heat-Maps* in Abbildung 4.2, Teil C zeigen einen deutlichen Unterschied in der Spaltspezifität der Endoprotease GluC in den beiden Puffersystemen. Die Mehrheit der gemessenen durchschnittlichen geometrischen Mittelwerte der TMT-Flächenverhältnisse der einzelnen *unique* Spaltsequenzen überschreitet den Wert 6,0 (bzw. 0,167) nicht. Dies bedeutet, dass sie nicht ausschließlich in nur einem der beiden Puffersysteme generiert wurden, sondern lediglich zu einem größeren Anteil.

4.4.2 Untersuchung des Einflusses des Puffersystems auf die Spaltspezifität der Endoprotease GluC mittels *reversed* Markierung

Insgesamt wurden in dem *reversed* Markierungsexperimentes zur Validierung des Q-PICS Ansatzes weniger TMT-markierte Peptide und daraus resultierende *unique* Spaltsequenzen

identifiziert als in dem normalen Markierungsexperiment. Die prozentuale Verteilung dieser Spaltsequenzen zu TEAB-, HEPES- oder beiden Puffersystemen stimmt jedoch sehr gut überein (siehe Abbildung 4.3, Teil B). Auch der durchschnittliche Mascot *ion score* unterscheidet sich praktisch nicht in den beiden Markierungsexperimenten (siehe Tabelle 4.1). Bezüglich der Quantifizierung der TMT-markierten Peptide stellt man fest, dass die Abweichung des durchschnittlichen geometrischen Mittelwertes von dem Erwartungswert 1,0 mit <7% für das *reversed* Markierungsexperiment etwas höher lag (im Vergleich zu <5%). Auch die durchschnittliche Standardabweichung lag mit 26% etwas höher (im Vergleich zu 22%). Dies stellt jedoch keine drastische Verschlechterung dar. Der Grenzwert, ab dem ein quantitativer Unterschied zwischen verschiedenen biologischen Zuständen (in diesem Fall die Klassifizierung zu einem oder beiden Puffersystemen) feststellbar ist, muss für jedes Experiment erneut berechnet werden (siehe Tabelle 4.2).

Die drei *unique* Spaltsequenzen, die bereits beispielhaft für das normale Markierungsexperiment gezeigt wurden (siehe Abbildung 4.2, Teil A), wurden über ihre TMT-Flächenverhältnisse zu den gleichen Puffersystemen wie zuvor klassifiziert.

Auch wenn in dem *reversed* Markierungsexperiment insgesamt weniger *unique* Spaltsequenzen identifiziert wurden (346 *unique* Spaltsequenzen klassifiziert zu TEAB- und 431 zu HEPES-Puffer), so ist die Anzahl immer noch in guter Übereinstimmung mit den Ergebnissen des ursprünglichen PICS Ansatzes (Schilling and Overall, 2008). Auch der Unterschied der Spaltspezifität der Endoprotease GluC in den beiden Puffersystemen wird durch die beiden *Heat-Maps* in Abbildung 4.3, Teil C deutlich abgebildet.

Die Reproduzierbarkeit zwischen dem normalen und dem *reversed* Markierungsexperiment wird durch den *scatter plot* in Abbildung 4.5 wiedergegeben. Insgesamt wurden 154 *unique* Spaltsequenzen nur in dem normalen, 80 *unique* Spaltsequenzen nur im *reversed* und 405 *unique* Spaltsequenzen in beiden Markierungsexperimenten identifiziert. Durch den Labeltausch der TMT-Reporterionenkanäle für die beiden Puffersysteme in den unterschiedlichen Markierungsexperimenten erhält man im Idealfall für jedes TMT-Flächenverhältnis des normalen Markierungsexperimentes den reziproken Wert im *reversed* Markierungsexperiment. Dies wird z. B. durch die drei in den Abbildungen 4.2, Teil A und 4.3, Teil A gezeigten Spaltsequenzen deutlich. Auch die in Abbildung 4.4 dargestellten MS/MS-Spektren des Peptides LYPGRPLAIALD des Proteins Pyruvatkinase 1 zeigen deutlich die umgekehrte Verteilung der TMT-Reporterionensignale für die beiden Markierungsexperimente.

Multipliziert man die beiden durchschnittlichen geometrischen Mittelwerte der TMT-Flächenverhältnisse der beiden Markierungsexperimente einer *unique* Spaltsequenz, so errechnet sich der Erwartungswert 1,0. Der größte Teil der 405 *unique* Spaltsequenzen (68%) liegt um den Erwartungswert 1,0 in einem Bereich von 0,8 bis 1,2. Dies stimmt gut mit der allgemein für isobare *tags* erwarteten Standardabweichung von 20% überein (Dayon, *et al.*, 2008, Gluckmann, *et al.*, 2007, Ogata, *et al.*, 2007, Ross, *et al.*, 2004, Tiberti, *et al.*, 2010). Die übrigen 32% der *unique* Spaltsequenzen, die diesen Bereich überschreiten, weichen jedoch nicht viel von dieser Standardabweichung ab. Dies demonstriert die hohe Reproduzierbarkeit des Q-PICS Ansatzes. Wie zuvor beschrieben, ist auch zu bedenken, dass Quantifizierungen mit isobaren *tags* i. d. R. letztendlich auf Proteinebene berechnet werden und somit meist eine größere Anzahl an quantitativen Messwerten pro errechneten geometrischen Mittelwert vorliegt. Es ist somit nicht überraschend, dass man auf Peptid-Ebene eine etwas größere Standardabweichung erhält.

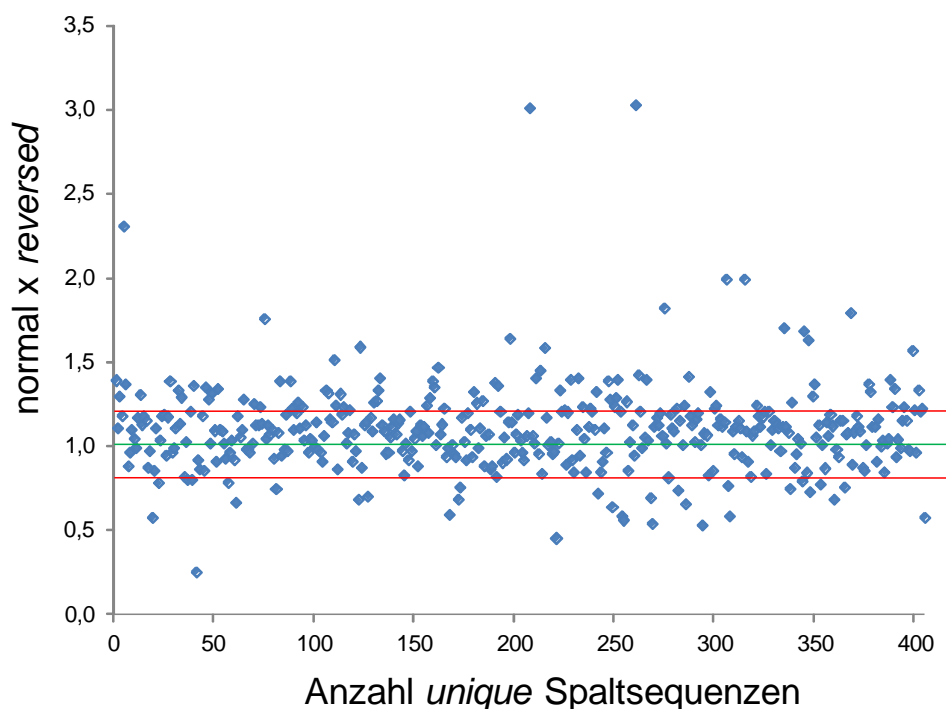


Abb. 4.5: Auftragung der multiplizierten durchschnittlichen geometrischen Mittelwerte der TMT-Flächenverhältnisse für jede der 405 *unique* Spaltsequenzen, die sowohl im normalen als auch im *reversed* Markierungsexperiment identifiziert wurden. Die Multiplikation der beiden TMT-Flächenverhältnisse entspricht im Idealfall der Multiplikation eines Wertes mit seinem reziproken Wert und errechnet sich dadurch zu dem Erwartungswert 1,0 (grüne Linie). Durch die rote Linie ist die allgemein akzeptierte Abweichung für isobare *tags* von 20% dargestellt.

4.5 Zusammenfassung und Ausblick

Insgesamt können mit dem Q-PICS Ansatz bis zu sechs verschiedene TMT-Reporterionenkanäle verwendet werden. Diese können dazu genutzt werden, sechs interne technische Replikate in nur einem LC-MALDI-TOF/TOF MS/MS Experiment zu untersuchen. Das quantitative Potential des Ansatzes ermöglicht allerdings auch den Vergleich verschiedener biologischer Zustände, in diesem Fall den Einfluss zweier verschiedener Puffersysteme auf die Spaltspezifität der Endoprotease GluC.

Das eigentliche Ziel dieses Experimentes, das quantitative Potential des Q-PICS Ansatzes zu belegen, wurde erreicht. Es wurden 405 Spaltsequenzen sowohl in einem normalen als auch in einem *reversed* Markierungsexperiment identifiziert. Über die TMT-Flächenverhältnisse wurde eine Klassifizierung durchgeführt, in welchem der beiden Puffersysteme (TEAB- und HEPES-Puffer) die Hydrolyse zur Generierung der Spaltsequenzen erfolgt ist. Fasst man nun die Ergebnisse der beiden Markierungsexperimente zur Untersuchung des Einflusses verschiedener Puffersysteme auf die Spaltspezifität der Endoprotease GluC zusammen, so wird die These von Sorensen et al. (1990) bestätigt. GluC weist eine höhere Affinität für Glu-X gegenüber Asp-X Peptidbindungen unabhängig von dem verwendeten Puffersystem auf. Der inhibitorische Effekt von Bicarbonatpuffer auf die GluC Gesamtaktivität resultiert dennoch in einer scheinbar ausschließlichen Affinität für Glu-X Peptidbindungen in diesem Puffer.

Durch das *reversed* Markierungsexperiment konnte zusätzlich die hohe Reproduzierbarkeit des Q-PICS Ansatzes gezeigt werden. Diese und die Robustheit der Quantifizierung sind v. a. auf die Verwendung dreier technischer Replikate in Kombination mit drei internen technischen Replikaten zurückzuführen. Die wohl größte Herausforderung des Q-PICS Ansatzes beruht auf der Quantifizierung auf Peptid-Ebene. So wurden im Vergleich zu Studien, die auf Quantifizierungen auf Proteinebene basieren, leicht erhöhte, durchschnittliche Standardabweichungen erreicht. Diese änderten jedoch nichts an dem quantitativen Potential des Q-PICS Ansatzes.

Theoretisch könnte man an Stelle von zwei auch drei biologische Zustände (z. B. ein weiteres Puffersystem) miteinander vergleichen. Dann könnte man für jeden dieser biologischen Zustände immer noch ein zusätzliches internes technisches Replikat einsetzen. Im Vergleich zu den Experimenten mit nur einem Puffersystem, in denen Parameter der LC-MALDI-

TOF/TOF MS/MS Methode optimiert wurden, wurden in dem quantitativen Experiment weniger Spaltsequenzen identifiziert. Die Gründe hierzu wurden bereits diskutiert.

Die Einführung eines weiteren biologischen Zustandes würde zu einer weiteren Erhöhung der Probenkomplexität bei gleichzeitiger Verringerung der Probenmenge pro einzelne Spaltsequenz führen. Die Signalintensitäten der Vorläuferionen würden also ebenfalls kleiner und damit die Identifizierungen über die Fragmentspektren schlechter werden. Man sollte also in einem solchen Fall das eindimensionale HPLC-Trennsystem auf eine zweidimensionale Trennung erweitern und eine höhere Probenmenge als die bisher verwendeten 2,5µg einsetzen.

Die Anzahl an quantitativen Messwerten pro Spaltsequenz könnte auch durch zusätzliche technische Replikate erhöht werden. Hierbei ist jedoch insbesondere auf die daraus resultierenden erhöhten Messzeiten zu achten und ob diese Erhöhung mit einer verbesserten robusten Quantifizierung einhergeht.

Nachdem das quantitative Potential des Q-PICS Ansatzes nun gezeigt wurde, sollten die unterschiedlichsten Fragestellungen mit diesem Ansatz analysiert werden können. So könnte man beispielsweise eine Protease unter Verwendung verschiedener Inhibitoren miteinander vergleichen, oder den Einfluss verschiedener Bedingungen des Verdaus wie Temperatur, Inkubationszeiten (oder wie gezeigt: Puffersysteme) untersuchen.

Bezüglich verschiedener Inkubationszeiten könnte man auch so genannte *time-course*-Experimente erstellen und damit kinetische Untersuchungen durchführen. Somit könnte beispielsweise analysiert werden, welche Spaltstellen durch eine Protease besonders schnell hydrolysiert werden, weil sie beispielsweise besonders gut zugänglich für die verwendete *target*-Protease sind. Es gibt jedoch einige Bedingungen bei solchen *time-course*-Experimenten auf die geachtet werden muss. So ist ein entscheidender Punkt, dass die Aktivität der *target*-Protease bei der Probennahme unmittelbar und quantitativ gestoppt werden muss. Wird die Aktivität in der entnommenen Probe nicht gestoppt läuft die Spaltung der Substrate auch nach der Probennahme weiter ab und eine Unterscheidung der verschiedenen Zeitpunkte ist nicht möglich. Das Abstoppen der Protease-Aktivität kann dabei auf verschiedene Arten erfolgen. So können beispielsweise spezifische Proteaseinhibitoren zugegeben werden (Dean and Overall, 2007, Gregorius, *et al.*, 2013) oder Chymotrypsin durch Harnstoff inaktiviert werden (Lathia *et al.*, 2010). Bei der Zugabe solcher Reagenzien zur Deaktivierung der Enzyme ist jedoch unbedingt darauf zu achten, dass sie mit den nachfolgenden Analysentechniken oder Probenvorbereitungsschritten kompatibel sind. Schlüter *et al.* nutzten die Immobilisierung der *target*-Proteasen an *beads*, so dass die

enzymatische Reaktion automatisch gestoppt wird, sobald ein Probenaliquot aus dem Reaktionsgemisch entfernt wird (Schluter *et al.*, 2003). Dadurch umgehen sie das Problem der Zugabe von Reagenzien.

Einen weiteren wichtigen Punkt stellt die eigentliche Quantifizierung dar. Hier können z. B. die Signalintensitäten des gespaltenen Substrats mit einem internen Standard im gleichen MS-Spektrum miteinander verglichen werden (Schluter, *et al.*, 2003, Tholey *et al.*, 2006). Eine weitere Möglichkeit zur Quantifizierung auf MS-Ebene ist die Markierung von Peptiden mittels DOTA (siehe Kapitel 1.2). Durch die Verwendung verschiedener Metalle sowie eines internen Produkt-/Substratstandards wurden *time-courses* der Neurotensin-Proteolyse ausgehend von 8 unterschiedlichen Substratkonzentrationen parallel analysiert (Gregorius, *et al.*, 2013). Das so genannte *multiplexing* wird auch durch die Verwendung isobarer *tags* wie iTRAQ ermöglicht. Dies wurde z. B. für die Analyse von Substraten der Metalloprotease MMP2 gezeigt (Dean and Overall, 2007). Aber auch andere quantitative MS-Techniken wie das *single reaction monitoring* (SRM) wurden bereits zur Erstellung solcher *time-course*-Experimente verwendet (Agard *et al.*, 2012).

Die Anwendung genomischer und proteomischer Analysentechniken zur Untersuchung von Proteasen und Proteasen-Substraten, das so genannte Degradom, stellt insbesondere für die Proteomik eine große Herausforderung dar (Doucet *et al.*, 2008, Lopez-Otin and Overall, 2002, Overall, *et al.*, 2004). In diesem Zusammenhang kann der Q-PICS Ansatz als eine weitere Analysetechnik hinzugefügt werden. So könnte man eine gut definierte Peptidbibliothek mit einem Zellsystem und einer hochregulierten bzw. überexprimierten Protease (z. B. ADAM 10 oder 17) inkubieren und somit die Oberflächenaktivität dieser Protease untersuchen. Eine Unterscheidung der Peptide der Peptidbibliothek von Peptiden, die letztendlich aus dem Zellsystem stammen, könnte beispielsweise über eine Isotopenmarkierung erfolgen. So könnte man zur Herstellung der Peptidbibliothek eine isotopenmarkierte Zellkultur verwenden.

In Hinblick auf Proteasen spielen beispielsweise bei Tumoren, häufig auch posttranslationale Modifikationen wie Phosphorylierungen und damit Kinasen/Phosphatasen eine wichtige Rolle (Lopez-Otin and Hunter, 2010). Während es in dieser Veröffentlichung eher um die regulatorischen Effekte von Proteasen durch Phosphorylierungen bzw. von Kinasen durch proteolytische Prozessierung geht, könnte der Q-PICS Ansatz hier in einem anderen Zusammenhang eingesetzt werden. So könnte man eine Peptidbibliothek beispielsweise mit Proteinkinase A behandeln und somit Phosphorylierungen einführen. Auch eine chemische Modifizierung zur Einführung solcher Phosphorylierungen wäre denkbar. Durch den Q-PICS

Ansatz könnte dann der Einfluss von Phosphorylierungen auf die Spaltspezifität verschiedener Proteasen auf Substratebene untersucht werden, also ob die Hydolyse einer Peptidbindung auch dann noch stattfindet, wenn sich eine Phosphorylierung in ihrer Umgebung befindet. Diese könnte Auswirkungen auf die Substratbindung im katalytischen Zentrum der Protease haben und somit einen Einfluss auf die Spaltspezifität ausüben. Dieser Einfluss könnte anschließend mit dem Q-PICS Ansatz auch quantitativ analysiert werden. Die eigentliche Phosphorylierung ist in der mittels MS/MS identifizierten Spaltsequenz nicht notwendigerweise enthalten, abhängig davon, ob sie N- oder C-terminal der Spaltstelle liegt. In einem solchen Experiment müsste deshalb sichergestellt werden, welche Phosphorylierungsstellen auch wirklich phosphoryliert vorliegen.

Kapitel V: Schnittstellen-Identifizierung in ARTC2.2 mittels LC-MALDI-TOF/TOF MS/MS

5.1 Einleitung

Das Ziel dieses Teilprojektes war die Identifizierung der C-terminalen Spaltstelle (*cleavage site*) innerhalb des Proteins ARTC2.2. Es handelt sich hierbei um eine GPI-verankerte Ekto-ADP-Ribosyltransferase. Diese werden auf den Oberflächen von murinen T-Zellen und Makrophagen exprimiert (Hong *et al.*, 2007, Hong *et al.*, 2009, Koch-Nolte *et al.*, 1996). Durch Aktivierung mittels NAD^+ führt ARTC2.2 eine Arginin-spezifische ADP-Ribosylierung des Ionenkanals P2X7 und anderer Zelloberflächenproteine durch (Adriouch *et al.*, 2008, Seman *et al.*, 2003). Diese ADP-Ribosylierung aktiviert wiederum eine Oberflächenprotease, die das *shedding* der Ektodomäne von L-Selektin (CD62L) steuert (Scheuplein *et al.*, 2009). Die ARTC2-Enzyme werden dann aber vermutlich selbst über eine Metalloprotease proteolytisch prozessiert. Die Spaltstellen sind derzeit jedoch unbekannt.

Die Aminosäuresequenz des hier verwendeten ARTC2.2 ist in Abbildung 5.1 dargestellt. Der vermutliche Bereich der Metalloproteasen-Spaltung ist hellblau gekennzeichnet. ARTC2.2 trägt zwei potentielle N-Glykosylierungsstellen Asn90 und Asn260 (Prof. Dr. Fritz Koch-Nolte per mündliche Mitteilung). Das hier untersuchte ARTC2.2 trägt ein N-terminales FLAG-tag mit der Peptidsequenz N-DYKDDDDK-C (monoisotopische Masse: 1012,39 Da). Es kann also über den kommerziell verfügbaren M2-mAb Antikörper aufgereinigt werden.

Die ARTC2-Enzyme sind C-terminal über einen GPI-Anker kovalent mit der Zellmembran verbunden (Koch-Nolte, *et al.*, 1996). Im Endoplasmatischen Retikulum werden GPI-verankerte Proteine zunächst als Vorläufer-Proteine exprimiert, die eine hydrophobe C-Terminale *pre-anchor* Sequenz (PAS) enthalten. Diese besteht aus 20-37 Aminosäuren und enthält die Spaltstelle, an welcher dann der GPI-Anker angebracht wird (Low, 1987). Die vermutete PAS des ARTC2.2 ist in Abbildung 5.1 rot gekennzeichnet. Das Vorläufer-Protein wird anschließend posttranslational modifiziert, indem ein Teil der PAS durch den GPI-Anker ersetzt wird (Ferguson and Williams, 1988). Der GPI-Anker sollte also innerhalb des rot gekennzeichneten Bereichs der ARTC2.2-Sequenz lokalisiert sein. Er kann z. B. durch die bakterielle Phospholipase C (PI-PLC) gespalten werden. Die Strukturformel des GPI-Ankers, sowie die Spaltstelle durch die Phospholipase sind Abbildung 5.2 zu entnehmen.

M A L P V T A L L L P L A L L L H A A R
P **D Y K D D D D K** R S L A V P F M L D M
A P N A F D D Q Y E G C V E D M E K K A
P Q L L Q E D F N M N E E L K L E W E K
A E I N W K E I K **N** S T S Y P A G F H D
F H G T A L V A Y T G N L A I D F N R G
V R D F K K S P D N F H Y K A F H Y Y L
T R A V Q L L N D Q G C S L V Y R G T K
V M F E Y T G K G S V R F G Q F S S S S
L T K R V A L S S N F F S N H G T L F I
I R T C L G V N I K E F S S F P R E E E
V L I P G Y E V Y H K V T A Q N D N G Y
N E I F L D S P E R K K S N F N C **F Y** **V**
G S A Q T V N I D F S I S G S R E S C V
S L F L V V L L G L L V Q Q L T L A E P

FLAG-TAG

Potentielle Glykosylierungsstelle

Bereich der Metalloprotease-Spaltung

PAS (hydrophobic pre-anchor sequence, meist 20-37 AS)

Abb. 5.1: Aminosäuresequenz des analysierten ARTC2.2. Das N-terminale FLAG-tag ist gelb unterlegt, die beiden potentiellen Glykosylierungsstellen Asn90 und Asn260 grün. Der vermutete Bereich der Metalloprotease-Spaltung ist in blau, die pre-anchor Sequenz (PAS) in rot markiert (Quelle: Prof. Dr. Fritz Koch-Nolte, Institut für Immunologie Universitätsklinikum Hamburg-Eppendorf).

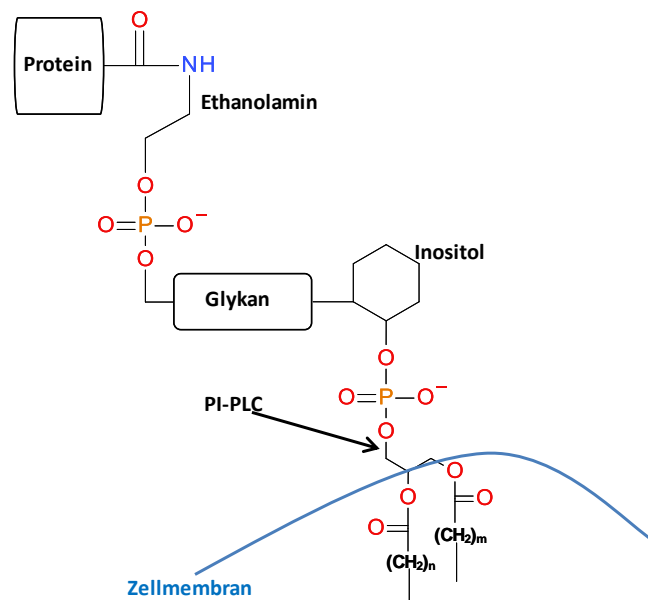


Abb. 5.2: Schematische Darstellung eines über einen GPI-Anker mit der Zell-Membran verbundenen Proteins und der darin enthaltenen PI-PLC-Spaltstelle.

Die Aufgabenstellung dieses Projektes beruhte auf der Identifizierung der C-terminalen Metalloprotease-Spaltstelle des ARTC2.2-Enzyms. Die Spaltung erfolgt dabei vermutlich über ADAM 10 oder ADAM17 bzw. TACE (Daten nicht gezeigt). Das proteolytisch prozessierte ARTC2.2-Enzym wird im Folgenden als ARTC2.2tace bezeichnet.

Die Identifizierung der C-terminalen Anlagerungsstelle des GPI-Ankers wurde zurückgestellt und in diesem Projekt nicht bearbeitet.

Die Aufreinigung des proteolytisch prozessierten ARTC2.2tace wurde bis zur Auftrennung über ein 10% Novex Fertigel in der Arbeitsgruppe von Prof. Dr. Fritz Koch-Nolte vom Institut für Immunologie am Universitätsklinikum Hamburg-Eppendorf durchgeführt.

5.2 Material und Methoden

Entionisiertes Wasser (18.2 M Ω *cm) wurde mit einem arium®611VF System (Sartorius, Göttingen, Deutschland) hergestellt und für alle Experimente verwendet. Acetonitril (E Chromasolv), Glu¹-Fibrinopeptid B, Dithiothreitol ($\geq 99.5\%$), Pepsin, GluC und Iodacetamid (Sigma Ultra) wurden von Sigma-Aldrich (Steinheim, Deutschland) bezogen. Ammoniumbicarbonat ($\geq 99.5\%$), Trifluoressigsäure und α -Cyano-4-hydroxymizinsäure (CHCA, $\geq 99\%$) wurden von Fluka und *Sequencing grade modified* Trypsin von Promega (Madison, WI, USA) erworben. Das zur MS-Kalibrierung verwendete 4700 Proteomics Analyzer Standard Kit wurde von AB SCIEX (Darmstadt, Deutschland) und 95% ¹⁸O-markiertes Wasser von Cambridge Isotope Laboratories, Inc. Andover, MA USA bezogen.

Probenvorbereitung

ARTC2.2tace wurde aus dem Überstand von 10⁹ transient transfizierten HEK Zellen nach Behandlung mit NAD und ATP isoliert. Die Reinigung erfolgte über eine 1 ml M2-Säule. Die Elution erfolgte mit 100 mM Glycin, pH 2,5, Neutralisation durch Zugabe von 100 μ l 1 M Tris pH 9 und Umpufferung in PBS über eine PD-10 Säule, wobei BSA als Carrierprotein zugegeben wurde. Die Probe wurde dann über einen Zentrifugationsfilter (*cutoff* 10.000 Da) auf ca. 120 μ l eingengt. Ein Viertel der Präparation wurde an 15 μ l einer Amino Link matrix (Pierce) mit kovalent gebundenem anti-ART2.2 *nanobody* (*Lama single domain antibody*) gebunden, gewaschen, mit SDS-haltigem endoF Puffer versetzt und nach Zugabe von endoF über Nacht bei 37°C inkubiert. Das Protein wurde mit 4x SDS-PAGE Probenpuffer ohne DTT 15 min bei 70°C von der Matrix gelöst, und über ein 10% Novex Fertigel größenfraktioniert (siehe Abbildung 5.3). Die Färbung des Gels erfolgte mit Colloidal Coomassie, die

Entfärbung mit entionisiertem Wasser. Die Banden wurden anschließend ausgeschnitten und in Eppendorfgefäße überführt.

In-Gel Verdau

Die, wie zuvor beschrieben, behandelten Gelstücke wurden in 100 µl Wasser aufgenommen und für 5 min bei 37°C und 600 U/min im Thermomixer (Eppendorf AG, Hamburg, Deutschland) gewaschen und der Überstand verworfen. Anschließend wurden die Gelstücke in 100 µl Wasser/Acetonitril 50:50 (v/v) für 5 min bei 37°C und 600 U/min entfärbt und der Überstand erneut verworfen. Die Reduktion der Probe erfolgte durch Zugabe von 100 µl 10 mM DTT-Lösung für 60 min bei 56°C und 600 U/min im Thermomixer. Der Überstand wurde anschließend verworfen. Zum Waschen des Reaktionsgemisches wurden für 5 min bei 37°C und 600 U/min im Thermomixer 100 µl Wasser zugegeben und der Überstand verworfen. Um vorhandene Cysteinseitenketten mittels Carbamidomethylierung zu alkylieren, wurden 100 µl 55 mM Iodacetamid-Lösung zugegeben und für 30 min bei RT im Dunkeln inkubiert. Der Überstand wurde verworfen. Um überschüssige Reagenzien zu entfernen, wurden 100 µl Wasser/Acetonitril 50:50 (v/v) zugegeben und für 5 min bei 37°C und 600 U/min im Thermomixer inkubiert. Der Überstand wurde verworfen. Diese Waschprozedur wurde fünfmal wiederholt. Es wurden 100 µl Acetonitril zugegeben und für 5 min bei 37°C und 600 U/min im Thermomixer inkubiert. Der Überstand wurde verworfen und der Schritt noch einmal wiederholt. Die Gelstücke wurden anschließend für 25 min an der Raumluft getrocknet. Für den Verdau der Probe P1 (siehe Abbildung 5.3) wurden 20 µg Trypsin in 40 µl 1 M HCL in ¹⁸O-markiertem Wasser gelöst, zu je 10 µl aliquotiert und anschließend mit jeweils 290 µl 40 mM Ammoniumbicarbonat-Puffer mit 95% ¹⁸O-markiertem Wasser pH 8,5 versetzt. Die Probe P1 wurde mit der ¹⁸O-markierten Trypsin-Lösung versetzt und über Nacht bei 37°C und 500 U/min im Thermomixer inkubiert. Zusätzlich wurde mit Probe P2 ein Verdau mit der Endoprotease GluC durchgeführt (siehe Abbildung 5.3). Hierzu wurden 50 µg GluC in 100 µl 95% ¹⁸O-markiertem Wasser gelöst und zu je 10 µl aliquotiert. Eines dieser Aliquote wurde mit 290 µl 40 mM Ammoniumbicarbonat in dem ¹⁸O-markierten Wasser pH 8,5 aufgefüllt. Der Verdau wurde nach Zugabe von 100 µl dieser GluC-Lösung über Nacht bei 37°C, mit 300 U/min im Thermomixer durchgeführt.

Trennung mittels IP-RP-HPLC bei pH 2,1

Für die Trennung der In-Gel Verdaue mittels IP-RP-HPLC bei pH 2,1 wurden jeweils 20 µl der wie zuvor beschrieben vorbereiteten Proben mit einem U3000 nanoHPLC-System

(Dionex, Idstein, Deutschland) über eine 50 µl Probenschleife injiziert und auf einer Acclaim PepMap100 C18 *Trap*-Säule (300 µm I.D. x 10 mm, 5 µm, Dionex, Idstein, Deutschland) aufkonzentriert. Die *Trap*-Säule wurde 8 min mit 0,1%iger (v/v) wässriger Trifluoressigsäure-Lösung mit 3% (v/v) Acetonitril und einer Flussrate von 30 µl/min mit Peptiden beladen. Die Analyten wurden dann auf eine Acclaim PepMap100 C18 Trennsäule (75 µm I.D. x 150 mm, Dionex, Idstein, Deutschland) gespült. Die Trennung wurde unter Verwendung eines Gradienten aus Eluent A (0,05% (v/v) wässrige Trifluoressigsäure) und Eluent B (0,04% (v/v) wässrige Trifluoressigsäure in 80% Acetonitril) durchgeführt, der Tabelle 5.1 zu entnehmen ist. Die Flussrate betrug 0,3 µl/min bei einer Säulentemperatur von 30°C. Die Detektion erfolgte mittels eines UV-Detektors mit Z-förmiger Detektorzelle mit einem Detektorvolumen von 3 nl und einer Wellenlänge von 214 nm.

Tab. 5.1: Zeitlicher Verlauf des Gradienten für die Trennung der Peptide der In-Gel-Verdaue von ARTC2.2tace mittels IP-RP-HPLC.

Zeit [min]	Eluent A [%]	Eluent B [%]
0,0	95	5
8,0	95	5
39,0	50	50
44,0	5	95
49,0	5	95
50,0	95	5
69,0	95	5

Nach Elution von der Säule wurden die Peptide ab der fünfzehnten Minute im Verhältnis 1:4 (v/v) mit MALDI-Matrix (α -Cyano-4-hydroxymethylsäure) durch einen Probot Mikrofraktionssammler (LC Packings, Dionex, Amsterdam, Niederlande) über ein T-Stück mittels Spritzenpumpe gemischt und alle 15 s auf ein MALDI-Target (Opti-TOF™ LC MALDI Insert, AB SCIEX, Darmstadt, Deutschland) gespottet. Das Spotten erfolgte für 50 min und die Matrix-Stockkonzentration betrug 3 mg/ml in 70% (v/v) Acetonitril/0.1% (v/v) wässriger Trifluoressigsäure und enthielt ca. 5 nM Glu¹-Fibrinopeptid B zur internen Kalibrierung.

Massenspektrometrische Analyse mittels MALDI-TOF/TOF MS

Zur massenspektrometrischen Analyse der In-Gel Verdaue und Datenprozessierung (GPS-Software, Version 3.6, AB SCIEX, Darmstadt, Germany) wurde ein MALDI-TOF/TOF 5800

Massenspektrometer (AB SCIEX, Darmstadt, Deutschland) verwendet. MS-Messungen wurden im *positive ion mode* im Massenbereich von 800 – 4000 m/z und einer Akkumulation von 1000 Laserschüssen durchgeführt. Die MS-Kalibrierung erfolgte mittels Standardpeptiden (4700 Proteomics Analyzer Mass Standard Kit, AB SCIEX, Darmstadt, Deutschland), welche den erforderlichen Messbereich abdecken, in Form einer *default calibration*. Zusätzlich wurde während der LC-MALDI MS-Analysen mit Glu¹-Fibrinopeptid B als internem Standard kalibriert. Die Liste der *precursor* für die MS/MS-Analyse wurde automatisch durch die Software des Massenspektrometers unter Verwendung der folgenden Parameter erstellt: Minimales S/N-Verhältnis 100; *precursor* Massentoleranz: ± 200 ppm; das Signal musste nur in einem Spot detektiert werden, um als *precursor* akzeptiert zu werden und maximal wurden 10 *precursor* pro Spot für die MS/MS-Analyse ausgewählt. Die MS/MS-Spektren wurden mit 1 kV mittels CID, unter Akkumulation von 2000 Laserschüssen generiert. Die MS/MS-Kalibrierung wurde mit Glu¹-Fibrinopeptid B aus dem 4700 Proteomics Analyzer Mass Standard Kit (AB SCIEX, Darmstadt, Deutschland) durchgeführt.

Proteinidentifizierung mittels Datenbankanalyse

Die Rohdaten wurden unter Verwendung der TS2Mascot Software (Matrix Sciences, London, England) in *mascot generic format* (mgf)-Dateien umgewandelt. Dabei wurden folgende Einstellungen verwendet: Massenbereich von 60 Da bis zu der Masse des *precursors* minus 35 Da; minimales S/N-Verhältnis von 10, wobei nur die monoisotopischen Signale exportiert wurden. Die Mascot Datenbanksuche wurde mit der Mascot 2.2.04 Software gegen die Swiss-Prot FASTA-Datenbank mittels des Taxonomie-Filters *Mus musculus* (26. Januar 2012, 16345 Sequenzen) durchgeführt. Diese enthielt zusätzlich die in Abbildung 5.1 aufgeführte Aminosäuresequenz von ARTC2.2, sowie die Aminosäuresequenz der jeweils verwendeten Protease. Folgende Suchparameter wurden eingestellt: Enzym: semi Trypsin bzw. semi GluC, wobei keine C- oder N-terminale Orientierung berücksichtigt wurde; erlaubte ausgelassene Schnittstellen (*missed cleavages*): 2; feste Modifizierungen: Carbamidomethylierung (C); variable Modifizierungen: Label 18O (1) (C-Term), Label 18O (2) (C-Term), Deamidierung (NQ), Deamidierung 18O (NQ) und Oxidierung (Methionin); die Massentoleranz für die *precursor* wurde auf ± 50 ppm und die der MS/MS-Fragmentionen auf $\pm 0,5$ Da festgesetzt. Der Vertrauensbereich für die Proteinidentifizierung wurde auf $\geq 95\%$ ($p < 0,05$) eingestellt. Der *identity threshold* der Peptide betrug für Trypsin einen Mascot *ion score* von 51 und für GluC von 52. Dadurch wurde eine falsch-positiv Rate (FDR) $< 3\%$ erhalten.

5.3 Ergebnisse

Ziel des Projektes war die Identifizierung der C-terminalen Spaltstelle (*cleavage site*) innerhalb des Proteins ARTC2.2, das durch eine Metalloprotease (vermutlich ADAM 10 oder ADAM17 bzw. TACE) proteolytisch prozessiert wird.

In Abbildung 5.3 ist das mit Colloidal Coomassie gefärbte Gel der Auftrennung in zwei technischen Replikaten des Proteins ARTC2.2tace aus HEK-Zellen nach der Behandlung mit endoF gezeigt.

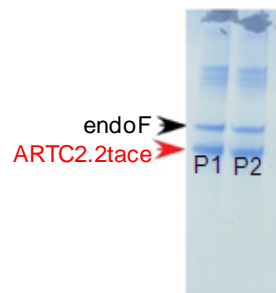


Abb. 5.3: Gelauf trennung der Probe des Proteins ARTC2.2tace. Es wurden zwei technische Replikate (P1 und P2) angefertigt. Das zur Abspaltung der Glykosylierung verwendete Protein endoF konnte innerhalb der Banden von ART2.2tace abgetrennt werden.

Die Strategie zur Identifizierung der Metalloprotease-Spaltstelle innerhalb des Proteins ARTC2.2 ist in Abbildung 5.4 dargestellt. Insgesamt wurden zwei verschiedene Proteasen für jeweils einen In-Gel Verdau verwendet. So wurde die Gelbande, die das Zielprotein ARTC2.2tace enthielt einmal tryptisch (Abbildung 5.3; P1) und einmal mit GluC (Abbildung 5.3; P2) proteolytisch prozessiert. Dabei wurde als Lösungsmittel isotopenmarkiertes H_2^{18}O verwendet. Durch den Verdau mit Trypsin bzw. GluC werden (ein oder) zwei neue Sauerstoffatome auf der C-terminalen Seite des neugebildeten Peptides eingeführt. Wird der Verdau nun in H_2^{18}O durchgeführt, so erhält man Peptide, die an ihrem neugebildeten C-Terminus mit zwei ^{18}O -Atomen isotopenmarkiert sind (Schnolzer, *et al.*, 1996).

Nur das Peptid mit dem C-terminalen Ende, welches aus der Metalloproteasen-Spaltung resultiert, kann die herkömmlichen ^{16}O -Isotope besitzen, da zum Zeitpunkt dieser Spaltung kein H_2^{18}O vorliegt und somit auch keine ^{18}O -Isotope in das Peptid eingebaut werden können. Eine Zuordnung, ob ein identifiziertes Peptid durch den In-Gel Verdau entstanden ist oder nicht ist damit also möglich. Alle Peptide, die C-terminal mit einem oder zwei ^{18}O -Isotopen markiert sind müssen aus dem In-Gel Verdau stammen.

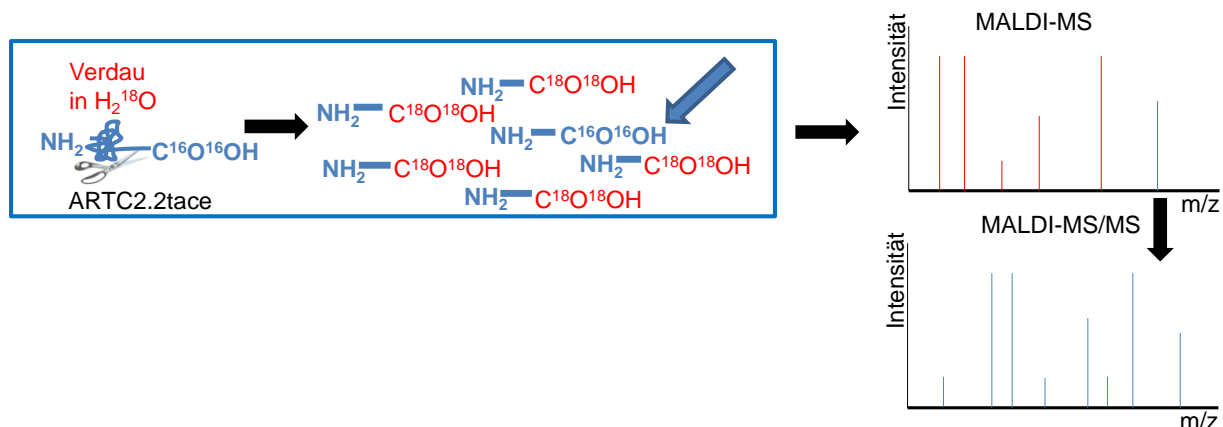


Abb. 5.4: Schematische Darstellung der Strategie zur Identifizierung der C-terminalen Metalloprotease-Spaltstelle innerhalb von ARTC2.2. Der Verdau von ART2.2tace erfolgt in $H_2^{18}O$. Dadurch werden, in Abhängigkeit des Reaktionsmechanismus der verwendeten Protease, ein oder zwei ^{18}O -Atome in den C-Terminus der Spaltsequenz eingebaut (Schnolzer, *et al.*, 1996).

Für die Datenbanksuche wurde eine Swiss-Prot Datenbank verwendet, die durch die Taxonomie Einstellung *Mus musculus* beschränkt wurde. Unter dieser Einstellung sind die beiden Proteine NAR2A_Mouse und NAR2B_Mouse enthalten, die eng mit dem Protein ARTC2.2 verwandt sind. Abbildung 5.5 zeigt das Alignment der Aminosäuresequenzen dieser eng verwandten Proteine mit der Sequenz des Proteins ARTC2.2.

```

A MALPVTALLLP LALLLHAARPDYKDDDDKRS LAVPFMLDMAPNAFDDQYEG CVEDMEKKA 60
B --MPSNNFKFFLTWWL TQQVTG----- LAVPFMLDMAPNAFDDQYES CVEDMEKKA 49
C --MPSNNFKFFLTWWL TQQVTG----- LAVPFMLDMAPNAFDDQYEG CVEDMEKKA 49
  : * . : : * : * . . *****
A PQLLQEDFNMNEELKLEWEKAEINWKEIKNSTSYYPAGFHDFHGTALVAYTGNLAIDFNRG 120
B PQLLQEDFNMNEELKLEWEKAEINWKEIKNSTSYYPAGFHDFHGTALVAYTGNLAIDFNRA 109
C PQLLQEDFNMNEELKLEWEKAEIKWKEIKNCMSYPAGFHDFHGTALVAYTGN IHRSLNEA 109
  *****: *****. *****: . : * . .
A VRDFKKS E DNFHYKAFHYL TRAVQLLNDQGC SLVYRGTKVMFEYTGKGSVRF GQFSSSS 180
B VRDFKKS E DNFHYKAFHYL TRAVQLLNDQGC SLVYRGTKVMFEYTGKGSVRF GQFSSSS 169
C TRDFKKS E DNFHYKAFHYL TRALQLLSDQGRS VYRGTNVRF RYTGKGSVRF GHFASSS 169
  . * : * * . * . *****: * * * * * * * * * * * * * * * * * * * * * * * *
A LTKRVALSSNFFSNHGT LFIIRTCLGVNIKEFSSFPREEEVLIPGYEVYHKVTAQNDNGY 240
B LTKRVALSSNFFSNHGT LFIIRTCLGVNIKEFSSFPREEEVLIPGYEVYHKVTAQNDNGY 229
C LNRSVATSSPFFENGQGT LFIIKTCLGAHIKHC SYTHEEEVLIPGYEVFHKVKTQSVERY 229
  * . : * * * * * . : *****: *****: * * . : * * * * * * * * * * * * *
A NEIFLDSPERKKS NFNCFYNGSAQTVMIDFSISGSRRESCVSLFLVLLGLLVQQLTLAEP 300
B NEIFLDSPERKKS NFNCFYNGSAQTVMIDFSISGSRRESCVSLFLVLLGLLVQQLTLAEP 289
C IQISLDSPEKRRKSNFNCFYSGSTQAANVSS--LGSRESCVSLFLVLLGLLVQQLTLAEP 287
  : * * * * * : * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
    
```

Abb. 5.5: Alignment der Sequenz von ARTC2.2 (A) mit der Sequenz von NAR2B_MOUSE (B) und NAR2A_MOUSE (C). Das Alignment wurde auf <http://www.uniprot.org/> mit der dort zur Verfügung stehenden Software erstellt.

Während NAR2B_MOUSE und ARTC2.2 kaum Sequenzunterschiede aufweisen, sind diese im Vergleich zu NAR2A_MOUSE viel ausgeprägter. Dies spielt insbesondere für den potentiellen Bereich der TACE-Spaltstelle eine wichtige Rolle (siehe Abbildung 5.1), da hier die Sequenzunterschiede ebenfalls sehr ausgeprägt sind. Das Alignment der drei Proteine zeigt also, dass es Sequenzunterschiede gibt. Diese Unterschiede können letztendlich dazu führen, dass eine Unterscheidung der drei Proteine über sogenannte *unique* Peptide möglich ist. Unter *unique* Peptiden versteht man solche Peptide, die für ein bestimmtes Protein einzigartig sind.

5.3.1 Ergebnisse des tryptischen Verdaus des Proteins ARTC2.2ace

Die Ergebnisse der LC-MALDI MS/MS Analyse und der anschließenden Datenbanksuche sind in Abbildung 5.6 dargestellt und die identifizierten Peptide in Tabelle 5.3 aufgelistet. Das Protein NAR2A_MOUSE wurde nicht identifiziert. Trotz der hohen Sequenzübereinstimmung zwischen NAR2B_MOUSE und ARTC2.2, wurde letzteres über *unique* Peptide identifiziert.

ARTC2.2, score: 1782, SC: 63%, MS-Massengenauigkeit: 5ppm

```

1  MALPVTALLL PLALLLHAAR PDYKDDDDKR SLAVPFMLDM APNAFDDQYE
51  GCVEDMEKKA PQLLQEDFNM NEELKLEWEK AEINWKEIKN STSYPAGFHD
101 FHGTALVAYT GNLAIDFNRG VRDFKKSPDN FHYKAFHYLL TRAVQLLNDQ
151 GCSLVYRGTK VMFEYTGKGS VRFQQFSSSS LTKRVALSSN FFSNHGTLFI
201 IRTCLGVNIK EFSSEFPREEE VLIPGYEVYH KVTAQNDNGY NEIFLDSPER
251 KKSFNCFYN GSAQTVNIDE SISGSRESCV SLFLVVLLGL LVQQLTLAEP

```

Abb. 5.6: Ergebnisse der Datenbankrecherche in der Swiss-Prot Datenbank mit Taxonomie *Mus musculus* zusätzlich der Aminosäuresequenzen der Proteasen Trypsin und GluC sowie des in Abbildung 5.1 abgebildeten ARTC2.2. SC = *sequence coverage*, der *score* bezieht sich auf die Proteinidentifizierung.

Zur Verdeutlichung der unterschiedlichen Aminosäure-Modifikationen innerhalb der identifizierten Peptide wurde eine Farbcodierung eingeführt (siehe Tabelle 5.2).

Mit dieser Codierung werden die modifizierten Aminosäuren innerhalb einer Peptidsequenz hervorgehoben. Da einige Peptide mit unterschiedlichen Modifikationen identifiziert wurden, soll dies die Unterscheidung der Peptide vereinfachen.

Tab. 5.2: Farbcodierung der unterschiedlichen Aminosäuren-Modifikationen.

Modifikation	Farbcodierung
Carbamidomethylierung	blau
Deamidierung mit einem ^{18}O -Isotop	orange
Deamidierung mit einem ^{16}O -Isotop	rot
Markierung am C-Terminus mit einem ^{18}O -Isotop	grün
Deamidierung mit einem ^{16}O -Isotop und Markierung am C-Terminus mit einem ^{18}O -Isotop	grau

Insgesamt wurden drei Peptide identifiziert, deren C-Terminus innerhalb der potentiellen Metalloproteasen-Spaltstelle liegt. Das Peptid KSNFNC**CFY**NGSAQ**TV** enthält eine Carbamidomethylierung an der Cysteinseitenkette (Cys6) und ist am Asn9 deamidiert. Da es C-terminal jedoch eine ^{18}O -Markierung enthält, handelt es sich dabei nicht um die Metalloprotease-Spaltstelle, sondern um ein Peptid, das durch den tryptischen Verdau generiert wurde.

Das Peptid SNFN**CFY**NGSAQ**TV**NIDF wurde mit zwei verschiedenen Modifikationen identifiziert. So wurde es einmal als SNFN**CFY**NGSAQ**TV**NIDF mit einer Carbamidomethylierung an Cys5 und einer Deamidierung des Asn8 sowie als SNFN**CFY**NGSAQ**TV**NIDF mit einer Carbamidomethylierung an Cys5, einer Deamidierung des Asn8 sowie einer Deamidierung mit einem ^{18}O -Isotop an Asn4 zugeordnet.

Die Deamidierung mit dem leichten ^{16}O -Isotop aller drei identifizierten Peptide entspricht einer der potentiellen Glykosylierungsstellen (Asn260) von ARTC2.2 (siehe Abbildung 5.1). Auch die zweite potentielle Glykosylierungsstelle wurde innerhalb des Peptides EIK**N**STSY**PAG**FHDFHG**TALV** identifiziert. Dabei wurde ein Mascot *ion score* von 98 erreicht und die Glykosylierungsstelle in Form des deamidierten Asn4 festgestellt. Das Peptid enthielt zusätzlich eine ^{18}O -Isotopenmarkierung am C-Terminus. Es stammt also aus dem tryptischen In-Gel Verdau. Wie zuvor beschrieben, erfolgt durch die Deglykosylierung mittels endoF eine Deamidierung (Tarentino and Plummer, 1994). Die identifizierten Deamidierungen geben also einen starken Hinweis darauf, dass das Protein an diesen Stellen (Asn90 und Asn260) tatsächlich glykosyliert vorlag. Die Deamidierung mit einem ^{18}O -Isotop muss während des tryptischen Verdau stattgefunden haben, da dieser in H_2^{18}O durchgeführt wurde. Beide verschieden modifizierten Formen des Peptides SNFN**CFY**NGSAQ**TV**NIDF weisen keine ^{18}O -Markierung am C-Terminus auf. Es muss sich dabei also um die gesuchte Metalloprotease-Spaltstelle handeln. Die Spaltung scheint nach dem Phe270 zu erfolgen. Dies sollte nun durch den Verdau mit einer alternativen Protease bestätigt werden.

Tab. 5.3: Auflistung der identifizierten Peptide der Datenbankrecherche in der Swiss-Prot Datenbank mit Taxonomie *Mus musculus*, zusätzlich der Aminosäuresequenzen der Protease Trypsin sowie des in Abbildung 5.1 abgebildeten ARTC2.2. Alle Cysteine sind carbamidomethyliert. Die Peptide, deren C-Terminus im Bereich der potentiellen Metalloproteasen-Spaltstelle liegt, sind grau unterlegt.

Molekulargewicht [Da]	Mascot ion score	Position P1	Aminosäuresequenz	Modifikationen
1836,71	74	N	AFDDQYEGCVEDMEK	Label:18O(1) (C-term)
1852,70	70	N	AFDDQYEGCVEDMEK	Label:18O(1) (C-term); Oxidation (M)
1919,92	81	K	APQLLQEDFNMNEELK	Label:18O(1) (C-term)
1738,89	129	R	AVQLLNDQGCSLVYR	Label:18O(2) (C-term)
1739,86	75	R	AVQLLNDQGCSLVYR	Deamidated (NQ); Label:18O(2) (C-term)
1741,87	93	R	AVQLLNDQGCSLVYR	Deamidated:18O(1) (NQ); Label:18O(2) (C-term)
1357,68	76	V	AYTGNLAIDFNR	Label:18O(2) (C-term)
1705,84	81	R	EEEVLPYGYVYHK	Label:18O(1) (C-term)
2556,24	101	K	EFSSFPREEVLPYGYVYHK	Label:18O(1) (C-term)
2293,09	98	K	EIKNSTSYPAGFDFHGTALV	Deamidated (NQ); Label:18O(1) (C-term)
1454,78	52	N	FFSNHGTLFIIR	Label:18O(2) (C-term)
1189,59	75	R	FGQFSSSSLTK	Label:18O(1) (C-term)
1347,69	55	R	FGQFSSSSLTKR	Label:18O(2) (C-term)
959,57	58	N	HGTLFIIR	Label:18O(2) (C-term)
2048,01	129	K	KAPQLLQEDFNMNEELK	Label:18O(1) (C-term)
2064,01	112	K	KAPQLLQEDFNMNEELK	Label:18O(1) (C-term); Oxidation (M)
2067,00	114	K	KAPQLLQEDFNMNEELK	Deamidated (NQ); Label:18O(2) (C-term); Oxidation (M)
2733,36	144	K	KAPQLLQEDFNMNEELKLEWEK	Label:18O(1) (C-term)
2749,35	89	K	KAPQLLQEDFNMNEELKLEWEK	Label:18O(1) (C-term); Oxidation (M)
1738,75	62	K	KSNFNCFYNGSAQTV	Deamidated (NQ); Label:18O(1) (C-term)
2609,03	114	F	MLDMAPNAFDDQYEGCVEDMEK	Label:18O(1) (C-term)
1639,68	84	K	NSTSYPAGFDFHGT	Deamidated (NQ); Label:18O(1) (C-term)
1823,80	76	K	NSTSYPAGFDFHGTAL	Deamidated (NQ); Label:18O(1) (C-term)
1922,87	80	K	NSTSYPAGFDFHGTALV	Deamidated (NQ); Label:18O(1) (C-term)
2156,97	65	K	NSTSYPAGFDFHGTALVAY	Deamidated (NQ); Label:18O(1) (C-term)
2097,87	102	K	SNFNCFYNGSAQTVNIDF	Deamidated (NQ)
2100,86	77	K	SNFNCFYNGSAQTVNIDF	Deamidated (NQ); Deamidated:18O(1) (NQ)
1160,64	68	F	SNHGTLFIIR	Label:18O(2) (C-term)
2086,92	102	V	TAQNDNGYNEIFLDSPER	Deamidated:18O(1) (NQ); Label:18O(1) (C-term)
1123,58	64	Y	TGNLAIDFNR	Label:18O(2) (C-term)
2026,08	109	R	VALSSNFFSNHGTLFIIR	Label:18O(2) (C-term)
2029,07	88	R	VALSSNFFSNHGTLFIIR	Deamidated:18O(1) (NQ); Label:18O(2) (C-term)
975,46	60	K	VMFEYTGK	Label:18O(1) (C-term)
2185,01	57	K	VTAQNDNGYNEIFLDSPER	Label:18O(2) (C-term)
2184,97	67	K	VTAQNDNGYNEIFLDSPER	2 Deamidated (NQ); Label:18O(1) (C-term)
2185,99	86	K	VTAQNDNGYNEIFLDSPER	Deamidated (NQ); Label:18O(2) (C-term)
2186,97	64	K	VTAQNDNGYNEIFLDSPER	2 Deamidated (NQ); Label:18O(2) (C-term)
2186,97	58	K	VTAQNDNGYNEIFLDSPER	2 Deamidated (NQ); Label:18O(2) (C-term)
2188,98	63	K	VTAQNDNGYNEIFLDSPER	Deamidated (NQ); Deamidated:18O(1) (NQ); Label:18O(2) (C-term)
2188,98	58	K	VTAQNDNGYNEIFLDSPER	Deamidated (NQ); Deamidated:18O(1) (NQ); Label:18O(2) (C-term)
2188,98	80	K	VTAQNDNGYNEIFLDSPER	Deamidated (NQ); Deamidated:18O(1) (NQ); Label:18O(2) (C-term)
2190,98	83	K	VTAQNDNGYNEIFLDSPER	2 Deamidated:18O(1) (NQ); Label:18O(2) (C-term)
2313,10	63	K	VTAQNDNGYNEIFLDSPERK	Label:18O(2) (C-term)
2316,09	87	K	VTAQNDNGYNEIFLDSPERK	Deamidated:18O(1) (NQ); Label:18O(2) (C-term)
2317,07	75	K	VTAQNDNGYNEIFLDSPERK	Deamidated (NQ); Deamidated:18O(1) (NQ); Label:18O(2) (C-term)
2319,08	85	K	VTAQNDNGYNEIFLDSPERK	2 Deamidated:18O(1) (NQ); Label:18O(2) (C-term)

5.3.2 Ergebnisse des Verdaus mittels der Endoprotease GluC des Proteins ARTC2.2tace

Ziel der nun folgenden Experimente war es mit Hilfe einer alternativen Protease die mit Trypsin lokalisierte Metalloprotease-Spaltstelle zu bestätigen. Um festzustellen, welche Protease sich hierzu am besten eignet, wurden zuerst verschiedene *in silico* Verdaue mit Hilfe der PeptideMass Software durchgeführt. Chymotrypsin konnte beispielsweise nicht verwendet werden, da im Falle eines Verdaus von ARTC2.2tace mit dieser Protease auf Grund ihrer

Spaltspezifität nach dem endständigen Phe270 eine Hydrolyse erfolgen würde. Die Auswahl fiel deshalb auf GluC in Ammoniumbicarbonat-Puffer. Wie in der Literatur beschrieben (Sorensen, *et al.*, 1991) und in den Experimenten in Kapitel 4 bestätigt wurde, sollte die Endoprotease GluC unter diesen Bedingungen eine scheinbar ausschließliche Spaltspezifität nach Glutamat besitzen. Dabei sollte das Peptid RKKSNFNCFYNGSAQTVNIDF generiert werden. Der Verdau wurde ebenfalls in H₂¹⁸O durchgeführt, um eine Aussage treffen zu können, welche Peptide durch Metalloprotease- und welche durch GluC-Spaltung entstanden waren. GluC führt, ebenso wie Trypsin, zum Einbau von einem bzw. zwei ¹⁸O-Atomen und damit zu einem Massenshift von 2 Da bzw. 4Da (Schnolzer, *et al.*, 1996).

Die Ergebnisse der beiden Datenbanksuchen für den Verdau von ARTC2.2tace mittels der Endoprotease GluC sind in Abbildung 5.7 und Tabelle 5.4 dargestellt.

ARTC2.2, score: 334, SC: 19%, MS-Massengenauigkeit: 5ppm

```

1 MALPVTALLL PLALLLHAAR PDYKDDDDKR SLAVPFMLDM APNAFDDQYE
51 GCVEDMEKKA PQLLQEDFNM NEELKLEWEK AEINWKEIKN STSYPAGFHD
101 FHGTALVAYT GNLAIDFNRG VRDFKKSPDN FHYKAFHYL TRAVQLLNDQ
151 GCSLVYRGTK VMFEYTGKGS VRFGQFSSSS LTKRVALSSN FFSNHGTLFI
201 IRTCLGVNIK EFSSFPREE VLIPGYEVYH KVTAQNDNGY NEIFLDSPER
251 KKSNNFNCFYN GSAQTVNIDF SISGSRESCV SLFLVLLGL LVQQLTLAEP

```

Abb. 5.7: Ergebnisse der Datenbankrecherche in der Swiss-Prot Datenbank mit Taxonomie *Mus musculus*, zusätzlich der Aminosäuresequenzen der Proteasen Trypsin und GluC sowie des in Abbildung 5.1 abgebildeten ARTC2.2. SC = *sequence coverage*. Der *score* bezieht sich auf die Proteinidentifizierung.

Tab. 5.4: Auflistung der identifizierten Peptide der Recherche in der Swiss-Prot Datenbank mit Taxonomie *Mus musculus* zusätzlich der Aminosäuresequenzen der Protease GluC sowie des in Abbildung 5.1 abgebildeten ART2.2tace. Alle Cysteine sind Carbamidomethyliert. Die Peptide, deren C-Terminus im Bereich der potentiellen TACE-Spaltstelle liegt, sind grau unterlegt.

Molekulargewicht [Da]	Mascot ion score	Position P1	Aminosäuresequenz	Modifikationen
1430,73	82	E	DMEKKAPQLLQE	Label:18O(1) (C-term)
1055,62	69	E	KKAPQLLQE	Label:18O(1) (C-term)
1317,51	69	D	MAPNAFDDQYE	Label:18O(1) (C-term); Oxidation (M)
1479,68	55	E	RKKSNNFNCFYN	Deamidated:18O(1) (NQ)
2510,16	57	E	RKKSNNFNCFYNGSAQTVNIDF	Deamidated (NQ)
2513,15	71	E	RKKSNNFNCFYNGSAQTVNIDF	Deamidated (NQ); Deamidated:18O(1) (NQ)
1754,76	75	E	VYHKVTAQNDNGYNE	Deamidated (NQ); Deamidated:18O(1) (NQ)
1755,78	88	E	VYHKVTAQNDNGYNE	Deamidated:18O(1) (NQ); Label:18O(1) (C-term)
1757,75	65	E	VYHKVTAQNDNGYNE	Deamidated (NQ); 2 Deamidated:18O(1) (NQ)
1758,77	91	E	VYHKVTAQNDNGYNE	2 Deamidated:18O(1) (NQ); Label:18O(1) (C-term)

Das Protein NAR2A_MOUSE wurde, wie in dem Experiment des tryptischen Verdau zuvor, nicht identifiziert. Da keine *unique* Peptide identifiziert wurden, kann eine eindeutige

Unterscheidung zwischen NAR2B_MOUSE und ART2.2tace in diesem Experiment nicht getroffen werden. Insgesamt wurden mit der Endoprotease GluC wesentlich weniger Peptide identifiziert als in dem Experiment mit dem tryptischen Verdau. Dies resultiert auch in einer geringeren Sequenzabdeckung. Es wurde kein Peptid identifiziert, das die potentielle Glykosylierungsstelle an Asn90 enthielt.

Obwohl der GluC Verdau weniger effektiv als der tryptische Verdau war, wurden erneut drei Peptide identifiziert, deren C-Terminus im Bereich der vermuteten Metalloproteasen-Spaltstelle liegt. Das Peptid RKKS^NFN^CF^YN enthält eine Carbamidomethylierung des Cys8 und ist laut Datenbanksuche am Asn11 mit einem ¹⁸O-Isotopdeamidiert. Diese Zuordnung ist jedoch falsch. Es muss sich dabei um das Peptid RKKS^NFN^CF^YN handeln, welches eine Carbamidomethylierung des Cys8 und eine Deamidierung von Asn11 mit einem ¹⁶O-Isotop sowie eine ¹⁸O-Isotopenmarkierung am C-Terminus enthält. Dies wird im Kapitel 5.4 ausführlich erläutert. Es kann sich bei diesem Peptid damit um keine Metalloprotease-Spaltstelle handeln, sondern um ein Peptid, das durch den GluC In-Gel Verdau generiert wurde.

Das durch den *in silico* Verdau vorhergesagte Peptid RKKS^NFN^CF^YNGSAQTVNIDF wurde ebenfalls mit zwei verschiedenen Modifikationen identifiziert. So wurde es einmal als RKKS^NFN^CF^YNGSAQTVNIDF mit einer Carbamidomethylierung an Cys8 und einer Deamidierung des Asn11 sowie als RKKS^NFN^CF^YNGSAQTVNIDF mit einer Carbamidomethylierung an Cys8, einer Deamidierung des Asn11 sowie einer Deamidierung mit einem ¹⁸O-Isotop an Asn7 zugeordnet. Die Deamidierung mit dem leichten ¹⁶O-Isotop aller drei identifizierter Peptide entspricht der potentiellen Glykosylierungsstelle Asn260 von ART2.2 (siehe Abbildung 5.1). Damit wird der Hinweis bekräftigt, dass das Protein an dieser Stelle tatsächlich glykosyliert vorliegt. Die Deamidierung mit einem ¹⁸O-Isotop während des GluC Verdau erfolgt ebenfalls an der gleichen Aminosäure, wie während des tryptischen Verdau. Die beiden verschieden modifizierten Formen des Peptides RKKS^NFN^CF^YNGSAQTVNIDF weisen, wie erwartet, keine ¹⁸O-Markierung am C-Terminus auf. Es muss sich dabei also erneut um die gesuchte Metalloprotease-Spaltstelle handeln, die nach dem Phe270 erfolgt. Damit bestätigt das Experiment des GluC Verdau die Identifizierung der Metalloprotease-Spaltstelle.

5.4 Diskussion

In den in diesem Kapitel beschriebenen Experimenten sollte die unbekannte Spaltstelle einer Metalloprotease innerhalb des Proteins ARTC2.2, also der C-Terminus dieses Proteins identifiziert werden. Wie in Kapitel 1.5 beschrieben, gibt es verschiedene Möglichkeiten den C-Terminus eines Proteins z. B. mittels COFRADIC (Van Damme, *et al.*, 2010) oder C-TAILS (Schilling, *et al.*, 2010) zu identifizieren. Diese Methoden haben aber auch Nachteile bzw. Schwierigkeiten (Rogers and Overall, 2013). Der COFRADIC-Ansatz ist anfällig für den Verlust von Probe während den Fraktionierungen mittels SCX-Vortrennung oder RP-HPLC. Außerdem ist er mit >50 Fraktionen pro Probe sehr messintensiv und damit zeitaufwendig. In Experimenten mittels C-TAILS ist es ein Problem eine quantitative Markierung der C-Termini zu erhalten. Aus diesen Gründen wurde sich für eine andere Strategie entschieden, durch die in einem Verdau C-terminal mit einem bzw. zwei ¹⁸O-Atomen markierte Peptide zu generieren. Dadurch sollte eine Unterscheidung von dem gesuchten unmarkierten C-Terminus möglich sein. Da jedoch das C-terminale Ende des unmarkierten Peptides (siehe Abbildung 5.4) unbekannt ist, wurde als Enzymspezifität für die Datenbanksuche semi Trypsin bzw. semi GluC angegeben. Unter Verwendung dieser Einstellung müssen Peptide nur an einem Ende, also entweder N- oder C-terminal, die für das Enzym vorhergesagte Spaltspezifität aufweisen.

Insgesamt wurden zwei verschiedene Proteasen (Trypsin und GluC) für den In-Gel Verdau des Proteins ARTC2.2tace verwendet. Dies sollte die Sicherheit der korrekten Identifizierung der Metalloprotease-Spaltstelle erhöhen.

Über den tryptischen Verdau wurde das Protein ARTC2.2 mit einem sehr hohen *score* von 1782 und einer guten Sequenzabdeckung von 63% identifiziert. Dabei erfolgte die Zuordnung über *unique* Peptide. Eine Unterscheidung des ARTC2.2 wurde somit sowohl von dem fast Sequenzgleichen Protein NAR2B_MOUSE als auch von dem eng verwandten NAR2A_MOUSE getroffen. Insgesamt wurden drei Peptide identifiziert, deren C-Terminus innerhalb des potentiellen Bereichs der TACE-Spaltstelle liegt (siehe Abbildung 5.5).

In Abbildung 5.8 sind das MS- und das MS/MS-Spektrum des Peptides KSNFNC**CFY**NGSAQ**TV** gezeigt. Das Peptid wurde mit einem Mascot *ion score* von 62 sicher zugeordnet. Durch die Zuordnung der b- und y-Ionen handelt es sich um ein Peptid, dessen C-Terminus eine ¹⁸O-Markierung enthält.

Es ist bekannt, dass Peptide, die ein Aspartat enthalten, an dieser Aminosäure besonders gut fragmentieren, wenn sich N-Terminal ein Lysin befindet (Jorgensen *et al.*, 2005). Deshalb

weist das b_9 -Ion (das Signal nach dem Bindungsbruch C-terminal zum Aspartat) die höchste Intensität im MS/MS-Spektrum auf.

Durch die ^{18}O -Markierung des C-Terminus steht fest, dass es sich bei dem Peptid nicht um eine Metalloprotease-Spaltstelle, sondern um ein tryptisch gespaltenes Peptid handelt, auch wenn die Spaltung nicht entsprechend der tryptischen Spaltspezifität nach einem Lysin oder Arginin in Position P1 erfolgt ist. Insgesamt wurden fünf Peptide identifiziert, die N- oder C-Terminal eine Spaltstelle enthalten, die nach einem Valin in Position P1 generiert wurden. Diese enthalten alle die ^{18}O -Markierung am C-Terminus und können daher nicht aus der Metalloprotease-Spaltung stammen.

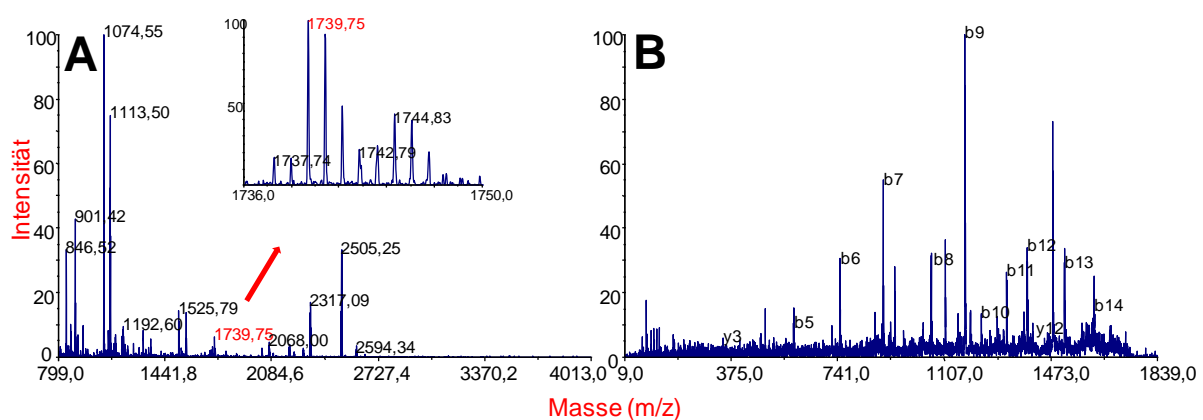


Abb. 5.8: A) MS-Spektrum des tryptischen Peptides KSNFNCFYNGSAQTV ($m/z = 1739,75$), das eine Carbamidomethylierung an Cys6, eine Deamidierung an Asn9 sowie eine ^{18}O -Markierung am C-Terminus trägt. Das m/z -Verhältnis des Peptides ist rot markiert und der m/z -Bereich dieses Signals ist in Vergrößerung dargestellt. B) MS/MS-Spektrum und Zuordnung der Signale des gleichen Peptides. Beide Spektren wurden mit einem MALDI-TOF/TOF Massenspektrometer im *positive ion mode* generiert.

Die beiden Zuordnungen der unterschiedlichen Modifikationen des Peptides SNFNCFYNGSAQTVNIDF führten letztendlich zur Identifizierung der Metalloprotease-Spaltstelle nach Phe270. Die zugehörigen MS- und MS/MS-Spektren sind in Abbildung 5.9 und Abbildung 5.10 dargestellt. Das Peptid SNFNCFYNGSAQTVNIDF mit einer Carbamidomethylierung an Cys5 und einer Deamidierung des Asn8 wurde in Spot 372 mit einem sehr hohen Mascot *ion score* von 102 identifiziert und scheint keine oder nur eine geringe Überlappung mit dem Peptid SNFNCFYNGSAQTVNIDF im selben Spot aufzuweisen (Abbildung 5.9, Teil A). Dieses Peptid mit einer Carbamidomethylierung an Cys5, einer Deamidierung des Asn8 sowie einer Deamidierung mit einem ^{18}O -Isotop an Asn4 wurde in Spot 373 mit einem hohen Mascot *ion score* von 77 identifiziert. Die zusätzliche

Deamidierung an Asn4 muss während des tryptischen Verdaus stattgefunden haben, da nur zu diesem Zeitpunkt H_2^{18}O zum Einsatz kam. Deamidierungen sind bekannte Modifikationen der Aminosäuren Asparagin und Glutamin, die bevorzugt an Asparagin ablaufen (Reubsæet *et al.*, 1998). Dabei wird die jeweils freie Carbonsäuregruppe gebildet. Aus Glutamin wird demnach Glutamat und aus Asparagin Aspartat gebildet. Dies führt jeweils zu einem Massenshift von +1 Da. Findet die Deamidierung jedoch in H_2^{18}O statt, so führt dies zu einem Massenshift von +3 Da. Damit sind die Deamidierungen im MS/MS-Spektrum (in Abhängigkeit der Qualität des MS/MS-Spektrums) unterscheidbar.

Im MS-Spektrum (siehe Abbildung 5.10) erkennt man noch ein kleines Signal für das Peptid SNFNCFYNGSAQTVNIDF. In dem verwendeten MALDI-TOF/TOF Massenspektrometer werden die *precursor* für die MS/MS-Analyse durch den *timed ion selector* (TIS) ausgewählt. Dieser ist jedoch nicht in der Lage, die beiden unterschiedlichen Peptidionen für die MS/MS-Analyse komplett voneinander zu trennen. Die zusätzlichen Signale des Peptides SNFNCFYNGSAQTVNIDF im MS/MS-Spektrum des Peptides SNFNCFYNGSAQTVNIDF könnten zu dem etwas geringeren Mascot *ion score* geführt haben. Auch wenn sich bei diesen Peptiden kein Lysin am C-Terminus befindet, ist die bevorzugte Fragmentierung nach einem Aspartat in Form des hohen Signals des b_{16} -Ions beobachtbar. Dennoch werden beide Peptide über die b- und y-Ionenserien sicher zugeordnet. Es liegt also keine ^{18}O -Markierung der beiden C-Termini vor. Die gesuchte Metalloprotease-Spaltstelle innerhalb des Proteins ARTC2.2 liegt also nach Phe270.

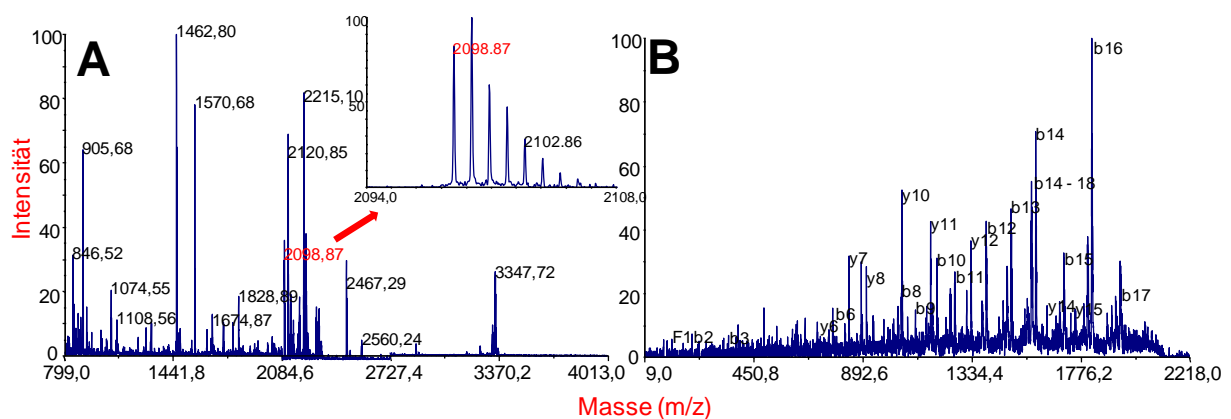


Abb. 5.9: A) MS-Spektrum des tryptischen Peptides SNFNCFYNGSAQTVNIDF ($m/z = 2098,87$), das eine Carbamidomethylierung an Cys6 und eine Deamidierung an Asn9 trägt. Das m/z -Verhältnis des Peptides ist rot markiert und der m/z -Bereich dieses Signals ist in Vergrößerung dargestellt. B) MS/MS-Spektrum und Zuordnung der Signale des gleichen Peptides. Beide Spektren wurden mit einem MALDI-TOF/TOF Massenspektrometer im *positive ion mode* generiert.

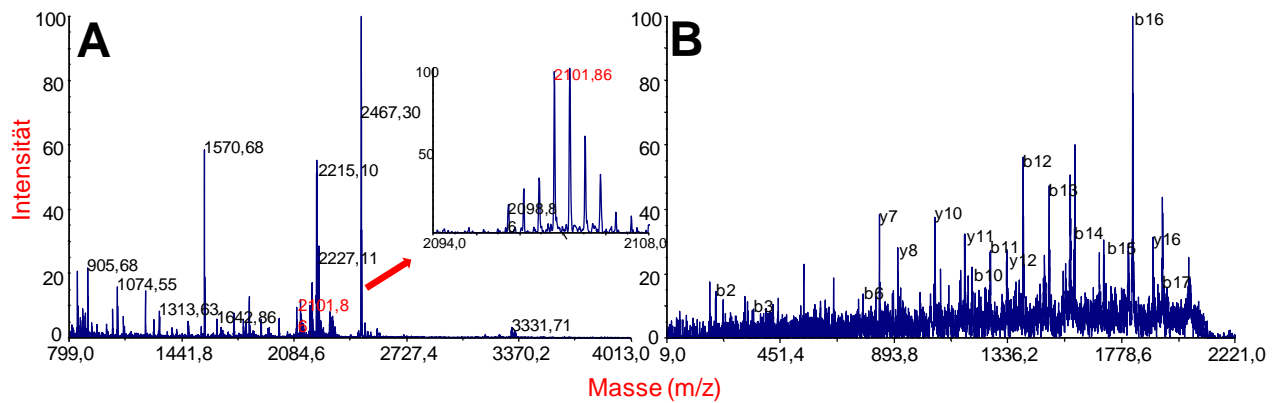


Abb. 5.10: **A)** MS-Spektrum des tryptischen Peptides SNFNCFYNGSAQTVNIDF ($m/z = 2101,86$), das eine Carbamidomethylierung an Cys6, eine Deamidierung an Asn9 sowie eine Deamidierung mit einem schweren ^{18}O -Isotop an Asn5 trägt. Das m/z -Verhältnis des Peptides ist rot markiert und der m/z -Bereich dieses Signals ist in Vergrößerung dargestellt. **B)** MS/MS-Spektrum und Zuordnung der Signale des gleichen Peptides. Beide Spektren wurden mit einem MALDI-TOF/TOF Massenspektrometer im *positive ion mode* generiert.

In den Experimenten zur Validierung der Metalloprotease-Spaltstelle wurde GluC als alternative Protease verwendet. Als Reaktionspuffer wurde Ammoniumbicarbonat gewählt, um eine Spaltung nach Asn269 eine Aminosäure vor der Metalloproteasen-Spaltstelle weitestgehend zu verhindern. Im Vergleich zu dem tryptischen Verdau wurde ARTC2.2 in der Swiss-Prot Datenbank nur mit einem niedrigen *score* von 334 und einer sehr niedrigen Sequenzabdeckung von 19% identifiziert. Eine Unterscheidung zu NAR2B_MOUSE war in diesem Fall nicht möglich. Dies war auf Grund der Ergebnisse des tryptischen Experimentes jedoch auch nicht erforderlich. Es wurden nur Peptide nahe des N- und C-Terminus identifiziert, der Kern des Proteins scheint weitestgehend intakt zu bleiben. Dies könnte v. a. daran liegen, dass GluC mit 36,3 kDa deutlich größer ist, als Trypsin mit 24,4 kDa. Dadurch sollte es Trypsin leichter fallen in das Gel einzudringen, um den Verdau durchführen zu können. Die Protease sollte proteolytisch aktiv gewesen sein, da GluC selbst mit einer Sequenzabdeckung von 63% identifiziert wurde. Bei der hier verwendeten Endoprotease GluC handelt es sich um ein ungeschütztes Protein, das nicht wie das von Promega hergestellte Trypsin durch TPCK-Behandlung weitestgehend vor Autoproteolyse geschützt ist. Dies könnte ein weiterer Grund für die schlechtere Sequenzabdeckung sein, da GluC sich vorwiegend selbst verdaut haben könnte.

In Abbildung 5.11 sind das MS- und das MS/MS-Spektrum des Peptides RKKSNNCFYN mit einer Carbamidomethylierung des Cys8 und einer Deamidierung des Asn11 gezeigt. Es

sollte laut Datenbanksuche fälschlicherweise am Asn11 mit einem ^{18}O -Isotop deamidiert sein. Eine Unterscheidung dieses Peptides von dem tatsächlichen Peptid RKKSNFNCFYN mit einer Carbamidomethylierung des Cys8, einer Deamidierung des Asn11 mit einem ^{16}O -Isotop sowie einer ^{18}O -Isotopenmarkierung am C-Terminus, ist nur über Fragmente der Seitenkette des Asn11 möglich. Die a-, b-, c- und x-, y-, z-Ionenserien sind für beide Peptidmodifizierungen identisch. Da es sich bei dem Aspartat jedoch um die potentielle Glykosylierungsstelle Asn260 handelt (siehe Abbildung 5.1), die in allen anderen identifizierten Peptiden, die diese Aminosäuresequenz enthalten, mit einer Deamidierung mit einem ^{16}O -Isotop identifiziert wurden, sollte dies nicht zusätzlich mit einem ^{18}O -Isotop stattgefunden haben.

Vorstellbar wäre jedoch, dass es sich bei dem Peptid ursprünglich um das unmarkierte C-terminale Ende von ARTC2.2tace handelt, bei dem eine chemische Austauschreaktion (*back exchange*) von Sauerstoffatomen stattgefunden hat. Es ist bekannt, dass ein solcher Austausch, durch Säure katalysiert stattfinden kann (Schnolzer, *et al.*, 1996). Da der Verdau in Ammoniumbicarbonat jedoch bei einem leicht basischen pH-Wert von 8,5 durchgeführt wurde, ist dies sehr unwahrscheinlich. Wie zuvor beschrieben, sollte die Endoprotease GluC unter der Verwendung von Bicarbonatpuffern eine scheinbare Affinität besitzen, auf der C-terminalen Seite von Glutamat Peptidbindungen zu hydrolysieren. Es ist also eher wahrscheinlich, dass eine Hydrolyse des durch Deamidierung gebildeten Aspartats durch die Endoprotease GluC erfolgt ist. So wurde auch das Peptid MAPNAFDDQYE identifiziert (siehe Tabelle 5.4), das ebenfalls durch die Hydrolyse einer Peptidbindung C-terminal zu einem Aspartat gebildet wurde.

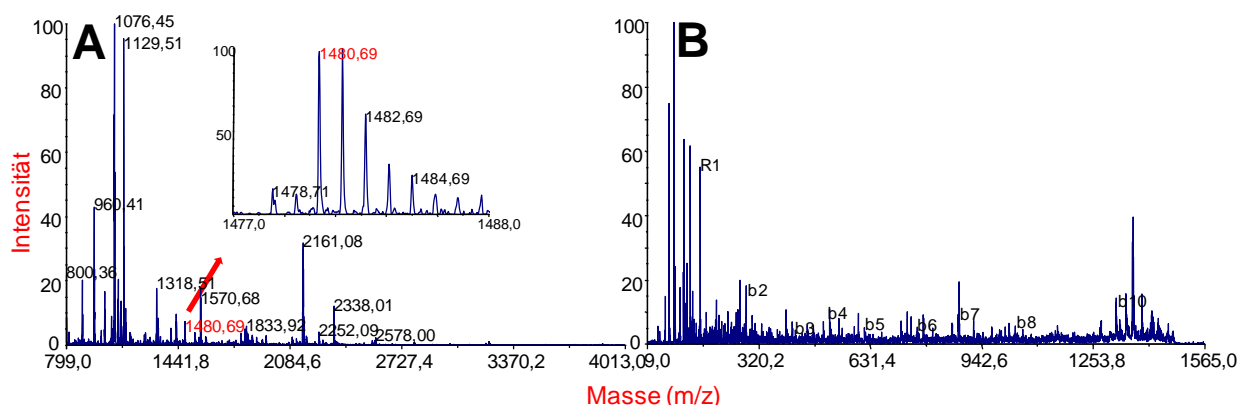


Abb. 5.11: A) MS-Spektrum des tryptischen Peptides RKKSNFNCFYN ($m/z = 1480,68$) mit einer Carbamidomethylierung des Cys8 und einer Deamidierung des Asn11 mit einem ^{18}O -Isotop bzw. des tatsächlichen Peptides RKKSNFNCFYN ($m/z = 1480,68$) mit einer Carbamidomethylierung des Cys8 einer Deamidierung des Asn11 mit einem ^{16}O -Isotop

sowie einer ^{18}O -Markierung am C-Terminus. Das m/z -Verhältnis des Peptides ist rot markiert und der m/z -Bereich dieses Signals ist in Vergrößerung dargestellt. **B)** MS/MS-Spektrum und Zuordnung der Signale des gleichen Peptides. Beide Spektren wurden mit einem MALDI-TOF/TOF Massenspektrometer im *positive ion mode* generiert.

Es ist also davon auszugehen, dass tatsächlich das Peptid RKKSNFNCFYN mit einer Carbamidomethylierung des Cys8 einer Deamidierung des Asn11 mit einem ^{16}O -Isotop sowie einer ^{18}O -Isotopenmarkierung am C-Terminus identifiziert wurde.

Leider wurde kein Peptid GSAGTVNIDF identifiziert. Dieses hätte die Metalloprotease-Spaltstelle nach Phe270 zusätzlich bestätigen können.

Wie in dem Experiment zuvor mit dem tryptischen Verdau, wird die Metalloprotease-Spaltstelle nach Phe270 auch in dem Experiment des GluC Verdau über die Zuordnungen der unterschiedlichen Modifikationen eines Peptides identifiziert. In diesem Falle handelt es sich um das Peptid RKKSNFNCFYNGSAQTVNIDF, das bereits zuvor in einem *in silico* Verdau vorhergesagt wurde. Die zugehörigen MS- und MS/MS-Spektren sind in Abbildung 5.12 und Abbildung 5.13 dargestellt. Das Peptid RKKSNFNCFYNGSAQTVNIDF mit einer Carbamidomethylierung an Cys8 und einer Deamidierung des Asn11 wurde in Spot 100 mit einem Mascot *ion score* von 57 identifiziert. Das Peptid mit der gleichen Aminosäuresequenz RKKSNFNCFYNGSAQTVNIDF mit einer Carbamidomethylierung an Cys8, einer Deamidierung des Asn11 sowie einer Deamidierung mit einem ^{18}O -Isotop an Asn7 wurde zusätzlich in Spot 99 mit einem höheren Mascot *ion score* von 71 identifiziert.

Betrachtet man die jeweiligen MS-Spektren, so sieht man, dass beide Peptide von dem anderen überlappt werden, auch wenn dieser Effekt in Spot 99 weniger stark ausgeprägt zu sein scheint. Das MS/MS-Spektrum des Peptides RKKSNFNCFYNGSAQTVNIDF (siehe Abbildung 5.12, Teil B) wird v. a. durch zwei sehr intensive Signale dominiert. Das dominante Signal des b_{11} -Ions entsteht wieder durch die besonders ausgeprägte Fragmentierung nach einem Aspartat (dem deamidierten Asn11), wenn sich N-Terminus ein Lysin befindet (Jorgensen, *et al.*, 2005). Dies konnte bereits in dem MS/MS-Spektrum des Peptides KSNFNCFYNGSAQTV beobachtet werden (siehe Abbildung 5.8). Betrachtet man nun das Spektrum des Peptides RKKSNFNCFYNGSAQTVNIDF, so scheint sich der Effekt noch zu verstärken, wenn sich ein Arginin anstatt eines Lysins am N-Terminus befindet. Die Bindungsspaltung C-terminal des Aspartats scheint dann so stark bevorzugt zu werden, dass andere Bindungsspaltungen unterdrückt werden. Gu *et al.* konnten dieses Phänomen mit einfach geladenen Oligopeptiden mittels *surface induced dissociation* (SID) ebenfalls beobachten (Gu *et al.*, 1999).

Das Signal mit einem m/z -Verhältnis von 878,3 konnte nicht zugeordnet werden. Dabei handelt es sich höchstwahrscheinlich um das b_7 -ion des Peptides RKKS~~NF~~NC~~FY~~NGSAQTVNIDF (siehe Abbildung 5.13, Teil B), das durch den TIS bei der *precursor*-Selektion nicht abgetrennt werden konnte. Bei diesem Peptid ist der Effekt der verstärkten Bindungsspaltung nach einem Aspartat in drei Fällen zu beobachten. Durch die beiden sehr starken Signale des b_7 - (Deamidierung des Asn7 mit einem ^{18}O -Isotop) und b_{11} -Ions (Deamidierung des Asn11 mit einem ^{16}O -Isotop) lassen sich auch hier beide Deamidierungen wieder sehr gut nachweisen. Das Signal des b_{20} -Ions, das ebenfalls durch die Fragmentierung nach einem Aspartat erfolgt, ist hier weniger stark ausgeprägt.

Sowohl das Experiment mit dem tryptischen als auch das Experiment mit dem GluC Verdau konnten also eine Metalloprotease-Spaltstelle nach Phe270 nachweisen. Zusätzlich wurden in den Experimenten an beiden potentiellen Glykosylierungsstellen des Proteins ARTC2.2 (Asn90 und Asn260) Deamidierungen identifiziert. Da das aufgereinigte Protein mit der Glykosidase endoF behandelt wurde, gibt dies einen starken Hinweis darauf, dass das Protein an diesen Stellen tatsächlich glykosyliert vorlag.

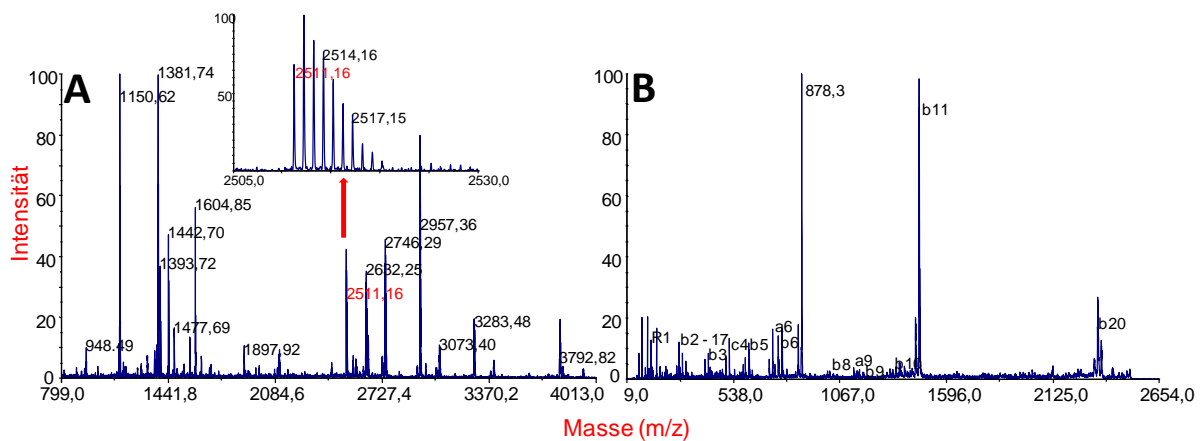


Abb. 5.12: A) MS-Spektrum des tryptischen Peptides RKKS~~NF~~NC~~FY~~NGSAQTVNIDF ($m/z = 2511,16$) mit einer Carbamidomethylierung an Cys8 und einer Deamidierung des Asn11. Das m/z -Verhältnis des Peptides ist rot markiert und der m/z -Bereich dieses Signals ist in Vergrößerung dargestellt. B) MS/MS-Spektrum und Zuordnung der Signale des gleichen Peptides. Beide Spektren wurden mit einem MALDI-TOF/TOF Massenspektrometer im *positive ion mode* generiert.

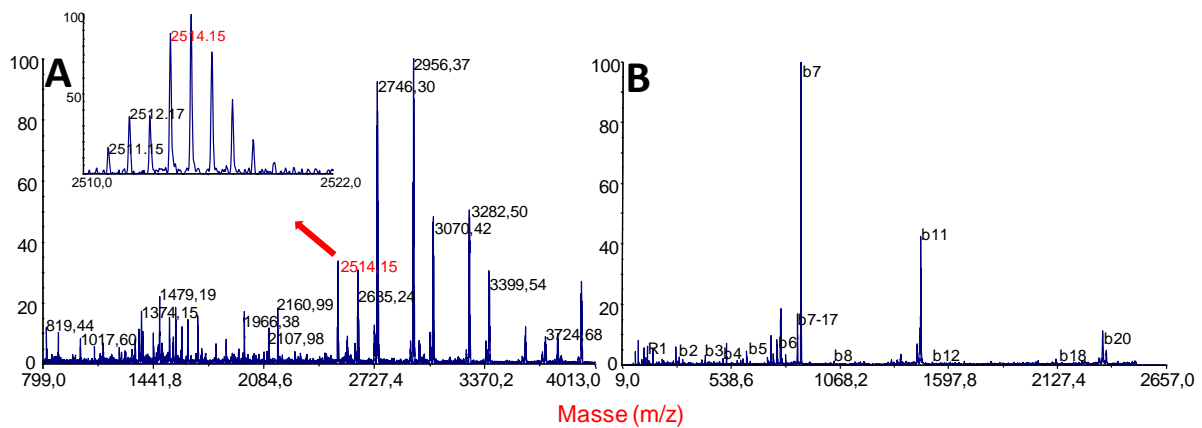


Abb. 5.13: **A)** MS-Spektrum des tryptischen Peptides RKKSNFNCFYNGSAQTVNIDF ($m/z = 2514,15$) mit einer Carbamidomethylierung an Cys8, einer Deamidierung des Asn11 sowie einer Deamidierung mit einem ^{18}O -Isotop an Asn7. Das m/z -Verhältnis des Peptides ist rot markiert und der m/z -Bereich dieses Signals ist in Vergrößerung dargestellt. **B)** MS/MS-Spektrum und Zuordnung der Signale des gleichen Peptides. Beide Spektren wurden mit einem MALDI-TOF/TOF Massenspektrometer im *positive ion mode* generiert.

5.5 Zusammenfassung und Ausblick

Das Ziel, die Metalloprotease-Spaltstelle innerhalb des Proteins ARTC2.2 zu identifizieren, wurde durch die in diesem Kapitel beschriebenen Experimente erreicht. So führte die Identifizierung mehrerer Peptide, sowohl durch tryptischen als auch durch GluC Verdau zur Bestimmung der Metalloprotease-Spaltstelle nach Phe270. Die Tatsache, dass es sich dabei um keine Spaltstelle handelt, die durch die verwendeten Proteasen generiert wurde, wurde durch die Durchführung der Proteolyse in ^{18}O -markiertem Wasser belegt. Nur die Metalloprotease-Spaltstelle kann unter diesen Bedingungen keine ^{18}O -Markierung am C-Terminus enthalten.

So konnte bei anderen Peptiden, deren C-Terminus ebenfalls im Sequenzbereich der potentiellen Metalloprotease-Spaltstelle liegen (siehe Abbildung 5.1), eine ^{18}O -Markierung nachgewiesen werden. Damit ist sichergestellt, dass diese Peptide durch den in-Gel Verdau mit der jeweils verwendeten Protease prozessiert wurden.

Zusätzlich wurden sowohl für die potentielle Glykosylierungsstelle Asn90 als auch Asn260 Peptide identifiziert, innerhalb deren Sequenz diese Glykosylierungsstellen deamidiert vorlagen. Nach der Behandlung mit der Glykosidase endoF liegen glykosylierte Asparagine als deglykosylierte Aspartate vor. Es findet neben der Deglykosylierung also auch eine Deamidierung statt (Tarentino and Plummer, 1994). Da an den potentiellen

Glykosylierungsstellen eine Deamidierung festgestellt wurde, stellt dies einen starken Hinweis dar, dass sie tatsächlich glykosyliert vorlagen.

Auch wenn diese Ergebnisse eindeutig sind, gibt es prinzipiell Möglichkeiten, weitere Experimente zu ihrer Absicherung durchzuführen. So könnte man weitere MS- bzw. MS/MS-Messungen mit einem LTQ-Orbitrap Massenspektrometer durchführen. Durch die höhere Massengenauigkeit des Orbitrap-Massenanalysators könnte man noch genauere Aussagen sowohl über die *precursor* innerhalb der MS-, als auch über die b- und y-Ionenserien innerhalb der MS/MS-Spektren treffen. Zusätzlich könnte man versuchen die HPLC-Auftrennung der in-Gel Verdaue mit einem flacheren Gradienten zu verbessern und somit eventuell eine bessere *precursor*-Selektion zu erreichen.

Eine weitere Möglichkeit stellt die Analyse synthetisch hergestellter Peptide dar, die die gleichen Sequenzen besitzen, wie die hier identifizierten. Auf diese Weise könnte man das MS/MS-Fragmentierungsverhalten vergleichen und somit eine weitere Aussage bezüglich der Sicherheit ihrer Identifizierung treffen.

Kapitel VI: Optimierte *Data Conversion* für die quantitative Proteomik, basierend auf isobarer stabiler Isotopenmarkierung und LC-MALDI-TOF/TOF MS/MS

6.1 Einleitung

Wie in Kapitel 1.2 beschrieben, gehören Reagenzien wie TMT (Thompson, *et al.*, 2003) oder iTRAQ (Ross, *et al.*, 2004) zu den am häufigsten verwendeten innerhalb der (relativen) quantitativen Proteomik (Aggarwal *et al.*, 2006, Treumann and Thiede, 2010).

Die Bearbeitung der Rohdaten stellt einen der wichtigsten Schritte innerhalb der bioinformatischen Analyse einer solchen quantitativen Strategie dar. Diese wird entweder mit kommerzieller oder sogenannter *open-source* Software durchgeführt. SEQUEST (Eng, *et al.*, 1994), der ParagonTM Algorithmus (Shilov *et al.*, 2007), MASCOT (Perkins, *et al.*, 1999), die OpenMS Plattform (Kohlbacher *et al.*, 2007, Sturm *et al.*, 2008) oder die Trans-Proteomic Pipeline (Keller *et al.*, 2005) sind nur einige Beispiele solcher Programme.

Während kommerzielle Programme häufig nur mit den Rohdaten des Massenspektrometers des gleichen Herstellers kompatibel sind, können *open-source* Programme oft auch Peaklisten in Formaten wie mgf (Perkins, *et al.*, 1999) oder mzML (Deutsch, 2008) bearbeiten. Das sogenannte *peak picking* stellt einen Teil der Prozessierung der Rohdaten dar. Es wird meistens über Algorithmen durchgeführt, die in der Software des Massenspektrometers enthalten sind.

Die Peaklisten werden dann in der Datenbanksuche sowohl für die Identifizierung als auch die Quantifizierung der Peptide genutzt. Zusätzlich besteht in der Datenbanksuche die Möglichkeit, eine Isotopenkorrektur durchzuführen, bei der die Reinheit der Isotopen bei der Herstellung der isobaren *tags* berücksichtigt wird. Hierzu werden Korrekturfaktoren von den Herstellern benötigt. In einer Reihe von Studien wurden verschiedene Methoden entwickelt, die sich auf die statistische Analyse der quantitativen Daten beziehen (Breitwieser *et al.*, 2011, Karp, *et al.*, 2010, Keshamouni *et al.*, 2006, Tan, *et al.*, 2008). Diese sind jedoch nicht Teil der hier gezeigten Experimente.

In einer Studie von Mahoney *et al.* wurde ein Problem beschrieben, iTRAQ-Reporterionen zu detektieren, wenn das *Launch Peaks to Mascot tool* innerhalb der 4000 Series Explorer Software (AB SCIEX, Darmstadt, Deutschland) verwendet wird (Mahoney *et al.*, 2011). Hier

scheinen iTRAQ-Reporterionen, die eine geringe Intensität aufweisen immer dann nicht in der Peakliste aufzutauchen, wenn sie auf ein iTRAQ-Reporterion mit einer hohen Intensität im MS/MS-Spektrum folgen. Das Problem wurde durch ein *in-house tool* gelöst. Durch dieses werden fehlende iTRAQ-Reporterionen in der mgf-Datei durch die Spektrum Datenbank des Massenspektrometers ergänzt. Eine genauere Erklärung oder zusätzliche Daten bezüglich dieses Problems wurden in der Veröffentlichung jedoch nicht gezeigt.

Das Problem der fehlenden Reporterionen muss auf Basis des *peak picking* Prozesses begründet sein. Es ist damit unabhängig von der späteren Datenbearbeitung, wie der Isotopenkorrektur oder statistischen Ansätzen. Um zu untersuchen, ob das Problem unabhängig von dem verwendeten isobaren *tag* besteht, wurde ein 15-Proteinstandard mit den verschiedenen isobaren *tags* iTRAQ 4-plex, iTRAQ 8-plex und TMT 6-plex derivatisiert und mittels LC-MALDI-TOF/TOF MS/MS mit einem 5800 Massenspektrometer (AB SCIEX, Darmstadt, Deutschland) untersucht. Die dabei generierten Rohdaten wurden mit drei verschiedenen Programmen (LPtMT, TS2Mascot und MS DATA Converter) in mgf-Dateien umgewandelt und die daraus resultierenden Peaklisten auf fehlende Reporterionen untersucht. Dabei wurde festgestellt, dass das Problem der fehlenden Reporterionen unabhängig von dem verwendeten isobaren *tag* bei allen drei Programmen auftritt. Verwendet man jedoch das TS2Mascot Programm unter der Einstellung, dass nicht nur die monoisotopischen Signale, sondern alle Signale exportiert werden (*monoisotopic only (MIO) disabled*), so scheint das Problem behoben zu sein. Diese Einstellung ist bei den beiden anderen Programmen nicht möglich, führt letztendlich allerdings auch zu einer geringeren Anzahl an Identifizierungen und zu geringeren Mascot *ion scores* der Peptide. Der Grund hierfür ist die Tatsache, dass unter dieser Einstellung zwar alle Signale exportiert werden, in der Mascot Datenbanksuche jedoch nur die monoisotopischen Signale zugeordnet werden können. Man erhält in der mgf-Datei für jedes MS/MS-Spektrum also eine Reihe zusätzlicher Signale, die zu einer schlechteren Identifizierung, also einem geringeren *score* führen.

Um diesem Problem entgegenzuwirken, wurde in Zusammenarbeit mit (Bart HJ van den Berg (Institut für Experimentelle Medizin, Christian-Albrechts Universität zu Kiel siehe Anhang Kapitel 6.I Perl-Skript RICmerge) ein Perl-Skript entwickelt. Dieses Programm ergänzt fehlende Signale (Flächen) der Reporterionen aus den mgf-Dateien der TS2Mascot Software unter Verwendung der *MIO enabled* Option, durch die Flächen in den mgf-Dateien unter Verwendung der *MIO disabled* Option. Mit den so entstandenen *merged* mgf-Dateien ist dann eine korrekte Identifizierung der Peptide über die monoisotopischen Signale, bei gleichzeitiger korrekter Quantifizierung über alle Reporterionensignale möglich.

6.2 Material und Methoden

Die Proteine des 15-Proteinstandards (Tabelle 6.1), Acetonitril (E Chromasolv), α -Cyano-4-hydroxy-zimtsäure (CHCA, $\geq 98\%$) und Glu¹-Fibrinopeptid B wurden von Sigma-Aldrich (Steinheim, Deutschland) bezogen.

Tab. 6.1: Informationen bezüglich der 15 Proteine des 15-Proteinstandards. Proteine, die in unterschiedlichen Verhältnissen in Mischung A zu B vorliegen, sind farblich markiert.

Protein (Herkunft)	Molekulargewicht [kDa/mol]	Konzentration in Mischung A [pmol/ μ l]	Konzentration in Mischung B [pmol/ μ l]	Verhältnis A:B
Albumin (Rind)	69,3	16,7	16,7	1:1
Carboanhydrase (Rind)	29,1	12,9	12,9	1:1
β -Lactoglobulin (Rind)	18,3	13,1	13,2	1:1
Transferrin (Mensch)	77,0	10,1	10,1	1:1
Catalase (Rind)	59,9	6,4	6,5	1:1
α -Lactalbumin (Rind)	16,2	24,4	24,3	1:1
β -Casein (Rind)	25,1	30,0	30,0	1:1
Ubiquitin (Rind)	18,5	21,1	21,0	1:1
PhosphorylaseB (Hase)	97,3	15,0	15,0	1:1
Lysozym (Huhn)	16,2	4,8	47,3	1:10
Ribonuklease (Rind)	16,5	47,7	4,9	10:1
α -Casein (Rind)	24,5	3,4	16,8	1:5
Cytochrom c (Pferd)	11,8	32,3	6,6	5:1
α -Crystallin A* (Rind)	19,9	18,27	36,0	1:2
Myoglobin (Pferd)	59,9	11,7	5,8	2:1
*enthält α -Crystallin B als Verunreinigung				

Entionisiertes Wasser (18,2 M Ω *cm) wurde mit einem arium®611VF System (Sartorius, Göttingen, Deutschland) hergestellt und für alle Experimente verwendet. Isopropanol (Chromasolv), Triethylammoniumbicarbonat (1M, puriss. P.a.), S-Methyl-methanthiosulfonat (MMTS, $\geq 98\%$), Tris(2-carboxyethyl)phosphin-Hydrochlorid (TCEP $\geq 98\%$) und Trifluoressigsäure (ca. 98%) wurden von Fluka, *Sequencing grade modified* Trypsin von Promega (Madison, WI, USA) erworben. Das zur MS-Kalibrierung verwendete 4700 *Proteomics Analyzer Standard Kit* sowie die beiden isobaren tags iTRAQTM *Reagents 4plex and 8plex Applications Kit*, wurden von AB SCIEX (Darmstadt, Deutschland) bezogen. Das *TMTsixplexTM Label Reagent Set* wurde von Thermo Fisher Scientific (Bremen, Deutschland) erworben.

Nasschemische Methoden - Probenvorbereitung

Um einen klassischen *shotgun*-Analyseansatz zu simulieren, wurde der 15-Proteinstandard in zwei verschiedenen Mischungen (A und B, siehe Tabelle 6.1 und Abbildung 6.1) verwendet. Sechs der 15 Proteine wurden in unterschiedlichen Verhältnissen in den beiden Mischungen A und B angesetzt, um unterschiedlich exprimierte Proteine abzubilden (2:1, 5:1, 10:1, 1:2, 1:5 and 1:10). Die Mischungen A und B wurden je zu zwei internen technischen Replikaten aufgeteilt und zu einer Konzentration von 2,5 mg/ml in 0,5 M Triethylammoniumbicarbonat-Puffer gelöst, so dass insgesamt vier verschiedene Proben zu je 200 µl (entspricht je 500 µg) erhalten wurden. Die intakten Proteine wurden mit *Sequencing grade modified* Trypsin (Promega, Verhältnis Protease:Protein 1:50 (m/m)) verdaut. Die Inkubation erfolgte über Nacht bei 37°C und 350 U/min im Thermomixer. Die Reduktion der Disulfidbrücken wurde jeweils durch Zugabe von 40 µl Reduzierungs-Reagenz (50 mM Tris-(2-carboxyethyl)phosphin, TCEP) und 1 h Inkubationszeit im Thermomixer bei 60°C und 350 U/min durchgeführt. Um eine Rückreaktion der reduzierten Cysteine zu Cystin zu verhindern, wurde eine Alkylierung durch Zugabe von 20 µl Sulphydryl-Blockierungs-Reagenz (200 mM Methylmethylthiomethylsulfoxid, MMTS) durchgeführt. Die Inkubationszeit betrug 30 min bei RT im Dunkeln.

Je 25 µl aller vier Proben wurden anschließend nach Herstellerangaben mit den drei verschiedenen isobaren *tags* markiert, wobei folgende Reporterionenkanäle verwendet wurden: Für iTRAQ 4-plex und 8-plex Reporterionenkanal 114 und 115 für Mischung A; 116 und 117 für Mischung B bzw. für TMT Reporterionenkanal 126 und 127 für Mischung A; 128 und 129 für Mischung B. Alle Proben wurden dann komplett eingengt und mit entionisiertem Wasser zu einer Konzentration von 1 µg/µl gelöst. Nach Vereinigung der Proben im Verhältnis 1:1:1:1 wurden die sie mit 0,1% wässriger Trifluoressigsäurelösung 1:128 (v/v) zu einer Endkonzentration von 250 fmol/µl verdünnt.

Trennung mittels IP-RP-HPLC bei pH 2,1

Die Auftrennung der mit den drei verschiedenen isobaren *tags* markierten Proben erfolgte mit einem U3000 nanoHPLC-System (Dionex, Idstein, Deutschland). Über eine 20 µl Probenschleife wurden je 2 µl Probe injiziert und auf einer Acclaim PepMap100 C18 *Trap*-Säule (300 µm I.D. x 10 mm, 5 µm, Dionex, Idstein, Deutschland) aufkonzentriert. Die *Trap*-Säule wurde 10 min mit 0,1%iger (v/v) wässriger Trifluoressigsäure-Lösung mit 3% (v/v) Acetonitril mit einer Flussrate von 30 µl/min mit Peptiden beladen. Die Auftrennung der Analyten erfolgte dann auf einer Acclaim PepMap100 C18 Trennsäule

(75 µm I.D. x 250 mm, Dionex, Idstein, Deutschland) über den Gradienten, der Tabelle 6.2 zu entnehmen ist. Zur Erstellung des Gradienten wurden die Lösungsmittel Eluent A (0.05% (v/v) wässrige Trifluoressigsäure) und Eluent B (0,04% (v/v) wässrige Trifluoressigsäure in 80% (v/v) Acetonitril) verwendet, wobei die Flussrate 0,3 µl/min, bei einer Säulentemperatur von 30°C betrug. Die Detektion erfolgte mittels eines UV-Detektors mit Z-förmiger Detektorzelle mit einem Detektorvolumen von 3 nl bei einer Wellenlänge von 214 nm. Alle LC-Auftrennungen wurden in Form von fünf technischen Replikaten durchgeführt, wobei jeweils 500 fmol (2 µl) an Probe injiziert wurden. Die von der Säule eluierten Peptide wurden ab der 29. Minute 1:3 (v/v) mit MALDI-Matrix (CHCA) über ein T-Stück mittels Spritzenpumpe gemischt und alle 15 s auf ein MALDI-Target (Opti-TOF™ LC MALDI Insert, AB SCIEX, Darmstadt, Deutschland) gespottet. Das Spotten erfolgte durch einen Probot Mikrofraktionssammler (Dionex, Idstein, Deutschland) für 85 min und die Matrix-Stockkonzentration betrug 3 mg/ml (in 70% (v/v) Acetonitril/0.1% (v/v) wässriger Trifluoressigsäure) und enthielt ca. 5 nM Glu¹-Fibrinopeptid B zur internen Kalibrierung.

Tab. 6.2: Zeitlicher Verlauf des Gradienten für die Trennung der mit den drei unterschiedlichen isobaren *tags* markierten Probe des 15-Proteinstandards mittels IP-RP-HPLC.

Zeit [min]	Eluent A [%]	Eluent B [%]
0,0	95	5
10,0	95	5
95,0	40	60
105,0	5	95
110,0	5	95
111,0	95	5
129,0	95	5

Massenspektrometrische Analyse mittels MALDI-TOF/TOF MS/MS

Die massenspektrometrischen Analysen der mittels nanoHPLC-System aufgetrennten Proben und Datenprozessierung (GPS-Software, Version 3.6, AB SCIEX, Darmstadt, Germany) wurde mit einem MALDI-TOF/TOF 5800 Massenspektrometer (AB SCIEX, Darmstadt, Deutschland) durchgeführt. Alle MS-Spektren wurden im *positive ion mode* in einem Massenbereich von 800 – 4000 *m/z* durch die Akkumulation von 1000 Laserschüssen generiert. Die MS-Kalibrierung erfolgte mittels Standardpeptiden (4700 Proteomics Analyzer Mass Standard Kit, AB Sciex, Darmstadt, Deutschland), welche den erforderlichen

Messbereich abdecken, in Form einer *default calibration*. Zusätzlich wurde während der LC-MALDI-Analysen mit Glu¹-Fibrinopeptid B als internem Standard kalibriert.

Die Auswahl der *precursor* für die MS/MS-Analyse erfolgte durch die AB SCIEX 4000 Series Explorer Software (Darmstadt, Deutschland) unter Verwendung der folgenden Parameter: Minimales S/N-Verhältnis 20; *precursor* Massentoleranz: ± 200 ppm; das Signal musste nur in einem Spot detektiert werden, um als *precursor* akzeptiert zu werden; maximal 25 *precursor* pro Spot, wobei zuerst die *precursor* mit der höchsten Intensität fragmentiert wurden ("*strongest precursor first*"). Alle MS/MS-Spektren wurden im *positive ion mode* mit 1 kV Kollisionsenergie mittels CID mit einem Druck von 4×10^{-6} Torr durch die Akkumulation von 3000 Laserschüssen generiert. Die MS/MS-Kalibrierung wurde mit Glu¹-Fibrinopeptid B aus dem 4700 Proteomics Analyzer Mass Standard Kit (AB SCIEX, Darmstadt, Deutschland) in Form einer *default calibration* durchgeführt.

Protein- bzw. Peptididentifizierung mittels Datenbankanalyse

Die Rohdaten wurden unter Verwendung drei verschiedener Software-Tools in *mascoat generic format* (mgf)-Dateien umgewandelt. Für das *Launch Peaks to Mascot tool* (LPtMt, AB SCIEX, Darmstadt), Deutschland wurden folgende Einstellungen verwendet: Massenbereich von 65 Da bis zu der Masse des *precursors* minus 35 Da; minimales S/N-Verhältnis von 10, minimale Fläche 100, maximal 20 Peaks pro 200 Da und maximal 65 Peaks pro *precursor*. Für die TS2Mascot Software (Matrix Science, London, England) wurden die Parameter Massenbereich von 65 Da bis zu der Masse des *precursors* minus 35 Da und minimales S/N-Verhältnis von 10 eingestellt. Hierbei wurden zwei verschiedene mgf-Dateien generiert, wobei einmal nur die monoisotopischen Signale (*monoisotopic only* (MIO) *enabled*) und einmal alle Signale (MIO *disabled*) exportiert wurden. Für die MS DATA Converter Software (AB SCIEX, Darmstadt, Deutschland) wurde die Option "*proteinpilot*" gewählt. Nach Herstellerangaben ist dies identisch mit den Peaks, die durch die Protein Pilot Software (AB SCIEX, Darmstadt, Deutschland) im *centroid mode* als mgf-Datei erstellt werden. Hier ist es nicht möglich weitere Parameter innerhalb dieses Modus zu wählen.

Zusätzlich wurde das Perl-Skript RICmerge entwickelt, das fehlende Signale (Flächen) der Reporterionen aus den mgf-Dateien der TS2Mascot Software unter Verwendung der MIO *enabled* Option durch die Flächen in den mgf-Dateien unter Verwendung der MIO *disabled* Option ergänzt (siehe Anhang Kapitel 6.I Perl Skript RICmerge). Diese mgf-Dateien wurden dann als *merged*-Dateien bezeichnet.

Die Datenbanksuche wurde mit der Mascot 2.2.04 Software gegen eine selbsterstellte Datenbank, die die Aminosäuresequenzen der 15 Proteine enthielt, durchgeführt. Trypsin wurde als Enzym eingestellt und die maximal zulässige Anzahl an ausgelassenen Schnittstellen (*missed cleavages*) betrug 2. Die Massentoleranz für die *precursor* wurde auf ± 25 ppm und die der MS/MS-Fragmentionen auf $\pm 0,2$ Da festgesetzt. Als variable Modifikationen wurden Oxidation (M) und Markierung isobares *tag* (Y), als feste Modifikationen Methylthio (C) sowie Markierung isobares *tag* (N-Term und K) eingestellt. Isobares *tag* steht dabei entweder für iTRAQ 4-plex bzw. 8-plex oder TMT. Der Vertrauensbereich für die Proteinidentifizierung wurde auf $\geq 95\%$ ($p < 0,05$) und der *identity threshold* der Peptide auf einen *ion score* von 20 festgesetzt.

Die quantitative Auswertung der Daten wurde ebenfalls mit der Mascot Software durchgeführt, wobei die iTRAQ- bzw. TMT-Faktoren für die Isotopenkorrektur nach Angaben von AB SCIEX (Darmstadt, Deutschland) bzw. von Thermo Fisher Scientific (Bremen, Deutschland) für die Charge MC155137 verwendet wurden. Für alle drei verschiedenen isobaren *tags* wurden die Flächen der einzelnen Reporterionen durch die Flächenwerte des Reporterionenkanals 1 geteilt (114 für iTRAQ 4-plex bzw. 8-plex und 126 für TMT, siehe Abbildung 6.1). Um statistische Berechnungen durchzuführen, wurden die daraus resultierenden Reporterionen-Flächenverhältnisse, zusammen mit den Ergebnissen der Mascot Datenbanksuche, als csv-Format exportiert. Mathematische Ausreißer der Reporterionen-Flächenverhältnisse wurden, wie von Tiberti *et al.* (2010) beschrieben, mittels der *outlier* Software identifiziert und von der weiteren Analyse ausgeschlossen (http://www.sediment.uni-goettingen.de/staff/dunkl/software/out_lier.html).

Insgesamt wurden dabei vier verschiedene Ausreißertests durchgeführt (Grubbs Test, Dixon Test, IQR Test und Gauss g-Test). Ein Reporterionen-Flächenverhältnis wurde von der weiteren Analyse ausgeschlossen, wenn einer der vier Tests ein positives Ergebnis lieferte. Die geometrischen Mittelwerte, sowie die zugehörige Standardabweichung der TMT-Flächenverhältnisse, wurden nach Tan *et al* errechnet (2008).

6.3 Ergebnisse

Die Vorgehensweise zur Simulation eines typischen *bottom-up* Experimentes ist in Abbildung 6.1 gezeigt.

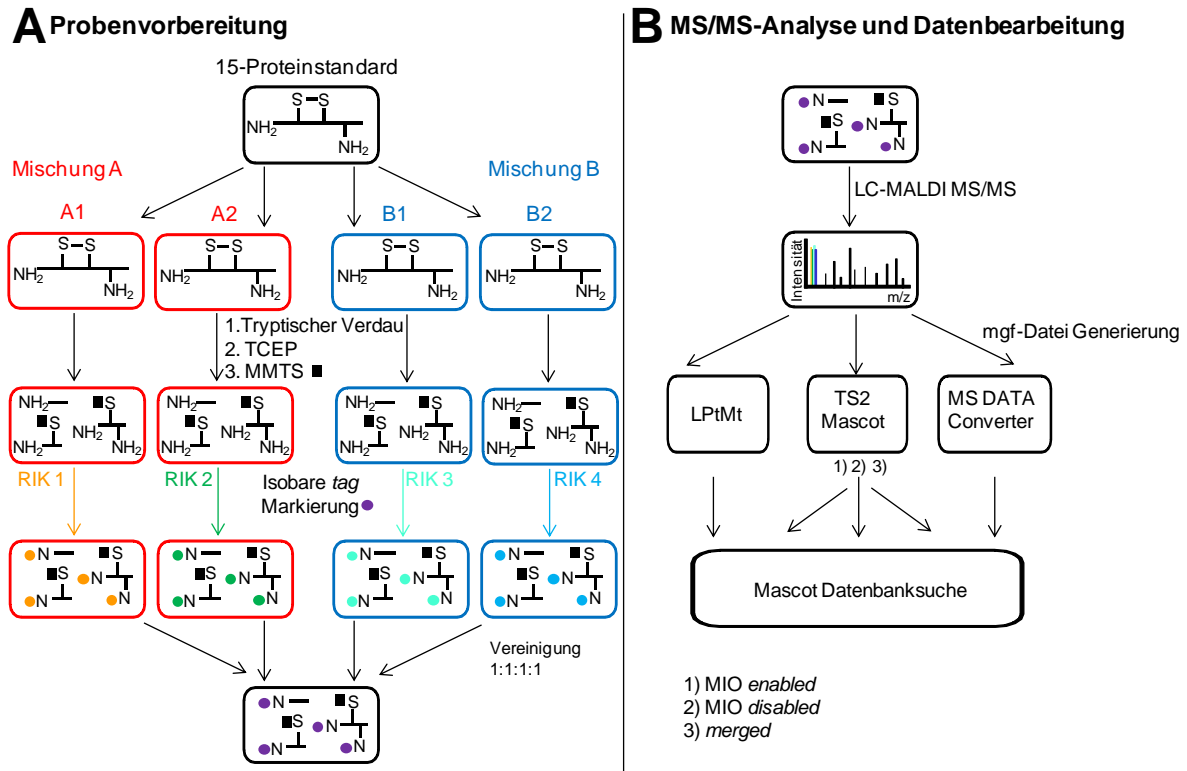


Abb. 6.1: Experimenteller Aufbau zur Simulation eines typischen *bottom-up* Experimentes. **A)** Probenvorbereitung: Ein 15-Proteinstandard wurde in zwei verschiedenen Mischungen (**Mischung A** und **Mischung B**) verwendet, in dem sechs Proteine in verschiedenen Verhältnissen gemischt wurden (2:1, 5:1, 10:1, 1:2, 1:5 und 10:1). Beide Mischungen wurden in zwei internen technischen Replikaten verwendet. Insgesamt kamen damit vier Proben zum Einsatz. Diese wurden zuerst tryptisch verdaut: Nach Reduzierung der Disulfidbrücken und Alkylierung der freien Cysteine wurden die tryptischen Spaltpeptide mit einem der Reporterionenkanäle (RIK) der drei verschiedenen isobaren tags iTRAQ 4-plex, iTRAQ 8-plex und TMT 6-plex markiert (RIK 1 = iTRAQ 4-plex und iTRAQ 8-plex 114 oder TMT 126; RIK 2 = iTRAQ 4-plex und iTRAQ 8-plex 115 oder TMT 127; RIK 3 = iTRAQ 4-plex und iTRAQ 8-plex 116 oder TMT 128 und RIK 4 = iTRAQ 4-plex und iTRAQ 8-plex 117 oder TMT 129). **B)** MS/MS-Analyse und Datenbearbeitung: Nach der LC-MALDI-TOF/TOF MS/MS Analyse wurden mgf-Dateien mit drei verschiedenen *data conversion* Programmen (LPtMt, TS2Mascot und MS DATA Converter) durchgeführt. Insgesamt wurden fünf technische Replikate untersucht. LPtMt = *Launch Peaks to Mascot tool*, MIO = *monoisotopic peaks only*.

In Abbildung 6.2 ist die Anzahl an Identifizierungen mit den drei verschiedenen isobaren tags markierter Peptide für die verschiedenen *data conversion* Programme gezeigt. Die Einstellungen für die Datenbanksuchen wurden dabei immer konstant gehalten, so dass

Unterschiede nur auf den verschiedenen *peak picking* Methoden in den unterschiedlichen Programmen begründet sein können.

Die höchste Anzahl an isobar markierten Peptiden wurde mit dem iTRAQ 4-plex Reagenz identifiziert, gefolgt von TMT 6-plex. Die wenigsten Identifizierungen wurden mit dem iTRAQ 8-plex Reagenz erreicht. Die unterschiedliche Anzahl an Identifizierungen wurde als unabhängig von dem verwendeten *data conversion* Programm beobachtet. Bezüglich der *data conversion* Programme wurden die meisten isobar markierten Peptide mit dem LPtMt identifiziert, während die wenigsten mit der TS2Mascot Software unter der Verwendung MIO *disabled* zugeordnet wurden. Die unterschiedliche Anzahl an Identifizierungen war dabei unabhängig von dem isobaren *tag*, mit dem die Peptide markiert waren (6% – 17% weniger). Für die MS DATA Converter und die TS2Mascot Software unter der Verwendung MIO *enabled* ist der Unterschied zum LPtMt mit 0% - 6% bzw. 2% - 5% weniger stark ausgeprägt (siehe Abbildung 6.2, Teil A). Das gleiche gilt für den durchschnittlichen Mascot *ion score*. Dieser erreichte ebenfalls für das LPtMt, unabhängig von dem verwendeten isobaren *tag*, die höchsten Werte. Für die TS2Mascot Software unter der Verwendung MIO *disabled* ist der Unterschied zum LPtMt mit 12% - 16% wieder am stärksten ausgeprägt. Für die Einstellung MIO *enabled* bzw. die MS DATA Converter Software hingegen war er nur 1% - 4% bzw. 3% - 4% geringer.

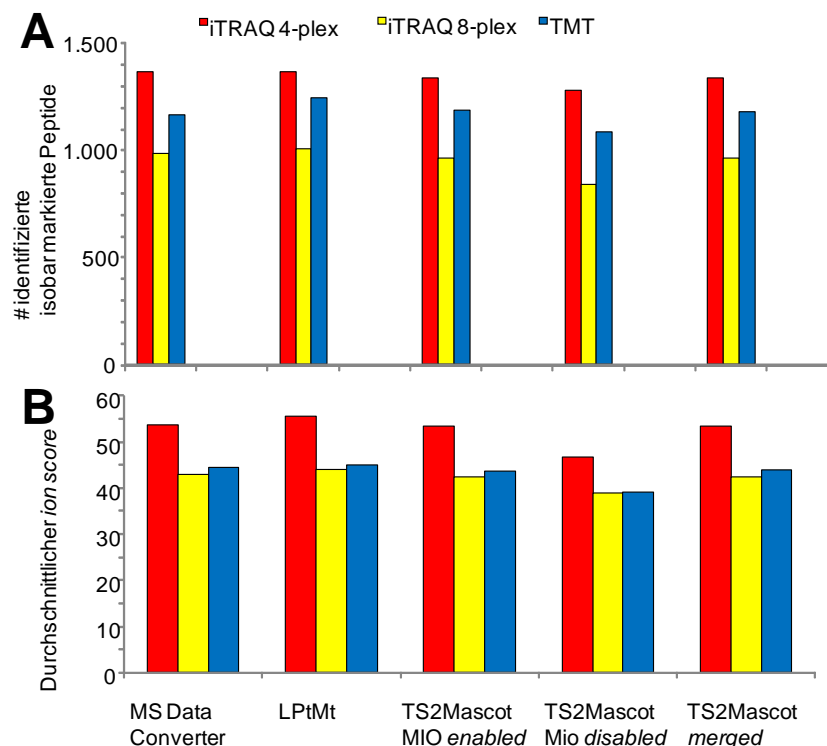


Abb. 6.2: A) Anzahl der identifizierten isobar-markierten Peptide für die verschiedenen *data conversion* Programme. B) Durchschnittlicher Mascot *ion score* für alle isobar-

markierten Peptide für die verschiedenen *data conversion* Programme. Insgesamt wurden fünf technische Replikate untersucht. LPtMt = *Launch Peaks to Mascot tool*, MIO = *monoisotopic peaks only*.

Um die Reporterionenkanal-Verhältnisse zu berechnen, wurde für alle drei verschiedenen isobaren *tags* der Reporterionenkanal 1 als Teiler verwendet (114 für iTRAQ 4-plex und iTRAQ 8-plex, 126 für TMT 6-plex). Da von beiden Mischungen jeweils ein zusätzliches internes technisches Replikat zum Einsatz kam (A1 und A2 bzw. B1 und B2, siehe Abbildung 6.1), sollte das Problem der fehlenden Reporterionensignale für den Reporterionenkanal 3 auftreten (116 für iTRAQ 4-plex und iTRAQ 8-plex, 128 für TMT 6-plex). Die Signale der Reporterionenkanäle 1 und 2 bzw. 3 und 4 sollten jeweils gleich groß sein, da es sich hierbei um die internen technischen Replikate handelt. Da das Problem nur dann zu beobachten ist, wenn ein Reporterionensignal mit niedriger Intensität auf ein Reporterionensignal mit hoher Intensität folgt, sollten die Proteine, die in den Mischungen A und B im Verhältnis 1:1 vorliegen, ebenfalls nicht betroffen sein. Die Proteine, die in Mischung B in einer größeren Konzentration als in Mischung A vorliegen (Lysozym, α -Casein und α -Crystallin), sollten auch keine fehlenden Reporterionensignale in der *mgf*-Datei aufweisen. Damit beschränkt sich die Anzahl der Proteine, die von dem Problem der fehlenden Reporterionensignale betroffen sein können auf Ribonuklease, Cytochrom c und Myoglobin.

Die Anzahl an identifizierten isobar markierten Peptiden und die Anzahl an Signalen des Reporterionenkanals 3, die durch die verschiedenen *data conversion* Programme in die *mgf*-Dateien exportiert wurden, sind in Tabelle 6.3 abgebildet.

Für Myoglobin (Verhältnis in A:B: 2:1) wurden zwischen 67 und 75 iTRAQ 4-plex markierte Peptide identifiziert. Für iTRAQ 8-plex markierte Peptide betrug die Anzahl an identifizierten Peptiden zwischen 43 und 50; für TMT 6-plex markierte Peptide zwischen 58 und 72. Auch die Anzahl der identifizierten Peptide über die TS2Mascot Software unter der Einstellung MIO *enabled* und den *merged mgf*-Dateien ist nahezu identisch. Lediglich im Fall der TMT markierten Peptide wurde mit den *merged mgf*-Dateien ein Peptid weniger (66 im Vergleich zu 67 identifizierten TMT 6-plex markierte Peptide) identifiziert. Nahezu alle Signale des Reporterionenkanals 3 wurden in die *mgf*-Dateien exportiert und nur wenige Peaks wurden nicht korrekt annotiert (0% - 4%). Dieses Ergebnis war sowohl unabhängig von dem verwendeten isobaren *tag* als auch von der verwendeten *data conversion* Software. Hier war also kein *peak picking* Problem beobachtbar (siehe Tabelle 6.3, Teil A).

Für Cytochrom c (Verhältnis in A:B: 5:1) betrug die Anzahl an fehlenden Peaks des Reporterionenkanals 3 zwischen 44% und 81% für das LPtMt, den MS DATA Converter und

die TS2Mascot Software unter der Einstellung MIO *enabled* (siehe Tabelle 6.3, Teil B). Für die Experimente mit iTRAQ 4-plex wurden mit dem MS DATA Converter, mit dem LPtMt und mit der TS2Mascot Software unter der Einstellung MIO *enabled* jeweils 35 von 78 Werten des Reporterionenkanals 3 (44%) korrekt zugeordnet. Für iTRAQ 8-plex waren die erhaltenen Werte mit 29 von 52 (56%), 27 von 54 (50%) und 27 von 53 (51%) für die 3 verschiedenen *data conversion* Programme ebenfalls vergleichbar. Von den TMT 6-plex markierten Peptiden wurden mit 12 aus 64 (18%), 12 aus 63 (19%) und 13 aus 60 (21%) die wenigsten Werte des Reporterionenkanals 3 korrekt zugeordnet. Im Gegensatz dazu wurde praktisch kein *peak picking* Problem für TS2Mascot unter der Einstellung MIO *disabled* beobachtet. Hier wurde lediglich der Peak des Reporterionenkanals 3 eines mit iTRAQ 8plex markierten Peptides nicht in die mgf-Datei exportiert. Das gleiche gilt für die Zuordnung in den *merged* mgf-Dateien. Hier wurden in den Datenbanksuchen jedoch mehr Peptide (2 Peptide mehr für iTRAQ 4-plex, 10 Peptide mehr für iTRAQ 8-plex und 5 Peptide mehr für TMT 6-plex) als für TS2Mascot unter der Einstellung MIO *disabled* identifiziert.

Tab 6.3: Anzahl der identifizierten isobar markierten Peptide der Proteine **A)** Myoglobin, **B)** Cytochrom c und **C)** Ribonuklease. Die Erwartungswerte der Reporterionenkanal-(RIK-) Verhältnisse für die Proteine betragen **A)** 0,5; **B)** 0,2 und **C)** 0,1. Zusätzlich ist die Anzahl an Signalen des RIK 3 aufgelistet, die durch das jeweilige *data conversion* Programm annotiert wurden. Insgesamt wurden fünf technische Replikate analysiert. LPtMt = *Launch Peaks to Mascot* tool, MIO = *monoisotopic peaks only*.

isobares tag	MS Data Converter		LPtMt		TS2Mascot MIO <i>enabled</i>		TS2Mascot MIO <i>disabled</i>		TS2Mascot <i>merged</i>	
	# identifizierter markierter Peptide	# Werte RIK 3	# identifizierter markierter Peptide	# Werte RIK 3	# identifizierter markierter Peptide	# Werte RIK 3	# identifizierter markierter Peptide	# Werte RIK 3	# identifizierter markierter Peptide	# Werte RIK 3
A Myoglobin										
iTRAQ 4-plex	75	74	69	68	69	68	67	67	69	69
iTRAQ 8-plex	49	48	50	48	50	49	43	43	50	50
TMT 6-plex	68	68	72	72	67	67	58	58	66	66
B Cytochrom c										
iTRAQ 4-plex	78	35	78	35	78	35	76	76	78	78
iTRAQ 8-plex	52	29	54	27	53	27	43	42	53	52
TMT 6-plex	64	12	63	12	60	13	53	53	58	58
C Ribonuklease										
iTRAQ 4-plex	7	0	7	0	7	0	7	6	7	6
iTRAQ 8-plex	6	3	7	3	5	1	3	3	5	5
TMT 6-plex	6	0	11	0	8	0	6	6	8	8

Für Ribonuklease (Verhältnis in A:B: 10:1) wurden unter Verwendung des MS DATA Converter, des LPtMt und der TS2Mascot Software unter der Einstellung MIO *enabled*, für iTRAQ 4-plex und TMT 6-plex keine Peaks des Reporterionenkanals 3 in die mgf-Dateien

exportiert (siehe Tabelle 6.3, Teil C). Lediglich für iTRAQ 8-plex wurden Peaks des Reporterionenkanals 3 korrekt annotiert, wobei 3 von 6 (50%) Werten des Reporterionenkanals 3 mit dem MS DATA Converter, 3 von 7 (43%) mit dem LPtMt und 1 von 5 (20%) mit der TS2Mascot Software unter der Einstellung MIO *enabled* korrekt zugeordnet wurden. Unter Verwendung der TS2Mascot Software mit der Einstellung MIO *disabled* sowie unter Verwendung des *RICmerge.pl* Programms wurden nahezu alle Peaks des Reporterionenkanals 3 in die mgf-Dateien exportiert. Lediglich 0% - 14% der Signale wurden nicht korrekt durch das *data conversion* Programm annotiert. In den Datenbanksuchen wurden jedoch jeweils 2 Peptide mehr für iTRAQ 8-plex und TMT 6-plex als mit TS2Mascot unter der Einstellung MIO *disabled* identifiziert.

Beispielhaft ist in Abbildung 6.3 der Einfluss der fehlenden Peaks des Reporterionenkanal 3 auf das Ergebnis der Quantifizierung für Cytochrom c und Ribonuklease unter Verwendung des isobaren *tags* iTRAQ 4-plex gezeigt.

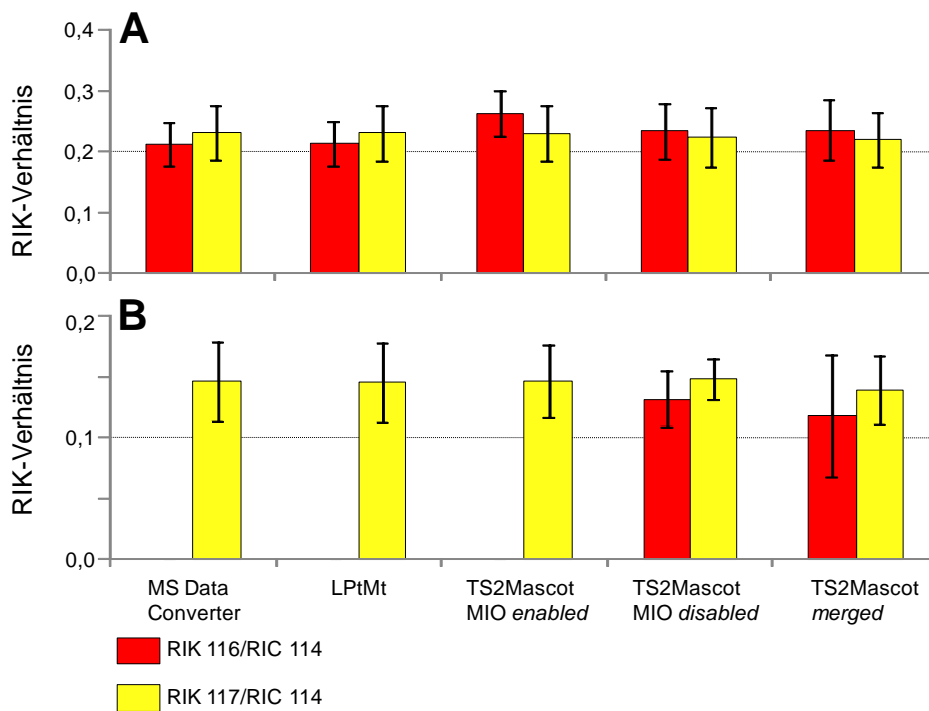


Abb. 6.3: Durchschnittliche geometrische Mittelwerte und Standardabweichungen für iTRAQ 4-plex Reporterionenkanal- (RIK-) Verhältnisse der beiden Proteine **A)** Cytochrom c und **B)** Ribonuklease. Die Erwartungswerte für die RIK-Verhältnisse der beiden Proteine betragen **A)** 0,2 und **B)** 0,1 und sind durch eine schwarze-gestrichelte Linie gekennzeichnet. Insgesamt wurden fünf technische Replikate analysiert. LPtMt = *Launch Peaks to Mascot* tool, MIO = *monoisotopic peaks only*.

Der Erwartungswert für das Verhältnis der Reporterionenkanäle (RIK 3/RIK 2) betrug 0,2 für Cytochrom c und 0,1 für Ribonuklease.

Für Cytochrom c sind keine ausgeprägten Unterschiede für die durchschnittlichen geometrischen Mittelwerte der Reporterionenkanal-Verhältnisse feststellbar. Die unterschiedliche Anzahl an Peaks des Reporterionenkanals 3 und die daraus resultierenden wenigeren Werte an Reporterionenkanal-Verhältnissen (RIK 3/RIK 2) scheint hier keinen großen Einfluss auf die Quantifizierung auszuüben.

Im Gegensatz dazu war für Ribonuklease über den Reporterionenkanal 3 der isobaren *tags* iTRAQ 4-plex und TMT 6-plex keine Quantifizierung möglich, da unter Verwendung des LPTMt, des MS DATA Converter und der TS2Mascot Software unter der Einstellung MIO *enabled* keine Peaks dieses Reporterionenkanals exportiert wurden (siehe auch Tabelle 6.3, Teil C). Hier wird also der negative Einfluss des *peak picking* Problems auf die Quantifizierung sehr deutlich.

Die Peaks der Reporterionenkanäle 3 von α -Crystallin A (Verhältnis von A:B: 1:2), α -Casein (Verhältnis von A:B: 1:5) und Lysozym (Verhältnis von A:B: 1:10) wurden mit allen *data-conversion* Programmen korrekt annotiert.

6.4 Diskussion

Ziel der hier gezeigten Experimente war nicht der qualitative und/oder quantitative Vergleich der unterschiedlichen isobaren *tags* oder der verschiedenen *data conversion* Programme. Vielmehr sollte eine Analyse des Problems fehlender Reporterionensignale bei der *mgf*-Datei Generierung in Abhängigkeit des verwendeten isobaren *tags* und *data conversion* Programms durchgeführt werden. Für einen qualitativen und/oder quantitativen Vergleich zwischen den drei verschiedenen Markierungsreagenzien ist die Anzahl an Proteinen bzw. identifizierten Peptiden nicht ausreichend, um eine aussagekräftige Statistik zu führen.

Dennoch sei hier darauf hingewiesen, dass die Ergebnisse bezüglich der Anzahl an Identifizierungen gut mit den Beobachtungen in LC-ESI MS/MS Experimenten an einem LTQ-Orbitrap Massenspektrometer übereinstimmen (Pichler *et al.*, 2010). Auch in diesen Experimenten wurden unter Verwendung unterschiedlich komplexer Proben die meisten Peptide mit dem isobaren *tag* iTRAQ 4plex, gefolgt von TMT 6-plex und die wenigsten mit iTRAQ 8-plex identifiziert. Auf die Gründe, die in der Veröffentlichung beschrieben werden soll an dieser Stelle jedoch nicht weiter eingegangen werden.

Die Tatsache, dass mit der TS2Mascot Software unter Verwendung der MIO *disabled* Einstellung, die wenigsten Peptide identifiziert wurden (siehe Abbildung 6.2) lässt sich einfach begründen. So wurden durch diese Einstellungen nicht nur die monoisotopischen Peaks in die mgf-Datei exportiert, sondern alle Signale. Dies hat den Vorteil, dass mit allen Signalen auch diejenigen der Reporterionen in die mgf-Datei exportiert werden. In einer Mascot Datenbanksuche können jedoch nur die monoisotopischen Signale zugeordnet werden. Dadurch enthalten die MS/MS-Spektren in diesen mgf-Dateien eine höhere Anzahl an Peaks, die nicht zugeordnet werden können und erniedrigen damit den durchschnittlichen Mascot *ion score* sowie die Anzahl an Identifizierungen (siehe Abbildung 6.2)

In Abbildung 6.4 ist beispielhaft das MS/MS-Spektrum des Peptides YPNCA¹¹⁴YK aus dem Protein Ribonuklease gezeigt. Es wurde mit iTRAQ 4-plex markiert. Alle vier Signale der Reporterionenkanäle 114, 115, 116 und 117 sind deutlich in dem MS/MS-Spektrum abgebildet. Dennoch wurde der Peak für den Reporterionenkanal 116 nicht durch die Software des Massenspektrometers annotiert.

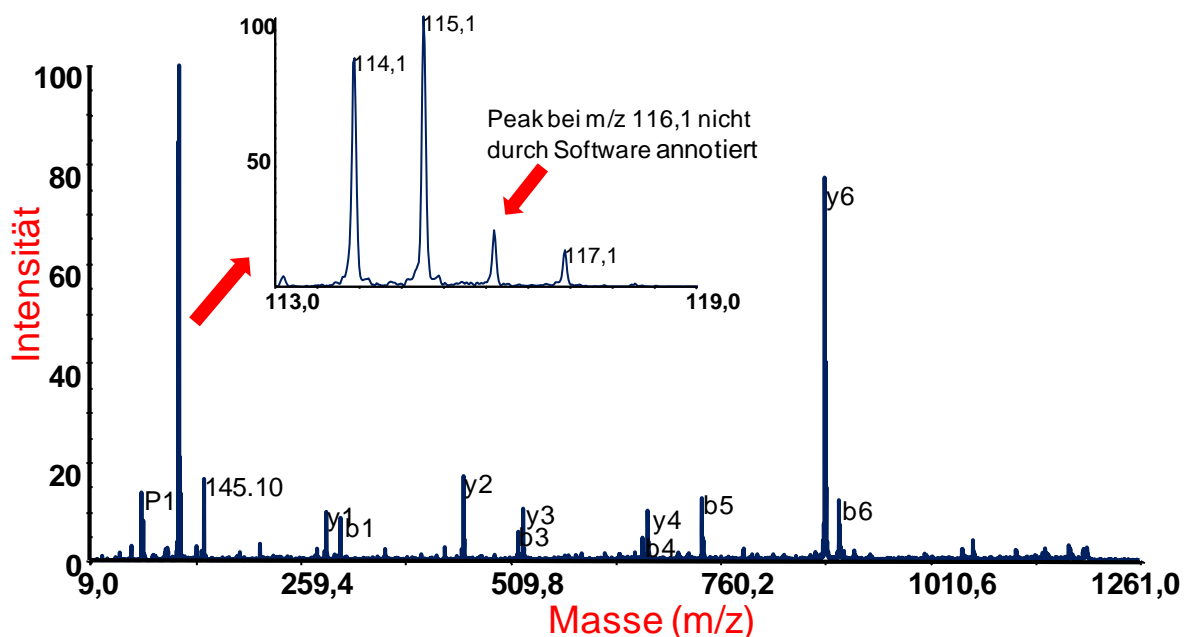


Abb. 6.4: MS/MS-Spektrum des Peptides YPNCA¹¹⁴YK ($m/z = 1192,57$) aus dem Protein Ribonuklease, mit iTRAQ 4-plex Markierung. Das Spektrum wurde mit einem MALDI-TOF/TOF Massenspektrometer im *positive ion mode* generiert. Der m/z -Bereich, der die Signale der Reporterionen enthält, ist in Vergrößerung dargestellt. Das Signal des iTRAQ-Reporterions 116 wurde durch die Software des Massenspektrometers nicht annotiert.

Eine Erklärung für das *peak picking* Problem könnte die Berechnung der Isotopenverteilung, das so genannte *de-isotoping* sein. So berechnet die 4000 Series Explorer Software (AB SCIEX, Darmstadt, Deutschland) die Isotopenverteilung auf Basis der Formel C_6H_5NO . Diese stimmt als Minimum für alle Aminosäuren überein. Die gleiche Formel wird für die Peaks der Reporterionen verwendet. Diese weisen jedoch eine andere atomare Zusammensetzung auf. Die atomare Zusammensetzung wirkt sich auf Grund der unterschiedlichen Anteile an Isotopen eines Atoms jedoch stark auf das Intensitätsverhältnis der unterschiedlichen Isotope eines Analyten aus. Damit ist also das Verhältnis der Intensitäten des M+1-Isotops, M+2-Isotops usw. betroffen. Die Berechnung der Isotopenverteilung sollte deshalb für die Reporterionen und die Fragmentationen der Peptide, auf unterschiedlichen Formeln basierend, berechnet werden. Möglicherweise berechnet die Software das Signal mit dem m/z -Verhältnis 115,1 als das erste Isotop des Reporterionenkanals 115 und das Signal mit dem m/z -Verhältnis von 116,1 auf Grund der falschen Formel als das zweite (M+1)-Isotop dieses Reporterionenkanals. Da nur die monoisotopischen Signale *gepickt* werden, hätte dies zur Folge, dass dieses Signal, das eigentlich zu dem Reporterionenkanal 116 gehört, nicht annotiert wird.

Nur unter Verwendung der TS2Mascot Software unter der Einstellung MIO *disabled* wurde das Signal in die mgf-Datei exportiert. Dies war nur möglich, da dann nicht nur die monoisotopischen, sondern alle Signale annotiert werden. Dies würde auch den Umstand erklären, warum nur die Reporterionen, die eine geringe Intensität aufweisen nicht korrekt annotiert werden, die auf ein Reporterion mit einer hohen Intensität im MS/MS-Spektrum folgen. Für die Proteine α -Crystallin A (Verhältnis von A:B: 1:2), α -Casein (Verhältnis von A:B: 1:5) und Lysozym (Verhältnis von A:B: 1:10) wurden deshalb mit allen *data-conversion* Programmen alle Peaks des Reporterionenkanals 3 korrekt annotiert. Hier folgt auf ein Reporterion mit einer geringen Intensität (Reporterionenkanal 2) ein Reporterion mit einer hohen Intensität (Reporterionenkanal 3). Damit kann die Software unter Verwendung der Formel C_6H_5NO das hohe Signal nicht fälschlicherweise als M+1-Isotop des niedrigen Signals interpretieren und die Annotierung der Signale erfolgt korrekt.

Betrachtet man nun die unterschiedlichen Ergebnisse bezüglich der Quantifizierung der unterschiedlichen *data conversion* Programme, so stellt man fest, dass die TS2Mascot Software unter der Einstellung MIO *enabled* ebenfalls nahezu alle Peaks des Reporterionenkanals 3 in die mgf-Datei exportiert. Das Problem besteht hier jedoch darin, dass diese Einstellung, wie zuvor beschrieben, schlechtere Mascot *ion scores* und damit weniger Identifizierungen zur Folge hat.

Insgesamt scheint der Einfluss der fehlenden Peaks des Reporterionenkanals 3 sehr unterschiedlich zu sein. In der Literatur wird der Effekt der sogenannten *underestimation* (Ow, *et al.*, 2009) für auf iTRAQ-basierte Studien beschrieben. Der Grund scheint v. a. an Untergrundsignalen zu liegen, die während der MS/MS-Messung nicht entfernt werden können. Dadurch tendieren die Verhältnisse der Reporterionen zu dem Wert 1. Es werden also Verhältnisse berechnet, die weniger von diesem Wert 1 abweichen, als ihr tatsächliches Verhältnis ergeben sollte. Die in dieser Arbeit gemessenen Verhältnisse der Reporterionenkanäle (RIK 3/RIK 2) für die Proteine Cytochrom c und Ribonuklease stimmen allerdings mit den Erwartungswerten von 0,2 bzw. 0,1 am Beispiel für iTRAQ 4-plex sehr gut überein (siehe Abbildung 6.3). Obwohl hier zum Teil sehr viele Peaks des Reporterionenkanals 3 und damit auch Werte für die Quantifizierung fehlten, sind praktisch keine Unterschiede für die durchschnittlichen geometrischen Mittelwerte der Reporterionenkanal-Verhältnisse (RIK 3/RIK 2) erkennbar. Trotz der fehlenden quantitativen Werte war deren Anzahl anscheinend immer noch ausreichend, um korrekte geometrische Mittelwerte zu berechnen. Reduziert man jedoch z. B. die Anzahl an technischen Replikaten pro biologischen Zustand, so reduziert sich zusätzlich zur Anzahl an quantitativen auch die Anzahl der qualitativen Werte. Außerdem muss man bedenken, dass in diesen Experimenten ein 15-Proteinstandard und kein Vollproteom als Probe verwendet wurde. Dies hätte zu einer viel höheren Probenkomplexität geführt, womit pro Protein wohl weniger Peptide identifiziert worden wären.

Deshalb könnte man davon ausgehen, dass das Problem der nicht korrekt annotierten Signale der Reporterionenkanäle in *large-scale* quantitativen (Voll-) Proteomstudien einen weit mehr ausgeprägten Effekt ausüben. Dies ist sicherlich nicht der Fall, wenn für ein Protein genügend quantitative Werte vorhanden sind (wie hier bei Cytochrom c). Diese können sowohl aus der Anzahl an biologischen oder technischen Replikaten als auch aus einer hohen Anzahl an identifizierten Peptiden stammen.

Es gibt aber auch Fälle, in denen z. B. durch eine geringe Probenmenge nur wenige oder gar keine Replikate gemessen werden können und somit auch nur wenige quantitative Werte vorliegen. Dies trifft auch auf kleine Proteine zu, von denen nur eine geringe Anzahl an (z. B. tryptischen) Peptiden erhalten werden kann. Zusätzlich sollten nur proteotypische Peptide, also Peptide, die nur aus einem bestimmten Protein stammen können, für die Quantifizierung herangezogen werden (Picotti *et al.*, 2007). Dadurch verringert sich zwar die Anzahl an quantitativen Daten, aber falsche Ergebnisse, die durch das Mitteln quantitativer Daten

verschiedener Isoformen oder (posttranslational) modifizierten Proteinen entstammen, können verhindert werden.

Auch in Experimenten wie in Kapitel 3 und Kapitel 4, in denen die Quantifizierung nur über einzelne Peptide erfolgt, kann das Problem der nicht korrekt annotierten Reporterionen ein großes Problem darstellen. Die Ergebnisse für Ribonuklease, bei denen unter Verwendung des LPTMt, des MS DATA Converter und der TS2Mascot Software unter der Einstellung MIO *enabled* keine Peaks des Reporterionenkanals 3 für iTRAQ 4-plex und TMT 6-plex exportiert wurden (siehe Tabelle 6.3, Teil C), verdeutlichen dieses Problem noch einmal zusätzlich.

Zusammenfassend kann man sagen, dass unter Verwendung der *merged mgf*-Dateien für die Datenbanksuche ein *tool* entwickelt wurde, dass sowohl die Identifizierung auf Basis der monoisotopischen Peaks als auch die Quantifizierung über nahezu alle Signale der Reporterionenkanäle ermöglicht. Damit wurde dem *peak picking* Problem der *data conversion* Programme erfolgreich entgegengewirkt.

6.5 Zusammenfassung und Ausblick

Die hier gezeigten Daten des Experimentes mit einem 15-Proteinstandard belegen, dass ein Problem bestehen kann, wenn LC-MALDI-TOF/TOF MS/MS Rohdaten auf Basis einer Quantifizierung über isobare Isotopenmarkierung in *mgf*-Dateien exportiert werden. So werden Signale eines Reporterionenkanals mit geringer Intensität nicht in die *mgf*-Datei exportiert, wenn sie auf ein Signal eines Reporterionenkanals mit einer hohen Intensität folgen. Die Experimente belegen, dass dieses Problem auch unabhängig von dem verwendeten isobaren *tag* besteht.

Insgesamt wurden die drei verschiedenen isobaren *tags* iTRAQ 4-plex, iTRAQ 8-plex und TMT 6-plex und die drei *data conversion* Programme LPTMt, TS2Mascot und MS DATA Converter untersucht. Ein Schwellenwert, ab dem dieses Problem zu beobachten ist, wurde nicht bestimmt. Dies ist für die hier vorgestellten Experimente aber auch nicht entscheidend.

Lediglich unter der Verwendung der TS2Mascot Software mit der Einstellung *monoisotopic peaks only* (MIO) *disabled* wurde das Problem der fehlenden Signale der Reporterionen nicht beobachtet. Hier wurden nahezu alle Signale der Reporterionen in die *mgf*-Datei exportiert. Unter dieser Einstellung sind alle Signale der MS/MS-Spektren, sowohl der Reporterionen als auch der Peptidfragmentationen, in der *mgf*-Datei enthalten. In einer Mascot Datenbanksuche werden allerdings nur die monoisotopischen Signale zugeordnet. Durch die erhöhte Anzahl an

Signalen der MS/MS-Spektren in der mgf-Datei, die nicht zugeordnet werden können, wird der durchschnittliche Mascot *ion score* erniedrigt und damit verbunden weniger Peptide identifiziert.

Die meisten isobar markierten Peptide wurden durch die mgf-Dateien des LPtMt identifiziert. Auch wenn das hier gezeigte Experiment eigentlich einen Vergleich der verschiedenen isobaren *tags* und der verschiedenen *data conversion* Programme auf Grund der geringen Komplexität zulässt, so sei hier dennoch darauf hingewiesen, dass auch die Möglichkeit besteht, die fehlenden Signale der Reporterionen in diesen mgf-Dateien zu ergänzen. Sollte man also generell über das LPtMt mehr Peptide identifizieren, so könnte man dies ebenfalls mit der verbesserten Quantifizierung kombinieren.

Darüber hinaus wurde das hier entwickelte Perl-Skript auch für die Experimente der Untersuchungen des Einflusses verschiedener Puffer auf die Spaltspezifität der Endoprotease GluC mittels des Q-PICS Ansatzes verwendet (Kapitel 4).

Literaturverzeichnis

- Adriouch, S.; Bannas, P.; Schwarz, N.; Fliegert, R.; Guse, A. H.; Seman, M.; Haag, F.; Koch-Nolte, F., ADP-ribosylation at R125 gates the P2X7 ion channel by presenting a covalent ligand to its nucleotide binding site. *FASEB J* **2008**, *22*, 861-9.
- Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198-207.
- Agard, N. J.; Mahrus, S.; Trinidad, J. C.; Lynn, A.; Burlingame, A. L.; Wells, J. A., Global kinetic analysis of proteolysis via quantitative targeted proteomics. *Proc Natl Acad Sci U S A* **2012**, *109*, 1913-8.
- Aggarwal, K.; Choe, L. H.; Lee, K. H., Shotgun proteomics using the iTRAQ isobaric tags. *Brief Funct Genomic Proteomic* **2006**, *5*, 112-20.
- Ahrends, R.; Pieper, S.; Kuhn, A.; Weisshoff, H.; Hamester, M.; Lindemann, T.; Scheler, C.; Lehmann, K.; Taubner, K.; Linscheid, M. W., A metal-coded affinity tag approach to quantitative proteomics. *Mol Cell Proteomics* **2007**, *6*, 1907-16.
- Anderson, N. L.; Taylor, J.; Scandora, A. E.; Coulter, B. P.; Anderson, N. G., The TYCHO system for computer analysis of two-dimensional gel electrophoresis patterns. *Clin Chem* **1981**, *27*, 1807-20.
- Andren, P. E.; Emmett, M. R.; Caprioli, R. M., Micro-electrospray: zeptomole/attomole per microliter sensitivity for peptides. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 867-69.
- Arntzen, M. O.; Koehler, C. J.; Treumann, A.; Thiede, B., Quantitative proteome analysis using isobaric peptide termini labeling (IPTL). *Methods Mol Biol* **2011**, *753*, 65-76.
- auf dem Keller, U.; Schilling, O., Proteomic techniques and activity-based probes for the system-wide study of proteolysis. *Biochimie* **2010**, *92*, 1705-14.
- auf dem Keller, U.; Overall, C. M., CLIPPER: an add-on to the Trans-Proteomic Pipeline for the automated analysis of TAILS N-terminomics data. *Biol Chem* **2012**, *393*, 1477-83.
- auf dem Keller, U.; Prudova, A.; Eckhard, U.; Fingleton, B.; Overall, C. M., Systems-level analysis of proteolytic events in increased vascular permeability and complement activation in skin inflammation. *Sci Signal* **2013**, *6*, rs2.
- Baeumlisberger, D.; Arrey, T. N.; Rietschel, B.; Rohmer, M.; Papisotiriou, D. G.; Mueller, B.; Beckhaus, T.; Karas, M., Labeling elastase digests with TMT: informational gain by identification of poorly detectable peptides with MALDI-TOF/TOF mass spectrometry. *Proteomics* **2010**, *10*, 3905-9.
- Banks, J. F., High-sensitivity peptide mapping using packed-capillary liquid chromatography and electrospray ionization mass spectrometry. *J Chromatogr A* **1996**, *743*, 99-104.
- Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* **2007**, *389*, 1017-31.
- Bantscheff, M.; Boesche, M.; Eberhard, D.; Matthieson, T.; Sweetman, G.; Kuster, B., Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer. *Mol Cell Proteomics* **2008**, *7*, 1702-13.

- Barrett, A. J.; McDonald, J. K., Nomenclature: protease, proteinase and peptidase. *Biochem J* **1986**, 237, 935.
- Beavis, R. C.; Chait, B. T., Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. *Rapid Commun Mass Spectrom* **1989a**, 3, 432-5.
- Beavis, R. C.; Chait, B. T., Matrix-assisted laser-desorption mass spectrometry using 355 nm radiation. *Rapid Commun Mass Spectrom* **1989b**, 3, 436-9.
- Beavis, R. C.; Chaudhary, T.; Chait, B. T., Alpha-cyano-4-hydroxycinnamic acid as a matrix for matrix assisted laser desorption mass spectrometry. *Org. Mass Spectrom.* **1992**, 156-158.
- Becker-Pauly, C.; Barre, O.; Schilling, O.; Auf dem Keller, U.; Ohler, A.; Broder, C.; Schutte, A.; Kappelhoff, R.; Stocker, W.; Overall, C. M., Proteomic analyses reveal an acidic prime side specificity for the astacin metalloprotease family reflected by physiological substrates. *Mol Cell Proteomics* **2011**, 10, M111 009233.
- Biemann, K., Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol* **1990**, 193, 886-7.
- Binossek, M. L.; Naegler, D. K.; Becker-Pauly, C.; Schilling, O., Proteomic Identification of Protease Cleavage Sites Characterizes Prime- and Non-prime Specificity of Cysteine Cathepsins B, L, and S. *J Proteome Res* **2011**, 10, 5363-5373.
- Bond, J. S.; Barrett, A. J., Degradation of fructose-1,6-bisphosphate aldolase by cathepsin B. *Biochem J* **1980**, 189, 17-25.
- Bredemeyer, A. J.; Lewis, R. M.; Malone, J. P.; Davis, A. E.; Gross, J.; Townsend, R. R.; Ley, T. J., A proteomic approach for the discovery of protease substrates. *Proc Natl Acad Sci USA* **2004**, 101, 11785-90.
- Breitwieser, F. P.; Muller, A.; Dayon, L.; Kocher, T.; Hainard, A.; Pichler, P.; Schmidt-Erfurth, U.; Superti-Furga, G.; Sanchez, J. C.; Mechtler, K.; Bennett, K. L.; Colinge, J., General statistical modeling of data from protein relative expression isobaric tags. *J Proteome Res* **2011**, 10, 2758-66.
- Broder, C.; Arnold, P.; Vadon-Le Goff, S.; Konerding, M. A.; Bahr, K.; Muller, S.; Overall, C. M.; Bond, J. S.; Koudelka, T.; Tholey, A.; Hulmes, D. J.; Moali, C.; Becker-Pauly, C., Metalloproteases meprin alpha and meprin beta are C- and N-procollagen proteinases important for collagen assembly and tensile strength. *Proc Natl Acad Sci U S A* **2013**, 110, 14219-24.
- Brown, R. S.; Lennon, J. J., Mass resolution improvement by incorporation of pulsed ion extraction in a matrix-assisted laser desorption/ionization linear time-of-flight mass spectrometer. *Anal Chem* **1995**, 67, 1998-2003.
- Bunger, M. K.; Cargile, B. J.; Ngunjiri, A.; Bundy, J. L.; Stephenson, J. L., Jr., Automated proteomics of *E. coli* via top-down electron-transfer dissociation mass spectrometry. *Anal Chem* **2008**, 80, 1459-67.
- Chen, H. S.; Rejtar, T.; Andreev, V.; Moskovets, E.; Karger, B. L., High-speed, high-resolution monolithic capillary LC-MALDI MS using an off-line continuous deposition interface for proteomic analysis. *Anal Chem* **2005**, 77, 2323-31.

- Chen, J. G.; Weber, S. G.; Glavina, L. L.; Cantwell, F. F., Electrical double-layer models of ion-modified (ion-pair) reversed-phase liquid chromatography. *J Chromatogr A* **1993**, 656, 549-76.
- Colby, S. M.; King, T. B.; Reilly, J. P.; Lubmann, D. M., Improving the resolution of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry by exploiting the correlation between ion position and velocity. *Rapid Commun Mass Spectrom* **1994**, 8, 865-868.
- Corthals, G. L.; Wasinger, V. C.; Hochstrasser, D. F.; Sanchez, J. C., The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis* **2000**, 21, 1104-15.
- Cox, J.; Mann, M., Is proteomics the new genomics? *Cell* **2007**, 130, 395-8.
- Creighton, T. E., *Proteins: Structures and molecular properties*, 2nd ed. W.H. Freeman and Company: New York, 1993.
- Dayon, L.; Hainard, A.; Licker, V.; Turck, N.; Kuhn, K.; Hochstrasser, D. F.; Burkhard, P. R.; Sanchez, J. C., Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Anal Chem* **2008**, 80, 2921-31.
- Dayon, L.; Pasquarello, C.; Hoogland, C.; Sanchez, J. C.; Scherl, A., Combining low- and high-energy tandem mass spectra for optimized peptide quantification with isobaric tags. *J Proteomics* **2010a**, 73, 769-77.
- Dayon, L.; Turck, N.; Scherl, A.; Hochstrasser, D. F.; Burkhard, P. R.; Sanchez, J. C., From relative to absolute quantification of tryptic peptides with tandem mass tags: application to cerebrospinal fluid. *Chimia (Aarau)* **2010b**, 64, 132-5.
- De Hoffmann, E.; Charette, J.; Stroobant, V., *Mass Spectrometry: Principles and Applications*. John Wiley & Sons: Paris, 1996.
- Dean, R. A.; Overall, C. M., Proteomics discovery of metalloproteinase substrates in the cellular context by iTRAQ labeling reveals a diverse MMP-2 substrate degradome. *Mol Cell Proteomics* **2007**, 6, 611-23.
- Deutsch, E., mzML: a single, unifying data format for mass spectrometer output. *Proteomics* **2008**, 8, 2776-7.
- Diamond, S. L., Methods for mapping protease specificity. *Curr Opin Chem Biol* **2007**, 11, 46-51.
- Doucet, A.; Butler, G. S.; Rodriguez, D.; Prudova, A.; Overall, C. M., Metadegradomics: toward in vivo quantitative degradomics of proteolytic post-translational modifications of the cancer proteome. *Mol Cell Proteomics* **2008**, 7, 1925-51.
- Doucet, A.; Kleifeld, O.; Kizhakkedathu, J. N.; Overall, C. M., Identification of proteolytic products and natural protein N-termini by Terminal Amine Isotopic Labeling of Substrates (TAILS). *Methods Mol Biol* **2011**, 753, 273-87.
- Doucet, A.; Overall, C. M., Broad coverage identification of multiple proteolytic cleavage site sequences in complex high molecular weight proteins using quantitative proteomics as a complement to edman sequencing. *Mol Cell Proteomics* **2011a**, 10, M110 003533.

- Doucet, A.; Overall, C. M., Amino-Terminal Oriented Mass Spectrometry of Substrates (ATOMS) N-terminal sequencing of proteins and proteolytic cleavage sites by quantitative mass spectrometry. *Methods Enzymol* **2011b**, 501, 275-93.
- Drapeau, G. R.; Boily, Y.; Houmard, J., Purification and properties of an extracellular protease of *Staphylococcus aureus*. *J Biol Chem* **1972**, 247, 6720-6.
- Dreisewerd, K., *The desorption process in MALDI*. *Chem Rev* **2003**, 103, 395-426.
- Eckhard, U.; Huesgen, P. F.; Brandstetter, H.; Overall, C. M., Proteomic protease specificity profiling of clostridial collagenases reveals their intrinsic nature as dedicated degraders of collagen. *J Proteomics* **2013**, doi: 10.1016/j.jprot.2013.10.004.
- Elias, J. E.; Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **2007**, 4, 207-14.
- Eng, J. K.; McCormack, A. L.; Yates III, J. R., An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **1994**, 5, 976-989.
- Enoksson, M.; Li, J.; Ivancic, M. M.; Timmer, J. C.; Wildfang, E.; Eroshkin, A.; Salvesen, G. S.; Tao, W. A., Identification of proteolytic cleavage sites by quantitative proteomics. *J Proteome Res* **2007**, 6, 2850-8.
- Ericson, C.; Phung, Q. T.; Horn, D. M.; Peters, E. C.; Fitchett, J. R.; Ficarro, S. B.; Salomon, A. R.; Brill, L. M.; Brock, A., An automated noncontact deposition interface for liquid chromatography matrix-assisted laser desorption/ionization mass spectrometry. *Anal Chem* **2003**, 75, 2309-15.
- Ferguson, M. A.; Williams, A. F., Cell-surface anchoring of proteins via glycosyl-phosphatidylinositol structures. *Annu Rev Biochem* **1988**, 57, 285-320.
- Fonovic, M.; Bogoy, M., Activity-based probes as a tool for functional proteomic analysis of proteases. *Expert Rev Proteomics* **2008**, 5, 721-30.
- Gan, C. S.; Chong, P. K.; Pham, T. K.; Wright, P. C., Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *J Proteome Res* **2007**, 6, 821-7.
- 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. *J Proteome Res* **2004**, 3, 958-64.
- Geng, X.; Regnier, F. E., Retention model for proteins in reversed-phase liquid chromatography. *J Chromatogr* **1984**, 296, 15-30.
- Gevaert, K.; Goethals, M.; Martens, L.; Van Damme, J.; Staes, A.; Thomas, G. R.; Vandekerckhove, J., Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat Biotechnol* **2003**, 21, 566-9.
- Gluckmann, M.; Fella, K.; Waidelich, D.; Merkel, D.; Kruff, V.; Kramer, P. J.; Walter, Y.; Hellmann, J.; Karas, M.; Kroger, M., Prevalidation of potential protein biomarkers in toxicology using iTRAQ reagent technology. *Proteomics* **2007**, 7, 1564-74.

- Goodlett, D. R.; Keller, A.; Watts, J. D.; Newitt, R.; Yi, E. C.; Purvine, S.; Eng, J. K.; von Haller, P.; Aebersold, R.; Kolker, E., Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation. *Rapid Commun Mass Spectrom* **2001**, *15*, 1214-21.
- Gregorius, B.; Schaumlöffel, D.; Hildebrandt, A.; Tholey, A., Characterization of metal-labelled peptides by matrix-assisted laser desorption/ionization mass spectrometry and tandem mass spectrometry. *Rapid Commun Mass Spectrom* **2010**, *24*, 3279-89.
- Gregorius, B.; Jakoby, T.; Schaumlöffel, D.; Tholey, A., Metal labeling for accurate multiplexed peptide quantification via matrix-assisted laser desorption/ionization mass spectrometry. *Anal Bioanal Chem* **2013**, *405*, 2735-41.
- Griffin, T. J.; Xie, H.; Bandhakavi, S.; Popko, J.; Mohan, A.; Carlis, J. V.; Higgins, L., iTRAQ reagent-based quantitative proteomic analysis on a linear ion trap mass spectrometer. *J Proteome Res* **2007**, *6*, 4200-9.
- Gu, C.; Somogyi, A.; Wysocki, V. H.; Medzihradszky, K. F., Fragmentation of protonated oligopeptides XLDVLQ (X=L, H, K or R) by surface induced dissociation: additional evidence for the 'mobile proton' model. *Anal Chim Acta* **1999**, *397*, 247-256.
- Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **1999**, *17*, 994-9.
- Hansen, L. C.; Sievers, R. E., Highly permeable open-pore polyurethane columns for liquid chromatography. *J Chromatogr A* **1974**, *99*, 123-133.
- Hayes, R. N.; Gross, M. L., Collision-induced dissociation. *Methods Enzymol* **1990**, *193*, 237-63.
- Hillenkamp, F.; Peter-Katalinic, J., *MALDI MS A Practical Guide to Instrumentation, Methods and Applications*. Wiley-VCH.: 2007.
- Hojilla, C. V.; Wood, G. A.; Khokha, R., Inflammation and breast cancer: metalloproteinases as common effectors of inflammation and extracellular matrix breakdown in breast cancer. *Breast Cancer Res* **2008**, *10*, 205.
- Hong, S.; Brass, A.; Seman, M.; Haag, F.; Koch-Nolte, F.; Dubyak, G. R., Lipopolysaccharide, IFN-gamma, and IFN-beta induce expression of the thiol-sensitive ART2.1 Ecto-ADP-ribosyltransferase in murine macrophages. *J Immunol* **2007**, *179*, 6215-27.
- Hong, S.; Schwarz, N.; Brass, A.; Seman, M.; Haag, F.; Koch-Nolte, F.; Schilling, W. P.; Dubyak, G. R., Differential regulation of P2X7 receptor activation by extracellular nicotinamide adenine dinucleotide and ecto-ADP-ribosyltransferases in murine macrophages and T cells. *J Immunol* **2009**, *183*, 578-92.
- Houmard, J.; Drapeau, G. R., Staphylococcal protease: a proteolytic enzyme specific for glutamoyl bonds. *Proc Natl Acad Sci U S A* **1972**, *69*, 3506-9.
- Hsieh, S.; Dreisewerd, K.; van der Schors, R. C.; Jimenez, C. R.; Stahl-Zeng, J.; Hillenkamp, F.; Jorgenson, J. W.; Geraerts, W. P.; Li, K. W., Separation and identification of peptides in single neurons by microcolumn liquid chromatography-matrix-assisted

- laser desorption/ionization time-of-flight mass spectrometry and postsource decay analysis. *Anal Chem* **1998**, 70, 1847-52.
- Hu, J.; Qian, J.; Borisov, O.; Pan, S.; Li, Y.; Liu, T.; Deng, L.; Wannemacher, K.; Kurnellas, M.; Patterson, C.; Elkabes, S.; Li, H., Optimized proteomic analysis of a mouse model of cerebellar dysfunction using amine-specific isobaric tags. *Proteomics* **2006**, 6, 4321-34.
- Huber, C. G.; Oefner, P. J.; Bonn, G. K., Rapid Analysis of Biopolymers on Modified Non-Porous Poly-styrene/divinylbenzene Particles. *Chromatographia* **1993**, 37, 653-658.
- Huber, C. G.; Kleindienst, G.; Bonn, G. K., Application of Micropellicular Polystyrene/divinylbenzene Stationary Phases for High-Performance Liquid Chromatography Electrospray-Mass Spectrometry of Proteins and Peptides. *Chromatographia* **1997**, 44, 438-448.
- Huesgen, P. F.; Alami, M.; Lange, P. F.; Foster, L. J.; Schroder, W. P.; Overall, C. M.; Green, B. R., Proteomic amino-termini profiling reveals targeting information for protein import into complex plastids. *PLoS One* **2013**, 8, e74483.
- Impens, F.; Colaert, N.; Helsens, K.; Ghesquiere, B.; Timmerman, E.; De Bock, P. J.; Chain, B. M.; Vandekerckhove, J.; Gevaert, K., A quantitative proteomics design for systematic identification of protease cleavage events. *Mol Cell Proteomics* **2010**, 9, 2327-33.
- International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome. In *Nature*, 2004/10/22 ed.; 2004; Vol. 431, pp 931-45.
- Jaskolla, T. W.; Karas, M., Compelling evidence for Lucky Survivor and gas phase protonation: the unified MALDI analyte protonation mechanism. *J Am Soc Mass Spectrom* **2011**, 22, 976-88.
- Jorgensen, T. J.; Bache, N.; Roepstorff, P.; Gardsvoll, H.; Ploug, M., Collisional activation by MALDI tandem time-of-flight mass spectrometry induces intramolecular migration of amide hydrogens in protonated peptides. *Mol Cell Proteomics* **2005**, 4, 1910-9.
- Karas, M.; Hillenkamp, F., Laser desorption/ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **1988**, 60, 2299-301.
- Karas, M.; Gluckmann, M.; Schafer, J., Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *J Mass Spectrom* **2000**, 35, 1-12.
- Karas, M.; Kruger, R., Ion formation in MALDI: the cluster ionization mechanism. *Chem Rev* **2003**, 103, 427-40.
- Karlsson, K. E.; Novotny, M., Separation efficiency of slurry-packed liquid chromatography microcolumns with very small inner diameters. *Anal Chem* **1988**, 60, 1662-5.
- Karp, N. A.; Huber, W.; Sadowski, P. G.; Charles, P. D.; Hester, S. V.; Lilley, K. S., Addressing accuracy and precision issues in iTRAQ quantitation. *Mol Cell Proteomics* **2010**, 9, 1885-97.
- Kelleher, N. L., Top-down proteomics. *Anal Chem* **2004**, 76, 197A-203A.

- Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R., Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **2002**, 74, 5383-5392.
- Keller, A.; Eng, J.; Zhang, N.; Li, X. J.; Aebersold, R., A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol Syst Biol* **2005**, 1, 2005 0017.
- Keshamouni, V. G.; Michailidis, G.; Grasso, C. S.; Anthwal, S.; Strahler, J. R.; Walker, A.; Arenberg, D. A.; Reddy, R. C.; Akulapalli, S.; Thannickal, V. J.; Standiford, T. J.; Andrews, P. C.; Omenn, G. S., Differential protein expression profiling by iTRAQ-2DLC-MS/MS of lung cancer cells undergoing epithelial-mesenchymal transition reveals a migratory/invasive phenotype. *J Proteome Res* **2006**, 5, 1143-54.
- Kessenbrock, K.; Plaks, V.; Werb, Z., Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell* **2010**, 141, 52-67.
- Kleifeld, O.; Doucet, A.; auf dem Keller, U.; Prudova, A.; Schilling, O.; Kainthan, R. K.; Starr, A. E.; Foster, L. J.; Kizhakkedathu, J. N.; Overall, C. M., Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat Biotechnol* **2010**, 28, 281-8.
- Kleifeld, O.; Doucet, A.; Prudova, A.; auf dem Keller, U.; Gioia, M.; Kizhakkedathu, J. N.; Overall, C. M., Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat Protoc* **2011**, 6, 1578-611.
- Klose, J., Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **1975**, 26, 231-43.
- Knochenmuss, R.; Zenobi, R., MALDI ionization: the role of in-plume processes. *Chem Rev* **2003**, 103, 441-52.
- Knochenmuss, R., Ion formation mechanisms in UV-MALDI. *Analyst* **2006**, 131, 966-86.
- Koch-Nolte, F.; Petersen, D.; Balasubramanian, S.; Haag, F.; Kahlke, D.; Willer, T.; Kastelein, R.; Bazan, F.; Thiele, H. G., Mouse T cell membrane proteins Rt6-1 and Rt6-2 are arginine/protein mono(ADPribosyl)transferases and share secondary structure motifs with ADP-ribosylating bacterial toxins. *J Biol Chem* **1996**, 271, 7686-93.
- Kocher, T.; Pichler, P.; Schutzbier, M.; Stingl, C.; Kaul, A.; Teucher, N.; Hasenfuss, G.; Penninger, J. M.; Mechtler, K., High precision quantitative proteomics using iTRAQ on an LTQ Orbitrap: a new mass spectrometric method combining the benefits of all. *J Proteome Res* **2009**, 8, 4743-52.
- Koehler, C. J.; Strozynski, M.; Kozielski, F.; Treumann, A.; Thiede, B., Isobaric peptide termini labeling for MS/MS-based quantitative proteomics. *J Proteome Res* **2009**, 8, 4333-41.
- Koehler, C. J.; Arntzen, M. O.; de Souza, G. A.; Thiede, B., An approach for triplex-isobaric peptide termini labeling (triplex-IPTL). *Anal Chem* **2013**, 85, 2478-85.
- Kohlbacher, O.; Reinert, K.; Gropl, C.; Lange, E.; Pfeifer, N.; Schulz-Trieglaff, O.; Sturm, M., TOPP--the OpenMS proteomics pipeline. *Bioinformatics* **2007**, 23, e191-7.

- Kussmann, M.; Lassing, U.; Sturmer, C. A.; Przybylski, M.; Roepstorff, P., Matrix-assisted laser desorption/ionization mass spectrometric peptide mapping of the neural cell adhesion protein neurolin purified by sodium dodecyl sulfate polyacrylamide gel electrophoresis or acidic precipitation. *J Mass Spectrom* **1997**, 32, 483-93.
- Kuzyk, M. A.; Ohlund, L. B.; Elliott, M. H.; Smith, D.; Qian, H.; Delaney, A.; Hunter, C. L.; Borchers, C. H., A comparison of MS/MS-based, stable-isotope-labeled, quantitation performance on ESI-quadrupole TOF and MALDI-TOF/TOF mass spectrometers. *Proteomics* **2009**, 9, 3328-3340.
- Lasaosa, M.; Delmotte, N.; Huber, C. G.; Melchior, K.; Heinzle, E.; Tholey, A., A 2D reversed-phase x ion-pair reversed-phase HPLC-MALDI TOF/TOF-MS approach for shotgun proteome analysis. *Anal Bioanal Chem* **2009**, 393, 1245-56.
- Lathia, U. S.; Ornatsky, O.; Baranov, V.; Nitz, M., Development of inductively coupled plasma-mass spectrometry-based protease assays. *Anal Biochem* **2010**, 398, 93-8.
- Lehmann, W. D., *Massenspektrometrie in der Biochemie* Spektrum Akademischer Verlag GmbH: Heidelberg, Berlin, Oxford, 1996.
- Lim, M. D.; Craik, C. S., Using specificity to strategically target proteases. *Bioorg Med Chem* **2009**, 17, 1094-100.
- Lopez-Otin, C.; Overall, C. M., Protease degradomics: a new challenge for proteomics. *Nat Rev Mol Cell Biol* **2002**, 3, 509-19.
- Lopez-Otin, C.; Hunter, T., The regulatory crosstalk between kinases and proteases in cancer. *Nat Rev Cancer* **2010**, 10, 278-92.
- Lottspeich, F.; Zorbas, H., *Bioanalytik*. Spektrum Akademischer Verlag GmbH: Heidelberg; Berlin, 1998.
- Lottspeich, F., Proteome Analysis: A Pathway to the Functional Analysis of Proteins. *Angew Chem Int Ed Engl* **1999**, 38, 2476-2492.
- Low, M. G., Biochemistry of the glycosyl-phosphatidylinositol membrane protein anchors. *Biochem J* **1987**, 244, 1-13.
- Mahoney, D. W.; Therneau, T. M.; Heppelmann, C. J.; Higgins, L.; Benson, L. M.; Zenka, R. M.; Jagtap, P.; Nelsestuen, G. L.; Bergen, H. R.; Oberg, A. L., Relative quantification: characterization of bias, variability and fold changes in mass spectrometry data from iTRAQ-labeled peptides. *J Proteome Res* **2011**, 10, 4325-33.
- Mamyrin, B. A., Laser Assisted Reflectron Time-of-Flight Mass Spectrometry. *Int. J. mass Spectrom. Ion Proc.* **1994**, 131, 1-19.
- Marino, G.; Huesgen, P. F.; Eckhard, U.; Overall, C. M.; Schroder, W. P.; Funk, C., Family-wide characterization of Matrix Metallo-proteinases from *Arabidopsis thaliana* reveals their distinct proteolytic activity and cleavage site specificity. *Biochem J* **2013**, doi:10.1042/BJ20130196.
- Matayoshi, E. D.; Wang, G. T.; Krafft, G. A.; Erickson, J., Novel fluorogenic substrates for assaying retroviral proteases by resonance energy transfer. *Science* **1990**, 247, 954-8.

- Melander, W. R.; el Rassi, Z.; Horvath, C., Interplay of hydrophobic and electrostatic interactions in biopolymer chromatography. Effect of salts on the retention of proteins. *J Chromatogr* **1989**, 469, 3-27.
- Meng, C. K.; Mann, M.; Fenn, J. B., Of protons or proteins. *Z. Phys. D Atom Mol. Clusters* **1988**, 10, 361-368.
- Menschaert, G.; Van Criekinge, W.; Notelaers, T.; Koch, A.; Crappe, J.; Gevaert, K.; Van Damme, P., Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events. *Mol Cell Proteomics* **2013**, 12, 1780-90.
- Mirgorodskaya, E.; Braeuer, C.; Fucini, P.; Lehrach, H.; Gobom, J., Nanoflow liquid chromatography coupled to matrix-assisted laser desorption/ionization mass spectrometry: sample preparation, data analysis, and application to the analysis of complex peptide mixtures. *Proteomics* **2005**, 5, 399-408.
- Nesvizhskii, A. I.; Vitek, O.; Aebersold, R., Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods* **2007**, 4, 787-97.
- Novotny, M., Recent advances in microcolumn liquid chromatography. *Anal Chem* **1988**, 60, 500A-510A.
- O'Farrell, P. H., High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **1975**, 250, 4007-21.
- Oberacher, H.; Premstaller, A.; Huber, C. G., Characterization of some physical and chromatographic properties of monolithic poly(styrene-co-divinylbenzene) columns. *J Chromatogr A* **2004**, 1030, 201-8.
- Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T., Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A* **1999**, 96, 6591-6.
- Ogata, Y.; Charlesworth, M. C.; Higgins, L.; Keegan, B. M.; Vernino, S.; Muddiman, D. C., Differential protein expression in male and female human lumbar cerebrospinal fluid using iTRAQ reagents after abundant protein depletion. *Proteomics* **2007**, 7, 3726-34.
- Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **2002**, 1, 376-86.
- Ong, S. E.; Mann, M., Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* **2005**, 1, 252-62.
- Ostresh, J. M.; Winkle, J. H.; Hamashin, V. T.; Houghten, R. A., Peptide libraries: determination of relative reaction rates of protected amino acids in competitive couplings. *Biopolymers* **1994**, 34, 1681-9.
- Overall, C. M.; Tam, E. M.; Kappelhoff, R.; Connor, A.; Ewart, T.; Morrison, C. J.; Puente, X.; Lopez-Otin, C.; Seth, A., Protease degradomics: mass spectrometry discovery of protease substrates and the CLIP-CHIP, a dedicated DNA microarray of all human proteases and inhibitors. *Biol Chem* **2004**, 385, 493-504.

- Overall, C. M.; Kleifeld, O., Tumour microenvironment - opinion: validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy. *Nat Rev Cancer* **2006**, *6*, 227-39.
- Ow, S. Y.; Salim, M.; Noirel, J.; Evans, C.; Rehman, I.; Wright, P. C., iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". *J Proteome Res* **2009**, *8*, 5347-55.
- Paes Leme, A. F.; Escalante, T.; Pereira, J. G.; Oliveira, A. K.; Sanchez, E. F.; Gutierrez, J. M.; Serrano, S. M.; Fox, J. W., High resolution analysis of snake venom metalloproteinase (SVMP) peptide bond cleavage specificity using proteome based peptide libraries and mass spectrometry. *J Proteomics* **2011**, *74*, 401-10.
- Pappin, D. J.; Hojrup, P.; Bleasby, A. J., Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* **1993**, *3*, 327-32.
- Patterson, A. W.; Wood, W. J.; Ellman, J. A., Substrate activity screening (SAS): a general procedure for the preparation and screening of a fragment-based non-peptidic protease substrate library for inhibitor discovery. *Nat Protoc* **2007**, *2*, 424-33.
- Patthy, M., Explanation of the selectivity differences between reversed-phase ion-pair chromatographic systems containing trifluoroacetate or heptafluorobutyrate as pairing ion. *J Chromatogr A* **1994**, *660*, 17-23.
- Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551-67.
- Pichler, P.; Kocher, T.; Holzmann, J.; Mazanek, M.; Taus, T.; Ammerer, G.; Mechtler, K., Peptide labeling with isobaric tags yields higher identification rates using iTRAQ 4-plex compared to TMT 6-plex and iTRAQ 8-plex on LTQ Orbitrap. *Anal Chem* **2010**, *82*, 6549-58.
- Pichler, P.; Kocher, T.; Holzmann, J.; Mohring, T.; Ammerer, G.; Mechtler, K., Improved precision of iTRAQ and TMT quantification by an axial extraction field in an Orbitrap HCD cell. *Anal Chem* **2011**, *83*, 1469-74.
- Picotti, P.; Aebersold, R.; Domon, B., The implications of proteolytic background for shotgun proteomics. *Mol Cell Proteomics* **2007**, *6*, 1589-98.
- Pieper, S.; Beck, S.; Ahrends, R.; Scheler, C.; Linscheid, M. W., Fragmentation behavior of metal-coded affinity tag (MeCAT)-labeled peptides. *Rapid Commun Mass Spectrom* **2009**, *23*, 2045-52.
- Pitteri, S. J.; Chrisman, P. A.; Hogan, J. M.; McLuckey, S. A., Electron transfer ion/ion reactions in a three-dimensional quadrupole ion trap: reactions of doubly and triply protonated peptides with SO₂*. *Anal Chem* **2005**, *77*, 1831-9.
- Poldermans, B., Proteolytic Enzymes. In: Proteolytic Enzymes in Industry: Production and Applications. In Gerhartz, W., Ed. VCH: Weinheim, Germany, 1990.
- Poreba, M.; Drag, M., Current strategies for probing substrate specificity of proteases. *Curr Med Chem* **2010**, *17*, 3968-95.

- Premstaller, A.; Oberacher, H.; Walcher, W.; Timperio, A. M.; Zolla, L.; Chervet, J. P.; Cavusoglu, N.; van Dorsselaer, A.; Huber, C. G., High-performance liquid chromatography-electrospray ionization mass spectrometry using monolithic capillary columns for proteomic studies. *Anal Chem* **2001**, *73*, 2390-6.
- Prudova, A.; auf dem Keller, U.; Butler, G. S.; Overall, C. M., Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics. *Mol Cell Proteomics* **2010**, *9*, 894-911.
- Rano, T. A.; Timkey, T.; Peterson, E. P.; Rotonda, J.; Nicholson, D. W.; Becker, J. W.; Chapman, K. T.; Thornberry, N. A., A combinatorial approach for determining protease specificities: application to interleukin-1beta converting enzyme (ICE). *Chem Biol* **1997**, *4*, 149-55.
- Rawlings, N. D.; Barrett, A. J.; Bateman, A., MEROPS: the peptidase database. *Nucleic Acids Res* **2010**, *38*, D227-33.
- Rawlings, N. D.; Barrett, A. J.; Bateman, A., MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **2012**, *40*, D343-50.
- Reid, G. E.; McLuckey, S. A., 'Top down' protein characterization via tandem mass spectrometry. *J Mass Spectrom* **2002**, *37*, 663-75.
- Reubsaet, J. L.; Beijnen, J. H.; Bult, A.; van Maanen, R. J.; Marchal, J. A.; Underberg, W. J., Analytical techniques used to study the degradation of proteins and peptides: physical instability. *J Pharm Biomed Anal* **1998**, *17*, 979-84.
- Righetti, P. G.; Castagna, A.; Antonucci, F.; Piubelli, C.; Cecconi, D.; Campostrini, N.; Antonioli, P.; Astner, H.; Hamdan, M., Critical survey of quantitative proteomics in two-dimensional electrophoretic approaches. *J Chromatogr A* **2004**, *1051*, 3-17.
- Riviere L. R.; Tempst, P., Enzymatic Digestion of Proteins in Solution. In *Current Protocols in Protein Science*, E., C. J., Ed. Wiley: Hoboken N.J., 1995; pp 11.1.1 - 11.1.19.
- Roepstorff, P.; Fohlman, J., Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* **1984**, *11*, 601.
- Rogers, L.; Overall, C. M., Proteolytic post translational modification of proteins: proteomic tools and methodology. *Mol Cell Proteomics* **2013**, doi: 10.1074/mcp.M113.031310.
- Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **2004**, *3*, 1154-69.
- Schechter, I.; Berger, A., On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **1967**, *27*, 157-62.
- Scheuplein, F.; Schwarz, N.; Adriouch, S.; Krebs, C.; Bannas, P.; Rissiek, B.; Seman, M.; Haag, F.; Koch-Nolte, F., NAD⁺ and ATP released from injured cells induce P2X7-dependent shedding of CD62L and externalization of phosphatidylserine by murine T cells. *J Immunol* **2009**, *182*, 2898-908.

- Schilling, O.; Overall, C. M., Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat Biotechnol* **2008**, 26, 685-94.
- Schilling, O.; Barre, O.; Huesgen, P. F.; Overall, C. M., Proteome-wide analysis of protein carboxy termini: C terminomics. *Nat Methods* **2010**, 7, 508-11.
- Schilling, O.; dem Keller, U. A.; Overall, C. M., Factor Xa subsite mapping by proteome-derived peptide libraries improved using WebPICS, a resource for proteomic identification of cleavage sites. *Biol Chem* **2011**, 392, 1031-7.
- Schluter, H.; Jankowski, J.; Rykl, J.; Thiemann, J.; Belgardt, S.; Zidek, W.; Wittmann, B.; Pohl, T., Detection of protease activities with the mass-spectrometry-assisted enzyme-screening (MES) system. *Anal Bioanal Chem* **2003**, 377, 1102-7.
- Schmidt, A.; Kellermann, J.; Lottspeich, F., A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **2005**, 5, 4-15.
- Schnolzer, M.; Jedrzejewski, P.; Lehmann, W. D., Protease-catalyzed incorporation of ¹⁸O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. *Electrophoresis* **1996**, 17, 945-53.
- Scigelova, M.; Makarov, A., Orbitrap mass analyzer--overview and applications in proteomics. *Proteomics* **2006**, 6 Suppl 2, 16-21.
- Seidler, J.; Zinn, N.; Boehm, M. E.; Lehmann, W. D., De novo sequencing of peptides by MS/MS. *Proteomics* **2010**, 10, 634-49.
- Seman, M.; Adriouch, S.; Scheuplein, F.; Krebs, C.; Freese, D.; Glowacki, G.; Deterre, P.; Haag, F.; Koch-Nolte, F., NAD-induced T cell death: ADP-ribosylation of cell surface proteins by ART2 activates the cytolytic P2X7 purinoceptor. *Immunity* **2003**, 19, 571-82.
- Shilov, I. V.; Seymour, S. L.; Patel, A. A.; Loboda, A.; Tang, W. H.; Keating, S. P.; Hunter, C. L.; Nuwaysir, L. M.; Schaeffer, D. A., The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol Cell Proteomics* **2007**, 6, 1638-55.
- Shirran, S. L.; Botting, C. H., A comparison of the accuracy of iTRAQ quantification by nLC-ESI MSMS and nLC-MALDI MSMS methods. *J Proteomics* **2010**, 73, 1391-403.
- Sinclair, J.; Metodieva, G.; Dafou, D.; Gayther, S. A.; Timms, J. F., Profiling signatures of ovarian cancer tumour suppression using 2D-DIGE and 2D-LC-MS/MS with tandem mass tagging. *J Proteomics* **2011**, 74, 451-65.
- Sorensen, S. B.; Sorensen, T. L.; Breddam, K., Fragmentation of proteins by *S. aureus* strain V8 protease. Ammonium bicarbonate strongly inhibits the enzyme but does not improve the selectivity for glutamic acid. *FEBS Lett* **1991**, 294, 195-7.
- Spengler, B.; Kirsch, D.; Kaufmaniv, R.; Jaeger, E., Peptide sequencing by matrix-assisted laser-desorption mass spectrometry. *Rapid Commun Mass Spectrom* **1992**, 6, 105-108.
- Staes, A.; Van Damme, P.; Helsens, K.; Demol, H.; Vandekerckhove, J.; Gevaert, K., Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). *Proteomics* **2008**, 8, 1362-70.

- Strupat, K.; Karas, M.; Hillenkamp, F., 2,5-Dihydroxy benzoic acid - a new matrix for laser desorption/ionisation mass-spectrometry. *Int. J. Mass Spectrom. Ion Processes* **1991**, 111, 89-102.
- Stump, M. J.; Fleming, R. C.; Gong, W.-H.; Jaber, A. J.; Jones, J. J.; Surber, C. W.; Wilkins, C. L., Matrix-assisted laser desorption mass spectrometry. *Appl. Spec. Rev.* **2002**, 37, 275-303.
- Sturm, M.; Bertsch, A.; Gropl, 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.
- Svec, F.; Frechet, J. M. J., Continuous rods of macroporous polymer as high-performance liquid chromatography separation media. *Anal Chem* **1992**, 64, 820-822.
- Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* **2004**, 101, 9528-33.
- 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. *Proc Natl Acad Sci U S A* **2002**, 99, 1774-9.
- Tan, H. T.; Tan, S.; Lin, Q.; Lim, T. K.; Hew, C. L.; Chung, M. C., Quantitative and temporal proteome analysis of butyrate-treated colorectal cancer cells. *Mol Cell Proteomics* **2008**, 7, 1174-85.
- Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y., Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **1988**, 2.
- Tarentino, A. L.; Plummer, T. H., Jr., Enzymatic deglycosylation of asparagine-linked glycans: purification, properties, and specificity of oligosaccharide-cleaving enzymes from *Flavobacterium meningosepticum*. *Methods Enzymol* **1994**, 230, 44-57.
- Tholey, A., MALDI Mass Spectrometry for Quantitative Proteomics - Approaches, Scopes and Limitations **2006**, Dagstuhl-Seminar on Functional Proteomics.
- Tholey, A.; Zabet-Moghaddam, M.; Heinzle, E., Quantification of peptides for the monitoring of protease-catalyzed reactions by matrix-assisted laser desorption/ionization mass spectrometry using ionic liquid matrixes. *Anal Chem* **2006**, 78, 291-7.
- Tholey, A.; Glückmann, M.; Seemann, K.; Karas, M., *In: Proteomics – Sample Preparation*. Wiley-VCH: 2008.
- Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K.; Hamon, C., Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **2003**, 75, 1895-904.
- Tiberti, N.; Hainard, A.; Lejon, V.; Robin, X.; Ngoyi, D. M.; Turck, N.; Matovu, E.; Enyaru, J.; Ndung'u, J. M.; Scherl, A.; Dayon, L.; Sanchez, J. C., Discovery and verification of osteopontin and Beta-2-microglobulin as promising markers for staging human African trypanosomiasis. *Mol Cell Proteomics* **2010**, 9, 2783-95.

- Treumann, A.; Thiede, B., Isobaric protein and peptide quantification: perspectives and issues. *Expert Rev Proteomics* **2010**, *7*, 647-53.
- Unlu, M.; Morgan, M. E.; Minden, J. S., Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **1997**, *18*, 2071-7.
- Van Damme, P.; Staes, A.; Bronsoms, S.; Helsens, K.; Colaert, N.; Timmerman, E.; Aviles, F. X.; Vandekerckhove, J.; Gevaert, K., Complementary positional proteomics for screening substrates of endo- and exoproteases. *Nat Methods* **2010**, *7*, 512-5.
- van den Berg, B. H.; Tholey, A., Mass spectrometry-based proteomics strategies for protease cleavage site identification. *Proteomics* **2012**, *12*, 516-29.
- van Domselaar, R.; de Poot, S. A.; Bovenschen, N., Proteomic profiling of proteases: tools for granzyme degradomics. *Expert Rev Proteomics* **2010**, *7*, 347-59.
- Viklund, C.; Svec, F.; Frechet, J. M. J., Monolithic, "Molded", Porous Materials with High Flow Characteristics for Separations, Catalysis, or Solid-Phase Chemistry: Control of Porous Properties during Polymerization. *Chem. Mater.* **1996**, *8*, 744-750.
- Vissers, J. P., Recent developments in microcolumn liquid chromatography. *J Chromatogr A* **1999**, *856*, 117-43.
- Walcher, W.; Toll, H.; Ingendoh, A.; Huber, C. G., Operational variables in high-performance liquid chromatography-electrospray ionization mass spectrometry of peptides and proteins using poly(styrene-divinylbenzene) monoliths. *J Chromatogr A* **2004**, *1053*, 107-17.
- Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **2001**, *19*, 242-7.
- Whetstone, P. A.; Butlin, N. G.; Corneillie, T. M.; Meares, C. F., Element-coded affinity tags for peptides and proteins. *Bioconjug Chem* **2004**, *15*, 3-6.
- Whitehouse, C. M.; Dreyer, R. N.; Yamashita, M.; Fenn, J. B., Electrospray interface for liquid chromatographs and mass spectrometers. *Anal Chem* **1985**, *57*, 675-9.
- Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J. C.; Yan, J. X.; Gooley, A. A.; Hughes, G.; Humphery-Smith, I.; Williams, K. L.; Hochstrasser, D. F., From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* **1996**, *14*, 61-5.
- Wilson, C. H.; Indarto, D.; Doucet, A.; Pogson, L. D.; Pitman, M. R.; McNicholas, K.; Menz, R. I.; Overall, C. M.; Abbott, C. A., Identifying natural substrates for dipeptidyl peptidases 8 and 9 using terminal amine isotopic labeling of substrates (TAILS) reveals in vivo roles in cellular homeostasis and energy metabolism. *J Biol Chem* **2013**, *288*, 13936-49.
- Wolters, D. A.; Washburn, M. P.; Yates, J. R., 3rd, An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **2001**, *73*, 5683-90.
- Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Matthews, D. E.; Yates, J. R., 3rd, Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal Chem* **2004**, *76*, 4951-9.

- Yates, J. R., 3rd; Eng, J. K.; McCormack, A. L.; Schieltz, D., Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal Chem* **1995**, 67, 1426-36.
- Yin, Q. Y.; de Groot, P. W.; de Jong, L.; Klis, F. M.; De Koster, C. G., Mass spectrometric quantitation of covalently bound cell wall proteins in *Saccharomyces cerevisiae*. *FEMS Yeast Res* **2007**, 7, 887-96.
- Zenobi, R.; Knochenmuss, R., Ion formation in MALDI mass spectrometry. *Mass Spectrom Rev* **1998**, 17, 337-366.
- Zhang, Y.; Ficarro, S. B.; Li, S.; Marto, J. A., Optimized Orbitrap HCD for quantitative analysis of phosphopeptides. *J Am Soc Mass Spectrom* **2009**, 20, 1425-34.
- Zhang, Y.; Askenazi, M.; Jiang, J.; Luckey, C. J.; Griffin, J. D.; Marto, J. A., A robust error model for iTRAQ quantification reveals divergent signaling between oncogenic FLT3 mutants in acute myeloid leukemia. *Mol Cell Proteomics* **2010**, 9, 780-90.
- Zhu, W.; Smith, J. W.; Huang, C. M., Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol* **2010**, 2010, 840518.
- Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W., Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **1998**, 120, 3265-66.

Anhang

Anhang Kapitel 3

Tab. 3.I: Liste der 50 tryptischen TMT-markierten Peptide, die in Form einer *inclusion list* verwendet wurden.

Peptidnummer	Protein	Masse des [m+H] ⁺ -ions [Da]	Aminosäuresequenz
1	BSA	925.4587	ACFAVE
2	BSA	1235.6686	ACFAVEGPK
3	Myoglobin (Pferd)	1126.5645	ADIAHGQGE
4	Transferrin (Mensch)	1281.7705	AGYFAVAVVK
5	β-Lactoglobulin B (Rind)	971.5914	AQSAPLR
6	BSA	1447.7791	CCDKPLLEK
7	Cytochrom C (Pferd)	1092.6792	DLIAYLK
8	BSA	1335.7999	DYLSLILNR
9	Transferrin (Mensch)	1168.6993	GGFVYIAGK
10	Ribonuklease A (Rind)	1258.6768	GNPYVPVHF
11	Ribonuklease A (Rind)	1373.701	GNPYVPVHFD
12	Ribonuklease A (Rind)	1630.8275	GNPYVPVHFDASV
13	Myoglobin (Pferd)	1107.698	IAAKYKE
14	Myoglobin (Pferd)	940.4915	IAGHGQGE
15	BSA	1333.7201	ICTLPDTEK
16	β-Lactoglobulin B (Rind)	871.6262	ILLQK
17	α-Lactalbumin (Rind)	907.5249	IMCVK
18	β-Lactoglobulin B (Rind)	789.4836	ISLLD
19	Lysozym (Huhn)	1220.6891	ITASVNCVK
20	Lysozym (Huhn)	861.536	LAAAMK
21	Myoglobin (Pferd)	893.511	LFRND
22	Myoglobin (Pferd)	1029.5492	LFTGHPE
23	Myoglobin (Pferd)	750.4202	LGFQG
24	Cytochrom C (Pferd)	977.6497	LIAYLK
25	Myoglobin (Pferd)	926.6196	LKPLAQ
26	β-Lactoglobulin B (Rind)	941.5861	LKPTPE
27	Transferrin (Mensch)	1044.5823	LLNQAQE
28	BSA	1141.6838	LLYYANK
29	BSA	1196.6855	NFVAFVDK
30	Catalase (Rind)	1009.6357	NIAGHLK
31	BSA	1416.7552	NLPPLTADFAE
32	β-Lactoglobulin B (Rind)	1463.7651	QSLACQCLVR
33	Ribonuklease A (Rind)	1562.8417	SLADVQAVCSQK
34	BSA	817.5032	SLVNR
35	Lysozym (Huhn)	1381.703	SNFNTQATNR
36	Ribonuklease A (Rind)	2110.925	SSTSAASSSNYCQMMK
37	β-Lactoglobulin B (Rind)	939.5591	TDYKK
38	α-Lactalbumin (Rind)	1548.7606	TQAIQNNDSSTE
39	Transferrin (Mensch)	820.5508	VAFVK
40	Catalase (Rind)	1471.8215	VLTGGGNPVGDK
41	Ribonuklease A (Rind)	1176.6611	VQAVCSQK
42	Transferrin (Mensch)	1133.6477	WSVNSVGK
43	Transferrin (Mensch)	1233.6379	YANCHLAR
44	BSA	1149.7271	YAVSVLLR
45	BSA	1409.8135	YGFQNALIVR
46	Lysozym (Huhn)	1292.7493	YGILQINSR
47	α-Lactalbumin (Rind)	1353.7719	YGLFQINNK
48	Cytochrom C (Pferd)	1020.5959	YLENPK
49	Cytochrom C (Pferd)	1176.7175	YLENPKK
50	Transferrin (Mensch)	951.5757	YVTAIR

Tab. 3.II: Liste der 50 chymotryptischen TMT-markierten Peptide, die in Form einer *inclusion list* verwendet wurden.

Peptidnummer	Protein	Masse des [m+H] ⁺ -ions [Da]	Aminosäuresequenz
1	Myoglobin (Pferd)	1126.5685	ADIAGHGQE
2	BSA	921.5687	AKDAFL
3	Myoglobin (Pferd)	1276.7128	AQGAMTKALE
4	β-Lactoglobulin B (Rind)	1032.6448	CAQKKI
5	β-Lactoglobulin B (Rind)	1345.8002	CAQKKIAE
6	BSA	1653.6945	CCAKDDPHACY
7	BSA	1049.5278	CCDKPL
8	BSA	1291.6541	CCDKPLLE
9	Cytochrom C (Pferd)	823.4702	DLIAY
10	β-Lactoglobulin B (Rind)	915.4531	EQCHI
11	Transferrin (Mensch)	1491.8	GCAPGSKKDSSL
12	Carboanhydrase (Rind)	990.6177	GLAVVGVF
13	BSA	1185.7369	GPKLVVSTQ
14	Myoglobin (Pferd)	822.5254	IAAKY
15	Myoglobin (Pferd)	1107.6973	IAAKYKE
16	Myoglobin (Pferd)	940.5024	IAGHGQE
17	BSA	1177.5857	ICTLPDTE
18	β-Lactoglobulin B (Rind)	1159.7015	IQKVAGTW
19	β-Lactoglobulin B (Rind)	1322.7607	IQKVAGTWY
20	BSA	929.5458	KDAIPE
21	Myoglobin (Pferd)	1762.0706	KFDKFKHLKTE
22	Myoglobin (Pferd)	1343.8497	KFKHLKTE
23	Cytochrom C (Pferd)	1033.735	KGKKIF
24	α-Lactalbumin (Rind)	946.5551	KLDQW
25	BSA	1085.7584	KQIKKQ
26	β-Lactoglobulin B (Rind)	1041.7166	KTKIPAV
27	β-Lactoglobulin B (Rind)	1188.7863	KTKIPAVF
28	α-Lactalbumin (Rind)	950.5861	KVGINY
29	α-Lactalbumin (Rind)	1136.6678	KVGINYW
30	Myoglobin (Pferd)	893.5034	LFRND
31	Myoglobin (Pferd)	750.4286	LGFQG
32	Myoglobin (Pferd)	926.6212	LKPLAQ
33	β-Lactoglobulin B (Rind)	941.5864	LKPTPE
34	BSA	1216.6725	NLPPLTADF
35	Cytochrom C (Pferd)	1108.6185	REDLIAY
36	Carboanhydrase (Rind)	1158.6287	RQSPVDID
37	Carboanhydrase (Rind)	1415.8079	RQSPVDIDTK
38	Catalase (Rind)	1593.806	SGSADTVRDPGRF
39	Ribonuklease A (Rind)	1191.6151	SLADVQAVC
40	Ribonuklease A (Rind)	865.5352	TAAKF
41	Ribonuklease A (Rind)	1110.5969	TGSSKYPN
42	Ribonuklease A (Rind)	1504.7242	TGSSKYPNCAY
43	Carboanhydrase (Rind)	1665.0363	TKAVVQDPALKPL
44	Cytochrom C (Pferd)	885.4538	TLMEY
45	β-Lactoglobulin B (Rind)	1019.508	VDDEALE
46	BSA	1157.6543	VEKDAIPE
47	Ribonuklease A (Rind)	805.441	VQAVC
48	Myoglobin (Pferd)	1742.9711	WQQVLNVWGKVE
49	α-Lactalbumin (Rind)	1042.5161	WVCTTF
50	Carboanhydrase (Rind)	1410.7246	YDDSQDKAVL

Anhang Kapitel 3.I Perl-Skript (TMT_SIXREP_IQR_Outlier v 2.0). Das Perl Skript wurde von Bart HJ van den Berg (Institut für Experimentelle Medizin, Christian-Albrechts Universität zu Kiel) entwickelt. Es dient dazu IQR-Tests zur Entfernung mathematischer Ausreißer innerhalb der Q-PICS Datenauswertung durchzuführen.

```
#! C:\perl\bin\perl
##### TMT_SIXREP_IQR_Outlier v 2.0
# Created by: Bart H.J. van den Berg (bvdberg@live.com)
# In support of manuscript:
# Quantitative protease cleavage site profiling using Tandem-Mass-Tag labeling and LC-
MALDI-TOF/TOF MS/MS analysis
# Jakoby, T; van den Berg, BHJ; Tholey, A.
# We reject the null hypothesis that there is no outlier when minimum < Q1 - 1.5*IQR or
maximum > Q3 + 1.5*IQR.
# To change the range, edit the values of the formulas in lines 121, 153, 185, 216 and 248
# To run the script, place the .pl file in the same folder with the .csv exported Mascot search
files. These need to be formatted
# as indicated in the example .csv output file. Take note of the different character formats on
Non-English configured software.
# For this script, the Mascot header information of the .csv output needs to be removed from
the .csv file.
# NOTE THAT THIS IS ONLY COMPATIBLE WITH TMT LABELING AND
QUANTITATION DONE VIA MASCOT!! THE EXPERIMENTAL DESIGN OF SIX
REPLICATES
# IS HARD-CODED INTO THE SCRIPT.
#####
use Acme::Tools;
use Cwd;
$dir = cwd;
@cvsfiles=();
opendir DH, $dir or die "Cannot open $dir: $!";
while ($inputfile = readdir DH){
if($inputfile =~ /\.csv$/ && $inputfile !~ /^_RESULTS\.csv$/ && $inputfile !~
/^_RESULTSSUMMARY\.csv$/){
$today = datestamp();
$outcsv = $inputfile;
$outcsv =~ s/.csv//;
$outcsv1 = $outcsv.'_'. $today.'_RESULTS.csv';
$outcsv2 = $outcsv.'_'. $today.'_RESULTSSUMMARY.csv';
##### CLEAN CSV #####
open(PRECSV, "$inputfile")|| die "cannot open CSV for clean-up. $!";
open(CLEANCSV, '>csvtemp')||die "Cannot create TEMP file for CSV cleaning.$!";
while(<PRECSV>){
if($_ =~ /^(\d+.*?){
if($_ =~ /^(d+.*?)(\.*OS\.*SV\.*\d+.*?)(\.*\.*\.*\n)$/){
$first = $1;
$section = $2;
$third = $3;
$section =~ s/,//g;
```

```

$fixed = $first.$section.$third;
print CLEANCSV $fixed;}
else{print CLEANCSV $_;}
else{print CLEANCSV $_;}
close CLEANCSV;
#####
##### PROCESSING DATA #####
open(CSVIN, 'csvtemp')|| die "cannot open CSV temp as input file";
@pre127126 = ();
@pre128126 = ();
@pre129126 = ();
@pre130126 = ();
@pre131126 = ();
@127126 = ();
@128126 = ();
@129126 = ();
@130126 = ();
@131126 = ();
$count = 0;
while(<CSVIN>){ $count++;
if($_ =~ /^prot_hit_num/){next;}
else{if($_ =~ /^\\d+.*\\/){
@line = split ("", $_);
unless (!$line[38]){push (@pre127126, $line[38]);}
unless (!$line[40]){push (@pre128126, $line[40]);}
unless (!$line[42]){push (@pre129126, $line[42]);}
unless (!$line[44]){push (@pre130126, $line[44]);}
unless (!$line[46]){push (@pre131126, $line[46]);}}}
close CSVIN;
# 127126
@127126 = sort{$a <=> $b} @pre127126; do{$nullhypmax127126 = 1;
>nullhypmin127126 = 1;
@pcts127126 = percentile([25, 50, 75], @127126);
$q1_127126 = $pcts127126[0];
$q2_127126 = $pcts127126[1];
$q3_127126 = $pcts127126[2];
$irq127126 = $q3_127126 - $q1_127126;
$irq15_127126 = 1.5 * $irq127126;
$maxIQR127126 = $127126[-1];
$minIQR127126 = $127126[0];
$testmin127126 = $q1_127126 - $irq15_127126;
$testmax127126 = $q3_127126 + $irq15_127126;
if($maxIQR127126 > $testmax127126){
>nullhypmax127126 = 2;
pop(@127126);}
elsif($minIQR127126 < $testmin127126){$nullhypmin127126 = 2;shift(@127126);}
until($nullhypmax127126 == 1 && $nullhypmin127126 == 1);
# 128126
@128126 = sort{$a <=> $b} @pre128126;
do{$nullhypmax128126 = 1; $nullhypmin128126 = 1;
@pcts128126 = percentile([25, 50, 75], @128126);

```

```

$q1_128126 = $pcts128126[0];
$q2_128126 = $pcts128126[1];
$q3_128126 = $pcts128126[2];
$irq128126 = $q3_128126 - $q1_128126;
$irq15_128126 = 1.5 * $irq128126;
$maxIQR128126 = $128126[-1];
$minIQR128126 = $128126[0];
$testmin128126 = $q1_128126 - $irq15_128126;
$testmax128126 = $q3_128126 + $irq15_128126;
if($maxIQR128126 > $testmax128126){$nullhypmax128126 =
2;pop(@128126);}elseif($minIQR128126 < $testmin128126){$nullhypmin128126 =
2;shift(@128126);} } until($nullhypmax128126 == 1 && $nullhypmin128126 == 1);
# 129126
@129126 = sort{$a <=> $b} @pre129126;
do{$nullhypmax129126 = 1;$nullhypmin129126 = 1;
@pcts129126 = percentile([25, 50, 75], @129126);
$q1_129126 = $pcts129126[0];
$q2_129126 = $pcts129126[1];
$q3_129126 = $pcts129126[2];
$irq129126 = $q3_129126 - $q1_129126;
$irq15_129126 = 1.5 * $irq129126;
$maxIQR129126 = $129126[-1];
$minIQR129126 = $129126[0];
$testmin129126 = $q1_129126 - $irq15_129126;
$testmax129126 = $q3_129126 + $irq15_129126;
if($maxIQR129126 > $testmax129126){$nullhypmax129126 = 2;pop(@129126);}
elseif($minIQR129126 < $testmin129126){$nullhypmin129126 = 2; shift(@129126);} }
until($nullhypmax129126 == 1 && $nullhypmin129126 == 1);
# 130126
@130126 = sort{$a <=> $b} @pre130126;
do{$nullhypmax130126 = 1;$nullhypmin130126 = 1;
@pcts130126 = percentile([25, 50, 75], @130126);
$q1_130126 = $pcts130126[0];
$q2_130126 = $pcts130126[1];
$q3_130126 = $pcts130126[2];
$irq130126 = $q3_130126 - $q1_130126;
$irq15_130126 = 1.5 * $irq130126;
$maxIQR130126 = $130126[-1];
$minIQR130126 = $130126[0];
$testmin130126 = $q1_130126 - $irq15_130126;
$testmax130126 = $q3_130126 + $irq15_130126;
if($maxIQR130126 > $testmax130126){$nullhypmax130126 = 2;pop(@130126);}
elseif($minIQR130126 < $testmin130126){$nullhypmin130126 = 2; shift(@130126);} }
until($nullhypmax130126 == 1 && $nullhypmin130126 == 1);
# 131126
@131126 = sort{$a <=> $b} @pre131126;do{$nullhypmax131126 = 1;$nullhypmin131126
= 1;
@pcts131126 = percentile([25, 50, 75], @131126);
$q1_131126 = $pcts131126[0];
$q2_131126 = $pcts131126[1];
$q3_131126 = $pcts131126[2];

```

```

$irq131126 = $q3_131126 - $q1_131126;
$irq15_131126 = 1.5 * $irq131126;
$maxIQR131126 = $131126[-1];
$minIQR131126 = $131126[0];
$testmin131126 = $q1_131126 - $irq15_131126;
$testmax131126 = $q3_131126 + $irq15_131126;
if($maxIQR131126 > $testmax131126){$nullhypmax131126 = 2;
pop(@131126);}
elseif($minIQR131126 < $testmin131126){
$nullhypmin131126 = 2;
shift(@131126);} } until($nullhypmax131126 == 1 && $nullhypmin131126 == 1);

@final127126 = sort @127126;
@final128126 = sort @128126;
@final129126 = sort @129126;
@final130126 = sort @130126;
@final131126 = sort @131126;
$final127126min = $final127126[0];
$final128126min = $final128126[0];
$final129126min = $final129126[0];
$final130126min = $final130126[0];
$final131126min = $final131126[0];
$final127126max = $final127126[-1];
$final128126max = $final128126[-1];
$final129126max = $final129126[-1];
$final130126max = $final130126[-1];
$final131126max = $final131126[-1];
##### END IQR TEST #####
open (CSVIN, 'csvtemp')|| die "cannot open CSV TEMP as 2nd input. $!";
open(OUTCSVDATA, ">$outcsv1")|| die "Cannot open xls OUTPUT file. $!";
while(<CSVIN>){if($_ =~ /^prot_hit_num/){print OUTCSVDATA $_;}
else{if($_ =~ /^^\d+.*$/){@line = split ("", $_);
if($line[38] eq ""){$line[38] = ";}
elseif($line[38] < $final127126min ||$line[38] > $final127126max){$line[38]=";}}
if($line[40] eq ""){$line[40] = ";}
elseif($line[40] < $final128126min ||$line[40] > $final128126max){$line[40]=";}}
if($line[42] eq ""){$line[42] = ";}
elseif($line[42] < $final129126min ||$line[42] > $final129126max){$line[42]=";}}
if($line[44] eq ""){$line[44] = ";}
elseif($line[44] < $final130126min ||$line[44] > $final130126max){$line[44]=";}}
if($line[46] eq ""){$line[46] = ";}
elseif($line[46] < $final131126min ||$line[46] > $final131126max){$line[46]=";}}
foreach $linepiece (@line){print OUTCSVDATA $linepiece.';'}} # CSV file read end}}
close DH;
close CSVIN;
close OUTCSVDATA;
print "\nDone. Press enter to exit...";
if(<STDIN>){unlink 'csvtemp'; exit;}
#####
##### SUBROUTINES #####
#####

```



```

sub
datestamp{($sec,$min,$hour,$day,$month,$yr19,@rest)=localtime(time);@rest=$sec=$min=$hour;$year=$yr19+1900;$month++;
if($day<10){$day='0'.$day;}
if($month<10){$month='0'.$month;}
date=$year.$month.$day;return $date;}
#####

```

Tab. 3.III: Liste der 50 TMT-markierten Peptide der tryptischen Peptidbibliothek des 10-Proteinstandards der *inclusion list*, die mittels LC-MALDI-TOF/TOF MS/MS identifiziert wurden (X).

Nummer	Protein	Aminosäure-sequenz	250 fmol	2,5 pmol
1	BSA	ACFAVE	X	X
2	BSA	ACFAVEGPK	X	X
3	Myoglobin (Pferd)	ADIAGHGQE	X	X
4	Transferrin (Mensch)	AGYFAVAVVK	X	X
5	β -Lactoglobulin B (Rind)	AQSAPLR	X	X
6	BSA	CCDKPLLEK	X	X
7	Cytochrom C (Pferd)	DLIAYLK	X	X
8	BSA	DYLSLILNR	X	X
9	Transferrin (Mensch)	GGFVYIAGK	X	X
10	Ribonuklease A (Rind)	GNPYVPVHF		X
11	Ribonuklease A (Rind)	GNPYVPVHFD	X	X
12	Ribonuklease A (Rind)	GNPYVPVHFDASV	X	X
13	Myoglobin (Pferd)	IAAKYKE	X	X
14	Myoglobin (Pferd)	IAGHGQE	X	X
15	BSA	ICTLPDTEK	X	X
16	β -Lactoglobulin B (Rind)	ILLQK	X	X
17	α -Lactalbumin (Rind)	IMCVK	X	X
18	β -Lactoglobulin B (Rind)	ISLLD	X	X
19	Lysozym (Huhn)	ITASVNCAK	X	X
20	Lysozym (Huhn)	LAAAMK	X	X
21	Myoglobin (Pferd)	LFRND	X	X
22	Myoglobin (Pferd)	LFTGHPE		
23	Myoglobin (Pferd)	LGFQG	X	X
24	Cytochrom C (Pferd)	LIAYLK	X	X
25	Myoglobin (Pferd)	LKPLAQ		
26	β -Lactoglobulin B (Rind)	LKPTPE	X	X
27	Transferrin (Mensch)	LLNQAE	X	X
28	BSA	LLYANK	X	X
29	BSA	NFVAFVVK	X	
30	Catalase (Rind)	NIAGHLK	X	X
31	BSA	NLPPLTADFAE	X	X
32	β -Lactoglobulin B (Rind)	QSLACQCLVR	X	X
33	Ribonuklease A (Rind)	SLADVQAVCSQK	X	X
34	BSA	SLVNR	X	X
35	Lysozym (Huhn)	SNFNTQATNR	X	X
36	Ribonuklease A (Rind)	SSTSAASSNYCNQMMK	X	X
37	β -Lactoglobulin B (Rind)	TDYKK	X	X
38	α -Lactalbumin (Rind)	TQAIQNNNDSTE	X	X
39	Transferrin (Mensch)	VAFVK	X	X
40	Catalase (Rind)	VLTTGGGNPVGDK	X	X
41	Ribonuklease A (Rind)	VQAVCSQK	X	X
42	Transferrin (Mensch)	WSVNSVGK	X	X
43	Transferrin (Mensch)	YANCHLAR	X	X
44	BSA	YAVSVLLR	X	X
45	BSA	YGFQNALIVR	X	X
46	Lysozym (Huhn)	YGILQINSR	X	X
47	α -Lactalbumin (Rind)	YGLFQINNK	X	X
48	Cytochrom C (Pferd)	YLENPK	X	X
49	Cytochrom C (Pferd)	YLENPKK	X	X
50	Transferrin (Mensch)	YVTAIR	X	X

Tab. 3.IV: Liste der 50 TMT-markierten Peptide der chymotryptischen Peptidbibliothek des 10-Proteinstandards der *inclusion list*, die mittels LC-MALDI-TOF/TOF MS/MS identifiziert wurden (X).

Nummer	Protein	Aminosäuresequenz	250 fmol	2,5 pmol
1	Myoglobin (Pferd)	ADIAGHGQE	X	X
2	BSA	AKDAFL	X	X
3	Myoglobin (Pferd)	AQGAMTKALE	X	X
4	β -Lactoglobulin B (Rind)	CAQKKI	X	X
5	β -Lactoglobulin B (Rind)	CAQKKIAE	X	X
6	BSA	CCAKDDPHACY	X	X
7	BSA	CCDKPL	X	X
8	BSA	CCDKPLLE	X	X
9	Cytochrom C (Pferd)	DLIAY	X	X
10	β -Lactoglobulin B (Rind)	EQCHI	X	X
11	Transferrin (Mensch)	GCAPGSKKDSSL		X
12	Carboanhydrase (Rind)	GLAVVGVF	X	X
13	BSA	GPKLVVSTQ	X	X
14	Myoglobin (Pferd)	IAAKY	X	X
15	Myoglobin (Pferd)	IAAKYKE	X	X
16	Myoglobin (Pferd)	IAGHGQE	X	X
17	BSA	ICTLPDTE	X	X
18	β -Lactoglobulin B (Rind)	IQKVAGTW	X	X
19	β -Lactoglobulin B (Rind)	IQKVAGTWY	X	X
20	BSA	KDAIPE	X	X
21	Myoglobin (Pferd)	KFDKFKHLKTE	X	X
22	Myoglobin (Pferd)	KFKHLKTE	X	X
23	Cytochrom C (Pferd)	KGKKIF	X	X
24	α -Lactalbumin (Rind)	KLDQW	X	X
25	BSA	KQIKKQ	X	X
26	β -Lactoglobulin B (Rind)	KTKIPAV	X	X
27	β -Lactoglobulin B (Rind)	KTKIPAVF	X	X
28	α -Lactalbumin (Rind)	KVGINY	X	X
29	α -Lactalbumin (Rind)	KVGINYW	X	X
30	Myoglobin (Pferd)	LFRND	X	X
31	Myoglobin (Pferd)	LGFGG	X	X
32	Myoglobin (Pferd)	LKPLAQ	X	X
33	β -Lactoglobulin B (Rind)	LKPTPE	X	X
34	BSA	NLPPLTADF	X	X
35	Cytochrom C (Pferd)	REDLIAY	X	X
36	Carboanhydrase (Rind)	RQSPVDID	X	X
37	Carboanhydrase (Rind)	RQSPVDIDTK	X	X
38	Catalase (Rind)	SGSADTVRDPGRF	X	X
39	Ribonuklease A (Rind)	SLADVQAVC	X	X
40	Ribonuklease A (Rind)	TAAAKF	X	X
41	Ribonuklease A (Rind)	TGSSKYPN	X	X
42	Ribonuklease A (Rind)	TGSSKYPNCAY	X	X
43	Carboanhydrase (Rind)	TKAVVQDPALKPL	X	X
44	Cytochrom C (Pferd)	TLMEY	X	X
45	β -Lactoglobulin B (Rind)	VDDEALE	X	X
46	BSA	VEKDAIPE	X	X
47	Ribonuklease A (Rind)	VQAVC	X	X
48	Myoglobin (Pferd)	WQQVLNVWGKVE	X	X
49	α -Lactalbumin (Rind)	WVCTTF	X	X
50	Carboanhydrase (Rind)	YDDSQDKAVL	X	X

Anhang Kapitel 4

Anhang Kapitel 4.I Perl-Skript (BufferComparison version 1.8.1). Das Perl Skript wurde von Bart HJ van den Berg (Institut für Experimentelle Medizin, Christian-Albrechts Universität zu Kiel) entwickelt. Das Skript dient dazu fehlende Signale (Flächen) der Reporterionen aus den mgf-Dateien der TS2Mascot Software unter Verwendung der MIO *enabled* Option durch die Flächen in den mgf-Dateien unter Verwendung der MIO *disabled* Option zu ergänzen.

```
#! C:\perl\bin\perl
#### BufferComparison version 1.8.1
# Created by: Bart H.J. van den Berg (bvdberg@live.com)
# In support of manuscript:
# Quantitative protease cleavage site profiling using Tandem-Mass-Tag labeling and LC-
MALDI-TOF/TOF MS/MS analysis
# Jakoby, T; van den Berg, BHJ; Tholey, A.
# We reject the null hypothesis that there is no outlier when minimum < Q1 - 1.5*IQR or
maximum > Q3 + 1.5*IQR.
# To change the range, edit the values of the formulas in line 580
# To run the script, place the .pl file in the same folder with the .csv exported Mascot search
files. These need to be formatted
# exactly as indicated in the example .csv output file. Important here is that the columns need
to match.
# Take note of the different character formats on Non-English configured software.
# NOTE THAT THIS IS ONLY COMPATIBLE WITH TMT-LABELING AND
QUANTITATION DONE VIA MASCOT!! THE LABELING STRATEGY IS NOW
HARD-CODED TO
# THREE TMT CHANNELS PER BUFFER, COMPARING TWO BUFFERS.
#####
#####
use Acme::Tools;
use Cwd;
$dir = cwd;
@cvsfiles=();
opendir DH, $dir or die "Cannot open $dir: $!";
```

```
while ($inputfile = readdir DH){if($inputfile=~ /\.csv$/ && $inputfile !~ /\_RESULTS\.csv$/
&& $inputfile !~ /\_RESULTSSUMMARY\.csv$/){$today = datestamp();
$outcsv = $inputfile;
$outcsv =~ s/.csv//;
$outcsv1 = $outcsv.'_'.$today.'_RESULTS.csv';
$outcsv2 = $outcsv.'_'.$today.'_RESULTSSUMMARY.csv';
##### CLEAN CSV #####
open(PRECSV, "$inputfile")|| die "cannot open CSV input for clean-up. $!";
open(CLEANCSV, '>csvtemp')||die "Cannot create TEMP file for CSV cleaning.$!";
while(<PRECSV){
if($_ =~ /^d+.*){if($_ =~ /^(d+\.*,*\.)("\.*OS\=.*SV\=d+.*")(\.*\\".*\n$)){ $first = $1;
$section = $2;$third = $3;$section =~ s//g;$fixed = $first.$section.$third;print CLEANCSV
$fixed;}
else{print CLEANCSV $_;}}
else{print CLEANCSV $_;}}
close CLEANCSV;
#####
#### IQR OUTLIER TEST for Buffer A and Buffer B #####
#### Buffer A
@pre127126 = ();
@pre128126 = ();
@pre128127 = ();
@final127126 = ();
@final128126 = ();
@final128127 = ();
#### Buffer B
@pre130129 = ();
@pre131129 = ();
@pre131130 = ();
@final130129 = ();
@final131129 = ();
@final131130 = ();
open(CSVIN1, 'csvtemp')|| die "Cannot open CSV temp as input file";
$flag = 0;while(<CSVIN1>){if($flag == 0){if($_ !~ /Protein hits/){next;}}
```

```
else{$flag = 1;}
if($_ =~ /Peptide matches/){last;}
if($_ =~ /^prot_hit_num/){ next;}
else{if($_ =~ /\^d+.*\/){@line = split ("", $_);unless (!$line[38]||$line[38]<= 0 ||$line[38] eq
'###'||$line[38] eq '---'){push (@pre127126, $line[38]);}
unless (!$line[40]||$line[40]<= 0 ||$line[40] eq '###'||$line[40] eq '---'){push (@pre128126,
$line[40]);}
unless (!$line[44]||$line[44]<= 0 ||$line[44] eq '###'||$line[44] eq '---'){push (@pre128127,
$line[44]);}
unless (!$line[68]||$line[68]<= 0 ||$line[68] eq '###'||$line[68] eq '---'){push (@pre130129,
$line[68]);}
unless (!$line[70]||$line[70]<= 0 ||$line[70] eq '###'||$line[70] eq '---'){push (@pre131129,
$line[70]);}
unless (!$line[74]||$line[74]<= 0 ||$line[74] eq '###'||$line[74] eq '---'){push (@pre131130,
$line[74]);}}}}
close CSVIN1;
@final127126 = sort(irqarray(@pre127126));
@final128126 = sort(irqarray(@pre128126));
@final128127 = sort(irqarray(@pre128127));
@final130129 = sort(irqarray(@pre130129));
@final131129 = sort(irqarray(@pre131129));
@final131130 = sort(irqarray(@pre131130));
# Buffer A outlier min and max
$final127126min = $final127126[0];
$final128126min = $final128126[0];
$final128127min = $final128127[0];
$final127126max = $final127126[-1];
$final128126max = $final128126[-1];
$final128127max = $final128127[-1];
print "$final127126min is min and $final127126max is max\n";
print "$final128126min is min and $final128126max is max\n";
print "$final128127min is min and $final128127max is max\n";
# Buffer B outlier min and max
$final130129min = $final130129[0];
```

```

$final131129min = $final131129[0];
$final131130min = $final131130[0];
$final130129max = $final130129[-1];
$final131129max = $final131129[-1];
$final131130max = $final131130[-1];
print "\n";
print "$final130129min is min and $final130129max is max\n";
print "$final131129min is min and $final131129max is max\n";
print "$final131130min is min and $final131130max is max\n";
print "\n";
open(CSVIN2, 'csvtemp')|| die "cannot open CSV temp as input file";
open(CSVOUT, '>csvtemp2')|| die "cannot open CSV2 temp as output file";
$there = 1;
while(<CSVIN2>){if($_ =~ /^prot_hit_num/){$there=2;}
else{if($_ =~ /\^d+.*\/ && $there == 2){@online = split(",", $_);for($i = 38; $i < 79;
$i+=2){if($oline[$i]<= 0 ||$oline[$i] eq '###' ||$oline[$i] eq '---'){ $oline[$i] = ";}
##### 127/126 Outlier
if($oline[38] > $final127126max || $oline[38] < $final127126min){ $oline[38]= ";}
##### 128/126 Outlier
if($oline[40] > $final128126max || $oline[40] < $final128126min){ $oline[40]= ";}
##### 128/127 Outlier
if($oline[44] > $final128127max || $oline[44] < $final128127min){ $oline[44]= ";}
##### 130/129 Outlier
if($oline[68] > $final130129max || $oline[68] < $final130129min){ $oline[68]= ";}
##### 131/129 Outlier
if($oline[70] > $final131129max || $oline[70] < $final131129min){ $oline[70]= ";}
##### 131/130 Outlier
if($oline[74] > $final131130max || $oline[74] < $final131130min){ $oline[74]= ";}
$comboline = join(",", @online);print CSVOUT $comboline;@online = ();} }
close CSVIN2;
close CSVOUT;
##### PROCESSING DATA #####
open(CSVIN3, 'csvtemp2')|| die "cannot open CSVtemp2 as input file";
open(OUTCSVDATA, ">$outcsv1")|| die "Cannot open csv OUTPUT file. $!";

```

```

open(TESTOUT, ">$outcsv2");
print TESTOUT "non-prime,Peptide,prime,modification,No.Buff A Ratios,Sum Buff A Ratios, AVERAGE Buff A Ratios, STDEV Buff A Ratios, RSD Buff A Ratios,";
print TESTOUT "No.Buff AB Ratios,Sum Buff AB Ratios, AVERAGE Buff AB Ratios, STDEV Buff AB Ratios, RSD Buff AB Ratios,";
print TESTOUT "No.Buff B Ratios,Sum Buff B Ratios, AVERAGE Buff B Ratios, STDEV Buff B Ratios, RSD Buff B Ratios,Ionscore Average,";
print TESTOUT "ionscore STDEV, Ionscore RSD\n";
%allinfo=();
while(<CSVIN3>){if($flag == 0){if($_ !~ /Protein hits/){print OUTCSVDATA $_;next;}
else{$flag = 1;}}
if($_ =~ /Peptide matches/){last;}
if($_ =~ /^prot_hit_num/){chomp $_;print OUTCSVDATA $_;}
else{if($_ =~ /\^(\d+.*\/){$negativeratioflag = 0;$info = $_;@endline = split(",", $_);$pep = $endline[28];$modinfo = $endline[31];$labelinfo = $endline[34];$labelinfo =~ /Label:\s(\d+)\s/i;$labelid = $1;$peaklistidinfo = $endline[35];$peaklistidinfo =~ /Peak\_List\_Id:\s(\d+)\s/i;$peaklistid = $1;$csvruninfo = $endline[36];$csvruninfo =~ /Run\_Id:\s(\d+)\s/i;$csvrunid = $1;$hashtag1 = $pep.'.'. $modinfo;$hashtag2 = $csvrunid.$labelid.$peaklistid;}
$comboline2 = join ('', @endline);unless($allinfo{$hashtag1}{$hashtag2}){$allinfo{$hashtag1}{$hashtag2}=$comboline2;}} # CSV file read endforeach $key1 (sort (keys %allinfo)){if(scalar keys %{$allinfo{$key1}} >= 2){$hashsize = scalar keys %{$allinfo{$key1}};$bufferatmtvaluecount = 0;$bufferbtmtvaluecount = 0;$bufferabmtvaluecount = 0;%buffera = ();%bufferab = ();%bufferb = ();%ionscore = ();foreach $key2 (sort (keys %{$allinfo{$key1}})){@finalline = split(",", $allinfo{$key1}{$key2});@finalline2 = split("_", $key1);$ionscore{$key2} = $finalline[23];$peptideinfo = $finalline[27].'.$finalline[28].'.$finalline[29].'.$finalline[31];
# Check Buffer A
unless($finalline[38] == ""){$bufferatmtvaluecount++;
$buffera{$bufferatmtvaluecount}=$finalline[38];}unless($finalline[40] == ""){$bufferatmtvaluecount++;$buffera{$bufferatmtvaluecount}=$finalline[40];}
unless($finalline[44] == ""){$bufferatmtvaluecount++;$buffera{$bufferatmtvaluecount}=$finalline[44];}

```

```

# Check Buffer AB
for($i=50;          $i<67;          $i=$i+2){unless($finalline[$i]          ==
"){$bufferabtotalvaluecount++;$bufferab{$bufferabtotalvaluecount}=$finalline}}
# $preab= keys %bufferab;m foreach $abkeyol (%bufferab)$postab= keys %bufferab;print
"\ntested and found $preab is pre and $postab is post\n";
#   Check   Buffer   Bunless($finalline[68]   ==   "){$bufferbtotalvaluecount++;n
$bufferb{$bufferbtotalvaluecount}=$finalline[68];}
unless($finalline[70]=="){ $bufferbtotalvaluecount++;$bufferb{$bufferbtotalvaluecount}=$final
line[70];}
unless($finalline[74]=="){ $bufferbtotalvaluecount++;$bufferb{$bufferbtotalvaluecount}=$final
line[74];}$peptideinfo          =
$finalline[27].!.$finalline[28].!.$finalline[29].!.$finalline[31];splice(@finalline,
79);$joinedarray = join ("", @finalline); $allinfo{$key1}{$key2} = $joinedarray;print
OUTCSVDATA "\n".$allinfo{$key1}{$key2};}
##### Geo mean calculation and geo StDev #####
##### Calculating mean sum of area perbuffer and %RDS #####
#####(standard deviation of array X / average of array X ) x 100 = %RDS          #####

# @splicensumbers = (78,77,76,75,72,71,48,47,46,45,42,41);
# while(@splicensumbers){
# $spliceindex = shift(@splicensumbers);
# splice(@finalline, $spliceindex,1);}
@aboutlier = ();
@finalab = ();
$preabtotal = keys %bufferab;
foreach $keyolab(keys %bufferab){push (@aboutlier, $bufferab{$keyolab});}
unless(scalar @aboutlier == 0){ @finalab = sort(irrarray(@aboutlier));
$finalabmin = $finalab[0];$finalabmax = $finalab[-1];}
foreach $abkeyol(keys %bufferab){if($bufferab{$abkeyol} > $finalabmax ||
$bufferab{$abkeyol} < $finalabmin){delete $bufferab{$abkeyol};}}
$bufferatotal= keys %buffera;$bufferabtotal= keys %bufferab;$bufferbtotal= keys
%bufferb;$bufferasumlnarea=0;$bufferabsumlnarea=0;$bufferbsumlnarea=0;foreach
$keya(sort(keys %buffera)){ $bufferalnsunarea += log($buffera{$keya});}
foreach $keyab(sort(keys %bufferab)){ $bufferablsumunarea += log($bufferab{$keyab});}

```

```

foreach $keyb(sort(keys %bufferb)){ $bufferblnsumarea += log($bufferb{$keyb});}
if($bufferatotal == 0){ $buffera{1} = 1; $bufferatotal= keys %buffera; $bufferasumlnarea=0;}
if($bufferabtotal == 0){ $bufferab{1} = 1; $bufferabtotal= keys
%bufferab; $bufferabsumlnarea=0;}
if($bufferbttotal == 0){ $bufferb{1} = 1; $bufferbttotal= keys %bufferb; $bufferbsumlnarea=0;}
$bufferaGEOlogR= $bufferalnsumarea / $bufferatotal;
$bufferabGEOlogR= $bufferablsumarea / $bufferabtotal;
$bufferbGEOlogR= $bufferblnsumarea / $bufferbttotal;
$bufferaGEOaverage= exp($bufferalnsumarea / $bufferatotal);
$bufferabGEOaverage=exp($bufferablsumarea / $bufferabtotal);
$bufferbGEOaverage=exp($bufferblnsumarea / $bufferbttotal);
$bufferaSDsum=0;
$bufferabSDsum=0;
$bufferbSDsum=0;
foreach $keya(sort(keys %buffera)){
$bufferaSDsum += ((log($buffera{$keya}) - $bufferaGEOlogR)**2);}
foreach $keyab(sort(keys %bufferab)){ $bufferabSDsum += ((log($bufferab{$keyab}) -
$bufferabGEOlogR)**2);}
foreach $keyb(sort(keys %bufferb)){ $bufferbSDsum += ((log($bufferb{$keyb}) -
$bufferbGEOlogR)**2);}
if($bufferatotal == 1){ $bufferaSDaverage = 0;}
else{ $bufferaSDaverage=$bufferaSDsum / ($bufferatotal-1);}
if($bufferabtotal == 1){ $bufferabSDaverage = 0;}
else{ $bufferabSDaverage=$bufferabSDsum / ($bufferabtotal - 1);}
if($bufferbttotal == 1){ $bufferbSDaverage = 0;}
else{ $bufferbSDaverage=$bufferbSDsum / ($bufferbttotal - 1);}
$bufferasd = ($bufferaSDaverage)**0.5;
$bufferabsd = ($bufferabSDaverage)**0.5;
$bufferbsd = ($bufferbSDaverage)**0.5;
print "$peptideinfo\t$bufferabGEOlogR\t$bufferabsd\n";
$bufferaGEOSD = exp($bufferaGEOlogR) * (exp($bufferasd) - 1);
$bufferabGEOSD = exp($bufferabGEOlogR) * (exp($bufferabsd) - 1);
$bufferbGEOSD = exp($bufferbGEOlogR) * (exp($bufferbsd) - 1);
$bufferarsdarea=($bufferaGEOSD / $bufferaGEOaverage)*100;

```

```

$bufferabrsdarea=($bufferabGEOSD / $bufferabGEOaverage)*100;
$bufferbrsdarea=($bufferbGEOSD / $bufferbGEOaverage)*100;
$ionscoresum = 0;
$ionscoretotal = %ionscore;foreach $keyis (keys %ionscore){$ionscoresum = $ionscoresum +
$ionscore{$keyis};}
$ionscoreaverage = $ionscoresum / $ionscoretotal;$ionscoresumcovar=0;
foreach $keyis (keys %ionscore){$ionscoresumcovar += ($ionscoreaverage-
$ionscore{$keyis})*2;}
$ionscorecovar= $ionscoresumcovar / $ionscoretotal;$ionscorestdev =
sqrt($ionscorecovar);$ionscorersdarea=($ionscorestdev / $ionscoreaverage)*100;
print OUTCSVDATA
",$bufferatotal,$bufferalnsumarea,$bufferaGEOaverage,$bufferaGEOSD,$bufferarsdarea";
print OUTCSVDATA
",$bufferabtotal,$bufferablsumarea,$bufferabGEOaverage,$bufferabGEOSD,$bufferabrsdar
ea";
printOUTCSVDATA
",$bufferbttotal,$bufferblnsumarea,$bufferbGEOaverage,$bufferbGEOSD,$bufferbrsdarea,$io
nscoreaverage, $ionscorestdev, $ionscorersdarea";
print TESTOUT "$peptideinfo";
print TESTOUT
",$bufferatotal,$bufferalnsumarea,$bufferaGEOaverage,$bufferaGEOSD,$bufferarsdarea";
print TESTOUT
",$bufferabtotal,$bufferablsumarea,$bufferabGEOaverage,$bufferabGEOSD,$bufferabrsdar
ea";
print TESTOUT
",$bufferbttotal,$bufferblnsumarea,$bufferbGEOaverage,$bufferbGEOSD,$bufferbrsdarea,$io
nscoreaverage, $ionscorestdev, $ionscorersdarea\n";
$peptideinfo=();
$bufferatotal=();
$bufferabtotal=();
$bufferbttotal=();
$bufferalnsumarea=();
$bufferablsumarea=();
$bufferblnsumarea=();

```

```
$bufferaGEOaverage=();
$bufferabGEOaverage=();
$bufferbGEOaverage=();
$bufferaGEOlogR=();
$bufferabGEOlogR=();
$bufferbGEOlogR=();
$bufferaSDsum=();
$bufferabSDsum=();
$bufferbSDsum=();
$bufferaSDaverage=();
$bufferabSDaverage=();
$bufferbSDaverage=();
$bufferasd=();
$bufferabsd=();
$bufferbsd=();
$bufferaGEOSD=();
$bufferabGEOSD=();
$bufferbGEOSD=();
$bufferarsdarea=();
$bufferabrsdarea=();
$bufferbrsdarea=();
$ionscoresum=();
$ionscoretotal=();
$ionscoreaverage=();
$ionscoresumcovar=();
$ionscorecovar=();
$ionscorestdev=();
$ionscorersdarea=();} # end of read keys %allinfo 2nd hash part} # end of read keys %allinfo
1st hash part} # if loop dir read for csv files} # while read for files in directory
##### CLOSE AND EXIT #####
print "\nScript completed. Press ENTER to exit....";
close DH;
close CSVIN3;
close OUTCSVDATA;
```

```

close TESTOUT;
if(<STDIN>){unlink csvtemp; unlink csvtemp2; exit;}
#####
##### SUBROUTINES #####
#####
sub datestamp{
($sec,$min,$hour,$day,$month,$yr19,@rest)=localtime(time);
@rest=$sec=$min=$hour;$year=$yr19+1900;$month++;if($day<10){$day='0'.$day;}
if($month<10){$month='0'.$month;}$date=$year.$month.$day;return $date;}
sub irqarray{
@irqarray = @_;
@sortedirq=();
@sortedirq = sort{$a <=> $b} @irqarray;
do{$nullhypmax= 1;$nullhypmin= 1;@pcts = percentile([25, 50, 75], @sortedirq);$q1 =
$pcts[0];$q2 = $pcts[1];$q3 = $pcts[2];$irq = $q3 - $q1;$irq15 = 1.5 * $irq;$maxIQR =
$sortedirq[-1];$minIQR = $sortedirq[0];$testmin = $q1 - $irq15;$testmax = $q3 + $irq15;
if($maxIQR > $testmax){$nullhypmax = 2;pop(@sortedirq);}
elseif($minIQR < $testmin){$nullhypmin = 2;shift(@sortedirq);}} until($nullhypmax == 1
&& $nullhypmin == 1);
@irqarray = ();@pcts = ();$q1 = ();$q2 = ();$q3 = ();$irq = ();$irq15 = ();$maxIQR =
();$minIQR = ();$testmin = ();$testmax = ();return @sortedirq;}
#####
BufferComparison version 1.8.1

```

Anhang Kapitel 6

Anhang Kapitel 6.I Perl-Skript (RICmerge). Das Perl Skript wurde von Bart HJ van den Berg (Institut für Experimentelle Medizin, Christian-Albrechts Universität zu Kiel) entwickelt. Es ergänzt fehlende Signale (Flächen) der Reporterionen aus den mgf-Dateien der TS2Mascot Software unter Verwendung der MIO *enabled* Option durch die Flächen in den mgf-Dateien unter Verwendung der MIO *disabled* Option. Diese mgf-Dateien wurden dann als sogenannte *merged* Dateien bezeichnet.

```
#! C:\perl\bin\perl
#### RICmerge.pl version 2.1
# Created by: Bart H.J. van den Berg (bvdberg@live.com)
# Generated with Perl version 5.10 and compatible for 5.10.XX and higher
# In support of manuscript:
# Improved reporter ion assignment of raw isobaric stable isotope labeled LC-MALDI-
    TOF/TOF MS/MS spectral data for quantitative proteomics
# Jakoby, T; Tholey, A; van den Berg, BHJ.
# This scripts is an extension on the TS2Mascot script from Matrix Science that covert
    ABSCEIX TOF/TOF data to .mgf peak lists.
# In order to use this script, two outputs need to be created with TS2MASCOT. One .mgf
    with the 'Monoisotopic Peaks' option enabled and
# one .mgf file with the function disabled (all peaks).
# To fill the missing reporter ion peaks into the final output, save both files with extentions as
    followed:
# 1) Monoisotopic peaks enabled: 'XXX_MIPonly.mgf' where XXX is any desired
    name.
# 2) Monoisotopic peaks disabled: 'XXX_allpeaks.mgf' where XXX is any desired
    name.
# The resulting .mgf file can then be used for the Mascot search.
# To change the mass accuracys for the reporter ions, change the following lines:
# For iTRAQ 4Plex:
# 121,126,131,136 241,244,247,250 463,466,469,472
# For iTRAQ 8Plex:
#147,152,157,162,167,172,177,182
```

290,293,296,299,302,305,308,311

528,531,534,537,540,543,546,549

For TMT 6Plex:

#193,198,203,208,213,218

362,365,368,371,374,377

629,632,635,638,641,644

ÚSED ACCURATE MASSES:

Masses iTRAQ 4plex:

114,11123

115,10826

116,11162

117,11497

Masses iTRAQ 8plex:

113,107873

114,111228

115,108263

116,111618

117,114973

118,112008

119,115363

121,122072

Masses TMT 6plex:

126.127725

127.131079

128.134433

129.137787

130.141141

131.138176

#####

#####

Use Cwd;

print "What labeling method did you use?\n1. iTRAQ 4-plex\n2. iTRAQ 8-plex\n3. TMT 6-plex\n\nPlease enter number: ";chomp(\$labeltype = <STDIN>);

```

do{unless($labeltype =~ /^[123]{1}$/ and $labeltype =~ /\d$/{print "\n$labeltype is not a
valid choice. This does not work.\n";print "Please enter number (Type 'exit' to exit program):
";chomp($labeltype=<STDIN>);exit if($labeltype eq "exit");}}until($labeltype =~
/^[123]{1}$/ and $labeltype =~ /\d$/{print "\nDo you want to apply a reporter ion signal
intensity treshold?\nPlease enter 'y' or 'n': ";chomp($answer = <STDIN>);
do{unless($answer =~ /^[yn]{1}$/i and $answer =~ /\w$/{print "\n$answer is not a valid
choice. This does not work.\n";print "Please enter 'y' or 'n' (Type 'exit' to exit program): ";
chomp($answer=<STDIN>);exit if($answer eq "exit");}}until($answer =~ /^[yn]{1}$/i and
$answer =~ /\w$/{if( $answer eq 'y' or $answer eq 'Y'){print "\nPlease enter a whole number
as treshold: ";chomp($treshold = <STDIN>);
do{unless($treshold >= 0 and $treshold =~ /\d*$/){print "\n$treshold is not a valid choice.
This does not work.\n";print "Please enter a whole number (Type 'exit' to exit program):
";chomp($treshold=<STDIN>);exit if($treshold eq "exit");}}until($treshold >= 0 and
$treshold =~ /\d*$/);}elseif( $answer eq 'n' or $answer eq 'N'){ $treshold = 0;}

```

```
##### GENERATE TMT ION HASH FROM ALL PEAK TEMP FILE #####
```

```
##### AND FIND THE M.I.P. MGF FILES #####
```

```
$dir = cwd;
```

```
@mipfiles=();
```

```
opendir DH, $dir or die "Cannot open $dir: $!";
```

```
while ($inputfile = readdir DH){
```

```
if($inputfile =~ /_allpeaks\.mgf$/{open(MGFIN, "$inputfile")|| die "cannot open MGF input
file";%masses=();$precursormass=();$runid=();%mgfdata=
```

```
();$mgfidentifier=();while(<MGFIN>){if($_ =~ /PEPMASS\=(.*)\n){ $precursormass =
$1;}elseif($_ =~ /MSMS\sJob\_Run\_Id:\s(\d*)\n/gi){ $runid=$1;}elseif($_ =~
/^\d*\.\d*)t(\d*\.\d*)\n){ $mzvalue=$1;$area=$2;
```

```
# iTRAQ 4-plex
```

```
if($labeltype == 1){if($mzvalue >= 114 && $mzvalue < 118){if($mzvalue >= 114.09 &&
$mzvalue < 114.12){if($area > $treshold){ $masses{$mzvalue}=$area;}}
```

```
elseif($mzvalue >= 115.09 && $mzvalue < 115.12){if($area >
$treshold){ $masses{$mzvalue}=$area;}}
```

```
elseif($mzvalue >= 116.09 && $mzvalue < 116.12){if($area >
$treshold){ $masses{$mzvalue}=$area;}}
```

```

elseif($mzvalue >= 117.09 && $mzvalue < 117.12){if($area >
$threshhold){$masses{$mzvalue}=$area;}}}
# iTRAQ 8-plex
elseif($labeltype == 2){if($mzvalue >= 113 && $mzvalue < 122){if($mzvalue >= 113.09 &&
$mzvalue < 113.12){if($area > $threshhold){$masses{$mzvalue}=$area;}}
elseif($mzvalue >= 114.09 && $mzvalue < 114.12){if($area >
$threshhold){$masses{$mzvalue}=$area;}}
elseif($mzvalue >= 115.09 && $mzvalue < 115.12){if($area >
$threshhold){$masses{$mzvalue}=$area;}}
elseif($mzvalue >= 116.09 && $mzvalue < 116.12){if($area >
$threshhold){$masses{$mzvalue}=$area;}}
elseif($mzvalue >= 117.09 && $mzvalue < 117.12){if($area >
$threshhold){$masses{$mzvalue}=$area;}}
elseif($mzvalue >= 118.09 && $mzvalue < 118.12){if($area >
$threshhold){$masses{$mzvalue}=$area;}}
elseif($mzvalue >= 119.09&& $mzvalue < 119.12){if($area >
$threshhold){$masses{$mzvalue}=$area;}}
elseif($mzvalue >= 121.10 && $mzvalue < 121.14){if($area >
$threshhold){$masses{$mzvalue}=$area;}}}}
# TMT 6-plex
elseif($labeltype == 3){if($mzvalue >= 126 && $mzvalue < 132){if($mzvalue >= 126.05 &&
$mzvalue < 126.25){if($area > $threshhold){$masses{$mzvalue}=$area;}}
elseif($mzvalue >= 127.05 && $mzvalue < 127.25){if($area >
$threshhold){$masses{$mzvalue}=$area;}}
elseif($mzvalue >= 128.05 && $mzvalue < 128.25){if($area >
$threshhold){$masses{$mzvalue}=$area;}}
elseif($mzvalue >= 129.05 && $mzvalue < 129.25){if($area >
$threshhold){$masses{$mzvalue}=$area;}}
elseif($mzvalue >= 130.05 && $mzvalue < 130.25){if($area >
$threshhold){$masses{$mzvalue}=$area;}}
elseif($mzvalue >= 131.05 && $mzvalue < 131.25){if($area >
$threshhold){$masses{$mzvalue}=$area;}}}}}}
elseif($_ =~ /\n$/){next;}
elseif($_ =~ /\^END IONS/gi){

```

```
# iTRAQ 4-plex
if($labeltype == 1){$flag114 = 1; $flag115 = 1; $flag116 = 1; $flag117 = 1; foreach
$checkkey (sort (keys %masses))
{if($checkkey >= 114.09 && $checkkey < 114.12){$flag114 = 2;}
if($checkkey >= 115.09 && $checkkey < 115.12){$flag115 = 2;}
if($checkkey >= 116.09 && $checkkey < 116.12){$flag116 = 2;}
if($checkkey >= 117.09 && $checkkey < 117.12){$flag117 = 2;}}
if($flag114 == 1){$masses{114} = 0;}
if($flag115 == 1){$masses{115} = 0;}
if($flag116 == 1){$masses{116} = 0;}
if($flag117 == 1){$masses{117} = 0;}
$mgfidentifier = "$precursormass\t$runid";foreach $sprintkey (sort(keys
%masses)){ $mgfdata{$mgfidentifier}{$sprintkey}=$masses{$sprintkey}; %masses=();$precur
sormass=();$runid=();$mgfidentifier=();}

# iTRAQ 8-plex
elsif($labeltype == 2){$flag113 = 1; $flag114 = 1; $flag115 = 1; $flag116 = 1; $flag117 = 1;
$flag118 = 1; $flag119 = 1; $flag121 = 1; foreach $checkkey (sort (keys %masses))
{if($checkkey >= 113.09 && $checkkey < 113.12){$flag113 = 2;}
if($checkkey >= 114.09 && $checkkey < 114.12){$flag114 = 2;}
if($checkkey >= 115.09 && $checkkey < 115.12){$flag115 = 2;}
if($checkkey >= 116.09 && $checkkey < 116.12){$flag116 = 2;}
if($checkkey >= 117.09 && $checkkey < 117.12){$flag117 = 2;}
if($checkkey >= 118.09 && $checkkey < 118.12){$flag118 = 2;}
if($checkkey >= 119.09 && $checkkey < 119.12){$flag119 = 2;}
if($checkkey >= 121.10 && $checkkey < 121.14){$flag121 = 2;}}
if($flag113 == 1){$masses{113} = 0;}
if($flag114 == 1){$masses{114} = 0;}
if($flag115 == 1){$masses{115} = 0;}
if($flag116 == 1){$masses{116} = 0;}
if($flag117 == 1){$masses{117} = 0;}
if($flag118 == 1){$masses{118} = 0;}
if($flag119 == 1){$masses{119} = 0;}
if($flag121 == 1){$masses{121} = 0;}
```

```

$mgfidentifier = "$precursormass\t$runid"; foreach $sprintkey (sort(keys
%masses)){ $mgfdata{$mgfidentifier}{$sprintkey}=$masses{$sprintkey};}%masses=();$prec
ursormass=();$runid=();$mgfidentifier=();}
# TMT 6-plex
elsif($labeltype == 3){ $flag126 = 1; $flag127 = 1; $flag128 = 1; $flag129 = 1; $flag130 = 1;
$flag131 = 1; foreach $checkkey (sort (keys %masses))
{if($checkkey >= 126.05 && $checkkey < 126.26){ $flag126 = 2;}
if($checkkey >= 127.05 && $checkkey < 127.26){ $flag127 = 2;}
if($checkkey >= 128.05 && $checkkey < 128.26){ $flag128 = 2;}
if($checkkey >= 129.05 && $checkkey < 129.26){ $flag129 = 2;}
if($checkkey >= 130.05 && $checkkey < 130.26){ $flag130 = 2;}
if($checkkey >= 131.05 && $checkkey < 131.26){ $flag131 = 2;}}
if($flag126 == 1){ $masses{126} = 0;}
if($flag127 == 1){ $masses{127} = 0;}
if($flag128 == 1){ $masses{128} = 0;}
if($flag129 == 1){ $masses{129} = 0;}
if($flag130 == 1){ $masses{130} = 0;}
if($flag131 == 1){ $masses{131} = 0;}
$mgfidentifier = "$precursormass\t$runid";foreach $sprintkey (sort(keys
%masses)){ $mgfdata{$mgfidentifier}{$sprintkey}=$masses{$sprintkey};}%masses=();$precu
rsormass=();$runid=();$mgfidentifier=();}}close MGFIN;}
elsif($inputfile =~ /\_MIPonly\,mgf$/ && $flag == 0){push(@mipfiles, $inputfile);} }
close DH;
print "\nTEMP HASH CREATED. CHECKING M.I.P. FILES NOW...\n";

##### GENERATE TMT ION HASH FROM TEMP FILE #####

foreach $mipfile (@mipfiles){
$mipoutput = $mipfile;
$mipoutput =~ s/.mgf/_RESULTS.mgf/i;
% mipmasses=();
open(MIPINPUT, $mipfile)|| die "Cannot open MIP file $mipfile because $!.";
open(MIPOUT, ">$mipoutput")|| die "Cannot open OUTPUT for MIP because $!.";

```

```

while(<MIPINPUT>){if($_ =~ /^PEPMASS\(.*\)n/){$mipprecursormass = $1;print
MIPOUT $_;}
elseif($_ =~ /MSMS\sJob\_Run\_Id\s(d*)n/gi){$miprunid=$1;print MIPOUT $_;}
elseif($_ =~ /\n$/){print MIPOUT $_;}
elseif($_ =~
/^(d*\.\d*)t(d*\.\d*)n/){$mipmzvalue=$1;$miparea=$2;$mipmasses{$mipmzvalue}=$mip
area;}
elseif($_ =~ /^END IONS/gi){
# iTRAQ 4-plex
if($labeltype == 1){$mipflag114 = 1; $mipflag115 = 1; $mipflag116 = 1; $mipflag117 = 1;
$mipidentifier = "$mipprecursormass\t$miprunid";foreach $mipcheckkey (sort (keys
%mipmasses))
{if($mipcheckkey >= 114.09 && $mipcheckkey < 114.12){$mipflag114 = 2;}
if($mipcheckkey >= 115.09 && $mipcheckkey < 115.12){$mipflag115 = 2;}
if($mipcheckkey >= 116.09 && $mipcheckkey < 116.12){$mipflag116 = 2;}
if($mipcheckkey >= 117.09 && $mipcheckkey < 117.12){$mipflag117 = 2;}}
if($mipflag114 == 1){foreach $check114 (keys % { $mgfdata{$mipidentifier} })
{if($check114 >= 114 && $check114 < 115 && $mgfdata{$mipidentifier}{$check114}>0)
{$mipmasses{$check114}=$mgfdata{$mipidentifier}{$check114};}}
if($mipflag115 == 1){foreach $check115( keys % { $mgfdata{$mipidentifier} })
{if($check115 >= 115 && $check115 < 116 && $mgfdata{$mipidentifier}{$check115}>0)
{$mipmasses{$check115}=$mgfdata{$mipidentifier}{$check115};}}
if($mipflag116 == 1){foreach $check116 (keys % { $mgfdata{$mipidentifier} })
{if($check116 >= 116 && $check116 < 117 && $mgfdata{$mipidentifier}{$check116}>0)
{$mipmasses{$check116}=$mgfdata{$mipidentifier}{$check116};}}
if($mipflag117 == 1){foreach $check117 (keys % { $mgfdata{$mipidentifier} })
{if($check117 >= 117 && $check117 < 118 && $mgfdata{$mipidentifier}{$check117}>0)
{$mipmasses{$check117}=$mgfdata{$mipidentifier}{$check117};}}
foreach $finalmass (sort{$a<=> $b}(keys %mipmasses)){print MIPOUT
"$finalmass\t$mipmasses{$finalmass}\n";}
print MIPOUT $_;%mipmasses=();}
# iTRAQ 8-plex

```

```

elseif($labeltype == 2){ $mipflag113 = 1; $mipflag114 = 1; $mipflag115 = 1; $mipflag116 = 1;
$mipflag117 = 1;$mipflag118 = 1;$mipflag119 = 1;$mipflag121 = 1;$mipidentifier =
"$mipprecursormass\t$miprunid";foreach $mipcheckkey (sort (keys %mipmasses))
{if($mipcheckkey >= 113.09 && $mipcheckkey < 113.12){ $mipflag113 = 2;}
if($mipcheckkey >= 114.09 && $mipcheckkey < 114.12){ $mipflag114 = 2;}
if($mipcheckkey >= 115.09 && $mipcheckkey < 115.12){ $mipflag115 = 2;}
if($mipcheckkey >= 116.09 && $mipcheckkey < 116.12){ $mipflag116 = 2;}
if($mipcheckkey >= 117.09 && $mipcheckkey < 117.12){ $mipflag117 = 2;}
if($mipcheckkey >= 118.09 && $mipcheckkey < 118.12){ $mipflag118 = 2;}
if($mipcheckkey >= 119.09 && $mipcheckkey < 119.12){ $mipflag119 = 2;}
if($mipcheckkey >= 121.10 && $mipcheckkey < 121.14){ $mipflag121 = 2;}}
if($mipflag113 == 1){foreach $check113 (keys % { $mgfdata{ $mipidentifier } })
{if($check113 >= 113 && $check113 < 114 && $mgfdata{ $mipidentifier }{ $check113 }>0)
{ $mipmasses{ $check113 }= $mgfdata{ $mipidentifier }{ $check113 };}}
if($mipflag114 == 1){foreach $check114( keys % { $mgfdata{ $mipidentifier } })
{if($check114 >= 114 && $check114 < 115 && $mgfdata{ $mipidentifier }{ $check114 }>0)
{ $mipmasses{ $check114 }= $mgfdata{ $mipidentifier }{ $check114 };}}
if($mipflag115 == 1){foreach $check115 (keys % { $mgfdata{ $mipidentifier } })
{if($check115 >= 115 && $check115 < 116 && $mgfdata{ $mipidentifier }{ $check115 }>0)
{ $mipmasses{ $check115 }= $mgfdata{ $mipidentifier }{ $check115 };}}
if($mipflag116 == 1){foreach $check116 (keys % { $mgfdata{ $mipidentifier } })
{if($check116 >= 116 && $check116 < 117 && $mgfdata{ $mipidentifier }{ $check116 }>0)
{ $mipmasses{ $check116 }= $mgfdata{ $mipidentifier }{ $check116 };}}
if($mipflag117 == 1){foreach $check117 (keys % { $mgfdata{ $mipidentifier } })
{if($check117 >= 117 && $check117 < 118 && $mgfdata{ $mipidentifier }{ $check117 }>0)
{ $mipmasses{ $check117 }= $mgfdata{ $mipidentifier }{ $check117 };}}
if($mipflag118 == 1){foreach $check118 (keys % { $mgfdata{ $mipidentifier } })
{if($check118 >= 118 && $check118 < 119 && $mgfdata{ $mipidentifier }{ $check118 }>0)
{ $mipmasses{ $check118 }= $mgfdata{ $mipidentifier }{ $check118 };}}
if($mipflag119 == 1){foreach $check119 (keys % { $mgfdata{ $mipidentifier } })
{if($check119 >= 119 && $check119 < 120 && $mgfdata{ $mipidentifier }{ $check119 }>0)
{ $mipmasses{ $check119 }= $mgfdata{ $mipidentifier }{ $check119 };}}
if($mipflag121 == 1){foreach $check121 (keys % { $mgfdata{ $mipidentifier } })
{if($check121 >= 121 && $check121 < 122 && $mgfdata{ $mipidentifier }{ $check121 }>0)

```

```

{$mipmasses{$check121}=$mgfdata{$mipidentifier}{$check121};}}
foreach $finalmass (sort{$a<=> $b}(keys %mipmasses)){print MIPOUT
"$finalmass\t$mipmasses{$finalmass}\n";}
print MIPOUT $_;%mipmasses=();}
# TMT 6-plex
elsif($labeltype == 3){$mipflag126 = 1; $mipflag127 = 1; $mipflag128 = 1; $mipflag129 = 1;
$mipflag130 = 1;$mipflag131 = 1;$mipidentifier = "$mipprecursormass\t$miprunid";foreach
$mipcheckkey (sort (keys %mipmasses))
{if($mipcheckkey >= 126.05 && $mipcheckkey < 126.26){$mipflag126 = 2;}
if($mipcheckkey >= 127.05 && $mipcheckkey < 127.26){$mipflag127 = 2;}
if($mipcheckkey >= 128.05 && $mipcheckkey < 128.26){$mipflag128 = 2;}
if($mipcheckkey >= 129.05 && $mipcheckkey < 129.26){$mipflag129 = 2;}
if($mipcheckkey >= 130.05 && $mipcheckkey < 130.26){$mipflag130 = 2;}
if($mipcheckkey >= 131.05 && $mipcheckkey < 131.26){$mipflag131 = 2;}}
if($mipflag126 == 1){foreach $check126 (keys %{$mgfdata{$mipidentifier}})}
{if($check126 >= 126 && $check126 < 127 && $mgfdata{$mipidentifier}{$check126}>0)
{$mipmasses{$check126}=$mgfdata{$mipidentifier}{$check126};}}
if($mipflag127 == 1){foreach $check127( keys %{$mgfdata{$mipidentifier}})}
{if($check127 >= 127 && $check127 < 128 && $mgfdata{$mipidentifier}{$check127}>0)
{$mipmasses{$check127}=$mgfdata{$mipidentifier}{$check127};}}
if($mipflag128 == 1){foreach $check128 (keys %{$mgfdata{$mipidentifier}})}
{if($check128 >= 128 && $check128 < 129 && $mgfdata{$mipidentifier}{$check128}>0)
{$mipmasses{$check128}=$mgfdata{$mipidentifier}{$check128};}}
if($mipflag129 == 1){foreach $check129 (keys %{$mgfdata{$mipidentifier}})}
{if($check129 >= 129 && $check129 < 130 && $mgfdata{$mipidentifier}{$check129}>0)
{$mipmasses{$check129}=$mgfdata{$mipidentifier}{$check129};}}
if($mipflag130 == 1){foreach $check130 (keys %{$mgfdata{$mipidentifier}})}
{if($check130 >= 130 && $check130 < 131 && $mgfdata{$mipidentifier}{$check130}>0)
{$mipmasses{$check130}=$mgfdata{$mipidentifier}{$check130};}}
if($mipflag131 == 1){foreach $check131 (keys %{$mgfdata{$mipidentifier}})}
{if($check131 >= 131 && $check131 < 132 && $mgfdata{$mipidentifier}{$check131}>0)
{$mipmasses{$check131}=$mgfdata{$mipidentifier}{$check131};}}
foreach $finalmass (sort{$a<=> $b}(keys %mipmasses)){print MIPOUT
"$finalmass\t$mipmasses{$finalmass}\n";}

```

```
print MIPOUT $_;%mipmasses=();}}
else{print MIPOUT $_;}}}
```

```
##### CLOSE AND EXIT #####
```

```
print "\nScript completed. Press ENTER to exit...";
```

```
if(<STDIN>){exit;}
```

```
#####
```

```
##### SUBROUTINES #####
```

```
#####
```

```
sub datestamp{($sec,$min,$hour,$day,$month,$yr19,@rest)=localtime(time);
```

```
@rest=$sec=$min=$hour;$year=$yr19+1900;$month++;if($day<10){$day='0'.$day;}if($month<10){$month='0'.$month;}$date=$year.$month.$day;return $date;}
```

```
#####
```

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und nur mit Hilfe der von mir angegebenen Quellen und Hilfsmittel unter Anleitung meiner akademischen Lehrer verfasst habe. Die vorliegende Arbeit entstand unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft und wurde bisher keiner anderen Universität zur Begutachtung vorgelegt.

Kiel, den

Thomas Jakoby-Gregorius

Danksagung

Vorab möchte ich die Gelegenheit nutzen, um allen Menschen zu danken, die mir in der Zeit der Dissertation mit Rat und Tat zur Seite gestanden haben. Die erfolgreiche Fertigstellung einer Dissertation beruht letztendlich auch auf der Hilfestellung vieler Personen.

Zuerst möchte ich meinen besonderen Dank an meinen Doktorvater Herrn Prof. Dr. Andreas Tholey richten. Dieser hat mir sowohl die Möglichkeit gegeben an dem Neuaufbau einer wissenschaftlichen Arbeitsgruppe teilzuhaben als auch sehr interessante wissenschaftliche Themenstellungen im Bereich der Proteomanalyse v. a. in Bezug auf die proteolytische Prozessierung von Proteinen und Peptiden zu bearbeiten. Die häufigen Diskussionen nicht zuletzt auch bis zu späteren Stunden auf SFB- oder IRTG-Seminaren haben wesentlich zum Gelingen der Arbeit beigetragen. Mit Freude betrachte ich auch die Entwicklung des Arbeitskreises vom „Drei-Mann-Betrieb“ in Saarbrücken bis hin zur Ausbildung einer größeren, stets wachsenden Gruppe in Kiel voller Postdocs und Doktoranden. Ich habe die Arbeitsgruppe mit einem lachenden und einem weinenden Auge verlassen und denke gerne an die alte Zeit zurück.

Ich danke Frau Prof. Lindhorst und Herrn Prof. Dr. Sönnichsen als mögliche Vertretung für die Bereitschaft meine Arbeit zu begutachten, auch wenn es während meiner Dissertation leider nicht zur gemeinsamen Bearbeitung von Projekten kam.

Ebenfalls möchte ich mich bei Herrn Prof. Dr. Christoph Becker-Pauly bedanken, der sich als 3. Prüfer in der Disputation zur Verfügung gestellt hat, leider war meine Zeit in Kiel vorbei, als seine begann.

Ein großer Dank sei ebenfalls an alle Korrekturleser gerichtet, die sich doch in einer großen Anzahl an meiner Arbeit ausgelassen haben. Hier sind Dr. Konstantin Schneider aus dem Arbeitskreis Heinzle, Marion Rohmer, Dominik Bäumlisberger und Benjamin Müller aus dem Arbeitskreis Karas, Christoph Alexander Wehe aus dem Arbeitskreis Karst, sowie Verena Abel, die als Geisteswissenschaftlerin v. a. einen Blick auf die deutsche Grammatik geworfen hat, zu nennen. Aber auch in den eigenen Reihen ist so mancher nicht um den Genuss meiner Arbeit herumgekommen. Hier waren es Dr. Thorsten Overath, Dennis Linke und meine

eigene Frau, Barbara Gregorius (ja, sie hat den Namen wirklich behalten), die ebenfalls den ein oder anderen Kommentar äußerten.

Natürlich möchte ich mich bei der gesamten Kieler Arbeitsgruppe sowohl für die tägliche gute Zusammenarbeit als auch für das freundschaftliche Zusammensein bedanken. Ich habe mich während der gesamten Zeit mit euch sehr wohl gefühlt und vermisse euch aus dem doch sehr bodenständigen Nordrhein-Westfalen, in dem ich (wir) nun gelandet sind. An dieser Stelle seien alle auch noch einmal persönlich genannt. Zuerst möchte ich mich bei meiner Arbeitskollegin Barbara Gregorius bedanken. War in der Danksagung meiner Diplomarbeit noch von der „alten Dame“ die Rede, so bin ich nun sehr glücklich, dass ich sie nun als mein angetrautes Ehefrau bezeichnen darf, das mir nicht nur im privaten, sondern v. a. auch im wissenschaftliche Alltag eine echte Hilfe und Ergänzung war. Ohne sie wäre die Dissertation unmöglich gewesen. Als nächstes sei Dr. Thorsten Overath genannt, der als erster Postdoc die Probezeit überstanden hatte und sich menschlich wie auch wissenschaftlich als echter Glücksgriff erwiesen hat. Auch die anderen Postdocs Dr. Chien-Wen Hung und Bart van den Berg seien an dieser Stelle erwähnt. Erstere kann man als die gute Seele der Arbeitsgruppe bezeichnen und letzterer hat mit seinen Skripten einen wesentlichen Teil zu meiner Arbeit beigetragen. An dieser Stelle sei dann auch passenderweise Dennis Linke genannt, der mittlerweile u. a. als „angewandter Informatiker“ seinen Platz in der Arbeitsgruppe gefunden hat und mit dem man auch privat gut zu Recht kommt. Jana Schlenk, Angela Holste und Joanna Tucher möchte ich ebenfalls danken, die auf ihre sympathischen Arten ihren positiven Anteil am sehr guten Betriebsklima beigetragen haben. Ich hoffe Jana kann mit ihrer nun vollen Stelle der Arbeitsgruppe (als echter Gewinn) noch lange erhalten bleiben, und dass meine Vorarbeiten Joana den Start etwas erleichtert haben (an dieser Stelle auch einen Gruß an den „Nürnberger“). Angela möchte ich natürlich auch für die schöne Woche in Frankreich und die regelmäßige Versorgung mit Gewürzen danken. Leider hatte ich mit Dr. Tomas Koudelka nur sehr kurz das Vergnügen, aber ich bin mir sicher, dass auch er seinen Beitrag zum guten Betriebsklima und der engen Zusammenarbeit liefert. Zuletzt (und hier sei darauf hingewiesen, dass die Reihenfolge der Auflistung keine Wertung enthält) möchte ich mich auch bei Christian Treitz bedanken, der zunächst als der „ewige Diplomand“ und dann auch als Doktorand den Einzug nach Kiel und in unsere Herzen fand. Ihr seid alle jederzeit herzlich in Nottuln willkommen und ich freue mich auf jeden Einsatz, der mich wieder nach Kiel und in euer Gefilde führt.

Auch die Leute aus dem SFB bzw. der IRTG sollen nicht vergessen werden, wobei v. a. Prof. Dr. Matthias Leippe, Prof. Dr. Ottmar Janßen, Prof. Dr. Fritz Koch-Nolte, Nina Hegemann, Jette Oberdörster, Benjamin Schönbeck und Dr. Jürgen Fritsch, genannt seien.

Dr. Stephan Phillipp möchte ich ebenfalls meinen Dank aussprechen, auch wenn unser Sofa seinen Ansprüchen leider nicht entsprechen konnte.

Bleiben noch die alten Freunde aus dem Saarland, die bei jedem Heimaturlaub für genügend Abwechslung von der Wissenschaft gesorgt haben um die Batterien neu aufzuladen.

Auch den Freunden in der neuen Heimat Nottuln, sowie den (neuen und alten) Arbeitskollegen bei Thermo Fisher Scientific sei für die Unterstützung in der Zeitspanne nach Kiel gedankt, die noch zum Abschluss der Dissertation von Nöten war.

Selbstverständlich wäre die Dissertation ohne die Unterstützung meiner Familie undenkbar gewesen. Hier möchte ich mich besonders bei meinen Eltern Heinz-Peter „Vadda“ Jakoby sowie Birgit „Mudda“ Jakoby bedanken, die mir nicht nur mein Studium, sondern auch vieles anderes erst ermöglicht haben. Auch meinem Bruder Dirk und seiner Frau Daniela Jakoby (ja, sie hat den richtigen Namen angenommen) möchte ich danken, da nicht zuletzt jedes herzliche Willkommen im Saarland auch zu gestärkter Motivation geführt hat. Hier sei natürlich auch Barbara wieder genannt, die mich nicht nur um Immanuel, sondern auch um Jolina-Marie bereichert hat, die wahrlich immer für genügend Abwechslung sorgen.

Lebenslauf

Name: Thomas Jakoby
Staatsangehörigkeit: deutsch
Geburtsdatum: 01. November 1981
Geburtsort: Saarbrücken

Ausbildung und beruflicher Werdegang:

Schulbildung: 1988 – 1992 Grundschule Scheidt
1992 – 2001 Albertus-Magnus-Gymnasium St. Ingbert

Schulabschluss: Abitur

Zivildienst: 08/2001 – 05/2002
Arbeiterwohlfahrt, Saarbrücken

Studium: WS 2002/2003 – SS 2008
Chemiestudium (Diplom) an der Universität des Saarlandes
Abschluss: Diplom-Chemiker

Auslandsaufenthalte: 11/2006 – 01/2007
Forschungsaufenthalt als akademischer Visitor im Chemistry Research Laboratory an der University of Oxford, UK

Promotionsarbeit (Prof. A. Tholey) - seit 11/2008; Universität des Saarlandes, innerhalb des Exzellenzclusters „Multimodal Computing and Interaction“
- ab 07/2009; Christian Albrechts Universität zu Kiel innerhalb des Exzellenzclusters „Inflammation at Interfaces“
- ab 07/2010; Christian Albrechts Universität zu Kiel innerhalb des Sonderforschungsbereichs 877 „Proteolysis as a Regulatory Event in Pathophysiology“

Beruf: seit 02/2012; Anstellung als Applikations-Spezialist für die Thermo Fisher Scientific GmbH, Idstein, Deutschland

Publikationsliste und Beiträge zu Kongressen

Teile dieser Dissertation finden sich in folgenden Veröffentlichungen wieder:

Jakoby T, van den Berg BHJ, Tholey A (2012). Quantitative protease cleavage site profiling using tandem mass tag labeling and LC-MALDI-TOF/TOF MS/MS analysis. *J Proteome Res*, **11**: 1812-1820.

Jakoby T, Tholey A, van den Berg BHJ (2012). Improved reporter ion assignment of raw isobaric stable isotope labeled LC-MALDI-TOF/TOF MS/MS spectral data for quantitative proteomics. *Rapid Commun Mass Spectrom*, **26**: 2777-2785.

Menzel S, Mikciewicz M, Jakoby T, Bannas P, Schwarz N, Rissiek B, Danquah W, Nissen M, Haag F, Meyer-Schwesinger C, Tholey A, Koch-Nolte F. Purinergic signaling triggers shedding of the ARTC2 ecto-enzymes and ADP-ribosylation of Interferon gamma, in preparation.

Mitwirkung an Veröffentlichungen zusätzlich zu den Teilen dieser Dissertation:

Lasaosa M, Jakoby T, Delmotte N, Huber CG, Heinzle E, Tholey A (2009). Rapid selection of peptide containing fractions in off-line two-dimensional HPLC in shotgun proteome analysis by screening with MALDI TOF MS. *Proteomics*, **9**: 4577-4581.

Krüger R, Tholey A, Jakoby T, Vogelsberger R, Lackner KJ (2011). Quantification of the Fabry marker LysoGB3 by tandem mass spectrometry. *Clin Chem Lab Med*, **49**: S764-S764.
Meeting Abstract

Krüger R, Tholey A, Jakoby T, Vogelsberger R, Mönnikes R., Rossmann H, Beck M, Lackner KJ (2012). Quantification of the Fabry Marker LysoGb3 in Human Plasma by Tandem Mass spectrometry. *J Chrom B*, **884**: 128-135

Philipp S, Jakoby T, Tholey A, Janssen O, Leippe M, Gelhaus C (2012). Cationic detergents enable the separation of membrane proteins of *Plasmodium falciparum*-infected erythrocytes by 2-D gel electrophoresis. *Electrophoresis*, **33**: 1120-1128.

Slawski M, Hussong R, Tholey A, Jakoby T, Gregorius B, Hildebrandt A, Hein M (2012). Isotope pattern deconvolution for peptide mass spectrometry by non-negative least squares/least absolute deviation template matching. *BMC Bioinformatics*, **13**: 291.

Gregorius B, Jakoby T, Schaumlöffel D, Tholey A (2013). Metal labeling for accurate multiplexed peptide quantification via matrix-assisted laser desorption/ionization mass spectrometry. *Anal Bioanal Chem*, **405**: 2735-2741

Gregorius B, Jakoby T, Schaumlöffel D, Tholey A (2013). Monitoring of protease catalyzed reactions by quantitative MALDI MS using metal labeling. *Anal Chem*, **85**: 5184-5190.

Beiträge zu Kongressen und Tagungen

Poster:

An alternative approach for quantitative shotgun proteomics: RP- x IP-RP-HPLC-MALDI-TOF/TOF MS in combination with the iTRAQ strategy.

34th International Symposium on High-Performance Liquid Phase Separations and Related Techniques, 28.06.2009-02.07.2009, Dresden, Deutschland (Einsatz als *Congress Assistant*).

An alternative approach for quantitative shotgun proteomics: RP- x IP-RP-HPLC-MALDI-TOF/TOF MS in combination with the iTRAQ strategy.

4th EU-Summer School in Proteomics Basics “High-Throughput Data Analysis and Statistics”, 01.08.2010-07.08.2010, Brixen, Italien.

Protease cleavage sites profiling using Tandem-Mass-Tag (TMT) labeling and MALDI-TOF/TOF MS/MS analysis.

IRTG-Symposium Proteases & Pathophysiology 26.09.2011-27.09.2011, Kiel, Deutschland