

# **Method Development for Qualitative, Quantitative and Computational Proteomics**

## **Dissertation**

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften

Dr.rer.nat.

der Mathematisch-Naturwissenschaftlichen Fakultät

der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Klaus Dennis Linke

Kiel, 2014



Erster Gutachter:

Prof. Dr. Andreas Tholey

Zweiter Gutachter:

Prof. Dr. Frank Sönnichsen

Tag der mündlichen Prüfung:

28.10.2014

Zum Druck genehmigt:

28.11.2014

gez.

Prof. Dr. Wolfgang J. Duschl, Dekan



## **Eidenstattliche Erklärung**

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Diese Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form in einem Verfahren zur Erlangung eines akademischen Grades vorgelegt.

---

Klaus Dennis Linke

## Zitat

***„Gehe nicht, wohin der Weg führen mag, sondern dorthin,  
wo kein Weg ist, und hinterlasse eine Spur.“***

(Johann Paul Friedrich Richter, 1763 – 1825)

# Danksagung

Die vorliegende Arbeit entstand am Institut für Experimentelle Medizin der Christian-Albrechts-Universität zu Kiel auf Anregung und unter der Leitung von

## **Herrn Prof. Dr. Andreas Tholey.**

Für die herzliche Aufnahme in seiner Arbeitsgruppe *Systematische Proteomics und Bioanalytik*, der Möglichkeit ein interessantes Gebiet der Proteomics zu erforschen und mich dabei in das Feld Bioinformatik einzuarbeiten, sowie die stetige Diskussionsbereitschaft und die mir gewährte Unterstützung möchte ich mich bei meinem Doktorvater herzlich bedanken.

Ebenso gilt mein Dank Herrn PD Dr. Norbert Reiling vom Forschungszentrum Borstel für die stetige Bereitschaft zur Diskussion von Resultaten sowie des Interesses an der Entwicklung und Verwendung alternativer Techniken in der quantitativen (Phospho-)Proteomforschung.

Ein besonderer Dank geht an die PostDocs der Arbeitsgruppe *Systematische Proteomics und Bioanalytik*. Speziell erwähnt seien Frau Dr. Chien-Wen Hung und Herr Dr. Thorsten Overath, die mir den Einstieg in die Proteomforschung erleichterten. Auch möchte ich Bart van den Berg für die Etablierung der Grundstruktur der IT danken. Weiterhin sei den Herren Dr. Tomas Koudelka und in ganz besonderem Maße Herrn Liam Cassidy für die stetige Unterstützung, gerade zum Ende der Dissertation, gedankt.

Nicht zu vergessen sind meine Doktoranden-Kollegen der ersten Stunden Barabara Gregorius, Dr. Thomas Jakoby-Gregorius, Christian Treitz; und natürlich auch den mir nachfolgenden Doktoranden Joanna Tucher, Prasath Somasundaram und Alexander Becker. Ich hoffe, dass ich den Wissensschatz, den ich mir während meiner Zeit als Doktorand aufbauen konnte, in einer ebenso großartigen Art und Weise als PostDoc am Institut für Experimentelle Medizin weitergeben kann, wie es einst Chien-Wen Hung und Thorsten Overath für mich taten. Last but not least möchte ich auch Jana Schlenk danken, die stets für den Nachschub an benötigten Chemikalien gesorgt hat.

Weiterhin möchte ich meiner Familie danken, die mir sehr dabei geholfen hat mir den Traum der Promotion zu ermöglichen.

# Widmung

Dem Linke-Clan

in Gedenken an

Oma und Opa

sie wären stolz



# Lebenslauf

## Persönliches

Vor- und Zuname	<b>Klaus <u>Dennis</u> Linke</b>
Geburtsort	<b>Wolfsburg, Deutschland</b>
Geburtstag	<b>12.08.1983</b>
Staatsangehörigkeit	<b>Deutsch</b>
Institut	<b>Universität Kiel - Institut für Experimentelle Medizin</b> AG Proteomics & Bioanalytik

## Berufserfahrung

Seit	<b>Universität Kiel - Institut für Experimentelle Medizin</b>
15/09/2013	AG Proteomics & Bioanalytik Wissenschaftlicher Mitarbeiter
15/09/2010	<b>Universität Kiel - Institut für Experimentelle Medizin</b>
-	AG Proteomics & Bioanalytik
15/09/2013	Doktorand
15/02/2010	<b>Universität Greifswald – Pharmazeutische Biotechnologie</b>
-	Master-Student
01/09/2010	<i>Isolierung und Charakterisierung von Bacillus Licheniformis Subpopulationen</i>
01/09/2008	<b>Fraunhofer Einrichtung für Marine Biotechnologie</b>
-	Laboringenieur
31/08/2009	
01/07/2008	<b>FHL Forschungsgruppe GmbH / FH Lübeck</b>
-	Laboringenieur
31/08/2008	
01/06/2007	<b>FH Lübeck – Institut für Technische Biochemie</b>
-	Diplomand
31/06/2008	<i>Pektine und Öle aus Sonnenblumen verschiedener Wachstumsstadien: Kombinierte Extraktionsverfahren</i>

## Studium

01/09/2009	<b>Hochschule Fresenius</b>
-	<i>Master of Science (Note: 1.8)</i>
31/06/2010	Bio- and Pharmaceutical Analysis (englisch)
07/2008	<b>Buchpreis für die jahrgangsbeste Leistung</b> im Fachbereich Angewandte Naturwissenschaften der FH Lübeck
06/2008	<b>Zeugnis zur</b> Ausbilder-Eignungsverordnung
01/03/2004	<b>Fachhochschule Lübeck</b>
-	<i>Diplom-Ingenieur (FH) (Note: 1.1)</i>
31/06/2008	Chemieingenieurwesen
15/09/2005	<b>Praxissemester:</b>
-	<i>Shell Global Solutions GmbH Deutschland, Hamburg</i>
31/03/2006	Abteilung „Industrial Lubricants“
01/08/2005	<b>Grundpraktikum: (je 6 Wochen)</b>
-	- Shell Global Solutions GmbH, Hamburg
01/07/2003	- EnerSys/ Hawker, New Port, Wales

# Veröffentlichungen

In Vereinbarung mit §7(2) Promotionsordnung (Satzung) der Mathematisch-Naturwissenschaftlichen Fakultät und Technischen Fakultät der Christian-Albrechts-Universität zu Kiel vom 31. August 2012:

Abschnitte dieser Dissertation sind bereits an anderer Stelle veröffentlicht:

Kapitel 2.1 (Unterkapitel Phosphopeptid Standard):

## **Optimized Fragmentation Conditions for iTRAQ labeled Phosphopeptides**

Dennis Linke, Chien-Wen Hung, Liam Cassidy, and Andreas Tholey

Journal of Proteome Research 2013 12 (6), 2755-2763

Kapitel 3.1:

## **Paleoproteomic study of the Iceman's brain tissue**

\*Frank Maixner, \*Thorsten Overath, \*Dennis Linke, Marek Janko, Gea Guerriero, Bart H. J. van den Berg, Bjoern Stade, Petra Leidinger, Christina Backes, Marta Jaremek, Benny Kneissl, Benjamin Meder, Andre Franke, Eduard Egarter-Vigl, Eckart Meese, Andreas Schwarz, Andreas Tholey, Albert Zink, Andreas Keller

\* authors contributed equally

Cell Mol Life Sci 2013 Oct 6;70(19):3709-22. Epub 2013 Jun 6.

Kapitel 3.2:

## **LC-MS Based Cleavage Site Profiling of the Proteases ADAM10 and ADAM17 Using Proteome-Derived Peptide Libraries.**

Joanna Tucher, Dennis Linke, Tomas Koudelka, Liam Cassidy, Claudia Tredup, Rielana Wichert, Claus Pietrzik, Christoph Becker-Pauly, Andreas Tholey

J Proteome Res 2014 Apr 17;13(4):2205-14. Epub 2014 Mar 17

Folgende Abschnitte befinden sich in der finalen Präparationsphasen zur Veröffentlichung:

Kapitel 2.2:

**Handling of nearly isobaric Labelings**

Dennis Linke, Harald Barsnes, Chien-Wen Hung, Marc Vaudel, Prasath Somasundaran, Andreas Tholey

Kapitel 2.3:

**Accurate Identification of iTRAQ tyrosine labeled Peptides**

Dennis Linke, Joanna Tucher, Andreas Tholey

Kapitel 2.4:

**Application of Multi-Proteases for targeted Phosphoprotein Quantification**

Dennis Linke, Tomas Koudelka, Alexander Becker, Andreas Tholey

Mitwirkung an weiteren Publikationen, die sich z.Z. noch in der Veröffentlichungsphase sind und nicht Bestandteil dieser Arbeit sind:

**Proteome Analysis in Schizosaccharomyces pombe: Amino Acid Biosynthesis and Membrane Composition as Target for Improved Protein Secretion**

Chien-Wen Hung, Konstantin Schneider, Dennis Linke, Tobias Klein, Uwe Anders, Olivier Claude, Matthias Bureik, Elmar Heinzle and Andreas Tholey

**Systems genomics evaluation of the SH-SY5Y neuroblastoma cell line as a model for Parkinson's disease**

Abhimanyu Krishna, Maria Biryukov, Christophe Trefois, Paul Antony, Rene Hussong, Jake Lin, Merja Heinäniemi, Gustavo Glusman, Sandra Koeglsberger, Olga Boyd, Bart van den Berg, Dennis Linke, Davig Huang, Kau Wang, Leroy Hood, Andreas Tholey, Reinhard Schneider, David Galas, Rudi Balling, Patrick May



## Zusammenfassung

Die Protein Phosphorylierung stellt eine wichtige posttranslationale Modifikation dar, die eine Vielzahl biologischer Prozesse reguliert. Um die dynamischen Kontrollmechanismen besser verstehen zu können, ist es wichtig phosphorylierte Proteine identifizieren und quantifizieren zu können. Die Anwendung isobarer Derivatisierungsstrategien (z.B. iTRAQ) zur quantitativen Phosphopeptid-Analyse erfordert (i) optimal eingestellte Peptid-Fragmentierungsbedingungen und (ii) speziell angepasste Algorithmen zur Identifizierung iTRAQ derivatisierter (phosphorylierter) Peptide. Weiterhin (iii) wird eine Vielzahl an *Peptide-Spectrum-Matches* bzw. Phosphopeptid spezifischer Sequenzen für eine akkurate Quantifizierung benötigt.

Diese drei Anforderungen wurden in einer Plattform vereinigt, um iTRAQ derivatisierte (Phospho)Proteine relativ zueinander quantifizieren zu können. So werden optimale Fragmentierungsbedingungen benötigt, um iTRAQ derivatisierte Phosphopeptide identifizieren und quantifizieren zu können. Dies schließt die Detektion von Peptid-Rückgrat Fragment-Ionen, spezifischer Phosphopeptid-Signale (z.B. Neutralverlust der Phosphat-Gruppe) und iTRAQ Reporter Ionen mit ein. Unterschiedliche Tandem MS Fragmentierungstechniken der LTQ Orbitrap Velos (*collision induced dissociation* (CID), *CID plus multistage activation*, and *higher energy collision dissociation* (HCD)) und ihre Fähigkeit diese Ionen zu erzeugen wurden untersucht. Hierbei ist anzumerken, dass die Identifizierung und Quantifizierung gegenläufige Ansprüche an die normalisierte Kollisionsenergie bei der (Phospho)Peptid-Fragmentierung stellen. Daher muss ein Kompromiss eingegangen werden, der beiden Anforderungen (Identifizierung und Quantifizierung) gerecht wird. Dafür wurde eine angepasste MS/MS Methode entwickelt. Diese verwendet CID ausschließlich zur Peptid-Identifizierung, währenddessen HCD sowohl zur Identifizierung als auch Quantifizierung angewandt wird. Trotz der dualen Fragmentierung eines Vorläufer-Iones ist die benötigte Messdauer mit LC-ESI MS/MS Experimenten vereinbar.

Zusätzlich zu den optimierten Fragmentierungsbedingungen werden auch speziell angepasste Algorithmen in der computergestützten Peptid-Identifizierung (z.B. mittels Datenbanksuchen) benötigt. Um z.B. die nicht genau berechnete Masse der Vorläufer-Ionen nach *iTRAQ-4Plex* Derivatisierung zu korrigieren, wurde ein Algorithmus entwickelt, der die Intensitäten der verschiedenen Reporter Ionen in die Berechnung miteinbezieht. Hierdurch konnte die berechnete Masse der Vorläufer-Ionen (im Mittel) um 4 ppm verbessert werden, wenn die verwendeten MS/MS Spektren ein sehr intensives iTRAQ<sub>114</sub> Reporter Ion aufweisen. Eigenhändig geschriebene Programme wurden auch verwendet, um die korrekte Peptid-Identifizierung iTRAQ derivatisierter Proben sicherzustellen. Hierfür wurde Nebenreaktion bei der iTRAQ Derivatisierung, die Modifizierung von Tyrosin mit iTRAQ, genauer untersucht. Es wurden zwei bisher nicht beschriebene Immonium-Ionen bei m/z 280.17 (iTRAQ-4Plex) und 440.28 (iTRAQ-8Plex) entdeckt, die aufgrund ihres spezifischen Charakters auch als zusätzliche Merkmale in der automatisierten Peptide-Identifizierung genutzt werden können.

Um die Plattform zur relativen Quantifizierung iTRAQ derivatisierter Peptide zu komplettieren, wurde ein neues Verfahren etabliert, welches sowohl die Identifizierung als auch Quantifizierung von Phosphopeptiden verbessert. Dies wurde durch das Zusammenführen eines Multi-Protease-Verdaus mit der iTRAQ-8Plex Derivatisierung erreicht. Hierdurch konnte die Anzahl quantifizierbarer Phosphopeptid relevanter Sequenzen

signifikant gesteigert werden. So wurden z.B. 16 (Multi-Protease Ansatz) statt nur vier (Trypsin) Phosphopeptid spezifische Sequenzen von  $\beta$ -Casein korrekt quantifiziert. Gleichzeitig war die Vereinbarkeit des Multi-Protease-Verdaus mit der SDS-PAGE, der iTRAQ Derivatisierung sowie der Phosphopeptid-Aufreinigung mittels  $\text{TiO}_2$  gewährleistet.

Im zweiten Abschnitt dieser Dissertation lag der Fokus auf der Entwicklung von Computerprogrammen für zwei (voneinander unabhängigen) Projekten, um eine schnelle und reproduzierbare Auswertung massenspektrometrischer Daten zu gewährleisten.

Das Ziel des ersten Projektes war es die Protein-Identifizierungen mehrerer Datenbanksuchen (SEQUEST, Mascot, OMSSA und X!Tandem) zu vereinen. Es stellte sich heraus, dass X!Tandem exzellent für die Auswertung massenspektrometrischer Daten einer 3500 Jahre alten Eis-Mumie (Ötzi) geeignet ist. Weiterhin wurde durch den kombinierten Ansatz von vier Datenbanksuchen nicht nur die Anzahl identifizierter Proteine erhöht, sondern es konnten auch 87% der identifizierten Proteine mit mehr als einer Datenbanksuche identifiziert werden.

In dem zweiten Projekt lag der Fokus auf der Entwicklung zweier Programme für die Auswertung von Q-PICS Daten im Vordergrund, um die Schnittstellenspezifität uncharakterisierter Proteasen identifizieren zu können. Die Programme beinhalten Funktionen zur Datenfilterung und zum Export der P1 – P5 Position in Relation zum neuen N-Terminus. Die entwickelten Programme wurden in einer Studie eingesetzt, um die zuvor berichtete Schnittstellenspezifität der Proteasen ADAM10 und ADAM17 zu bestätigen.

Schlussendlich kombinierte diese Dissertation unterschiedliche Bereiche der Proteomforschung, um neue Techniken zu entwickeln oder bereits bestehende massenspektrometrische Mess- oder Auswerte-Methoden zu verbessern. Es konnten beiden Ziele, (i) Etablierung einer Plattform zur relativen Quantifizierung und (ii) der Entwicklung neuer Computerprogramme zur automatisierten Datenanalyse, erfolgreich abgeschlossen werden.

## Abstract

Protein phosphorylation is an important posttranslational modification that plays a regulatory role within numerous biological processes. The simultaneous identification, localization, and quantification of phosphorylated proteins is vital for understanding this dynamic control mechanism. The application of isobaric labeling strategies, e.g., iTRAQ, for quantitative phosphopeptide analysis requires (i) optimal peptide fragmentation conditions, (ii) sophisticated computational proteomics algorithms to identify (phosphorylated) iTRAQ labeled peptides, and (iii) the ability to use more than one Peptide-Spectra-Match and phosphopeptide sequence to guarantee accurate phosphorylation specific quantification.

These three demands were combined into a platform to relatively quantify iTRAQ-4Plex labeled (phospho)proteins on a LTQ Orbitrap Velos. Optimal fragmentation conditions for identification and quantification of iTRAQ labeled phosphopeptides requires parallel monitoring of specific ions arising from peptide backbone fragmentation, neutral losses of phospho moieties, and cleaved iTRAQ tags. Tandem MS fragmentation techniques available in the LTQ Orbitrap Velos (*collision induced dissociation* (CID), *CID plus multistage activation*, and *higher energy collision dissociation* (HCD)) were examined for their ability to generate these ions. The optimal Normalized Collision Energy for quantification and identification of iTRAQ labeled (phospho)peptides show inverse demands, and a compromise between optimal parameters for each aspect is necessary. A MS/MS measurement protocol was established that involves CID measurement in ion trap for identification followed by HCD measurement in Orbitrap for parallel identification and quantification that satisfies the time requirements for LC-ESI MS/MS experiments.

In addition to optimize fragmentation conditions, automated peptide identification techniques via database search engines require computational, proteomics specific, algorithms. To avoid an imprecise *in-silico* precursor mass calculation after iTRAQ-4Plex labeling, an algorithm was compiled which factors in the intensity distribution of all four iTRAQ reporter ions in tandem MS spectra to calculate the precursor mass more accurately. The benefit of this algorithm was shown by significantly improving the mass accuracy (up to 4 ppm) for Peptide-Spectrum-Matches (PSMs) which were dominated by iTRAQ<sub>114</sub> reporter ions. *In-house* compiled computational proteomics tools were also used to enhance the validity of peptide identification after iTRAQ labeling. Here, the iTRAQ tyrosine modification, a frequently occurring side-reaction of iTRAQ labeling, was under special investigation. Two novel immonium ions, at  $m/z$  280.17 (iTRAQ-4Plex) and 440.28 (iTRAQ-8Plex), were observed which are highly typical for this modification. These ions can serve as additional identification features in search algorithms to enhance the identification confidence.

To complement the platform to relatively quantify phosphopeptides by isobaric labeling, a new procedure was established to improve phosphopeptide identification and quantification simultaneously. This was achieved by combining a multi-protease digest with iTRAQ-8Plex labeling, which significantly raised the number of quantifiable (phospho)peptide sequences (e.g., from 4 (trypsin only) to 16 (multi-protease) phosphopeptide related sequences for  $\beta$ -casein). In addition, the compatibility of the multi-protease approach with SDS-PAGE, iTRAQ labeling and phosphopeptide enrichment using TiO<sub>2</sub> was achieved successfully.



In the second part of this dissertation the development of computational proteomics tools to assist in reproducible and fast data-analysis was performed across two independent projects.

The aim of the first project was to develop and establish a computational workflow to combine protein identification results obtained from multi-database searches of SEQUEST, Mascot, OMSSA and X!Tandem. Among these, X!Tandem was found to perform excellent to identify proteins from the 3,500 years old Tyrolean iceman *Oetzi*. By combining the four search engines, the number of proteins identified increased significantly; additionally, it could be demonstrated that more than 87% of the proteins could be identified by more than one search engine.

For the second project, two computational proteomics tools for the Q-PICS approach were developed to identify cleavage site motifs of uncharacterized proteases. These tools covered the functions to filter peptide list and export the P1 – P5 position in relation to the new N-Terminus. The developed tools were used in a study to confirm the previously reported cleavage site specificities of the proteases ADAM10 and ADAM17.

In conclusion, this dissertation has combined various distinct aspects within proteomics to develop novel, or enhance already established, measurement and data processing procedures. To this end, the initial aims of (i) providing a platform for relative quantification and (ii) the development of computational tool set to assist data analysis, were successfully achieved.

# Abbreviations

(p)AGC	(predicted) automated gain control
[M-80 or [M-80/z]	precursor ion with neutral loss fragmentation of $\text{HPO}_3$
[M-98] or [M-98/z] or [M-P]	precursor ion with neutral loss fragmentation of $\text{H}_3\text{PO}_4$
[M]	precursor
[P]	phospho
1024-A; 1025-A	sample 1024 and 1025 analyzed by in-gel digestion
1024-B; 1025-B	sample 1024 and 1025 analyzed by in-solution digestion
AA	amino acid
AB	antibodies
AC	alternating current
ACN	acetonitrile
AP	alkaline phosphatase
APCI	atmospheric pressure ionization
APS	ammonium persulfate
AQUA	absolute quantification
BEMAD	$\beta$ -elimination michael addition
bF	big FASTA file
CE	collision energy
CETG	in-silico combined sample of chymotrypsin, elastase, GluC and trypsin
CF	combined fraction; non-phosphorylated + wash fraction + phosphorylated fraction
Ch	chymotrypsin
CID	collision induced dissociation
CID-IT	collision induced dissociation with analysis in ion trap
CIT-MSA(-IT)	collision induced dissociation with multi stage activation & analysis in ion trap
CLI	command line input
Da	Dalton
DC	direct current
DHB	dihydroxybenzoic acid
DIGE	differential gel electrophoresis
DNA	deoxyribonucleic acid
é	electron
ECD	electron capture dissociation
EDTA	ethylenediaminetetraacetic acid
El	elastase
ESI	electro spray ionization
Et-OH	ethanol
ETD	electron transfer dissociation
eV	electron volt
FA	formic acid

FAB	fast atom bombardment
FDR	false discovery rate
FT	fourier transformation
GA	glycolic acid
GAMBAS	GAMBAS almost means basic
GC	gas chromatography
GI; GluC	Staphylococcus aureus V-8 Protease
GPC	gel permeations chromatography
GUI	graphical user interface
HCD	higher collision induced dissociation
HE-CID	high energy collision induced dissociation
HEPES	4-(2-hydroxyethyl)-1-piperazineethansulfonic acid
HIC	hydrophobic chromatography
HILIC	hydrophilic interaction chromatography
HPC	high pressure cell
HPLC	high performance liquid chromatography
IAA	iodoacetamide
ICAT	isotope coded affinity tags
ICPL	isotope-coded protein label
ICR	ion cyclotron resonance
ID	identification
IDi	isotope dilution
IEC	ion exchange chromatography
IEF	isoelectric focusing
IMAC	immobilized metal affinity chromatography
IP	immuno precipitation
iPeptide <sub>n</sub>	n-terminal and c-terminal iTRAQ labeled peptide
IT	ion trap
iTRAQ	isobaric tags for relative and absolute quantitation
iX	immonium ion of X; X = any amino acid (including modifications)
LC	liquid chromatography
LE-CID	low energy collision induced dissociation
LIT	linear ion trap
LPC	low pressure cell
MA	mascot
MALDI	matrix assisted laser desorption/ionization
MCP	multi channel plate
Me-OH	methanol
MECi	multi enzyme combined iTRAQ labeling

min	minute
MOAC	metal oxide affinity chromatography
MOPS	3-(n-morpholine)propanesulfonic acid
MS	mass spectrometry
ms	milli seconds
MSA	multi stage activation
<i>M. tuberculosis</i>	<i>Mycobacterium tuberculosis</i>
MudPIT	multidimensional protein identification technology
m/z	mass-to-charge ratio
nanoLC	nano liquid chromatography
NCE	Normalized Collision Energy
NHS	n-hydroxysuccinimide
NI	normalized intensity
NP	non-phosphorylated fraction
NSAF	normalized spectral abundance factors
OM	OMSSA (Open Mass Spectrometry Search Algorithm)
P; P'	prime; non-prime (only chapter 3.2)
PA	phosphoric acid
PC	phosphoamidate chemistry
PD	proteome discoverer
Peptide <sub>p</sub> S	c-terminal phosphorylated serine
PF	phosphopeptide fraction
PICS	proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites
PLP	myelin proteolipid protein
PMF	peptide mass finger print
ppm	parts per million
PQD	pulsed q collision dissociation
pS	phosphoserine
PSD	post source decay
PSM(s)	Peptide-Spectrum-Match(es)
pT	phosphothreonine
PTM	posttranslational modifications
pX <sub>Number</sub>	phosphorylation at amino acid X at the numbers position within the peptide
pY	phosphotyrosine
Q	quadrupole
Q-PICS	quantitative protease cleavage site profiling using tandem-mass-tag labeling
QconCAT	quantification concatamer
Qq	double quadrupole
QqQ	triple quadrupole

RF	radio frequency
RGB	red green blue
RI	relative intensity
RP	reversed phase
rpm	rounds per minute
RT	retention time
RT	room temperature
S-Lens/ SRIG	stacked ring ion guide
s; sec	seconds
S/N	signal to noise ration
sample D; sample P	dephosphorylated and phosphorylated sample, respectively
SAX	strong anion exchange
SCX	strong cation exchange
SDS-PAGE	sodiumdodecylsulfate polyacrylamid gel electrophorese
SE	SEQUEST
SEM	secondary electron multiplier
sF	small FASTA file
SFB	Sonderforschungsbereich
SILAC	stable isotope labeling by/with amino acids in cell culture
SISCAPA	table isotope standard capture with anti-peptide antibodies
SPC	spectral counting
TCEP	tris-2-(carboxyethyl)phosphine
TEAB	triethylammonium bicarbonate
TEMED	tetramethylethylenediamine
TFA	trifluoroacetic acid
TIC	total ion chromatogram
TiO <sub>2</sub>	titanium dioxide
TMT	tandem mass tag
TOF	time of flight
Tr	trypsin
u	unit
UV	ultra violet
v/v	volume/volume
VB(A)	visual basic for applications
WF	wash fraction
XIC	extracted ion chromatogram
XT	X!Tandem
z	charge

# Table Of Contents

Eidenstattliche Erklärung .....	i
Zitat .....	ii
Danksagung .....	iii
Widmung .....	iv
Lebenslauf .....	v
Veröffentlichungen .....	vi
Zusammenfassung .....	x
Abstract .....	xii
Abbreviations .....	xiv
Table Of Contents .....	xviii
1 Introduction .....	1
1.1 Proteomics.....	1
1.1.1 Qualitative Proteomics.....	1
1.1.2 Quantitative Proteomics.....	3
1.1.2.1 MS based Quantification Strategies.....	4
1.1.2.2 MS/MS based Quantification Strategies.....	6
1.1.3 Phosphoproteomics.....	12
1.2 Mass Spectrometry.....	14
1.2.1 Soft Ionization Techniques.....	15
1.2.2 Peptide Fragmentation Techniques.....	17
1.2.2.1 Peptide Fragmentation.....	17
1.2.2.2 Fragmentation Techniques.....	18
1.2.3 Mass Analyzers.....	21
1.2.4 Data-Processing and Automated Data-Analysis.....	24
1.2.4.1 Identifying Peptides using the Search Engine SEQUEST.....	24
1.2.4.2 Alternative Peptide Identification Techniques.....	26
1.2.5 Phosphoproteomics by Mass Spectrometry.....	27
1.2.5.1 Phosphopeptide Identification by Mass Spectrometry.....	27
1.2.5.2 Phosphopeptide Quantification by Mass Spectrometry.....	29
1.3 Aim of this Study.....	31
2 Developing a Platform for relative Quantification using iTRAQ .....	33
2.1 iEM – iTRAQ Evaluation Method.....	33
2.1.1 Introduction.....	33
2.1.2 Material and Methods.....	35
2.1.2.1 Sample Preparation.....	35
2.1.2.2 Mass Spectrometry.....	40

2.1.2.3 Database Searches and Feature-Detection.....	41
2.1.3 Results.....	46
2.1.3.1 Phosphopeptide Standard.....	46
2.1.3.2 Ten Protein-mix.....	61
2.1.3.3 Biological Sample.....	66
2.1.4 Conclusion.....	73
2.2 n-iTRAQ – Handling of Nominal Isobaric Labelings.....	75
2.2.1 Introduction.....	75
2.2.2 Material and Methods.....	76
2.2.2.1 Chemicals.....	76
2.2.2.2 Sample Preparation.....	76
2.2.2.3 Mass Spectrometry and Database Searches.....	76
2.2.3 Results and Discussion.....	79
2.2.3.1 Monitoring ppm Deviations Introduced by Different iTRAQ Ratios.....	79
2.2.3.2 Variations of Precursor Mass Accuracy Window.....	82
2.2.3.3 Post Database Search Re-Calculation of Precursor Masses.....	83
2.2.4 Conclusion.....	88
2.3 Y-iTRAQ – Accurate Identification of iTRAQ Tyrosine Labeled Peptides.....	89
2.3.1 Introduction.....	89
2.3.2 Material and Methods.....	90
2.3.2.1 Chemicals.....	90
2.3.2.2 Sample Preparation and iTRAQ Labeling.....	90
2.3.2.3 Mass Spectrometry, Database Searches and Data Analysis.....	91
2.3.3 Results.....	95
2.3.3.1 Detection of iY-iTRAQ-4Plex Specific Immonium Ions.....	95
2.3.3.2 Y-iTRAQ Immonium Ion Formation in Dependence of Collision Energy.....	103
2.3.3.3 Detection of iY-iTRAQ-8Plex specific Immonium Ions.....	104
2.3.3.4 Influence of iTRAQ tyrosine Labeling on the LC Retention Time.....	107
2.3.3.5 N-terminal Tyrosine as a Special Variant of Y-iTRAQ Modifications.....	112
2.3.4 Conclusion.....	116
2.4 MECi – Targeted Phosphoprotein Quantification.....	117
2.4.1 Introduction.....	117
2.4.2 Material and Methods.....	118
2.4.2.1 Chemicals.....	119
2.4.2.2 Sample Preparation and Analytical Workflow.....	119
2.4.2.3 Mass Spectrometry.....	121
2.4.3 Results.....	125
2.4.3.1 Percolator and FDR Analysis to Identify iTRAQ-8Plex Labeled Peptides.....	125
2.4.3.2 The Advantage of a Multi-Protease Approach.....	128

2.4.3.3	Enrichment of iTRAQ-8Plex labeled Phosphopeptides.....	134
2.4.3.4	Phosphoprotein Quantification using iTRAQ-4x2Plex.....	136
2.4.4	Conclusion.....	148
2.5	Discussion and Outlook.....	149
3	Development of Computational Proteomics Pipelines for Automated Data Processing .....	153
3.1	Paleoproteomics – Studying the Iceman's Brain Tissue.....	153
3.1.1	Material and Methods.....	154
3.1.1.1	Sample Preparation.....	154
3.1.1.2	Mass Spectrometry.....	155
3.1.1.3	Database Searches.....	156
3.1.2	Results and Discussion.....	157
3.1.2.1	Protein Identification Using a Multi-Database Search Engine Approach.....	157
3.1.2.2	Protein Aging Processes.....	160
3.1.3	Conclusion.....	166
3.2	PCSP – Computational Proteomics Tools for Protease Cleavage Site Determination.....	167
3.2.1	Implementation of PeptideMerger.....	168
3.2.2	Implementation of AminoAcidFinder.....	169
3.2.3	Conclusion.....	172
3.3	Discussion and Outlook.....	173
4	Appendix .....	I
4.1	Chemicals and Instrumentation.....	II
4.2	Amino Acids.....	VI
4.3	cRAP FASTA File.....	VII
4.4	Supplemental Tables for Chapter 2.4.....	IX
4.5	Figure Index.....	XXI
4.6	Equation Index.....	XXVIII
4.7	Table Index.....	XXIX
4.8	Copyright.....	XXXI
4.8.1	Journal of Proteome Research.....	XXXI
4.8.2	Cellular and Molecular Life Sciences.....	XXXIII
4.9	References.....	XXXVII



# 1 Introduction

## 1.1 Proteomics

The human genome contains about 2.85 billion nucleotides interrupted by 341 gaps, which contain about 20 – 25,000 gene-coding proteins (1). The entire set of proteins expressed by a genome at a point in time and under specific conditions is termed the *proteome*. This term was coined by Marc Wilkins who described the proteome as “*The entire PROTEin complement expressed by a genOME*” (2). The analysis of proteomes is complicated by alterations and protein splicing events. For example, up to 500,000 different protein species are thought to exist in humans (3). Moreover, in contrast to the static genome, the proteome changes dynamically by re-synthesis, degradation or posttranslational modifications of proteins (4,5).

The technique to identify proteins of a cell or tissue and research their sequences, modifications and quantity is defined as *proteomics* (2,6). Since the introduction of soft-ionization techniques namely *Matrix-assisted Laser Desorption/Ionization* (MALDI; (7–9)) and *Electrospray Ionization* (ESI; (10)), the term proteomics is often used as a synonym for the analysis of proteins by mass spectrometry (MS).

### 1.1.1 Qualitative Proteomics

Qualitative proteomics by mass spectrometry plays an important role in identifying proteins and can be classified into two categories: *top-down* and *bottom-up*. Top-down proteomics involves the analysis of intact proteins (11), whereas peptides released from proteins by proteolytic digestion are measured in bottom-up approaches (12–15). The bottom-up process is also referred to as *shotgun* proteomics if a mixture of proteins is digested (15).

To confidently identify proteins, these or their respective digests are often chromatographically separated prior MS-analysis to reduce the number of co-ionizing analytes (16). This technology is referred to as *Liquid Chromatography Mass Spectrometry* (LC MS; (17)). In LC MS, the LC can be coupled online (e.g., for LC-ESI MS) or offline (e.g., for LC-MALDI MS). Combining two complementary LC separation techniques is referred to as *Two Dimensional Liquid Chromatography Mass Spectrometry* (2D-LC MS (18)). *Strong Cation-Exchange* (SCX) chromatography in the first and *Reversed Phase* (RP) chromatography in the second dimension are often combined (19). Alternatively, two RP separation (first high pH, second low pH) can also be connected (20,21). The technology, using more than one LC dimension and combining the peptides identified from the different dimensions to identify their respective proteins, is referred to as *Multi-Dimensional Protein Identification Technology* (MudPIT; (22)).

An alternative technique to identify proteins is *GeLC-MS/MS*. Here, intact proteins are first separated using SDS-PAGE (*Sodium Dodecylsulfate Polyacrylamide Gel Electrophoresis*). Afterwards, each gel lane is cut into slices (e.g., 10 pieces) and each gel slice is subjected to in-gel digestion followed by LC MS. The advantage of GeLC-MS/MS is to reduce the sample complexity on the protein level and peptides, of a protein, are (ideally) retained in a single fraction. Moreover, interfering substances (e.g., for protein resolubilization) are removed

## 1 Introduction

prior in-gel digestion (23–25).

Another gel-based technique is *Two-Dimensional Gel Electrophoresis* (2-DE) which combines the isoelectric focusing of intact proteins in the first dimension and SDS-PAGE in the second dimension. After the stained protein spots are excised, these are in-gel digested and analyzed by mass spectrometry (26,27). Usually, peptides obtained from proteolytic digest of 2-DE spots are directly measured by MS (direct infusion for ESI MS or spot based MALDI MS) without further LC separation.

One technique to identify proteins (e.g., obtained from 2-DE) is by *Peptide-Mass-Fingerprint(ing)* (PMF; (28)). In this technique, the unknown protein of interest is digested by endopeptidases into smaller peptides, whose masses are measured by mass spectrometry (MS full scan). The measured peptide masses are then compared to theoretical peptide masses from known proteins. The higher the number of matching peptide masses, the more confident is the protein identification. Identifying proteins by PMF does well with 2D-gel electrophoresis where the protein purity is high. Here, a high number of specific and unique peptides can be detected, which is required to accurately identify proteins (29). However, protein samples of low purity generate to complex MS full scans, in which protein specific signals might be suppressed and not be detected. Nowadays, PMF is almost replaced by a technique referred to as *Peptide-Spectrum-Match* (PSM), which is used for LC separated protein digests as well as 2-DE gel spots of higher complexity.

Contrary to PMF, Peptide-Spectrum-Matching uses the information of MS (peptide mass) and MS/MS spectra (peptide sequence specific tags) to identify peptides and their respective proteins. To generate MS/MS spectra, which are also referred to as *tandem MS spectra*, precursor ions of interest are isolated and induced to dissociation (see upper section in Figure 1-1). The resulting fragment ions (including non-dissociated precursor ions if still present) are then analyzed and recorded as tandem MS spectra (see upper right spectrum in Figure 1-1).

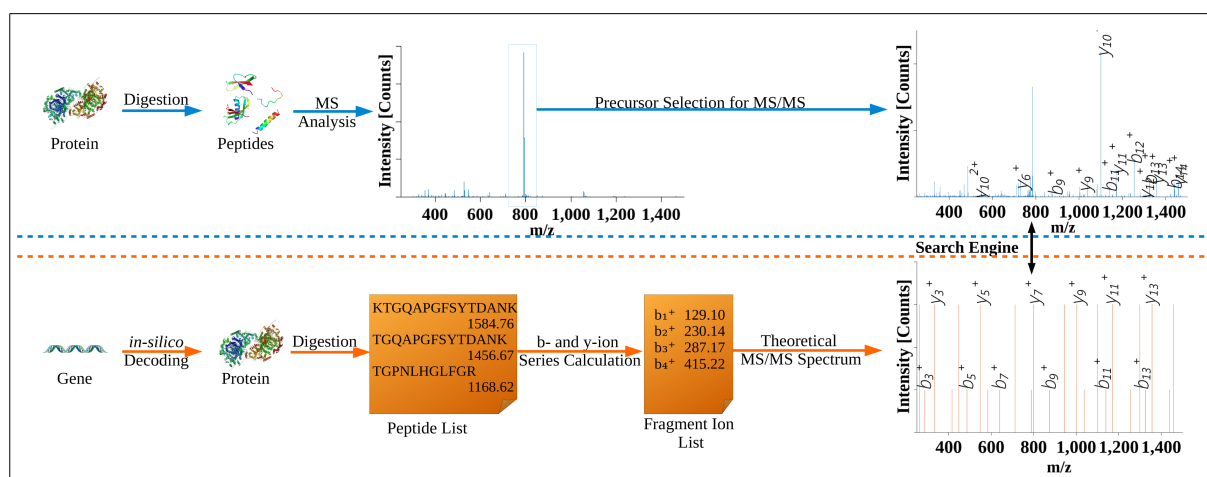


Figure 1-1: Mass spectrometry workflow describing the principle steps to obtain an experimental (upper) and theoretical (lower) spectrum. Both spectra are used for peptide identification by a search engine. Protein images used were published under the CC-BY-SA-3.0 and GNU Free Documentation License, respectively (30,31).

To finally identify peptide sequences by Peptide-Spectrum-Matching, the precursor masses and the specific fragmentation tag of each peptide is used in combination to their *in-silico* analogs. Independently of the

identification method (either by manual or computer assisted interpretation), the protein sequence is obtained by gene decoding. Afterwards, the protein is *in-silico* digested into smaller peptides (see Figure 1-1, Peptide List; MS analogs), and the corresponding peptides are *in-silico* fragmented; forming theoretical tandem MS spectra (see Figure 1-1, Fragment Ion List; MS/MS analogs). These theoretical tandem MS spectra are compared to the experimental MS/MS spectra, and a score is calculated based on the homology of both spectra. This scoring process is usually performed by so called database search engines, whose mode of operation will be described in chapter 1.2.4. Generating sequence specific fragment ions, which are necessary for peptide identification by PSM, and the different techniques commonly applied to induce peptide dissociation are explained in chapter 1.2.2.

### 1.1.2 Quantitative Proteomics

Quantitative proteomics offers tools to evaluate proteome wide changes of protein abundances under different biological conditions (32). Either the absolute amount of a protein species is determined or relative differences between two states are compared. At its starting point, quantitative proteomics was performed on 2D-gels by comparing spot sizes, shapes and intensities after staining. To compensate for experimental errors, a high number of multiple gels was required which increased the number of experiments performed (33,34). To overcome some experimental errors, *Differential Gel Electrophoresis* (DIGE) was introduced (35). In DIGE experiments, up to three biological states are labeled with unique amine reactive fluorescent dyes (36). Afterwards, the samples are combined and separated together on the same 2D-gel. Differences in protein abundances are visualized by different intensities of the fluorescent dyes after superimposing them. This information is used to identify differentially regulated proteins. As for all gel based techniques, there are limitations regarding the protein species that can be observed (e.g., very acidic or basic proteins are frequently lost during sample loading or separation). Although reference gels can be created by merging multiple gels, each protein (or spot) of interest needs to be extracted and identified separately by mass spectrometry (37).

Alternative gel-free techniques were developed which combine the identification and quantification process of proteins in a single workflow. These MS based techniques can be grouped into absolute and relative quantification, as illustrated in Figure 1-2. Absolute quantification involves the measurement of representative proteolytic peptides and stable-isotope labeled analogs in a single LC MS run (38) to determine the protein concentration and amount. In contrast to absolute quantification methods, signal intensities of two or more biological states are compared in relative quantification methods to recognize differentially regulated proteins. Comparing signal intensities can be performed on the MS or MS/MS level, either without the utilization of tags, by introducing (labeled) tags, or the addition of (labeled) internal standard peptides (39). Only relative quantification methods will be described in the next section, as absolute quantification was not performed in this dissertation.

## 1 Introduction

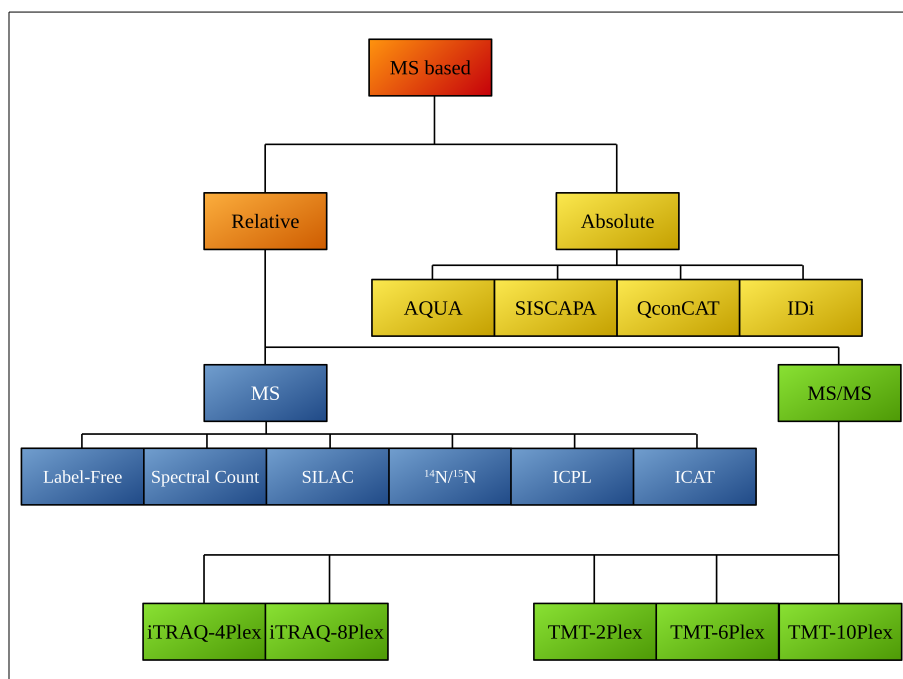


Figure 1-2: Summary of absolute (yellow) and relative (orange) quantification methods commonly applied in mass spectrometry based proteome research. Relative quantification can be further subdivided into MS (blue) and MS/MS (green) methods depending at which level relative quantification is performed. Abbreviations: AQUA (Absolute Quantification); SISCAPA (Stable Isotope Standard Capture with Anti-Peptide Antibodies); QconCAT (Quantification Concatamer); IDi (Isotope Dilution); SILAC (Stable Isotope Labeling by/with Amino Acids in Cell Culture); ICPL (Isotope-Coded Protein Label); ICAT (Isotope-Coded Affinity Tag); iTRAQ (Isobaric Tags for Relative and Absolute Quantitation); TMT (Tandem Mass Tags).

### 1.1.2.1 MS based Quantification Strategies

Relative MS based quantification strategies can be grouped into isotope labeling approaches or methods which do not require any labeling. Incorporation of stable isotopes *in-vivo* and *in-vitro* has frequently been utilized to quantify proteins from cell or tissue cultures or, from pull-down experiments from cell compartments (40).

*Protein Metabolic Labeling* (41) utilizes  $^{15}\text{N}$  labeled inorganic salts which are added as substrates to the growth medium for cell, bacterial or yeast cultures. By excluding all other nitrogen sources, all nitrogen atoms within the culture contain the  $^{15}\text{N}$  isotope (referred to as heavy sample). A second biological sample is feed exclusively with  $^{14}\text{N}$  (referred to as light sample). After combining both samples, different protein abundances are determined by comparing precursor ion signals on the MS level.

*Stable Isotope Labeling by Amino Acids in Cell Culture* (SILAC) involves the *in-vivo* incorporation of specific amino acids into proteins (42). The cell culture media lacks one or more essential amino acids, which are replaced by a non-radioactive, isotopically labeled form. Similar to protein metabolic labeling, up to three different samples are commonly compared in a quantitative proteomics workflow.

Another MS based quantification techniques is *Isotope-Coded Protein Label* (ICPL; (43)), which uses an

N-hydroxysuccinimide coupling reaction to bind differentially isotope-coded reagents to the amino group of N-termini and lysine side chains. After endopeptidase digestion, up to four different samples can be labeled, which appear as quadruplets in the acquired MS spectrum (44).

*Isotope Coded Affinity Tags* (ICAT; (45)) consist of three essential parts: a biotin affinity tag to isolate ICAT-labeled peptides, a linker which includes the stable isotopes and a reactive group with a specific activity to thiol-groups. Only peptides containing a labeled thiol-group can be used for quantification. This limits the application of ICAT in quantitative proteomics experiments.

All MS based quantification methods using stable isotope labeling increase the sample complexity on the MS level. This not only influences the ionization efficiency of peptides, but also requires specialized software tools for automated (data-dependent) precursor ion selection to exclude the repeating selection of the same precursor pair (46). Additionally, the identification of peptides is complex if only one isotope (e.g.,  $^{15}\text{N}$ ) is present in the MS spectra.

Some of these obstacles can be overcome by using relative quantification techniques which do not involve peptide labeling. One such technique is *Spectral Counting* (SPC; (47)). This semiquantitative technique measures the protein abundance based on the total number of identified MS/MS spectra (PSMs) for a particular protein. This technique is accepted as a practical approach for relative quantification. Although various statistical frameworks are required to improve abundance estimations (48). For example, larger proteins tend to contribute more spectra and peptides than smaller proteins do. To compensate for this effect, the *Normalized Spectral Abundance Factors* (NSAFs) can be calculated to correct this. Other major challenges for SPC are co-eluting peptides while precursor ion selection, the application of dynamic exclusion to increase the number of different precursor ions being selected and the FDR filtering of peptide identification (49).

Another quantitative method for relative quantification relies on the assumption that, under controlled conditions, peptides can be compared directly across different LC-ESI or LC-MALDI MS measurements. This is assisted by more robust chromatographic separation techniques and an increased scanning speed, resolution and mass accuracy of mass spectrometers (46). For quantification purposes, the *Extracted Ion Chromatogram* (XIC) of unique and razor peptides is calculated and used to explore differentially regulated proteins (50). This technique is referred to as *Label-Free Quantification*.

This approach is rather flexible as data-dependent or data-independent fragmentation techniques (e.g., MS<sup>n</sup>) can be applied (51,52). The major advantage is that no extra experimental steps (labeling and sample clean-up) are required. Moreover, quantification can be performed across a multiple of samples simultaneously.

As co-eluting peptides might influence the ionization efficiency or column aging processes might change the retention time of peptides, sophisticated computational approaches are required for data alignment of label-free experiments. Usually, these software tools also include data reduction algorithms (e.g., baseline removal, denoising and centroiding (53)). Software suites such as MaxQuant (54) integrate these processing features and directly links the quantification process to identified peptide and proteins.

An important factor for label-free quantification is that samples are measured independently of each other. This increases its flexibility but makes great demands on the LC MS instrumentation. For LC-ESI MS, changes

## 1 Introduction

in the *Total Ion Chromatogram* (TIC) intensities need to be compensated mathematically (55). This is a complex task if spray instabilities (the main factor for TIC changes) are observed. For LC-MALDI MS, laser adjustments during instrument calibration can influence the ionization efficiency of peptides, which in turns affects the observed intensity of peptides. Moreover, the analyte needs to be dried uniformly with the matrix to ensure reproducible ionization processes (56). Ideally for LC-MALDI MS, a peptide would only be present in exactly one fraction. As this is usually not observed, grouping of several MS spectra from different fractions is required to balance-out this effect (57). Due to the online-coupling of LC and analyte ionization, this correction procedure is not required for LC-ESI MS which facilitates the application of LC-ESI for label-free quantification.

All label-free experiments require a well designed experimental plan. This can include a measurement of samples in triplicates in tandem repeats ( $[A_1, B_1, C_1]$ ,  $[A_2, B_2, C_2]$ ,  $[A_3, B_3, C_3]$ ) or to shuffle samples randomly (46,58,59).

### 1.1.2.2 MS/MS based Quantification Strategies

The quantitative information of different (relative) protein abundances can not only be obtained from MS spectra, but also from MS/MS spectra. Predominantly two different, but chemically very similar, reagents are used: (i) iTRAQ which is an acronym for isobaric tags for relative and absolute quantification (60) and (ii) TMT (tandem mass tag; (61)). Both rely on isobaric labeling of peptides or proteins each with a specific tag.

TMT is available in a duplex, sixplex and tenplex form, whereas iTRAQ is available in a four- and eightplex form. The count correlates to the number of samples, which can be compared in a single experiment. Each of the reagents shares a basic structure in which by variation of isotopes, isobaric structures are gained. The basic structure of iTRAQ-4Plex, TMT-6plex and iTRAQ-8Plex, which are most frequently employed in quantitative proteomics (62), are shown in Figure 1-3.

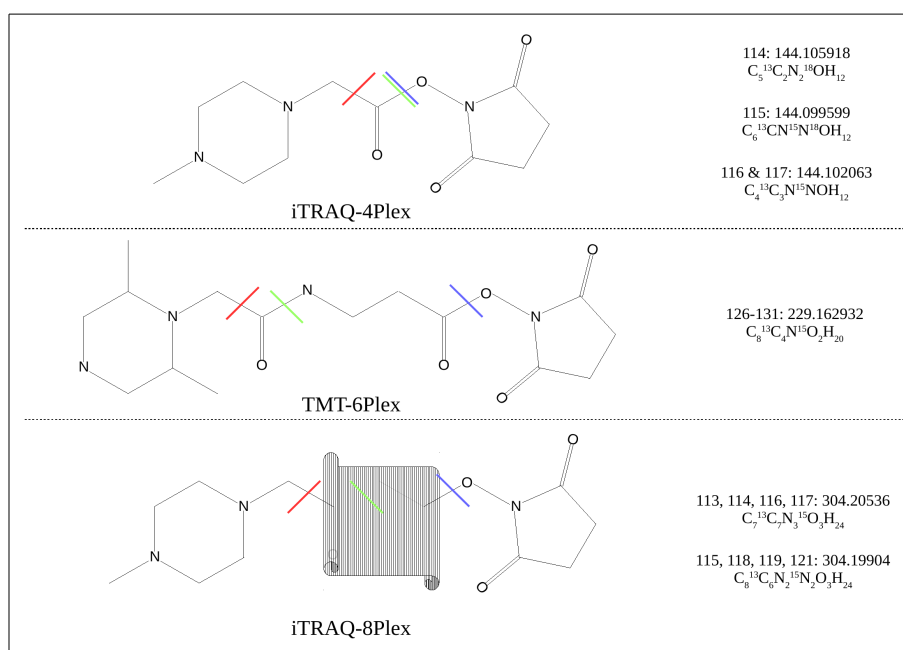


Figure 1-3: Chemical structure of iTRAQ-4Plex (upper), TMT-6Plex (middle) and iTRAQ-8Plex (lower) and their corresponding isotopes being used and their resulting masses, respectively (right side). Cleavage sites after

*MS/MS fragmentation are coded red and green for iTRAQ liberation and iTRAQ + balancer group dissociations which are also observable, respectively. The blue line indicates the leaving group while iTRAQ labeling. The exact structure for iTRAQ-8Plex is not published. Illustration is adapted by (63).*

iTRAQ and TMT consist of an amine-reactive group, a mass-normalization spacer and a reporter ion group (44,64), as illustrated in Figure 1-3. The purpose of the amine-reactive group is to bind iTRAQ and TMT reagents to the  $\alpha$ - (peptide N-terminus) and  $\epsilon$ -amino (lysine side chain) group of peptides. This amine-reactive group employs N-hydroxysuccinimide (NHS) chemistry (63), which is a relatively specific reaction that cannot be reversed and permits an almost complete labeling of primary amines with the amine-reactive group. In addition to primary amino groups, side reactions can be observed for serine, threonine and tyrosine (65). Especially the phenolic hydroxyl group of tyrosine is frequently derivatized at high pH values with up to 3% of total reactions (60). In contrast to primary amino groups, this reaction is reversible by hydrolysis (66).

The different reporter ion groups contain the quantitative information, whereas the mass normalizer group out-balances the peptide masses after isobaric labeling. Due to varying number of isotopes ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^{18}\text{O}$ ) in the reporter and mass normalizer group, the isobaric labeled peptides are indistinguishable on the MS level (isobaric mass) and appear as a single peak in MS scans (*MS Spectrum* in Figure 1-4).

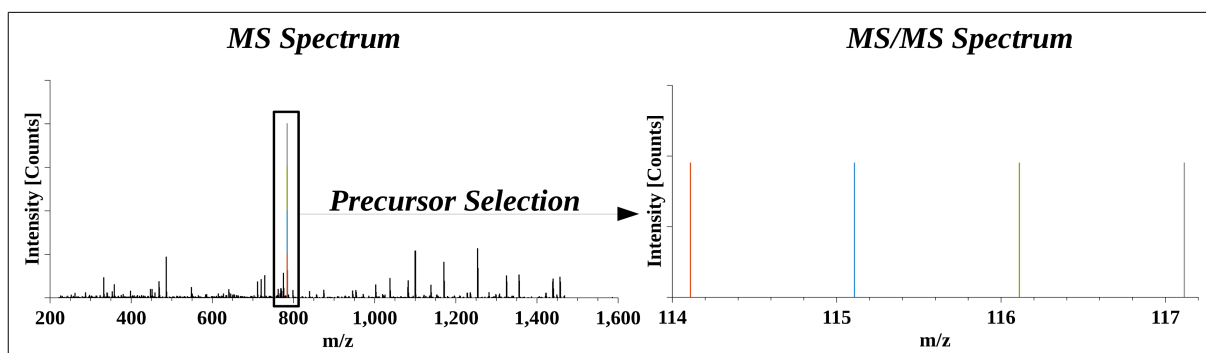


Figure 1-4: Representative MS (left) and MS/MS (right) spectrum showing the combinatorial effect of iTRAQ-4Plex multiplexing. The four samples are non-distinguishable in MS spectra (same precursor  $m/z$  due to iTRAQ reporter ion and balancer group). Upon fragmentation, the four iTRAQ labelings liberate unique iTRAQ reporter ions in the low  $m/z$  region (demonstrated by orange, blue, green and gray signals).

Yet upon precursor ion fragmentation of isobaric labeled peptides, the reporter ions are readily formed by cleavage at the carbonyl group of the piperazine ring (indicated with a red line in Figure 1-3; (63)) and are detected in the low  $m/z$  region in MS/MS spectra. As each reporter ions contains a different number of isotopes, which are out-balanced by the mass normalizer group in the undissociated precursor ion, the reporter ions are detected at different  $m/z$  values mostly separated by 1 Da (see *MS/MS Spectrum* in Figure 1-4).

Multiplexed peptides can be relatively quantified based on the reporter ion intensities (or areas), which correlate to the quantities of the differentially labeled peptides (Figure 1-4) and by extension, their corresponding proteins. For iTRAQ-4Plex, these reporter ions are detected at  $m/z$  114.1, 115.1, 116.1 and 117.1 for the reagents iTRAQ-114 to iTRAQ-117 (see *MS/MS Spectrum* in Figure 1-4). The balancer group is usually cleaved as a neutral fragment (cleavage between red and blue line in Figure 1-3) and is not detected (44). Fragmentation

## 1 Introduction

pathways leading to non informative fragment ions are suppressed for iTRAQ labeled peptides, and predominantly sequence specific signals are observed (65).

### The Fundamental Steps of Isobaric Labeling Workflows

Every iTRAQ and TMT labeling workflow is composed of nine different sample preparation steps. These are schematically illustrated in Figure 1-5: After cell lysis, protein extraction and concentration determination (summarized as “Protein Extraction” step), the biological samples are reduced and alkylated (being part of *Protein Digestion* step), as it is done for other quantitative proteomics techniques (e.g., label-free quantification). During these steps, primary amine free reagents have to be used in order to avoid cross-reactions with the iTRAQ reagents, such as tris-2(-carboxyethyl)phosphine (TCEP) for reduction and methyl methanethiosulfonate (MMTS) or iodoacetamid (IAA) for alkylation (65).

Typically, digestion is performed with trypsin, which cleaves almost exclusively after arginine and lysine (67). Digestion can be performed in buffers such as triethyl ammonium bicarbonate (TEAB), 4-(2-hydroxyethyl)-1-piperazineethansulfonic (HEPES), 3-(N-morpholine)propanesulfonic acid (MOPS) or phosphate buffers as these are compatible with iTRAQ labeling (free of primary amines).

For isobaric labeling (step 4), the protein digests of each sample are incubated with a unique isobaric reagent, e.g., iTRAQ-114, iTRAQ-115, iTRAQ-116 and iTRAQ-117 for iTRAQ-4Plex labeling. Either one (arginine) or two (lysine) isobaric labelings are attached to a tryptic peptide. Tyrosine labeling or missed cleavages after trypsin digestion can result in more than two isobaric labelings. However, this is less commonly observed than single or double isobaric labeling. To prevent cross-labeling of peptides with different isobaric labeling reagents prior sample multiplexing, pure water can be added to quench remaining isobaric reagents (68).

After the labeling reaction is completed, samples are mixed together (step 5) and can be used for further processing, such as peptide fractionation by LC and subsequent analysis by mass spectrometry (step 6 & 7). Upon precursor fragmentation, reporter ion signals are formed alongside peptide sequence specific signals in MS/MS spectra (as demonstrated in the combined schematic of *Peptide Fractionation* and *Tandem MS Analysis* (right side) in Figure 1-5). It is noteworthy, that a single MS/MS spectra contains qualitative (sequence specific fragment ions) and quantitative (reporter ions) information.

After data acquisition, peak areas (or intensity values depending on the applied scan mode) of the reporter ions are extracted, corrected for impurities (69), and peptide ratios are grouped into protein ratios (*Data Analysis*, step 8, Figure 1-5). ANOVA or students t-test are frequently used to validate the results (70).

The quantitative proteomics data can finally be used to highlight differentially regulated networks (*Results*, step 9, lower right section in Figure 1-5) or be used in combination to other techniques like metabolomics.



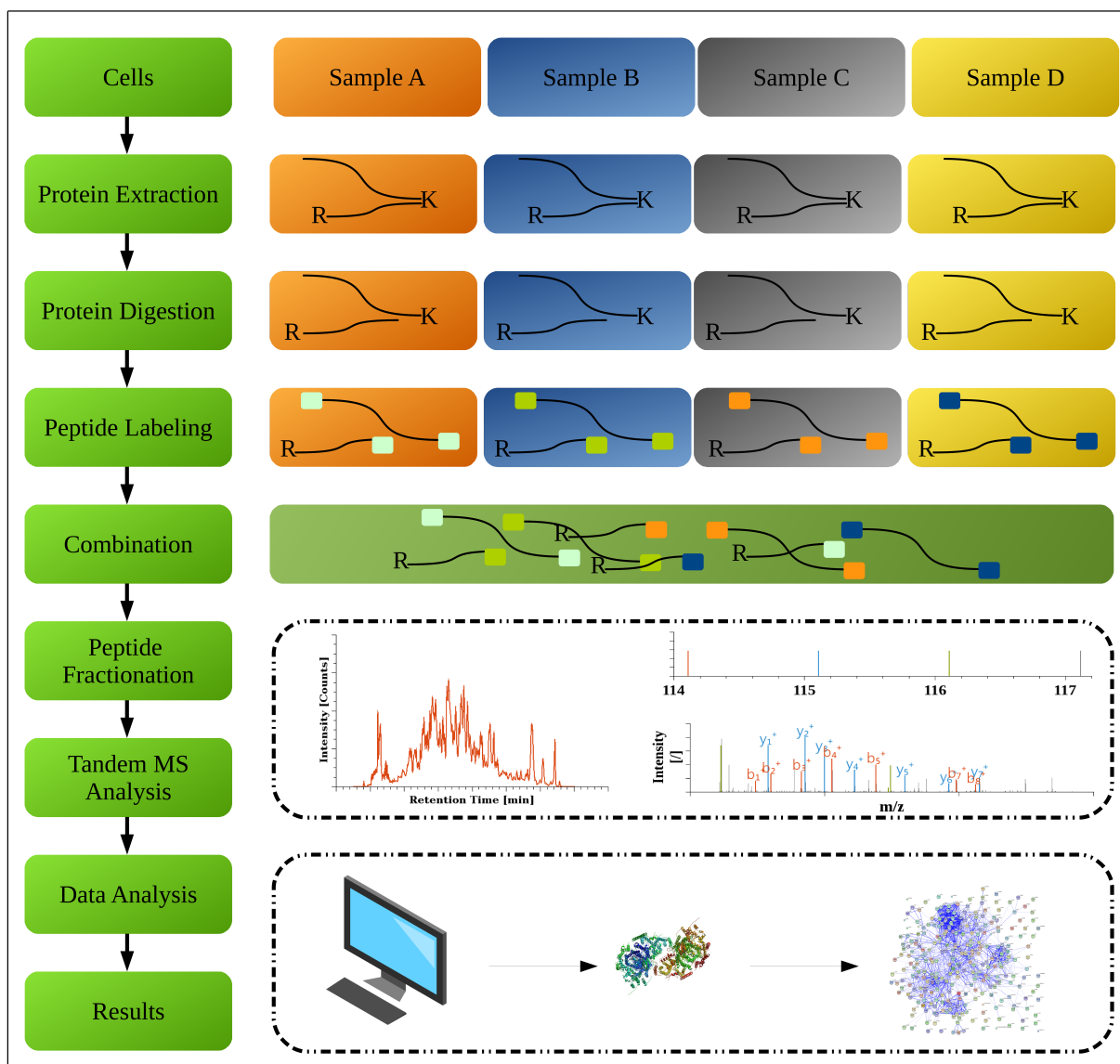


Figure 1-5: From iTRAQ labeled samples to biological pathways. In iTRAQ-4Plex experiments, the four samples are first reduced, alkylated and digested (usually trypsin), followed by iTRAQ labeling, each with a unique iTRAQ labeling reagent. After labeling, samples are combined and analyzed by 1D- or 2D-LC ESI or MALDI MS. Upon precursor ion fragmentation, reporter ions are formed together with sequence specific fragment ions. Intensities (or areas) of the iTRAQ reporter ions are extracted from MS/MS spectra. Protein identification and quantification results from database search are combined to calculate the significance of differentially abundant proteins. For visualization, protein networks are a common approach to highlight differentially regulated pathways (string-db; (71)). Workflow was adapted by (69) and further modified, images used were published under the CC-BY-SA-3.0 License (31).

## 1 Introduction

### **The Benefit and Drawback of Isobaric Labeling**

An advantage of iTRAQ and TMT labeling is the reduced impact of handling errors due to the early stage of sample combination compared to label-free approaches, which are only *in-silico* combined. Moreover, isobaric labeled protein digests can be pre-fractionated or enriched for specific peptides (such as phosphopeptides), which is a critical step for label-free quantification (72). Further, all peptides are labeled (if the N-terminus is unmodified) and can be used to quantify proteins. This improves quantitative reproducibility as all identified peptides are used to calculate the protein ratio rather than only a subset of peptides (e.g., cysteine labeled peptides in ICAT; (69,73)). iTRAQ can be performed on the peptide and protein level, which increases its flexibility of use (74). Moreover, iTRAQ and TMT do not show a significant retention time shift using reversed phase chromatography (75) due to the application of carbon, nitrogen and oxygen isotopes. In contrast to deuterium isotopes ( $^2\text{H}$ ), these isotopes do not influence the retention time (76,77).

iTRAQ and TMT quantification is exclusively performed on MS/MS level (Figure 1-4). Consequently, the sample complexity remains constant on MS level (*MS Spectrum* in Figure 1-4), which induces a signal amplification and increases the sensitivity of analyte ions (69). Theoretically, chemical noise is also reduced upon precursor fragmentation, offering a higher quantification accuracy compared to MS based quantification techniques (44,66).

A drawback of isobaric labeling using iTRAQ or TMT is that every precursor ion needs to be selected for fragmentation to receive iTRAQ reporter ion intensities. This increases the number of MS/MS spectra to be recorded compared to the *re-match between runs* feature for label-free approaches (78). Consequently, pre-fractionation of isobaric labeled peptides has to be performed, which reduces the advantage of sample multiplexing, as a higher number of fractions have to be measured (20,79,80).

Another drawback of isobaric labeling is that frequently only a single MS/MS spectrum is detected for a peptide species. This increases the need to measure technical replicas and makes demands on robust LC MS techniques. However, the overlap of identified proteins varies between technical and biological replicas. In a recently published study, only 35.4% of identified proteins have been found in each of the nine technical replicates being measured (81). This is often referred to as missing values and is prominently observed for peptides in low abundances and in data-dependent acquisition methods (81–83).

Beside low precursor ion intensities, also extreme ratios of protein abundances, diverse fragmentation upon precursor excitation and different ionization efficiencies of peptides can result in less robust quantification (77). Further, the MS instruments sensitivity and resolution in MS/MS mode directly affect the relative quantification, due to reporter ion detection in MS/MS spectra (70,83).

As MS/MS spectra contain qualitative and quantitative information, interfering substances in the low  $m/z$  region of reporter ions have to be avoided and must be excluded. Several interfering signals are known. Some of these are, the phenylalanine immonium ion's  $M+1$  isotope ( $m/z$  120.0 +1) which can interfere with iTRAQ<sub>121</sub> at  $m/z$  121.1 (low resolution MS/MS; (73)); the immonium ion of arginine ( $m/z$  115.0), which shows an  $m/z$  value close to iTRAQ<sub>115</sub> ( $m/z$  115.1) (74) and the  $y_1$  fragment ion from C-terminal proline (at 116.1 Da; (84)) that

shares the  $m/z$  value with the iTRAQ<sub>116</sub> reporter ion. However, using high resolution and high mass accuracy instruments (e.g., equipped with an Orbitrap mass analyzer (85)), it is possible to distinguish between interferences and reporter ions. The application of higher collision energies upon fragmentation assists to generate higher intensive reporter ion signals, while the signal intensity of interfering signals (e.g., contaminants) does not increase (69). Contaminant signals can have a dramatic impact on peptide and protein quantification regarding accuracy, precision and reproducibility, which can lead to miss-interpretation of data.

Co-eluting peptides, which have similar  $m/z$  values close to the target precursor ( $\leq 3\text{Da}$ ), can be co-isolated during precursor isolation. These ions will be co-fragmented (alongside the precursor ion of interest) and will also contribute to the iTRAQ reporter ion signals. This introduces a bias towards 1:1 ratios, a process often referred to as *underestimation* (73) and is, in particular, observed for precursor ions with lower intensity (59,77). To avoid these co-isolations, sample complexity has to be decreased, e.g., by two-dimensional liquid chromatography. However, this does not always lead to improved precursor selection and different techniques have been developed to overcome this problem. For example, proton transfer ion-ion reaction (PRT) can be used by which the charge of the precursor is reduced. This removes interfering substance at the observed  $m/z$  value. Consequently, the accuracy and precision of protein quantification increases. However, less peptides are identified as the sensitivity decreases (86). An alternative technique isolates the most intense ions of the MS/MS spectra and subsequently fragments them (MS/MS/MS analysis). The MS/MS spectrum is used for identification, whereas the MS<sup>3</sup> spectrum is used for quantification. Again, protein quantification accuracy increases, but less peptide are identified due to the prolonged cycle time. Moreover, specialized software is needed for data-analysis (87).

Overall, customized software is required for each MS and MS/MS quantitative workflow (e.g., SILAC, label-free or iTRAQ). Usually, a single software package cannot focus on all kind of quantification techniques, leading to a variety of different tools being published. More problematically, each quantification method requires its own algorithm (adapted to different artifacts created (64)), which needs to be developed individually (66). This is usually performed by different research groups, increasing the level of complexity (software language, cross-platform compatibility, statistical tests, noise and data filtering methods). In many cases, this hampers *inter-lab* comparison of results. To facilitate data sharing, defined data processing workflows (e.g., statistical tests) and standard formats (e.g., mzQuantML (88)) are required. A systematic overview, which compares the different quantification techniques, is shown in Table 1-1.

*Table 1-1: Overview of the main groups used to relatively quantify proteins. Metabolic labeling includes techniques such as SILAC. Chemical labeling on protein or peptide level cover techniques like iTRAQ, TMT or ICAT. For SPC, the number of identified spectra is summarized (MS/MS\*).*

Method	Labeling Level	Quantification Level	Application	Accuracy	Proteome Coverage	Linear Dynamic Range (logs)
Metabolic Labeling	Protein	MS	2-3 entities cell culture systems	+++	++	1-2
Chemical Labeling	Protein	MS or MS/MS	2-3 entities cell culture systems	+++	++	1-2
Chemical Labeling	Peptide	MS or MS/MS	2-8 entities cell culture systems	++	++	2
Label-Free	-	MS	multiple	+	+++	2-3
SPC	-	MS/MS*	multiple	+	+++	2-3

## 1 Introduction

### 1.1.3 Phosphoproteomics

*Posttranslational Modifications* (PTMs) are a fast method for cells to adapt to internal or external stimuli. In contrast to regulations at the genome and transcriptome level, the modification of proteins is a fast (within milliseconds to seconds) and more dynamically regulated mechanism (89).

Beside glycolization, ubiquitination and acetylation, phosphorylation is one of the most prominent PTMs, which can be grouped into four categories: (i) O-phosphomonoesters on the hydroxyl-group (predominantly: serine, threonine and tyrosine; rare: hydroxyproline (90)), (ii) N-phosphoramidates (arginine, lysine and histidine), (iii) S-phosphothioesters (cysteine) and (iv) phosphoanhydrides (glutamic and aspartic acid; (91,92).

Reversible protein phosphorylation is an important post translational modification in both eukaryotic and prokaryotic biological systems involved in various regulation mechanisms, e.g., catalytic activity changes, protein-protein and protein-DNA interactions, protein synthesis, gene expression, apoptosis; and survival, metabolism, division, growth and differentiation of cells (92–96). Upon phosphorylation, not only a negatively charged group is introduced: it can alter a protein's biochemical properties and cause conformational changes (94,97). Protein phosphorylation - dephosphorylation events are important for medical research, as a variety of diseases (e.g., liver cirrhosis or insulin resistance (98–102)) are a consequence of dysregulation in signaling networks.

The formation of phosphoproteins is catalyzed by kinases using high-energy cofactors (e.g., ATP and GTP). The dephosphorylation is catalyzed by phosphatases releasing orthophosphate (75,103). About 3% of the human gene coding regions are responsible for kinases and phosphatases (104). 517 kinases have been categorized based on public and proprietary genomic, complementary DNA, and expressed sequence tag (EST) sequences of the human genome (105).

Primarily serine (~90%, *pS*), threonine (~10%, *pT*) and tyrosine (<1%, *pY*) are phosphorylated in eukaryotes (106–110). These O-phosphomonoesters are acid-stable, making them (*pS*, *pT* and *pY*) the most commonly studied phosphorylations by mass spectrometry as they are compatible with treatments commonly used in proteomics (111,112). Newer studies using MS-based methods have revealed the importance of *pY*, changing the distribution to 86.4% (*pS*), 11.8% (*pT*) and 1.8% (*pY*), respectively (113,114).

Phosphorylation events occur in many cases in the sub stoichiometry level, making their detection and analysis in the presence of non-phosphorylated peptides challenging (92,115,116). To circumvent this problem, mass spectrometry compatible methods were established to enrich phosphopeptides, including antibody precipitation (IP; (117)), immobilized metal ion chromatography (IMAC; (118)), metal oxide affinity chromatography (MOAC; (119,120)) and HPLC based fractionation techniques. IP is a technique optimal for *pY*-peptides. IMAC and MOAC are also capable of enriching *pS*- and *pT*- peptides. The IMAC method is based on the selective affinity of trivalent ions, such as  $\text{Fe}^{3+}$  or  $\text{Ga}^{3+}$  to phosphate groups (118). The presence of highly acidic non-phosphorylated peptides decrease the specificity of enrichment, as these tend to bind to the IMAC material. However, lowering the pH below 1.9 neutralizes the acidic amino acids and improves the selectivity of IMAC towards phosphopeptides (121).

TiO<sub>2</sub>, the most prominent material for MOAC, is often applied in offline enrichment strategies either in a batch mode or via pre-packed columns (122,123). The addition of dihydroxybenzoic acid (DHB) or similar acids increases the selectivity of this phosphopeptide enrichment technique (124). As for IMAC, lowering the pH-value reduces co-purification of non-phosphorylated (highly acidic) peptides during enrichment. It has been shown that TiO<sub>2</sub> is tolerant to a variety of buffers and salts used during sample preparation, thereby increasing its application possibilities (125). The elution of phosphopeptides is typically performed under basic conditions (pH 9 to 11) using ammonia solutions. For singly phosphorylated peptides, the elution is reproducible. However, it is speculated that the elution of multi-phosphorylated peptides is difficult and less reproducible due to their high affinity towards TiO<sub>2</sub> (122).

Alternative strategies to analyze phosphopeptides involves the replacement of the phosphorylation into a different chemical group ( $\beta$ -elimination with michael addition, BEMAD; phosphoamidate chemistry, PC). To overcome some of the limitations observed for phosphopeptide enrichment technique (Table 1-2), combination of these can be applied to increase the number of identified phosphopeptides. A common method employs SCX fractionation prior IMAC or MOAC phosphopeptide enrichment (126–128). Recently, 7,000 unique phosphopeptides were identified using a 3D-LC-MS setup (high-pH RP, SAX and low-pH separation; (129)).

Table 1-2: Overview of phosphopeptide enrichment techniques commonly applied in proteomics (130). Abbreviation: [P] phospho.

Enrichment Technique	Function/ Interaction	Advantages	Disadvantages
Antibodies (AB)	selective enrichment by specific regions for the AB and antigen	<ul style="list-style-type: none"> <li>✓ Efficient for pY-peptides (more electron rich results in stronger binding)</li> </ul>	<ul style="list-style-type: none"> <li>✗ Expensive</li> <li>✗ limited by number of AB available</li> <li>✗ less specific for pS- and pT-peptides</li> </ul>
IMAC	Affinity of [P] for transition metal ions (Cu <sup>2+</sup> , Ni <sup>2+</sup> , Zn <sup>2+</sup> , Fe <sup>2+</sup> , Co <sup>2+</sup> , and other) based on coordinative bonds between electron donor (oxygen of [P]) and metal ion	<ul style="list-style-type: none"> <li>✓ Variation of metal ion influences specificity</li> </ul>	<ul style="list-style-type: none"> <li>✗ Variation of metal ion influences specificity</li> <li>✗ nitrogen &amp; sulfur (especially histidine) is enriched</li> <li>✗ co-purification of acidic peptides</li> <li>✗ pH dependent</li> <li>✗ loss of mono[P] at 0.1% TFA</li> <li>✗ bias towards multi[P]</li> </ul>
MOAC	Affinity of phosphate group for metal oxides	<ul style="list-style-type: none"> <li>✓ High selectivity</li> <li>✓ High robust towards a variety of buffers</li> <li>✓ Compatible with organic solvents</li> <li>✓ TA/ TFA minimizes acidic peptide isolation</li> </ul>	<ul style="list-style-type: none"> <li>✗ Acidic peptides are co-enriched</li> <li>✗ excluder (organic solvents) required</li> <li>✗ DHB induces high background signal &amp; ionization is suppressed</li> <li>✗ bias towards mono[P]</li> <li>✗ multi[P] bind to strong;</li> <li>✗ high pH required</li> </ul>
Precipitation	Addition of calcium salts forms insoluble calcium phosphate deposits	<ul style="list-style-type: none"> <li>✓ Efficiency comparable with IMAC</li> <li>✓ Less biased towards multi[P]</li> </ul>	<ul style="list-style-type: none"> <li>✗ Sample loss</li> <li>✗ Reproducibility</li> </ul>
Chemical derivatization	Removal of [P] by $\beta$ -elimination and addition of a suitable chemical group (michael-addition)	<ul style="list-style-type: none"> <li>✓ Removal of [P] circumvents neutral loss effects and increases ionization efficiency</li> <li>✓ Incorporation of label for quantification possible</li> </ul>	<ul style="list-style-type: none"> <li>✗ Sample loss</li> <li>✗ Side reactions</li> </ul>
SCX/ SAX	Retains cations (SCX) or anions (SAX) by interaction with stationary phase	<ul style="list-style-type: none"> <li>✓ [P]peptides elute earlier (SCX) or later (SAX) than non-[P] peptides</li> </ul>	<ul style="list-style-type: none"> <li>✗ Highly hydrophobic [P]peptides can be retained (SCX) leading to co-elution</li> <li>✗ Highly acidic non-[P]peptides co-elute</li> <li>✗ Miss cleavages lead to co-elution</li> </ul>
HILIC	[P] have a higher hydrophilicity and elute later than non-[P]peptides based on interactions with stationary phase	<ul style="list-style-type: none"> <li>✓ Compatible with most buffers used</li> </ul>	<ul style="list-style-type: none"> <li>✗ Low efficiency</li> <li>✗ Co-elution</li> </ul>

## 1 Introduction

### 1.2 Mass Spectrometry

Mass spectrometry (MS) has become the method of choice for identifying, quantifying and characterizing proteins at the molecular level. The MS instruments were built *in-house* and have evolved since the first commercial variant introduced in 1943 (Consolidated Engineering Corporation) into instruments which can be referred to as *black boxes* (131) due to their complexity and inter-disciplinary aspects (physics, electronic, informatics & related fields like chemistry and biology). Pioneer work has been done by Wien and Sir J.J. Thomson in the beginning of the twentieth century (132–134). In the 1940s and 1950s, the fundamental work of introducing Time of Flight (TOF; (135)) ion cyclotron resonance (ICR; (136)) and ion traps (IT; (137)) revolutionized MS. By coupling gas-chromatography to MS (GC-MS; (138,139)), analytes are separated prior to ionization. This significantly improves the analysis of complex samples. 25 years later, the introduction of soft ionization techniques [Fast Atom Bombardment (FAB; (140)), Electrospray Ionization (ESI; (10,141)) and Matrix-assisted Laser Desorption/ Ionization (MALDI; (7–9))] allowed researches to also study biologically complex samples, e.g., proteins and peptides. Here, tandem MS technique using fragmentation of precursor ions with a collision gas (introduced 1960s; (142)), in combination with the intact ionization of biological samples, has allowed researches to study a variety of scientific questions, e.g., measuring precisely the molecular mass ( $m/z$  value) of intact compounds, molecular structures like disulfide-linkages, reaction dynamics like enzyme kinetics and gas-phase reactions to name a few (131).

In the 21<sup>st</sup> century, the Orbitrap (Makarov, year 2000; (143)) and linear quadrupole ion trap (LIT; Hager, year 2002 (144)) were introduced. These, in combination to atmospheric pressure ionization sources (ESI, APCI) and commercialization allowed LC-ESI MS to become the method routinely applied to analyze protein samples (131). However, LC-MALDI MS is still a widely applied method. Due to its complementary nature to LC-ESI MS, an increasing number of identified peptides and proteins are observed by combining both techniques (145–147). In principle, all mass spectrometers are made up of five parts: (i) inlet system, (ii) ion source, (iii) mass analyzer, (iv) detector, and (v) data storage (external system), as it is outlined in Figure 1-6.

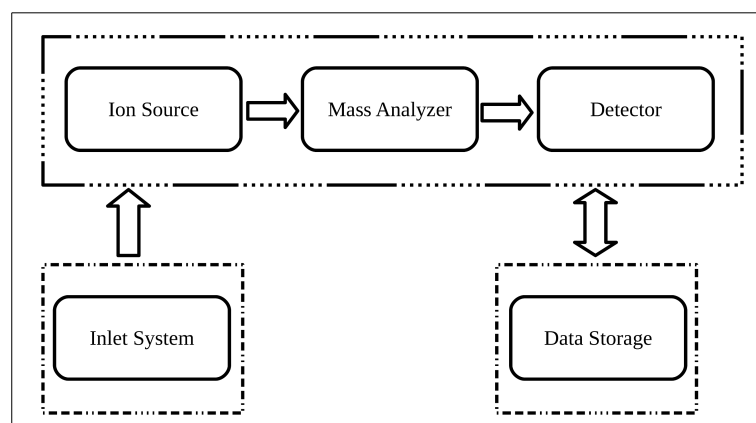


Figure 1-6: Schematic diagram of the fundamental parts of a mass spectrometer with the ion source, mass analyzer and detector being part of the instrument and the inlet system and data storage system which are attached by interfaces.

The inlet system (i) can be either under vacuum or under atmospheric pressure (148). Its purpose is to transfer samples into the gas phase and further into the mass spectrometer. Within the ion source (ii), analytes are ionized and further transferred into the mass analyzer (iii), where they are separated according to their mass to charge ratio ( $m/z$ ). Each separated ion is detected (e.g., by electron multiplier) which is recorded by the detector (iv) and stored in a data system (v).

### 1.2.1 Soft Ionization Techniques

Two soft ionization techniques (MALDI and ESI) paved the way for modern mass spectrometers (6). In MALDI experiments, the sample to be analyzed is firstly co-crystallized with a matrix compound (e.g., 2,5-dihydroxybenzoic acid). Due to its structure, this weak organic acids serves as proton donator and absorbs ultraviolet (UV) laser energy (149). By rapid laser heating, matrix molecules carrying analytes are transferred into the gas phase (6). As weak organic acids function as proton donators the analyte is positively ionized. If weak bases are used, the matrix accepts protons from the analyte leading to a negative sample ionization (149). By direct and efficient energy transfer from the matrix to the analyte molecule, the analyte ions can be measured intact. Further, measurements with a high mass accuracy and up to the sub-picomole concentration are possible (150–152).

During ionization, mostly singly charged precursor ions are observed (153). By changing the classical MALDI sample preparation methods, multiply charged precursor ions can also be detected (154–156). To receive stable and high signal-to-noise ratios, several MS spectra have to be accumulated. This is one of the major drawbacks for MALDI. Other factors negatively affecting MALDI MS are the shot-to-shot and sample-to-sample reproducibility, that is more critical for complex rather than simplified (pre-fractionated) samples (157).

For electrospray ionization (ESI), analytes are directly injected as liquids via the inlet system into the ion source. The application of voltage differences between a capillary and the ionization source of MS allows the solvent to assimilate charges. With increasing charge state, the droplet at the end of the tip becomes unstable, forming a so called Taylor cone (see Figure 1-7). This conical shape turns into a jet, which further fragments into a stream of highly charged droplets with a diameter of two to three  $\mu\text{m}$ . By attraction forces and evaporation of solvents, the size of each droplet shrinks and the charge to volume ratio increases. At a certain point, the surface tension cannot be hold (Rayleigh limit) and smaller droplets are formed (Coulomb explosion). By further evaporation of solvent, the Rayleigh limit is reached again and additional Coulomb explosions are observed. This process is repeated until solvent free analyte ions are formed. These ions are attracted to and introduced into the mass spectrometer (158,159). Upon analyte ionization, the formation of multiply charged ions is a common observation for ESI (in contrast to MALDI), and there are several theories and hypotheses to explain this, although the the process in not understood entirely (160–165).

## 1 Introduction

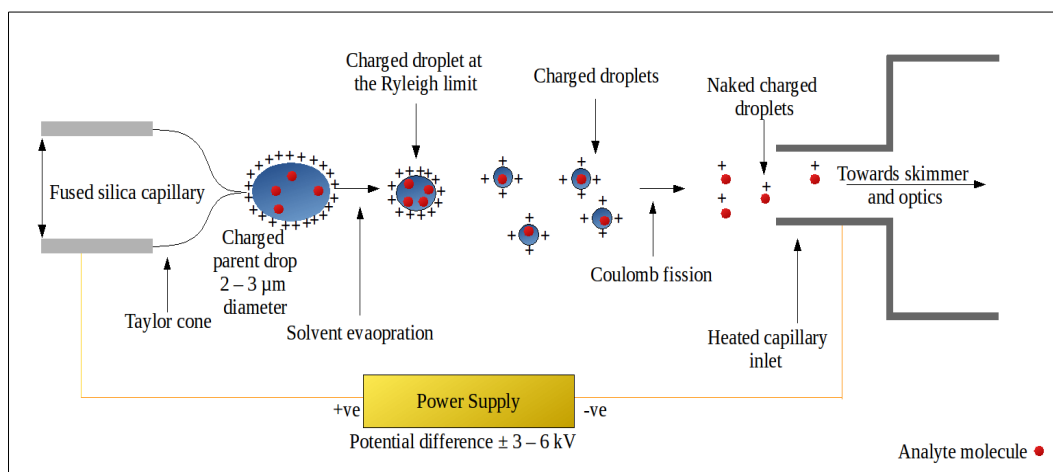


Figure 1-7: Ion formation during the ESI process. According to the polarity of the capillary, either positively or negatively charged analyte ions are obtained. Typically, molecules > 1000 Da are observed as multiple charged ions. Image is adapted by (158,159) and further modified.

### Precursor Ion Transfer in LTQ Orbitrap Velos

The LTQ Orbitrap Velos is one of the most used mass spectrometers worldwide (166). It is a hybrid LC-ESI MS instrument introduced in 2010 by Thermo Fisher Scientific that combines two different mass analyzers, that is the linear quadrupole ion trap and Orbitrap. This instrument was used in particular for MS measurement in this dissertation and some aspects of this instrument will be described in the following section.

To transfer ions from the inlet system via the ion source to the mass analyzers (Figure 1-8), an ion transfer tube is used to pass the barrier of atmospheric pressure (760 Torr, API side) to the stacked ring ion guide (SRIG, also known as S-lens) region with approx. 1 Torr (167,168). The S-lens captures and focuses ions into a tight beam, by applying a radio frequency (RF) which eliminates mass discrimination. The S-lens is made of 18 rings. Each ring diameter is smaller than the previous one in order to capture as many ions as possible, focus them simultaneously and transfer the ion beam into the multipole regime. Three multipoles are used to filter or select all ions to be transferred into the ion trap (High Pressure Cell; HPC): 1st bent square quadrupole (to remove neutral ions increasing detector noise), 2nd square quadrupole (transfer and further focusing) and 3rd round octapole (transfer into ion trap mass analyzer). The High Pressure Cell is used for ion trapping, isolation (if required) and excitation (fragmentation of precursor ions), whereas the Low Pressure Cell (LPC) is used for ion ejection towards the electron multiplier for detection, or towards the Orbitrap part of the hybrid nano-LC-ESI MS instrument.

One unique feature of this hybrid instrument is its capability to isolate and accumulate precursor ions for a specific time frame. This accumulation time is inversely linear to the precursor ion intensity. Its maximal value (in time scale) can be user defined. However, if a sufficient number of ions are isolated, the time regime for trapping is dynamically adjusted. This process is referred to as *Automatic Gain Control (AGC)*. This prevents the Orbitrap mass analyzer to be overfilled with ions, which can cause detector blindness. The AGC also decreases the cycle time for MS/MS events. Moreover, a higher number of low abundant precursor ions of a specific  $m/z$  value can be accumulated for fragmentation. The reproducibility and robustness of tandem MS spectra are



increased due to accumulation of (approximately) the same number of ions for fragmentation.

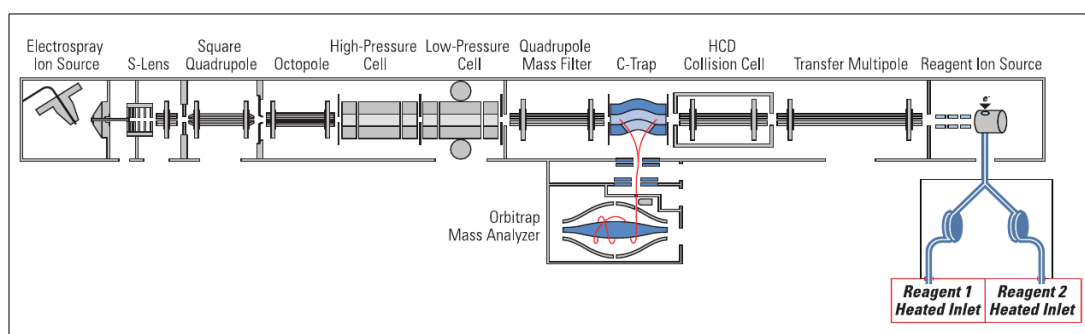


Figure 1-8: Schematic diagram of the LTQ Orbitrap Velos equipped with ETD. ©2009 Thermo Fisher Scientific

## 1.2.2 Peptide Fragmentation Techniques

Identifying peptides and proteins by Peptide-Spectrum-Matching (PSM) has evolved into the technique commonly applied in proteomics. In contrast to Peptide-Mass-Fingerprint (PMF), sequence specific signals upon precursor fragmentation are used to match tandem MS spectra and peptide sequences. To reveal the advantages and disadvantages of each fragmentation techniques commonly applied in proteomics, it is important to understand how peptides tend to fragment during precursor ion excitation.

### 1.2.2.1 Peptide Fragmentation

Ionized peptides in the gas phase, exposed to an external stimulus (mostly by collision with gas molecules), tend to dissociate in a controlled manner by cleaving the peptide between amino acid residues. This cleavage leads to predominantly six ion series (a-, b-, c- and x-, y-, z-ions), which are related to only three possible cleavage points within the peptide backbone; forming three ion-series pairs (a/x, b/y and c/z, respectively; see Figure 1-9; (84,169)). The fragment ions are termed a-, b-, c-ions if the charge remains at the N-terminal site during peptide dissociation. Contrary, if the charge remains at the C-terminal site, the ions are termed x-, y-, z-ions (see Figure 1-10).

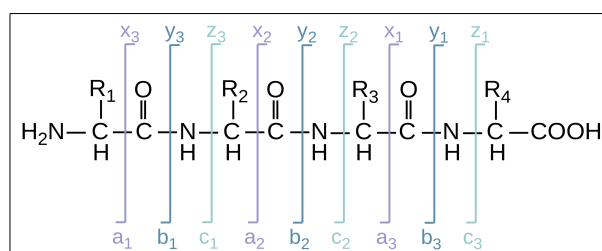


Figure 1-9: Nomenclature to describe peptide backbone fragmentation leading to a-, b-, c- and x-, y-, z-ions depending on the charge remaining N- or C-terminal. Illustration is adapted by (84,169).

The cleaving position in the peptide is indicated by lowercase numbering in right position of the peptide specific ions (e.g., a<sub>3</sub>-ion). If hydrogens are transferred to (or lost from) a fragment ion, this is indicated with an apostrophe in right (or left) position (not shown in Figure 1-10). Additional side chain fragmentation of peptides can be observed, to which the letters d, w and v are assigned (169). Other fragmentation pathways, caused by double backbone cleavage, are referred to as amino acylium, amino immonium and immonium ion, respectively

## 1 Introduction

(170–172). The former is usually a product of b- and y-type ion, the latter of a- and y-type cleavage. If the internal fragment only carries one amino acid, formed by a combination of a- and y-type cleavage, the resulting immonium ion is labeled with its corresponding one letter amino acid code. All or just a subset of the fragment ions described above are used for manual or automated peptide identification (see chapter 1.2.4).

The relative distribution of peptide sequence specific ions (e.g., b-, y- or immonium ions) in tandem MS spectra depends on several factors, such as the primary peptide sequence, the energy introduced while precursor fragmentation or the type of fragmentation applied (see chapter 1.2.2.2).

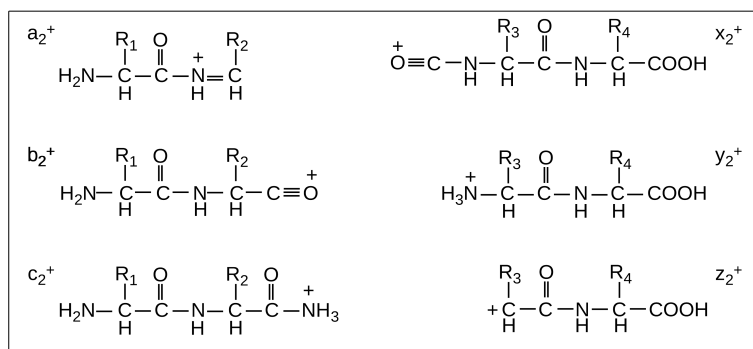


Figure 1-10: Structure of the commonly observed a-, b-, c- and x-, y- and z-ions after peptide fragmentation.

### 1.2.2.2 Fragmentation Techniques

Peptide fragmentation techniques are applied to obtain sequence specific signals which are used to identify peptides by Peptide-Spectrum-Matching. Presently, *Collision Induced Dissociation* (CID; (173)) is one of the most commonly applied fragmentation techniques in bottom-up proteomics.

CID can be further sub-grouped into high- and low-energy CID (HE-CID and LE-CID, respectively). HE-CID applies energy regimes of hundreds to several thousands eV and is more often used in sector field TOF-TOF instruments (166,174). The advantage of HE-CID is to monitor b- and y- beside a-, x- and immonium ions (175). These and the presence of specific indicator ions help to distinguished isobaric amino acids in tandem MS sequence ladders (e.g., leucine and isoleucine; (176)). Low-energy CID (< 200 eV) is the fragmentation techniques predominantly applied in proteomics experiments and can be further sub-grouped into (166,174,177–181):

- Resonant-excitation CID (CID-IT; 1 – 20 eV; performed in ion trap devices)
- Beam-type CID (1 – 200 eV; triple quadrupole (QqQ) or hybrid double quadrupole – Time of Flight instruments (QqTOF)) and
- Pulsed q collision dissociation (PQD; performed in ion trap devices)

Fragmentation of the peptide precursor ion is most commonly obtained by precursor isolation, its acceleration to a higher kinetic energy state allows it to collide with an inert gas (helium, nitrogen or argon; (181)). This inelastic collision converts the kinetic into internal vibrational energy. If the vibrational energy reaches a specific point, covalent bonds might dissociate. However, as the dissociation rate is slower compared to the redistribution of energy, the weakest bond in a peptide is favored to break first (referred to as *low-energy*

*pathway*). This phenomena is especially observed for CID-IT (resonant-excitation).

CID fragmentation of peptides follows the *mobile proton model* (171,182,183) in which a proton attached to a basic side (R or K) is mobilized after activation and transferred to a backbone heteroatom (e.g., terminal amino groups, amide oxygens, carbonyl oxygens, amide nitrogens or side chain groups) leading to charge-directed fragmentation (182,184). As more than one protonated species exists after activation, several position for peptide backbone cleavage are available. This process is monitored as a fragment ion series (sequence tag ladders) in the MS/MS spectra. For certain amino acids (e.g., proline) or modifications (e.g., phosphorylation and oxidation) a significant influence in peptide backbone fragmentation is reported. Beside peptide backbone cleavage, water or ammonia losses are frequently observed as competing fragmentation pathways (181).

If activation and deactivation processes are in competition (time scale based), this is referred to as *slow-heating* and typical for resonant excitation CID (CID-IT; (185)). B- and y-ions are mainly observed for CID-IT, as this represents the lowest energy pathway (CO-N bond cleavage of peptides). Low m/z b- and y-ions cannot be detected due to the low mass cut-off rule of ion traps (175,184,186,187).

As the energy introduced is inversely proportional to the mass of the precursor, less energy is used for conversion to internal energy with increasing mass. This can be overcome by changing the gas mass (using nitrogen instead of helium), or using CID fragmentation techniques applying higher energy such as beam type CID (e.g., in QqQ or QqTOF instruments; (188)). As the kinetic energy is retained, multiple (sequential) fragmentation processes are more likely to be observed by further energetic collisions (166,189,190). The fragmentation process is faster than the CID-IT in circumventing the problem of internal energy distribution to (predominantly) break the weakest bond available in sequence.

The advantage of monitoring the low m/z region led to a further beam-type CID fragmentation technique, *Higher Collision induced Dissociation* (HCD), being available for Orbitrap hybrid MS instruments. During the development of HCD, beam-type like fragmentation has been induced in the C-trap by changing the radio frequency voltages (2,500 instead of 1,500 Volts). This has allowed to retain a maximal number of ions. It has been observed that the energy required for fragmentation scales linearly with m/z values. This has led to a normalized value (*Normalized Collision Energy*, NCE) at m/z 1,000 (190)). HCD fragmentation has been further optimized by the integration of a dedicated octopole collision cell, which is located at the end of C-trap (*HCD Collision Cell* in Figure 1-8). A variety of collision gases (nitrogen, helium or argon) can be used. This increases the flexibility of HCD fragmentation, as different amounts of energy can be introduced. Recently, the limitation of acquiring fragment ions only in the Orbitrap mass analyzer has been overcome, allowing detection of fragment ions also in the low resolution ion traps after HCD activation (191,192). However, this is not the commonly applied method, as fragmentation processes outside the ion trap (octopole collision cell; (188)).

A further CID like fragment technique is the so-called *Pulsed Q Dissociation* (PQD), which overcomes the problem of ion detection in the low m/z region with ion traps (e.g., iTRAQ reporter ions (see chapter 1.1.2)). This is achieved by precursor activation with a high q value (0.6 – 0.8) for a short period of time using high amplitude resonance excitation pulse (~ 100  $\mu$ s). Only ions resonant to this excitation pulse are kinetically excited by absorption of energy. Before significant dissociation processes occur, the high value is kept constant

## 1 Introduction

(~ 100  $\mu$ s) allowing for conversion of kinetic into internal energy. Then, the q value is lowered by dropping RF amplitude, while activated ions undergo fragmentation. Due to the low q values, low mass ions are trapped, as well (177,193).

The pattern of tandem MS spectra of CID activated peptide precursor ions are influenced by several factors, such as amino acid composition, mass and charge state of the precursor ion and the time and energy introduced upon fragmentation. A disadvantage of CID is that labile peptide modifications can be lost. This can be overcome by applying other fragmentation techniques (194): *Electron Capture Dissociation* (ECD) and *Electron Transfer Dissociation* (ETD).

Both techniques allow peptide backbone fragmentation breaking (N-C $_{\alpha}$  bond), which leads to the formation of c- and z-ions (see Figure 1-10; (188)). Due to their nature, ECD and ETD are complementary techniques to CID (195). Recently, it has been reported that the sequence coverage is higher for ECD than for CID-IT (196,197). However, an intense signal originated from charge-reduced precursor ions limits its use, as sequence specific signals are suppressed and cannot (or barely) be used for peptide identification (198).

ECD is predominantly applied in *Fourier Transform Ion Cyclotron Resonance* (FTICR) mass spectrometers. In the initial process of ECD, protonated peptides assimilate an electron. The formed peptide cation then undergoes rearrangement processes and dissociates (Equation 1-1; (175,199)).



ETD is more commonly applied in ion traps. An electron is transferred from a radical anion (e.g., fluoranthene for the LTQ Orbitrap Velos) to the protonated peptide precursor ion. This results in fragmentation along the peptide backbone forming c- and z- ions (Equation 1-2; (199)). The ETD process is most effective for highly charged peptide precursor ions. Similar to ECD, a charge-reduced precursor ion often dominates the tandem MS spectra. To induce further fragmentation, a coupled ETD-CID event can be triggered on the precursor ions, which increases the peptide sequence coverage significantly (200). This combined fragmentation process is referred to as *ETD with supplemental activation* (ETD-sa)



The LTQ Orbitrap Velos (equipped with ETD support) offers the feature to create data-dependent decision-trees to use CID for low charged ions ( $z < 3^+$ ) and ETD for higher charged ions ( $z > 2^+$ ). Although this improves the sequence coverage, the total number of identified peptides does not increase significantly, and is less effective than double measurement using CID only (201,202). Similar results are reported for the combination of CID and HCD (177), although the search algorithms and their settings have an impact on the number of peptides identified (203). This is often not taken into account for fragmentation comparison.

### 1.2.3 Mass Analyzers

To separate and isolate precursor or fragment ions from surrounding ions, two basic classes of mass analyzers exist: (i) scanning and ion-beam devices (*Time-of-Flight* (TOF) and quadrupole (Q)), which are often coupled to the pulsed ionization technique MALDI and (ii) trapping devices, (IT, Orbitrap and FT-ICR) coupled online via continuous ESI (6).

TOF analyzers determine the flight time of ions through a free flight tube. This flight time is proportional to the  $m/z$  values. Principally after ionizing analytes, these are accelerated by an electric voltage to a specific kinetic energy  $E_{kin}$ . If the acceleration voltage and the flight length are known, the  $m/z$  value can be calculated by measuring the flight time:

$$E_{kin} = \frac{1}{2} \times mv^2 = z \times e \times U \quad \text{Equation 1-3}$$

$E_{kin}$  : kinetic energy  
 $m$  : mass of analyte ion  
 $v$  : ion velocity after acceleration  
 $z$  : charge state  
 $e$  : elementary charge  
 $U$  : acceleration voltage

The speed is linearly corrected with the flight length and time, respectively.

$$v = \frac{L}{t} \quad \text{Equation 1-4}$$

$L$  : length of drift tube  
 $t$  : total flight time

By combination, the  $m/z$  value can be calculated as:

$$\frac{m}{z} = \frac{2 \times e \times U \times t^2}{L^2} \quad \text{Equation 1-5}$$

Typically, the flight time ranges from  $\mu\text{s}$  to several 100  $\mu\text{s}$  with flight length from one to four meters (204). One of the main limitations of TOF instruments is the inability to measure tandem MS. Here, techniques like PSD and prompt fragmentation are applied, but the fragmentation and selection capabilities of ions are limited (205). This can be overcome if an other mass analyzer (e.g., quadrupole or TOF) is integrated, which functions as a precursor ion filter. Only ions of a specific  $m/z$  are then subjected to fragmentation. Upon fragmentation, the ions are accelerated towards the TOF analyzer and the drift time of each ion is measured. Often *Multi Channel Plates* (MCP) and *Secondary Emission Multiplier* (SEM) are used for ion detection, respectively (206).

In contrast to other mass analyzers, no increase in speed is observed for tandem MS. The only exception is if a large number of ions are created, transferred and detected. MALDI TOF/TOF creates reproducible spectra, but to obtain a high signal-to-noise ratio, several spectra have to be merged (205).

Quadrupole mass analyzers are build up of four rods with a parallel alignment and a hyperbolic profile. A combination of audio frequency (AC) and direct current (DC) voltage is applied to stabilize ions over the

## 1 Introduction

complete lengths (158). At counterpart rods, the same polarity (DC voltage) and phase (AC) are applied, whereas side-by-side rods have reverse polarity (DC) and a phase (AC), which is shifted by 180 degrees. While scanning, the DC and amplitude of alternating electrical field are increased and ions, passing the rods axially, begin to oscillate. As the ratio of DC and amplitude is kept constant, only ions of a specific  $m/z$  are stabilized; while other ions do not pass the complete length of the quadrupole and are filtered out. This also represents the biggest disadvantage of quadrupoles: the loss of ions generated.

Ion traps are often made up of three-dimensional (3D) quadrupole fields. Ions are trapped in the center and scanned out in two directions. Although the cycle time for trap filling and spectrum generation is short (207) their small volume limits the trapping efficiency of externally generated ions to 1-10% and the overall number of ions that can be trapped. Also, the dynamic range is limited as space-charging effects decrease spectra quality and the mass accuracy and resolution is limited even if only a small  $m/z$  window is monitored.

The development of *Linear Ion Traps* (LITs) has allowed the storage of a up to 10 times more ions compared to ITs (187). This has lead to improved ion statistic and injection efficiency, respectively (205). The injection is performed through an end cap and ejected from it as well (187) towards two detectors allowing to double the current detected. One disadvantage of ion traps operated in classical mode (non PQD, as described previously in chapter 1.2.2.2) is its low mass cut-off at approximately  $1/3^{\text{rd}}$  of the precursors ion mass.

LITs are versatile and fast mass analyzers, have the capability to perform  $MS^n$  experiments (tandem in-time), have high-duty cycles and a high sensitivity. These features make LITs best suited for bottom-up approaches. In hybrid mass spectrometers (e.g., LITs and Orbitrap mass analyzers combined), the linear ion trap is used for trapping, ion selection and ion reactions (6). While tandem MS experiments are performed in the LIT, ions can be further analyzed in the Orbitrap mass analyzer to improve spectra quality (208).

Orbitrap mass analyzers are a further development of the Kindon trap (209). It applies trapping of ions in a electrostatic field by an orbital motion around the spindle, as illustrated by the red arrows in Figure 1-11 (210). By subsequent *Fast Fourier Transformation*, time-domain signals are converted into mass-to-charge spectra.

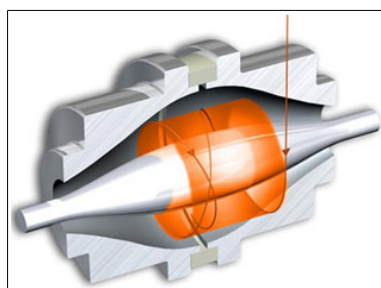


Figure 1-11: Cross section of the Orbitrap mass analyzer. ©2009 Thermo Fisher Scientific

The Orbitrap mass analyzer has some unique features such as high mass resolution (up to 150,000) and accuracy (2 ppm and 5 ppm for internal and external calibration, respectively), an  $m/z$  range of 6,000, a high space-charging, a dynamic range of four order of magnitude (210) and no relevant mass cut-off, as observed for ion trap devices (177,190,211,212).

The LTQ Orbitrap Velos combines a linear ion trap and an Orbitrap (Figure 1-12). To benefit from this

## 1.2 Mass Spectrometry

hybrid configuration, a MS snapshot is performed at a reduced resolution in the Orbitrap mass analyzer ( $R = 7,500$ ) to obtain a list of precursor ions ( $m/z$  value and signal intensity). While the full scan completes in the Orbitrap, precursor ions of interest are isolated, trapped and fragmented in the Low Pressure Cell of the linear ion trap. Afterwards, ions are transferred to the High Pressure Cell and scanned out from low to high  $m/z$  values. This is referred to as the high-low strategy, because MS spectra are recorded at high resolution and MS/MS spectra at low resolution in the LIT with a higher scanning speed and sensitivity compared to the Orbitrap mass analyzer (6).

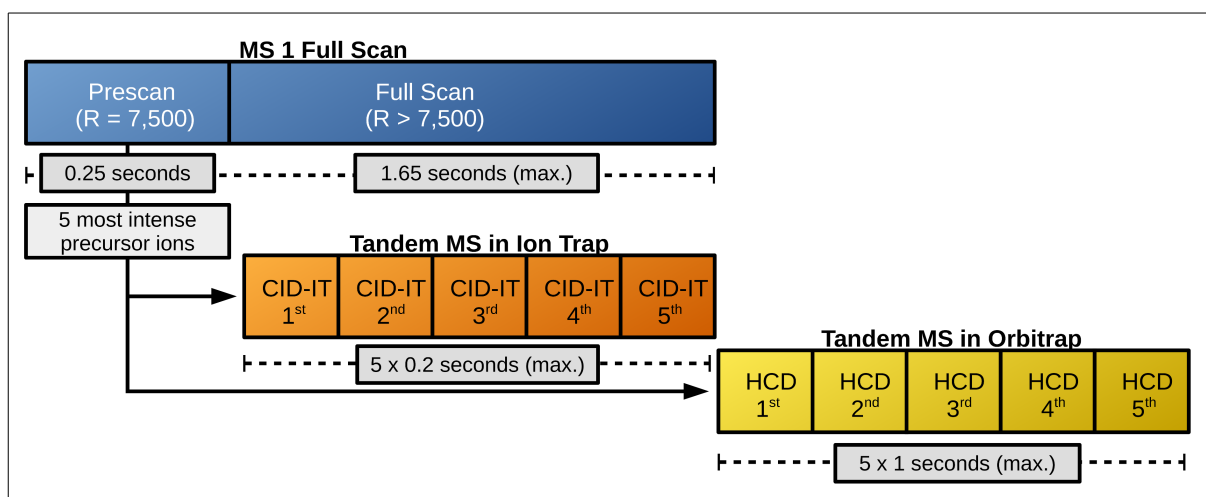


Figure 1-12: Schematic overview of a dual fragmentation method utilizing CID-FT and HCD. After a prescan, the five most intense precursor ions are being fragmented by CID and analyzed in the ion trap mass analyzer. In parallel, the full scan is completed in the Orbitrap. After both processes have finished, the same precursor ions are fragmented using HCD and subsequently analyzed in the Orbitrap mass analyzer.

To prevent overfilling of the linear ion trap and the Orbitrap mass analyzer, the number of ions to be selected for scanning or fragmentation are predicted from the full scan. This method is referred to as *predictive Automatic Gain Control* (pAGC). It is important to limit the number of ions to  $10^6$  (Orbitrap) and  $10^5$  (linear ion trap), respectively, as a higher number of ions causes space-charging effects by which the mass accuracy and resolution decreases (213).

The Orbitrap mass analyzer of the LTQ Orbitrap Velos can be operated in five different resolution modes: 7,500; 15,000; 30,000; 60,000 and 100,000. The different resolutions are obtained by varying the acquisition time for a full or tandem MS scan. The time scale for an MS scan ranges from 0.25 sec ( $R = 7,500$ ) to 1.9 sec ( $R = 100,000$ ). The acquisition of MS full scans with a resolution of 60,000 has become the gold-standard in proteomics and requires approximately 1 sec per scan (210).

The high mass accuracy of the Orbitrap mass analyzer has improved peptide identification upon database searches due to a higher specificity, and hence lowering the number of false positive identifications for the same number of true positives. Identifying modified or unexpected peptides is greatly facilitated due to the higher resolution and mass accuracy to exclude wrong annotations (214).

## 1 Introduction

### 1.2.4 Data-Processing and Automated Data-Analysis

The application of different fragmentation techniques (e.g., CID, HCD, PQD and ETD) generates spectra typically containing a high degree of redundant and or complementary information. Despite this, the various fragmentation techniques produce spectra with highly typical characteristics in terms of the presence and intensities of various sequence specific ions (e.g., a-ions). Manual interpretation and validation of these complex data is a very time consuming and error-prone process. To circumvent this, so called database search engines were developed. SEQUEST (215) and Mascot (216) are commercially distributed search engines, whereas OMSSA (217) and X!Tandem (218) are available as open-source software, which are developed by the National Center for Biotechnology Information (NCBI) and Global Proteome Machine Organization (219), respectively.

In contrast to Peptide-Mass-Fingerprint approaches (see chapter 1.1.1), which only use the precursor mass for identification rather than the MS/MS spectrum (28,220–222), database search algorithms are capable of identifying peptides also from complex peptide digests (shotgun proteomics) due to the incorporation of peptide sequence specific signals (e.g., b- and y-ions; see chapter 1.2.2.2)) of tandem MS spectra (223) .

These fragment ions signals can also be predicted *in-silico*. By supplying a FASTA file, in which the protein sequence is encoded, and using a specific set of features to digestion a protein *in-silico* (e.g., enzyme specificity), a list of peptide candidates including their specific fragment ions can be calculated. The compilation of such a list is a common approach for all database search engines. However, each search engine uses a different algorithm (also referred to as model) to identify and match peptides from tandem MS spectra. These algorithms can be sub-grouped into four groups (223):

- (i) descriptive models use a mechanistic prediction of peptide fragmentation, which is scored against the experimental spectrum.
- (ii) interpretative models are based on the assumption, that sequence ladders are generated. These, in combination to divide the spectra into three regions (unknown, sequence tag and other) are used to compute probabilities or correlation scores.
- (iii) stochastic models use a training set to evaluate the likelihood of measurement and fragmentation for peptide matching.
- (iv) statistical and probability models create relationships between sequences (*in-silico* created) and MS/MS spectra (experimental). Both are used to derive the probability and significance for a peptide identification.

#### 1.2.4.1 Identifying Peptides using the Search Engine SEQUEST

The search engine SEQUEST, a member of the descriptive model group, is the algorithm (predominantly) applied for data analysis in this dissertation. SEQUEST (215) has been designed to identify peptide from low-energy CID spectra (LE-CID; see chapter 1.2.2.2), which generates mainly b- and y-ions. Several pre-steps need to be taken before using this engines. This includes to calculate the corresponding a-, b- and y-fragment ions from the peptide precursor ion.



## 1.2 Mass Spectrometry

The first step that is directly performed by the search engine SEQUEST is to simplify the experimental MS/MS data (m/z values and intensities) by removing background signals. This is performed by eliminating all but the 200 – 500 most intensive signals. Additionally, within a 10 Da window around the precursor, signals are removed to circumvent wrong annotations. A pre-search (Sp-Score calculation) to obtain peptide candidates by linear sequence combination is performed using only b- and y-ions. The masses of b- and y-ions are calculated according to Equation 1-6 and 1-7.

$$b_n = \sum a_n + 1 \quad \text{Equation 1-6}$$

$$y_n = MW - \sum a_n \quad \text{Equation 1-7}$$

$a_n$  : mass of amino acid

$b_n$  : type-b ion

$y_n$  : type-y ion

The number of matched fragments ions ( $n_i$ ) and their intensities ( $i_m$ ) are summed. As peptides fragment ideally randomly by forming sequence letters, this can be interpreted as a *positive hit indicator* ( $\beta$ ), which is increased by 0.075 with each consecutive hit. The presence of immonium ions are scored with the addition of 0.15 points for each immonium ion present ( $\rho$ ). The sequence length (exactly: the number of predicted sequences in terms of b- and y-ions) is also taken into account to compensate for biases introduced.

$$S_p = (\sum i_m) n_i (1 + \beta) (1 + \rho) / n_t \quad \text{Equation 1-8}$$

$S_p$  : score

$n_i$  : number of matched fragment ions

$i_m$  : abundances of matched fragment ions

$\beta$  : component of score (continuity of an ion series)

$\rho$  : presence of immonium ions

$n_t$  : predicted number of sequence ions

For the Top 500 candidates obtained by Sp-Score calculation, a cross correlation is performed between the experimental (normalized) spectrum and the theoretically calculated spectra. To normalize the experimental spectrum, the precursor is removed and the spectrum is divided into 10 equally distributed parts. Within each part, the signals are normalized to a value of 50. The theoretically calculated spectrum is created by integration of b- and y-ions and magnified them to a value of 50. In addition to each b- and y-ion, two signals at  $\pm 1$  Da are inserted, respectively, and magnified to 25. Type a-ions, loss of water or ammonia are magnified to 10. Finally, by *Fourier Transformation* (FT), the cross-correlation is calculated and the obtained scores are normalized to 1.

Principally, the XCorr represents the sum of ions observed for both, the theoretically calculated and experimentally (but modified) MS/MS spectrum. For a detailed explanation of the cross-correlation, its dis- and advantages by using continuous fragment series for score calculation, the phenomenon to receive higher XCorrs for peptides with a higher number of amino acids, the assignment problem of charge states or spectra quality handling problems (background removing), please refer to these publications (224–227); as these topics are outside the scope of this work.

Mascot, X!Tandem or OMSSA are also widely applied database search engines. It is noteworthy, that each

## 1 Introduction

has its own and characteristic scoring system, but results of each search engine have to be validated to distinguish between true and false positive peptide identifications. This process is usually performed by a *False Discovery Rate* (FDR) calculation (228,229).

To calculate a FDR, a *decoy database* is created by reversing or randomly shuffle the original peptide sequences (230). Both databases (decoy & original) are then supplied for peptide identification by database search engines. The FDR validation is based on the assumption that PSMs *from* the decoy database are *false positive* results, and results *from* the original database are only true positive hits. Moreover, the number of false (true) positive PSMs is assumed to be equal to the number of decoy (original) PSMs. The FDR is then calculated by dividing the number of decoy hits (false positives) by the number of target hits (false and true positive hits). A detailed description about the FDR approaches can be found elsewhere (229).

An alternative method to validate peptide hits represent *Semi-Supervised Machine Learning Algorithms* (e.g., Percolator; (231)), which have been shown to increase the number of peptide identifications compared to FDR approaches (231). For PSM validation, they use a higher (but fixed) number of identification features instead of only one feature used in FDR approaches. For SEQUEST, 20 instead of one feature (XCorr) are used for peptide validation by the Percolator. This includes for example the Sp-Score, the precursor mass deviation, the peptide lengths or the fraction of matched by b- and y-ions (231). Further improvements and critical comparisons of the different validation approaches can be found elsewhere (232–236).

### 1.2.4.2 Alternative Peptide Identification Techniques

A newly upcoming approach for database searching is to supply a spectra library instead of a FASTA file. These library files are published by organizations such as *The Global Proteome Machine Organization* and contain spectra that have previously been confidently assigned to a particular peptide sequence.

The advantage of this spectral library database searches is that the shape of a spectrum is known *a priori* and does not need to be *in-silico* predicted, as it is common for FASTA file based database searches (see previous section). Predicting fragmentation patterns (appearance and intensity of sequence specific fragment ions) is generally a difficult process (237). Therefore, matching experimental and library spectra is more accurate and sensitive compared to FASTA file approaches, due to utilization of known intensities of expected fragment ions. Furthermore, the search algorithm seeks for unexpected modifications or amino acid substitutions if the corresponding spectrum is supplied by the library. One such search algorithm is X!Hunter (238). In analogy to Peptide-Spectrum-Matching, the process of identifying peptides from libraries is referred to as *Annotated Spectrum Library* (ASL) pattern matching.

The main disadvantage of spectral library searches is that only peptides which were identified before are integrated into the spectra library and can be searched for. This limits the number of peptides that can be identified. Further, fragmentation pattern vary between instruments due to different technical implementations, which can decrease the level of homology between experimental and library spectra. Another very important disadvantage is that no spectra exist in the library for all kind of static modification, e.g., protection of new N-termini after protein digestion is generally not included or supported. However, N-termini protection or modification is frequently required (60). Therefore, spectra library searches are less flexible than database

searches using FASTA files and cannot replace them entirely (239).

Contrary to database searches, *de-novo* sequencing is usually performed without spectra libraries or FASTA files and used to predict peptide sequences of uncharacterized proteins (240). Abandoning FASTA files or spectral libraries increases the search space and search time as various possible peptide sequences (plus modifications) have to be considered for identification. This can decrease the number of peptides identified, but can also increase the accuracy of PSMs simultaneously (241). For *de-novo* analysis, all potential peptide candidates of a specific precursor ion are calculated, *in-silico* fragmented, compared to the experimental spectra and scored. The scoring and validation process is similar to database search approaches.

The increasing resolution and mass accuracy of mass spectrometers (e.g., mass deviation for precursor ions < 10 ppm) has allowed to narrow the number of potential peptide sequences to be considered, which improves the sensitivity of *de-novo* peptide sequencing due to a limited search space (242). Algorithms performing *de-novo* sequencing are SHERENGA (243), Lutefisk97 (244) or PepNovo+ (245). Recently, a software (DenovoGUI (240)) was published which supports data-input via a front end to analyze data using PepNovo+.

However, *de-novo* peptide identification is not widely used (242). Low resolution and low mass accuracies of MS and, especially, tandem MS spectra hinder an effective identification process. Further, incomplete fragmentation, the plethora of fragmentation techniques which directly influence series directionality and cleavage abnormalities have not paved the way to replace peptide identification by database searches (either by FASTA files or spectra libraries) by *de-novo* sequencing (242,246).

### 1.2.5 Phosphoproteomics by Mass Spectrometry

The analysis of phosphopeptides is a formidable challenge due to its elaborate nature (short lifetime, sub-stoichiometric amounts and the existence of phosphorylation isoforms). Approximately 17% of the total amino acids content in eukaryotes represents serine, threonine and tyrosine residues. These are the main target of phosphorylation events in eukaryotes and it is estimated that 700,000 different potential phosphorylation sites might exist (247). These factors (especially the stoichiometric amounts of phosphopeptides) impacts directly tandem MS strategies which are used to fragment phosphopeptides (e.g., CID-IT and beam-type CID) to identify the sequence, to localize the exact site of phosphorylation and to quantify the phosphorylation, respectively.

#### 1.2.5.1 Phosphopeptide Identification by Mass Spectrometry

The phosphate group of phosphopeptides provides a low-energy pathway upon precursor ion excitation. This leads to a selective and preferred fragmentation at the phosphate group giving rise to non-sequence neutral loss signals, and depleting sequence specific signals (e.g., b- or y-ions) at the same time. This hampers the ability to identify and localize the phosphorylated amino acid exactly.

For pS-peptides, two neutral losses at 98 or 80 Da lower than the precursor are frequently detected in MS/MS spectra. These originate from the loss of  $H_3PO_4$  ([M-P]) and  $HPO_3$  ([M-P<sub>80</sub>]), respectively. These signals are often the major signals in tandem MS spectra of pS-peptides (CID-IT). pT-peptides show a lower extent of neutral loss events compared to pS-peptide. However, these signals can yet dominate tandem MS spectra (181).

## 1 Introduction

The neutral loss events observed for pS- and pT-peptides can be explained by several theories (181). The charge-remote  $\beta$ -elimination reaction (248,249) predicts a transfer of the hydrogen (located at the  $\alpha$ -carbon atom) towards the phosphate group, which allows the phosphoryl-moiety to leave the peptide while forming dehydroalanine (pS) and dehydroaminobutyric acid (pT), respectively. In contrast, the phosphate group is protonated directly in charge-directed fragmentation mechanisms (E2-elimination and  $S_N^2$ -neighboring group participation reactions (250,251)). E2-eliminations are characterized by abstraction of an acidic hydrogen on the  $\alpha$ -carbon by the neighboring carbonyl oxygen after the phosphate group has been protonated. In contrast, for  $S_N^2$ -neighboring reactions the phosphate group is protonated by a (protonated) neighboring amino acid, which induces a nucleophilic attack of the  $\beta$ -carbon by the amide carbonyl group onto the side chain of pS or pT.

However, in experiments to confirm one of the proposed mechanisms, different results have been obtained, so that neither theory can be excluded. Instead, a competition of both mechanisms (charge-remote and charge-directed) is speculated, where the latter seems to dominate (181).

For pY-peptides, the extent of [M-P] neutral loss events is less intense than [M-P<sub>80</sub>]. This is a consequence of the aromatic structure and its influence onto the C-O-P bond. For pY-peptides, the C-O bond is stronger, whereas the O-P bond is weaker compared to pS- and pT-peptides. This favors the loss of [M-P<sub>80</sub>]. So far, the exact mechanism for the neutral loss of HPO<sub>3</sub> for pY-peptides is not understood entirely. It has been shown, that protonated phosphotyrosine peptides do not undergo any neutral loss reaction in the absence of a basic residue (252–254). In special cases, a signal at 98 Da below the precursor ion can also be monitored for pY-peptides. However, this is more likely to incorporate a simultaneous loss of H<sub>2</sub>O alongside [M-P<sub>80</sub>]. An alternative theory suggests the transfer of the HPO<sub>3</sub> group from pY to an aspartic acid residue, which then cleaves H<sub>3</sub>PO<sub>4</sub> (254).

Peptides containing pT, pS, and pY are subjected to a range of neutral loss reactions, the extent of which is dependent on various factors such as (i) amino acid sequence, (ii) precursor ion charge state, (iii) proton mobility, (iv) collision energy applied, (v) and the fragmentation type (181). Additionally, the ratio of charge state versus the number of arginine, lysine and histidine (basic amino acids) influences the neutral loss event. If the total number of basic residues is smaller than the precursor charge state, the neutral loss occurs to a smaller extent (250). This may be explained by the additional proton, which is available favoring alternative fragmentation paths according to the *mobile proton* model. For beam-type CID, the extent of neutral loss events is lower compared to CID-IT (255). This is due to the fact, that peptide ions are activated within a short period of time. This does not allow rearrangement reactions and supports multiple cleavage events (256) as the energy is distributed faster and preventing neutral loss reactions at the same time; whereas slow heating approaches through resonant excitation favors low-energy pathways (CID-IT).

The intense neutral loss signals might compromise the ability to determine the exact site of phosphorylation. This includes the identification of site-determining ions, which are required to validate the correct localized phosphorylation site (251). This process is complicated due to in gas phase re-arrangement processes, in which the phosphate group is transferred to another serine, threonine or tyrosine upon fragmentation (257). It has been shown, that the re-arrangement is facilitated if a protonated arginine is presented and the phosphopeptide is fragmented by CID-IT (258).

## 1.2 Mass Spectrometry

Alternative techniques to determine the exact site of phosphorylations are MS<sup>3</sup> and multistage activation (MSA), which have also been used in *large-scale phosphoproteomics studies* (259–261). For MS<sup>3</sup> experiments, the product ions of the neutral loss events (mainly, [M-P] and [M-P<sub>80</sub>]) are isolated and subjected to excitation, which induces a higher degree of backbone cleavage events (262). MS<sup>3</sup> experiments are usually triggered automatically if neutral loss peaks are observed, to prevent an increase in overall cycle time (181). MSA, also referred to as *pseudo-MS<sup>3</sup>*, combines the advantages of MS/MS and MS<sup>3</sup> spectra (263). After MS/MS fragmentation, a subsequent activation step of the neutral loss fragment ion (e.g., [M-P]) is performed while trapping the remaining ions in the ion trap. Finally, all ions are scanned out, which results in a spectrum lacking the intense neutral loss signal. However, the benefit of MS<sup>3</sup> and MSA is depending on the sample complexity as it lowers the total number of MS/MS spectra that can be acquired in a specific time frame. In contrast to these techniques, high resolution data (for example acquired in Orbitrap mass analyzers) seem to outperform these techniques, as false-positive localizations are more likely to be excluded due to the improved mass accuracy and resolution of sequence specific signals in MS/MS spectra (181).

ETD and ECD are fragmentation techniques preventing the loss of labile groups, such as the phosphomoiety (194,264–267). However, ETD and ECD require a higher cycle time and highly charged precursors ions (density per mass) for efficient fragmentation. These represent conditions under which CID does not undergo the intense neutral loss events, hampering the use of ETD and ECD for phosphopeptide identification (181).

To complicate the situation of phosphopeptide analysis, the software used for automated interpretation of MS/MS spectra influences the identification and determination of the exact site of phosphorylation. For example, some tools can analyze only CID or ETD tandem MS data or favor one of them (268,269).

Search engines such as Mascot, SEQUEST, OMSSA or X!Tandem do not incorporate the mass accuracy of Orbitrap mass analyzers for score calculation. Moreover, none of these combines the complementary information of CID-IT, beam-type CID and ETD, which diminishes the accuracy of identification and phosphorylation site localization. MS-GF(DB) offers the possibility to combine CID and ETD spectra via p-values (269). However, no sequence tag prediction model is implemented which is adopted to the different fragmentation pattern observed for phosphopeptides upon precursor ion fragmentation.

### 1.2.5.2 Phosphopeptide Quantification by Mass Spectrometry

Quantitative proteomics approaches can assess relative changes in phosphorylation such as deregulated signaling in cancer (270). Isobaric labeling (iTRAQ and TMT) are frequently applied in combination to phosphopeptide enrichment and their analysis by high-accuracy mass spectrometers such as the LTQ Orbitrap XL. For this instrument, the collection of CID-IT (fast data acquisition) and HCD (no low mass cut-off) spectra for identification and quantification, respectively, has been reported recently (211). To combine the advantages of both, the low m/z region of HCD spectra have been merged with the CID-IT spectra by *in-house* written software. However, this approach does not take advantage of the high mass accuracy of HCD spectra to identify and localize phosphorylation sites.

## 1 Introduction

### **Further Needs and Obstacles in Quantitative Phosphoproteomics**

To further benefit from hybrid MS instruments, such as the LTQ Orbitrap Velos with its newly designed HCD fragmentation cell, optimized instrumentation and data interpretation methods are required to identify, localize the phosphorylation sites and accurately quantify phosphopeptides and their respective proteins. This includes to elaborate the application of the new HCD fragmentation cells of the LTQ Orbitrap Velos, as its fragmentation cell is now directly coupled to the C-trap and the Orbitrap mass analyzer to improve peptide fragmentation and transferring fragment ions to the Orbitrap (compared to the Orbitrap XL; (271)). Compared to the Orbitrap XL, the new linear ion trap (LIT) of the LTQ Orbitrap Velos uses a high pressure cell to fragment peptides, and a low pressure cell to read-out the resulting ions (271). Acquiring CID fragment spectra in the ion trap requires maximal 0.2 seconds, whereas those generated via HCD fragmentation requires up to one second (Orbitrap mass analyzer; values can be adjusted in Thermo Scientific software *TunePage*). In previous studies using the Orbitrap XL to quantify iTRAQ labeled peptides (63), the prolonged time, which is required to record HCD spectra, was only invested in quantifying peptides by generating intense iTRAQ reporter ions upon precursor fragmentation (at high NCE values of 75%). However, fragmenting phosphopeptides by HCD at well-balanced NCEs to simultaneously provide accurate identification and quantification is an important aspect in phosphoproteomics.

Using optimized fragmentation conditions to identify and quantify phosphorylated peptides is also important, as higher precursor charge states and more precursor charge state heterogeneities are observed upon isobaric labeling (272). Higher charge states can impair phosphopeptide fragmentation events such as the neutral loss of the phospho-moiety and the generation of reporter and sequence specific ions (62,272,273).

Beside optimizing instrumental MS parameters to fragment isobaric labeled phosphopeptides, quantitative phosphoproteomics studies suffer from iTRAQ modification masses, which are not completely isobaric (274); and various tandem MS spectra, which cannot be assigned to peptide sequences (275). Using high quality HCD MS/MS spectra obtained from the newly designed HCD fragmentation cell of the LTQ Orbitrap Velos in combination with sophisticated computational proteomics algorithms could, at least partially, solve these limitations to obtain more information about differentially regulated phosphoproteins.

A major drawback of quantitative phosphoproteomics studies using isobaric labeling such as iTRAQ or TMT is that their quantification relies on only a single peptide sequence; and frequently, only on a single PSM. In addition, differentially regulated phosphorylation events of a phosphorylated protein do not follow to the overall protein ratio (that is usually the median of all unique and quantifiable peptides), which complicates quantitative phosphoproteomics. Therefore, establishing methods to increase the number of quantifiable phosphopeptides is another important aspect in phosphoproteomics.

## 1.3 Aim of this Study

The analysis of protein phosphorylation is still a challenge in mass spectrometry based proteomics due to their low stoichiometry and the lability of the phosphate group upon CID fragmentation. The application of software tools to automatically identify and localize the exact site of phosphorylation events has been studied recently (276). However, the aspect of isobaric labeling to relatively quantify phosphopeptides, and the influence of the isobaric tag on phosphopeptide identification has not been considered in this study (276).

The aim of this dissertation (chapter 2) is to establish a platform to relatively quantify non- and phosphorylated peptides using iTRAQ on the LTQ Orbitrap Velos. For this purpose, different aspects of proteomics will be joined together to optimize experimental and computational processes for identification and quantification.

The LTQ Orbitrap Velos, which will be used throughout this dissertation, offers several techniques to induce precursor ion dissociation, such as CID or HCD. In the first step, the process parameters to efficiently fragment iTRAQ labeled peptides will be optimized. This includes the analysis of specific sequence ions to accurately identify (phospho)peptides. Additionally, the generation of intense iTRAQ reporter ions for quantification has to be ensured. These inverse and conflicting demands will be evaluate systematically by the application of CID and HCD fragmentation at different collision energy regimes (chapter 2.1).

The discrepancy to accurately calculate the *in-silico* precursor mass upon iTRAQ-4Plex labeling has been described recently (274). Its influence on database searches will be evaluated systematically in chapter 2.2 by analyzing the precursor mass accuracy after database searches at different iTRAQ reporter ions ratios. The final goal is to present an alternative algorithm to improve the precursor ion mass calculation of iTRAQ-4Plex labeled peptides.

The decreased number of peptide identifications after database searches is an often criticized aspect of relative quantification using iTRAQ labeling (63). Beside primary amines, iTRAQ can also react with hydroxyl-groups of serine, threonine and tyrosine. To study the influence on peptide fragmentation and peptide separation during reversed phase chromatography, iTRAQ tyrosine modified peptides will be used to identify unique features to validate peptide identifications (chapter 2.3).

To complement the establishment of a platform for relative quantification using iTRAQ, the optimized conditions of the previous sections (chapter 2.1 - 2.3) will be combined into a novel analytical workflow to relatively quantify phosphoproteins using iTRAQ (chapter 2.4). The new strategy focuses onto the parallel use of four proteases to receive a high number of unique and quantifiable (phospho)peptides. This high number of peptides is required due to the reduced ability to identify posttranslational modified peptides (such as phosphopeptides) after iTRAQ labeling (62). Further, quantitative PTM analysis requires stable and reproducible reporter ions ratios, as PTMs usually do not follow the overall quantitative protein profiling (62). An important aspect of this workflow is to ensure compatibility between SDS-PAGE, iTRAQ labeling and phosphopeptide enrichment. The advantages of this quantitative multi-protease strategy will be evaluated to accurately identify and relatively quantify phosphopeptides using a mix of three multi-phosphorylated proteins.

## 1 Introduction

In addition to establish a platform to relatively quantify (phospho)proteins (chapter 2), the development of computational proteomics tools for two side-projects to assist in reproducible and fast data-analysis is another aim of this dissertation (chapter 3).

The aim of the first project of chapter 3 is to develop and establish a computational workflow to combine protein identification results obtained from multi-database searches. The performance of each database search engine and the advantages of combining multi-database searches will be evaluated. At the end of the first project, the need of such a strategy to accurately identify proteins from an 5,300 years old wet-mummy, the Tyrolean iceman *Oetzi*, will be highlighted (chapter 3.1).

The second project of chapter 3 is related to identifying cleavage site motifs of uncharacterized proteases, which is still a challenging task in analytical proteomics. In traditional bottom-up proteomics approaches, only the peptide sequence C-terminal to the cleavage site is analyzed and identified by LC MS. The corresponding peptide sequence N-terminal to the cleavage site can only be extracted from FASTA files containing the entire proteome sequences of the investigated organism. However, manual cleavage site profiling is a time-consuming and error-prone process, which is further complicated by peptide sequence homologous, which have to be evaluate critically. The aim is of the second project of chapter 3 is to compile computational proteomics tools to filter out non-significant peptides sequences, check for sequence homologies and finally to export the relative abundance of each amino acid located next to the cleavage site (chapter 3.2).



## 2 Developing a Platform for relative Quantification using iTRAQ

Quantitative proteomics is a widely applied method of comparing two or more biological states to identify differentially abundant proteins. Beside SILAC (42), which has often been referred to as *gold-standard* for quantitative proteomics, several studies using iTRAQ for (phospho)peptide and (phospho)protein quantification have been published (255,272,273,277). However, a complete shotgun proteomics workflow for relative (phospho)protein quantification using isobaric tags depends on several different techniques, such as (i) sample preparation, (ii) labeling, (iii) LC MS analysis, (iv) data interpretation, (v) data validation and (vi) statistical analysis.

Chapter 2 will cover most of these topics, whereas the focus is set considerably onto LC MS analysis and data interpretation and their validation, starting with the optimization of MS acquisition conditions regarding phosphorylated and non-phosphorylated peptide identification and quantification (chapter 2.1); the refinement of precursor mass calculation upon iTRAQ-4Plex labeling (chapter 2.2); the discovery of unique features to enhance peptide identification of iTRAQ tyrosine labeled peptides (chapter 2.3); and finally, the development and establishment of a new workflow for relative quantification of phosphoproteins using multiple proteases in combination to SDS-PAGE and phosphopeptide enrichment (chapter 2.4). The topic (vi) statistical analysis represents the end point of each relative quantitative shotgun proteomics experiment and is highly depending on the sample source (e.g. whole cell lysate or SDS-PAGE purified proteins), its treatment (e.g., protease used for protein digestion) and experimental plan (e.g., iTRAQ labeling schema including technical and biological replica measurements). Therefore, this topic needs to be adapted to each sample individually, which is out of the scope of this dissertation.

### 2.1 iEM – iTRAQ Evaluation Method<sup>1</sup>

#### 2.1.1 Introduction

For phosphopeptide identification, both ESI and MALDI MS are commonly applied. For peptide fragmentation, CID-IT and beam-type CID have been used for a various of experiments, although CID-IT excitation forms an intense and characteristic neutral loss of  $H_3PO_4$  (observed as  $[M-98/z]^{z+}$  and referred to as  $[M-P]$  in the following). This process hampers backbone fragmentation and complicates fragment ion matching due to their depleted intensity. However, CID-IT is still a widely applied fragmentation method for phosphopeptide identification as it is a sensitive and fast scanning technique (ion trap) improving the total cycle time required for MS and MS/MS measurement.

CID-IT is limited by the  $1/3^{rd}$  rule, not allowing to observe low  $m/z$  fragments. These signals can be a certain class of potentially useful and interesting ions, e.g., immonium ion or iTRAQ reporter ions. Recently, a

---

1 Sections related to phosphopeptide analysis were published in J. Proteome Res. 2013, 12, 2755–2763

## 2 Developing a Platform for relative Quantification using iTRAQ

new beam-type fragmentation technique (HCD) was developed by Thermo Scientific and is nowadays an integral part of Orbitrap hybrid mass spectrometers. HCD does not suffer from the low-mass cutoff limitation, as fragmentation (dedicated collision cell), ion accumulation (C-Trap) and acquisition (Orbitrap mass analyzer) are performed in-space rather than in-time (190). Accordingly, isobaric reporter ions (e.g., iTRAQ-4Plex reporter ions at  $m/z$  114 - 117) can be acquired with a high mass accuracy which allows to discriminate them from other ions with similar  $m/z$  values. As CID-IT and HCD belong to LE-CID, their spectra are comparable, as they predominantly generate b- and y-ions (271).

Several studies have been published in which optimization procedures to identify and quantify phosphopeptide using iTRAQ labeling are described (255,273). It has been shown that both processes (identification and quantification) are highly depending on the precursor ion charge states. For doubly charged precursors, the identification process is demonstrated to be only optimal in a narrow and lower collision energy (CE) regime, whereas iTRAQ reporter ion intensity increases with elevated CE. However, these optimization of process parameters has been performed with the Orbitrap XL, whereas the LTQ Orbitrap Velos (equipped with ETD support) is used for LC-ESI MS analysis through this thesis. One of the primary differences (and advantages) between the Orbitrap XL and the LTQ Orbitrap Velos is that a dual pressure linear ion trap is utilized for the LTQ Orbitrap Velos. This separates the *capturing* and *detection* process of ions, which doubles the number of MS/MS spectra being acquired in the same time frame. Moreover, an integrated C-trap/HCD collision cell facilitates the ion transfer efficiency by directly coupling the HCD collision cell with the C-trap and the Orbitrap mass analyzer (271).

The goal of this study is to optimize fragmentation conditions applied on the LTQ Orbitrap Velos to (i) obtain an optimal sequence coverage for identification, (ii) detect phospho-specific fragment ions, (iii) allow for targeted phosphorylation site mapping using a phosphopeptide standard with *a priori* known phosphorylation sites, (iv) enhance identification results (PSMs and XCorr) of non-phosphorylated peptides and (iv) to generate reporter ions derived from isobaric labeling that fulfill the requirements for accurate quantification. As quantitative phosphoproteomics studies require datasets of both, non- and phosphorylated peptides for phosphoprotein quantification, the optimization process is also performed on two non-phosphorylated samples: the first is a ten protein-mix with approximately 260 unique tryptic peptides. The second sample is a peptide digest of human macrophages.

## 2.1.2 Material and Methods

Chemicals used in this study are summarized in chapter 4.1.

### 2.1.2.1 Sample Preparation

#### Phosphopeptide Standard

A phosphopeptide standard composed of synthetic peptides, containing single peptides in microtiter plate wells (Intavis, Cologne, Germany) was used in this study. Each lyophilized phosphopeptide was resuspended at a concentration of approximately 1.5 pmol/ $\mu$ L in ultrapure water. For online LC-ESI MS measurements, two different sample types were prepared. The first mixture (mix-1), used for in-depth analysis, consisted of 12 phosphopeptides carrying pS-, pT- and pY-residues (Table 2-1). The second mixture (mix-2), used for validation of observations obtained in the in-depth analysis, consisted of a complex mixture of 131 phosphopeptides (Table 2-2). Additionally, a selection of peptides from mix-1 were measured separately in offline mode (see Table 2-1).

*Table 2-1: Peptides used for the semi-complex mix-1 (online measurement) with a tandem CID-MSA (ion trap) and HCD (Orbitrap) method. After labeling, the ratio of the four different iTRAQ channels was 1:1:1:1.3 for 114:115:116:117. Peptides in bold were used for offline measurements with CID, CID-MSA and HCD in Orbitrap.*

Sequence	Phosphorylation Type	Phosphorylation Position	iTRAQ Labeling	Usage
<b>SRNSPLLER</b>	<b>pS</b>	<b>1</b>	<b>114</b>	<b>offline/online</b>
<b>VYELMR</b>	<b>pY</b>	<b>2</b>	<b>115</b>	<b>offline/online</b>
<b>VIEDNEYTAR</b>	<b>pY</b>	<b>7</b>	<b>114</b>	<b>offline/online</b>
<b>STFHAGQLR</b>	<b>pT</b>	<b>2</b>	<b>115</b>	<b>offline/online</b>
<b>ATSLPSLDTPGELR</b>	<b>pS</b>	<b>6</b>	<b>116</b>	<b>offline/online</b>
ADENYYK	pY	6	-	online
YMEDSTYYKASK	pYpY	7&8	-	online
GHLSEGLVTK	pS	4	-	online
GTVTPPPR	pT	3	-	online
VQTTPPPAVQGQK	pT	3	-	online
SRNSPLLER	pS	1	-	online
VYTDEVVTLWYR	pT	3	-	online

## 2 Developing a Platform for relative Quantification using iTRAQ

Table 2-2: Peptides in complex mixture (mix-2) used for online measurements with a tandem CID-MSA (ion trap) and HCD (Orbitrap) method. After labeling, the ratio of the two different iTRAQ channels was 1:0.75 for 114:116.

Sequence	Phosphorylation Site/ Position	Sequence	Phosphorylation Site/ Position
AGGKPSQSPSQEAAGE AVLGAK	pS/ 6	ADENYYK	pY/ 6
ALQKSPGPQR	pS/ 5	DGSLNQSSGYR	pY/ 10
AQSFDPNR	pS/ 3	DIYSTDYYR	pY/ 7
ATSLPSLDTPGELR	pS/ 3	EALPMDTEVYESPYA DPEEIRPK	pY/ 10
ATSPEAGGGGGALK	pS/ 3	EDAANNYAR	pY/ 7
AVGMPSPVSPK	pS/ 9	EPPPVVNYEEDAR	pY/ 8
DKSPSSLLEDAK	pS/ 3	EVGDYGQLHETEVL K	pY/ 5
DSPGIPPSAGAHQLFR	pS/ 8	FSDQAGPAIPTSNSYSK	pY/ 15
DSPGIPPSANAHQLFR	pS/ 2	GHGQPGADAIEKPFYV NVEFHHER	pY/ 14
ESKSSPRPTAEK	pS/ 4	GLPSDYGR	pY/ 6
ETTTSPKKYYLAEK	pS/ 5	GQEYLILEK	pY/ 4
FGESDTENQNNK	pS/ 4	HSWYHGPVSR	pY/ 4
FGSLTMDGGLR	pS/ 3	HTDDEMTGYVATR	pY/ 9
FTNSETAEHIAQGLR	pS/ 4	IKSYSFPK	pY/ 4
GGFFSSFMK	pS/ 6	KYSLTVAVK	pY/ 2
GHLSEGLVTK	pS/ 4	LIEDNEYTAR	pY/ 7
GRGSRDALVSGALEST K	pS/ 4	LSYYEYDFER	pY/ 4
GRRSPSPGNPSGR	pS/ 4	NSFNNAYYVLEGV- PHQLLPPEPPSPAR	pY/ 8
HSIAGIIRSPK	pS/ 9	SISLRYEGR	pY/ 6
ILSDVTHSAVFGVPASK	pS/ 3	TAGTSFMMTPYVVTR	pY/ 11
IPLIKSHNDFVAILDLPE GEHQYK	pS/ 6	TIYVRDPTSNK	pY/ 3
IQPAGNTSPR	pS/ 8	VIEDNEYTAR	pY/ 7
IQPSSPPPNNHNNHLFR	pS/ 4	VKEEGYELPYNPATDD YAVPPPR	pY/ 17
ISSLGSQAMQMER	pS/ 3	VSPSPTYR	pY/ 8
KTSPLNFK	pS/ 3	VYELMR	pY/ 2
LPLTRSHNNFVAILDLP EGEHQYK	pS/ 6	VYHYR	pY/ 2
LQPQEISPPPTANLDR	pS/ 7	WTAPESLAYNK	pY/ 9
LRSADSENALSVQER	pS/ 3	YATPQVIQAPGPR	pY/ 1
LTEERDGLNQSSGYR	pS/ 8	YELTGLPEQDR	pY/ 1
MMSLSQSR	pS/ 5	YIEDEDYYK	pY/ 7
MPSHEAR	pS/ 3	YMEDSTYYK	pY/ 7
NFSAAKSLLNK	pS/ 7	YSLTVAVK	pY/ 1
NGSLKPGSSHR	pS/ 8	YVLDDQYTSSVGSK	pY/ 7
NIDQSEFEGFSFVNSEF LKPEVK	pS/ 11	YVLDDQYTSSGAK	pY/ 7

## 2.1 iEM – iTRAQ Evaluation Method

Table 2-2 (continued): Peptides in complex mixture (mix-2) used for online measurements with a tandem CID-MSA (ion trap) and HCD (Orbitrap) method. After labeling, the ratio of the two different iTRAQ channels was 1:0.75 for 114:116. (continued)

Sequence	Phosphorylation Site/ Position	Sequence	Phosphorylation Site/ Position
RDSPPPPAR	pS/ 3	AYTHQVVTR	pT/ 3
RLSIIGPISR	pS/ 3	ETTTSPKKYYLAEK	pT/ 2
RLSSFVTK	pS/ 3	GTVTPPPR	pT/ 4
RLSSTSLASGHSVR	pS/ 3	LMTGDTYTAHAGAK	pT/ 6
RPGAAASGERDDRGPP ASVAALR	pS/ 7	LQTVHSIPLTINK	pT/ 3
RSMSPFRGPK	pS/ 4	NSQPNRYTNR	pT/ 8
RSPRPDHPGTPPHK	pS/ 2	RLEEPKPKVLTPE EQLADKLR	pT/ 12
SDGGVKPQSNNK	pS/ 9	STFHAGQLR	pT/ 2
SEGSPVLPHEPAK	pS/ 4	STGDPQGVIR	pT/ 2
SESPPPLSDPK	pS/ 3	STLVLHDLK	pT/ 2
SEVAVLSPEKAENDDT YKDDVNHQK	pS/ 7	TAPTPPKR	pT/ 4
SFNGSLKNVAVDELSR	pS/ 1	THLGTGMERSPGAME R	pT/ 1
SGGQRHSPLSQR	pS/ 7	TSPLNFK	pT/ 1
SINEKDYHSR	pS/ 1	TYTHEVVTLWYR	pT/ 3
SLESVLSLGPR	pS/ 1	VGSLTPPSSPK	pT/ 5
SLPAPQDNDFLSR	pS/ 1	VQTTTPPAVQGQK	pT/ 4
SNSTSSMSSGLPEQDR	pS/ 3	VYTHEVVTLWYR	pT/ 3
SPGPSSPKEPLLSR	pS/ 1	ARSRTPPSAPSQR	pSpS/ 3&8
SQERPTFYR	pS/ 1	GISRLESVLSLGRPT GGGSSPPEIR	pSpS/ 3&5
SQSDIFSR	pS/ 1	RLSSTSLASGHSVR	pSpS/ 3&4
SQSNPILGSPFFSHFDG QDSYAAAVR	pS/ 3	SFGSPNRAYTHQVVTR	pSpT/ 4&10
SQSTSEQEK	pS/ 1	TSSFAEPGGGGGGG GGPGGSASGPGGTGG GK	pSpS/ 2&3
SRNSPLLER	pS/ 4	TVSTSSQPEENVDR	pSpS/ 3&5
SRTPPSAPSQR	pS/ 1	IGEGTYGVVYK	pTpY/ 5&6
SSSFREMDGQPER	pS/ 2	NSFNNPAYVLEGV HQLLPPEPPSPAR	pYpY/ 8&9
SSSFREMENQPHK	pS/ 3	PKPSNPIYNPDEPIAF YAMGR	pYpY/ 8&18
SSSPTQYGLTK	pS/ 3	PTTGVLPSGNTLRVK	pTpT/ 2&3
STVAMMMHR	pS/ 1	YIEDEDYKASVTR	pYpY/ 7&8
SVTLPRDLQSTGR	pS/ 1	YMEDSTYYKASK	pYpY/ 7&8
SVTSMERK	pS/ 4	-	-
TGMGSGSAGKEGGPF K	pS/ 5	-	-
TVSTSSQPEENVDR	pS/ 3	-	-
VPASPLPLER	pS/ 4	-	-
VSGRTSPPLDR	pS/ 6	-	-

## 2 Developing a Platform for relative Quantification using iTRAQ

For nano-ESI (offline) measurements, individual phosphopeptides were dried by vacuum evaporation and resolubilized in 0.1% FA. To each sample, a specific iTRAQ reagent (prepared according to manufactures protocol) was added, gently stirred, and incubated for 1 h at room temperature. Samples were dried down, then resolubilized in labeling buffer (172 mM TEAB and 0.172 mM EDTA in 34% ethanol), and desalted via ZipTip C<sub>18</sub>. Cleaned samples were dried by vacuum evaporation and resolubilized in 3% ACN with 0.2% FA. For labeling of the semi-complex mixture (mix-1), peptides were combined and dried before being resolubilized in labeling buffer and split into four equal aliquots. For iTRAQ labeling, each of the four samples was mixed with one of the iTRAQ-4Plex reagents (114, 115, 116, 117), incubated at room temperature for 2 h, combined and then dried by vacuum evaporation. For LC-ESI MS measurements, the sample was reconstituted in loading buffer (3% ACN and 0.1% FA) prior to injection. For preparation of the complex mixture (mix-2), 10 µL of each of the 131 phosphopeptides were combined. The sample was split into equal aliquots and labeled with iTRAQ reagents 114 and 116, respectively. The iTRAQ labeling was performed according to manufactures instructions; subsequently, samples were dried by vacuum evaporation and resolubilized in loading buffer prior to LC-ESI MS injection and analysis.

### Ten protein-mix

The ten protein-mix contained the following proteins: cytochrome c, bovine serum albumin, carboanhydrase, β-casein, lysozyme, myoglobin, ribonuclease a, transferrin, α-Lactoglobulin and β-lactoglobulin (see Table 4-1 for further details).

Of each protein, 0.1 mg was used to prepare a protein stock solution (diluted in H<sub>2</sub>O), and 200 µg of the stock were dried by vacuum evaporation (speedvac, 30 min, 45°C). Samples were resolubilized in 40 µL iTRAQ-4plex dissolution buffer (supplied in the iTRAQ-4plex kit by AB SCIEX), reduced (1 h at 60°C with 2 µL of Reducing Agent supplied by AB SCIEX) and alkylated with IAA (200 mM IAA, room temperature, 30 min, dark). To prevent trypsin inhibition, the sample volume was increased with a further addition of 12 µL iTRAQ-4plex dissolution buffer. Proteins were digested in-solution using trypsin (4 µg dissolved in 8 µL iTRAQ-4plex dissolution buffer) overnight at 37°C under shaking.

For iTRAQ labeling, a new buffer was prepared (denoted as iTRAQ-4plex labeling buffer). It was composed of 160 µL H<sub>2</sub>O, 160 µL iTRAQ-4plex dissolution buffer and 80 µL ethanol (supplied in the iTRAQ-4plex kit). The digested sample was dried by vacuum evaporation and filled up with 200 µL of the freshly prepared iTRAQ-4plex labeling buffer resulting in a concentration of 1 µg/µL. To ensure complete rehydration, the samples were mixed for 15 min at room temperature. Meanwhile, the vials containing the iTRAQ-4plex labeling reagents were allowed to reach room temperature (5 min) and filled up with 70 µL of ethanol.

The digested and rehydrated protein sample was split into four aliquots and 25 µL of iTRAQ-4plex reagent was added to each. Labeling reaction took place for 1 h at room temperature followed by 1 h at 4°C. To obtain an optimal iTRAQ-4Plex labeling, the samples were stored at -20°C for 24 h prior use. Afterwards, the four iTRAQ-4plex labeled samples were dried by vacuum evaporation and each sample was filled up with 100 µL of 3% ACN and 0.1% TFA in deionized water resulting in a concentration of 0.25 µg/µL. For LC-ESI MS analysis, the samples were combined in a 1:1:1:1 ratio and 1.5 µL of sample was injected.

### **Biological Sample**

The biological sample was a composition of three human macrophage samples, which were supplied by the group of Norbert Reiling, Departement Mikrobielle Grenzflächenbiologie, Forschungszentrum Borstel, Germany. Two of these samples were infected with different *Mycobacterium tuberculosis* strains (2335 & 7761), whereas the third sample was not infected with a *M. tuberculosis* strain. After cell harvesting, all samples were stored at -80°C prior use. To disrupt cells and extract proteins, the samples were treated five times with freezing - ultra sonication intervals (freezing: at -80°C for 5 min; ultra sonication: 1 min intervals). Additionally, the samples were once heated up to 95°C for 5 min, followed by a further ultra sonication step (1 minute). To pellet the cell debris, samples were centrifuged (24,000 rpm, 15 min), and the supernatant of each sample was transferred to new Eppendorf tubes. Ice-cold (-20°C) acetone was added (six times the sample volume) to precipitate proteins (overnight, -20°C).

To remove acetone, samples were centrifuged for 45 min at 14,000 rpm (Thermo Scientific Heraeus Fesco 21). The obtained protein pellet of each biological sample was redissolved in 100 µL iTRAQ-4Plex buffer. After protein concentration determination using the Bradford-assay, 50 µg of each *Mycobacterium tuberculosis* infected sample was used for further processing. The control sample was used twice (control I & control II; each 50 µg).

Samples were dried by vacuum evaporation, reconstituted in 20 µL iTRAQ buffer and reduced with 2 µL reducing agent for 1 h at 60°C. After sample cooling to room temperature, 2 µL of alkylation agent were added (200 mM IAA, room temperature, 30 min, dark). Proteins were digested with trypsin (2 µg dissolved in 10 µL iTRAQ-4plex dissolution buffer) overnight at 37°C under shaking. Samples were labeled using an entire iTRAQ-4Plex reagent vial for each of the four samples according to manufacture protocol. The iTRAQ labeling schema is illustrated in Table 2-3. iTRAQ-4Plex reagent was allowed to react for 1h at room temperature, followed a further incubation at 4°C. Samples were stored at -20°C overnight to allow hydrolysis of remaining iTRAQ-4Plex reagent prior usage.

Table 2-3: Summary of the iTRAQ labeling applied for the biological sample.

<b>Sample</b>	<b>iTRAQ Reagent</b>
7761	ITRAQ-114
control I	ITRAQ-115
2336	ITRAQ-116
control II	ITRAQ-117

## 2 Developing a Platform for relative Quantification using iTRAQ

### 2.1.2.2 Mass Spectrometry

#### **Phosphopeptide Standard**

NanoLC-ESI MS experiments were performed on a Dionex U3000 (Dionex, Idstein, Germany) nano HPLC system coupled online to a LTQ Orbitrap Velos MS equipped with ETD (ThermoFisher Scientific, Bremen, Germany). Solvents used for LC separation at 300 nL/min were: solvent A: 0.05% FA, solvent B: 80% acetonitrile (ACN) and 0.04% FA. For sample loading and desalting on a Acclaim Pepmap C<sub>18</sub> 300 μm x 10 mm, 5 μm, 100 Å trap column, a solution of 3% ACN and 0.1% TFA was used (flowrate 30 μL/min). The semi-complex sample (mix-1) was separated using a linear 46 min gradient from 5% B to 50% B on a AcclaimPepMap 100 analytical column (3 μm, 75 μm x 150 mm, Dionex, Sunnyvale, CA), followed by a column wash step (10 min) with a linear gradient from 50% B to 95% B. Inter-run equilibration of column was set to 5 min. A tandem CID/HCD method was written using XCalibur (V 2.1): first, a full scan at a resolution of 30,000 with activated prescan was performed. In parallel, the 3 most intense precursors were selected for fragmentation using CID in the ion trap. CID parameters were set to default (35% NCE), with the exception that multistage activation was enabled. For HCD, resolution of Orbitrap measurements was set to 7,500. The activation time was 0.1 ms, default charge was set to 2<sup>+</sup>, wideband activation was disabled. Dynamic exclusion was enabled for all runs. The NCE was varied from 0 to 85% NCE in 5% increments for HCD resulting in a total of 18 LC runs. Each run was performed twice at each NCE value, however, in order to maximize the number of selected precursors two dynamic exclusion list settings were used. In the first run, the repeat count was set to 2 with a repeat duration of 45 sec, with precursors selected for MS/MS set for dynamic exclusion for 60 sec. In the second run, the repeat count was set to 3 with a repeat duration of 30 sec and precursors, selected for MS/MS, were set for dynamic exclusion for 30 sec.

For analysis of the complex mixture (mix-2) via HCD and CID, the NCE was varied between 0 and 85% NCE in 5% steps. The NCE of HCD was 30% higher than for CID until reaching a value of 55%; above this value, HCD increased from zero in 5% steps simultaneously with the increasing CID-NCE. Samples were separated using a linear 80 min gradient from 5% B to 60% B on a AcclaimPepMap 100 analytical column (3 μm, 75 μm x 150 mm, Dionex, Sunnyvale, CA) for the complex sample, followed by a column wash (up to 95% B) and equilibration. For offline nano-ESI MS measurements, a top 1 method was applied in which the same precursor was fragmented with CID, CID plus activated MSA (CID-MSA) and HCD with 0-85% NCE in 5% steps. The repeat count was set to 1 with a dynamic exclusion duration of 5 min. This ensured all precursors with a charge state higher than 1<sup>+</sup> and an intensity >500 counts were selected for fragmentation. The measurement was stopped manually after the precursor of interest was selected for NCE fragmentation series.



### **Ten protein-mix**

The same LC-ESI MS setup as for the phosphopeptide samples was used for the ten protein-mix. Samples were separated using a linear gradient from 5% B to 50% B in 51 min, followed by a column wash step (10 min) with a linear gradient from 50% B to 95% B. Inter run equilibration of the columns was set to 9 minutes. The HPLC flow rates were 30  $\mu$ L/min and 300 nL/min for the loading- and micro pump, respectively.

A tandem CID/HCD method was written using vendors software: first, a full scan at 60,000 resolution with activated prescan was performed. In parallel, the 5 most intense precursor ions were selected for fragmentation using CID in ion trap. CID parameters were set to default. For HCD, resolution of Orbitrap measurements was set to 7,500. The activation time was 0.1 ms, default charge was set to 2<sup>+</sup>, wideband activation was disabled. Dynamic exclusion was enabled for all runs. The repeat count was set to 3 with a repeat duration of 20 sec and precursors, selected for MS/MS, were set for dynamic exclusion for 30 sec. The NCE value was varied from 0 – 85% NCE in 5% increments for HCD and CID resulting in a total of 2 x 18 LC runs (2 technical injections of 1  $\mu$ L)

### **Biological Sample**

The same LC-ESI MS settings as for the 10 protein-mix were used, except that the LC gradient time was prolonged (linear increase of 5% B to 45% B in 146 min, followed by a column wash step for 13 min with a linear gradient from 45% B to 95% B). Inter run equilibration of the columns was set to 15 minutes.

### **2.1.2.3 Database Searches and Feature-Detection**

#### **Phosphopeptide Standard**

For database searches, Proteome Discoverer 1.2 (Thermo-Fisher, San Jose, CA) and the search algorithm SEQUEST was used. N-terminal and lysine- $\epsilon$ -amino iTRAQ were set as static modifications, and iTRAQ-4Plex labeling at tyrosine, phosphorylation of tyrosine, threonine and serine, and oxidation of methionine, were set as variable modifications. The precursor mass tolerance was set to 10 ppm with 0.02 Da fragment mass tolerance for MS/MS measured via Orbitrap and 0.5 Da fragment mass tolerance measured via ion trap. The database that was used contained all phosphopeptides from the phosphopeptide standard.

To evaluate the dependence of normalized collision energy on the XCorr, no FDR filtering was used for the offline measured samples. However, FDR filtering was applied for the semi- and complex samples (mix-1 and mix-2, respectively) based on a 1% FDR and search engine rank equal to 1.

In the first step of optimizing the process parameters, the offline measured samples were used. Here, SEQUESTs XCorr was used to evaluate identification success across the NCE range for the fragmentation methods applied. Moreover, the m/z values and intensities of each peptide, dependent on NCE and fragmentation method applied, were exported from XCalibur (Version 2.1) to Excel and processed using *in-house* VBA scripts for Excel (2010, 64 bit version). The script scans each spectra exported for the precursor ion (M + nH)<sup>nt</sup> (abbreviation: [M]), the neutral loss fragments ([M-P] and [M-P<sub>80</sub>]), the pY immonium ion, the iTRAQ reporter

## 2 Developing a Platform for relative Quantification using iTRAQ

ions at m/z 114 - 117, and the precursor fragment ions in charged reduced form ( $n - 1$ ) resulting from loss of iTRAQ. If a match (mass tolerance set to 0.01 Da) between the theoretical and experimental m/z value was obtained, the corresponding intensity was set in relation to the most intense fragment ion present in spectra Equation 2-1. These values were then used to generate the plots shown (Figure 2-5).

$$\text{Normalized Intensity (x)} = \frac{\text{Intensity (x)}[\text{Counts}]}{\text{Base Peak}[\text{Counts}]} \quad \text{Equation 2-1}$$

Moreover, ions of interest (x) were normalized (Equation 2-2) against cumulative counts for all reported ions within each spectrum (inter alia Figure 2-11).

$$\text{Relative Intensity (x)} = \frac{\text{Intensity (x)}[\text{Counts}]}{\sum_{y=100}^{y=2000} \text{Intensity (y)}[\text{Counts}]} \quad \text{Equation 2-2}$$

A computer-assisted approach using an *in-house* VBA script for the semi-complex sample (mix-1) was applied. Here, the m/z and ion intensity counts were exported from .mgf files created by Proteome Discoverer 1.2 (signal-to-noise filter was set to 0). To be independent of the identifications reported by SEQUEST for higher NCE values, the following criteria (information obtained by scan header in .mgf file) had to be fulfilled to recognize a precursor selected from MS/MS as a hit: precursor mass deviation  $\leq 10$  ppm, correct precursor charge state, and retention time deviation  $< 2.5$  min around individually predefined values (obtained by averaging the retention time and precursor mass across all LC runs). If all criteria were met, information for the fragment ions being searched for was exported to separate text files and then plotted (Figure 2-8, 2-10 & 2-11).

The approach described above was used to process the data of the complex sample (mix-2), except that the precursor fragment ions in charged reduced form ( $n - 1$ ) resulting from loss of iTRAQ was not monitored. However, due to the higher number of peptides present (about 20 times higher), this data set was used to correlate the identification success by SEQUEST against peptide identification and correct phosphorylation site mapping. Therefore, peptide identifications were filtered (FDR  $< 1\%$  and Rank = 1) and exported. This data set was further split into correctly and incorrectly mapped phosphorylation sites. Further, these sub- data sets were split into activation type (CID or HCD) and charge state ( $2^+$ ,  $3^+$ ). Afterwards, for each NCE value the corresponding number of peptide hits (having the same modification form) was summed and used to generate the plot of correctly and incorrectly identified modifications (Figure 2-7). Furthermore, the normalized intensities obtained from .mgf files for the  $2^+$  and  $3^+$  charged precursors for pS-, pT- and pY- peptides were grouped and the average values for the identified peptides calculated. The dependency of the normalized intensity to NCE was plotted for the precursor, neutral loss ([M-P] and [M-P<sub>80</sub>]), pY immonium ion, non-dissociated iTRAQ and the iTRAQ reporter ions 114 and 116 Figure 2-13).

## Ten protein-mix

For database search using Proteome Discoverer 1.4, a custom build workflow was utilized, which is shown in Figure 2-1.

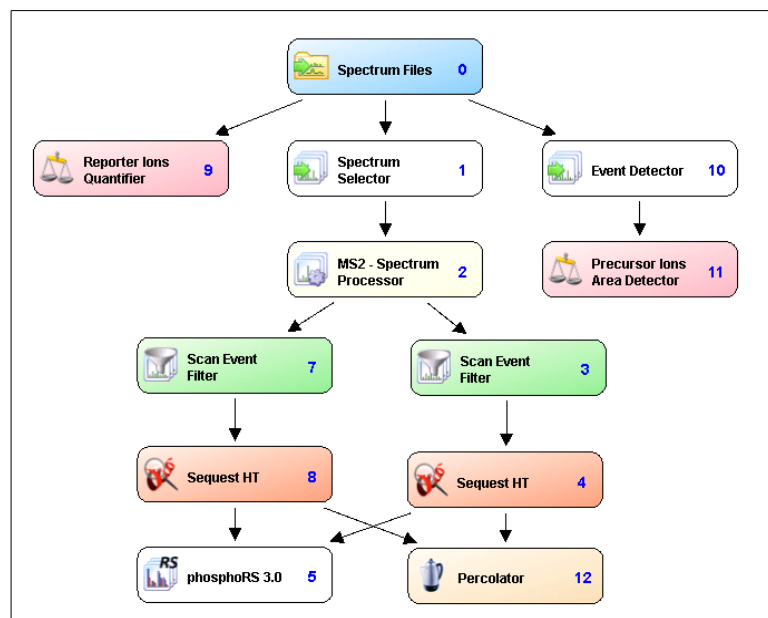


Figure 2-1: Workflow utilized for database search in Proteome Discoverer. The default setting of each node (e.g., Spectrum Files) was used if not otherwise specified: (0) Spectrum Files to load in the .raw files; (1) Spectrum Selector (pre-slicing and filtering of MS and MS/MS data); (2) MS2-Spectrum Processor (deisotope and deconvolute), (3) & (7) Scan Event Filter (to split into CID and HCD fragmentation data); (4) & (8) SEQUEST HT (search algorithm settings, see below); (5) phosphoRS 3.0 (an algorithm to assign phosphorylation sites); (9) Reporter Ions Quantifier (iTRAQ-4Plex (Thermo Scientific Instruments; Integration Tolerance 20 ppm; Most Confident Centroid as Integration Method; Additional Settings: Show the Raw Quan Values; Apply Value Corrections, use All Peptides enabled and deactivated Experimental Bias); (10) Event Detector (filter for precursor ion settings); (11) Precursor Ions Area Detector (reports XICs of identified peptides); and (12) Percolator (for peptide validation and filtering).

For database searches using the search engine SEQUEST HT, the reference proteome (see UniProt for further description) of *Escherichia coli* (without isoforms, downloaded December 2013) plus cRAP like proteins (see chapter 4.3), the 180 phosphopeptides of the phosphopeptide-mix and the proteins from the ten protein-mix were merged into an *in-house* build FASTA file (referred to as bF-FASTA; see chapter 2.4.2.3), with a total number of 4,526 proteins. Enzyme specificity was set to trypsin (full), with a minimum and maximum peptide length of 4 and 144, respectively. Precursor mass tolerance was set to 10 ppm, 0.3 Da for CID and 0.02 Da for HCD fragmentation ions. The maximum number of dynamic and equal modifications per peptide was limited to 12 and 8, respectively. Phosphorylation (S, T & Y) and oxidation (M) were set as variable modifications. Static modifications included iTRAQ-4Plex (peptide N-terminus and K), and carbamidomethylation (C). PSMs were filtered for high confidence (1% FDR) and exported as .xlsx file.

## 2 Developing a Platform for relative Quantification using iTRAQ

### *Feature-Detection for Identification*

The exported PSMs were processed using *in-house* build VBA (Excel 2013, 64 bit and GAMBAS Version 3.5.3) scripts. At first, the XCorr, the number unique and total peptides (same m/z value) were extracted. For peptides with equal m/z values ( $\Delta_{\max} = 0.1$  Da) and sequence, the corresponding XCorr for each NCE value (0 to 85%, 5% increments), precursor charge state and fragmentation type (CID, HCD) were averaged. Additionally, the corresponding number of PSMs was summed up for each peptide, fragmentation type and NCE value, respectively. The intermediate results were exported to a .txt file and further processed with GAMBAS (Version 3.5.3). In order to count the number of unique and totally identified peptides (PSMs), a master-list was created including each identified (unique) peptide (m/z, sequence and charge state dependent). According to this list, the features for each NCE value (if available or as blank if no match was found) was added to the master-list and exported as .txt files. For each charge state and fragmentation type, a separate list was created, which was simultaneously utilized as an input-file being fully compatible with QtiPlot (ProIndep Serv S.r.l., Version 0.9.8.9 svn 2288). Finally, the number of unique (x10 for sophisticated visualization) and total peptides (PSMs) were plotted (right y-axis) versus the applied NCE. Additionally, a Whiskers-Box-Plot with 25, 50 and 75% (box); max (upper *cross*) and min (lower *cross*); mean (*rectangle*); and 5 and 95% (*vertical lines*) was created for the XCorr (y-axis) as a function of the NCE value (see Figure 2-14).

### *Feature-Detection for Quantification*

For quantification, the same VBA workflow as applied for the phosphopeptide-mix was used to extract the iTRAQ reporter ion raw, normalized and relative intensities. The quantification features (normalized-, relative and absolute intensity) of each NCE value and each peptide groups (same m/z charge state and NCE, only HCD) were merged and plotted in a Whiskers-Box-Plot (Figure 2-16). Additionally, the median of the raw iTRAQ intensity of each charge state (2 – 4<sup>+</sup>) was calculated and plotted versus the NCE value applied (Figure 2-21).

To determine the average length of peptides identified as a function of NCE and charge state (2 – 4<sup>+</sup>; Figure 2-15), the number of PSMs for each unique peptide (HCD only and NCE value, respectively) was divided by the total number of PSMs observed (HCD and each NCE value, respectively). This value was multiplied with the length (amino acid number) of the monitored peptide (Equation 2-3). To get the averaged peptide length over all peptides, the sum of all peptides lengths was calculated (Equation 2-4).

$$\text{Peptide Length}(x) = \frac{\text{PSMs}(x)}{\sum \text{PSMs}(x)} \times \text{AA} \quad \text{Equation 2-3}$$

$$\text{Peptide Length} = \sum \text{Peptide Length}(x) \quad \text{Equation 2-4}$$

x : Peptide from the master-list  
AA: Number of amino acids for peptide (x)

### **Biological Sample**

The same database settings of the ten protein-mix were used for the biological sample (macrophages with and without infection of *Mycobacterium tuberculosis*). Solely, the FASTA file was exchanged by an *in-house* merged FASTA file of human and *M. tuberculosis* FASTA files. For both, the reference proteomes including protein isomers were downloaded from UniProt on 2013.02.11. The *taxon identifier* was 9606 and 1773 for human and *M. tuberculosis*, respectively.

Variable modification of phosphorylation (S, T, Y) was deactivated, as no phosphopeptide enrichment was preformed. Deamidation of asparagine and glutamine was activated, as these were common modification for *M. tuberculosis* infected human macrophages (278,279).

### *Feature-Detection for Identification*

Firstly, peptides retention times were averaged based on identical peptide features (sequence, modification, m/z, charge state) to create a master file containing (averaged) information about all peptides identified. This list was compared with identified peptides of each raw file (CID and HCD identifications were treated as identical). For every PSM, the averaged retention time from the master list and the retention time of each raw file were exported. Both values were plotted to check for retention time shifts.

For all files, a linear re-calibration was performed to re-calibrate the retention times. In order to re-calculate the retention times in the raw files, the raw files (binary) were exported into (text-based) .mgf files for CID and HCD activation, respectively. Each retention time within the .mgf file was read in, re-calibrated using its specific linear re-calculation term and exported as a new .mgf file. The re-calibration approach will be described in detail in the results chapter 2.1.3.3.

The re-calibrated .mgf files were finally used for database searches applying the same workflow as before. However, as CID and HCD .mgf files were used as input files, in which the information about the fragmentation type is not included, the database searches were run separately for CID and HCD, respectively. The .msf files of Proteome Discoverer were opened together. Finally, the PSMs were filtered (1%FDR) and exported to .txt files. These were processed as described in the ten protein-mix section.

### *Feature-Detection for Quantification*

The peptide identifications and .mgf files after re-calculation were used for processing. The same workflow as for the ten protein-mix was applied.

## 2 Developing a Platform for relative Quantification using iTRAQ

### 2.1.3 Results

The three different samples (phosphopeptide standard, ten protein-mix and biological sample) were independently analyzed to extract identification and quantification features.

For the phosphopeptide standard, the detection of phosphorylation specific signals (e.g., neutral loss) in dependence of the NCE was an important aspect for peptide identification. Additionally, the intensity of iTRAQ reporter ions was linked to the collision energy to determine the optimal NCE trade off to identify and quantify phosphopeptides.

The ten protein-mix was used to establish and validate the newly *in-house* compiled tools. Moreover, it was used to characterize the number of unique peptides, PSMs and the XCorr distribution via different collision energies applied for non-phosphorylated peptides.

The biological sample (infected human macrophages) was a highly complex sample in terms of MS/MS spectra recorded per LC-ESI MS run, including a variety of different precursor ion intensities, peptide sequence lengths and amino acid compositions. This complex sample was used to validate the previously findings of the phosphopeptide- and ten protein-mix.

#### 2.1.3.1 Phosphopeptide Standard

For simultaneous identification and quantification of phosphopeptides using isobaric labeling approaches, different fragmentation events have to be induced: (i) the fragmentation of peptide backbone to form sequence specific (e.g., b- and y-ions) and phospho-specific ions (e.g., pY immonium ions at  $m/z$  216.04 and low intensity neutral loss derived signals), and (ii) the formation of isobaric labeling derived reporter ions with sufficient ion statistics to guarantee accurate quantification results. These events undergo different fragmentation mechanisms, which, together with the different energies of the cleaved chemical bonds, lead to different energy requirements. The aim was to determine at which Normalized Collision Energies the different events occur and to find optimal MS/MS conditions to satisfy the needs of all events. For this purpose, nano-ESI and LC-ESI MS experiments using CID, CID plus multistage activation (CID-MSA), and HCD were performed using iTRAQ labeled synthetic phosphopeptides. The focus was set to precursor charge states  $2^+$  and  $3^+$  as these ions contributed to at least 80% of all spectra recorded in the phosphopeptide data sets.

#### Influence of NCE on Peptide Sequence Assignment

In the first step of optimizing and analyzing the NCEs required for optimal phosphopeptide identification, three different experimental setups were tested: CID, CID-MSA and HCD. For this purpose, five iTRAQ labeled phosphopeptides from mix-1 were individually analyzed by offline nano-ESI-Orbitrap MS with NCE values ranging from 0 to 85%. Application of high resolution and high mass accuracy Orbitrap MS to acquire peptide MS/MS spectra derived from CID, CID-MSA, and HCD enabled to compare directly the fragmentation patterns among three different fragmentation types; this overcomes limitations accompanied with the use of low resolution ion traps.

## 2.1 iEM – iTRAQ Evaluation Method

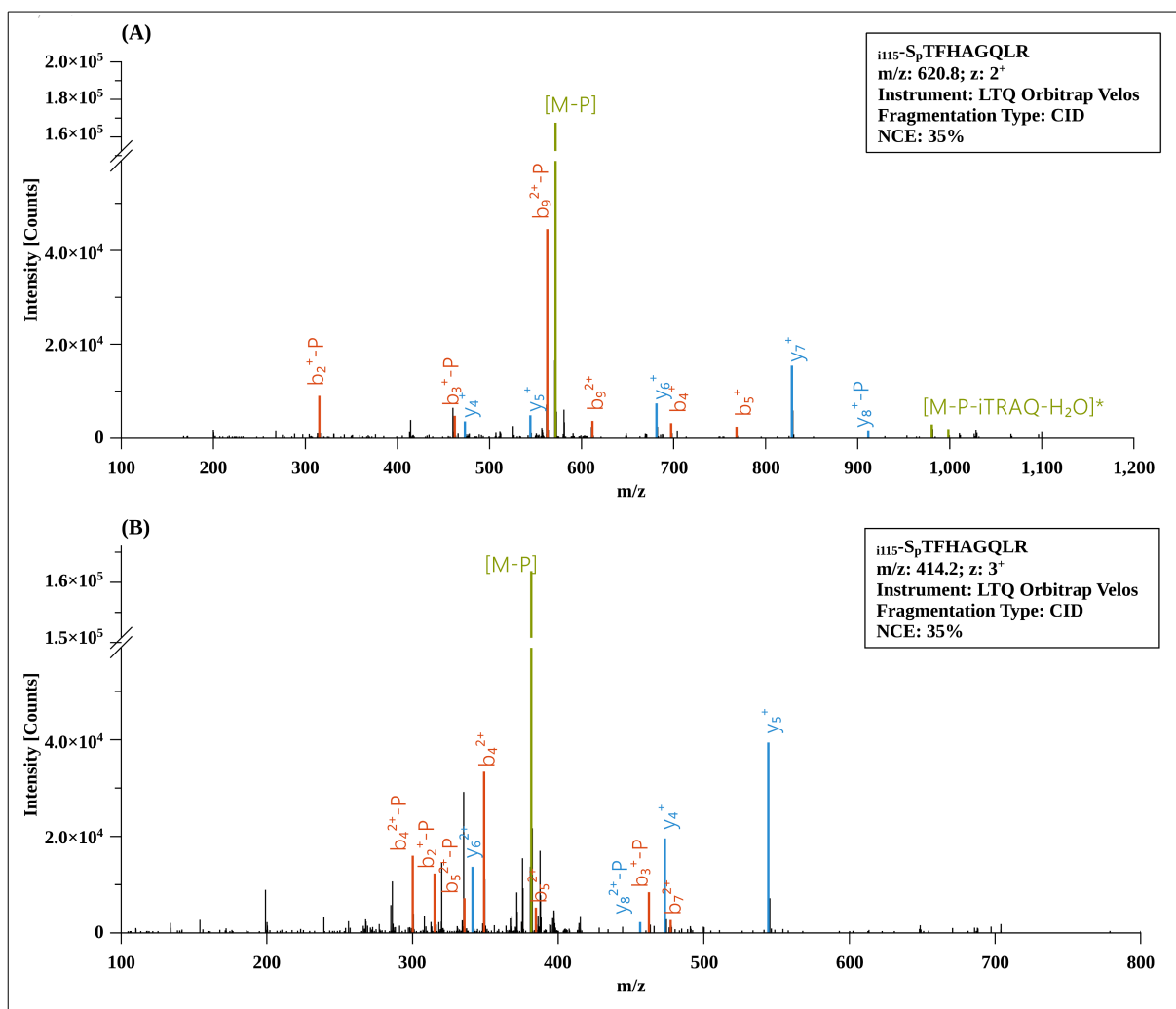


Figure 2-2: Offline MS/MS measurements in Orbitrap of the iTRAQ labeled phosphopeptide  $i_{115}\text{-S}_p\text{TFHAGQLR}$ . (A) CID, 2<sup>+</sup> charged precursor at  $m/z$  620.8; (B) CID, 3<sup>+</sup>-charged precursor at  $m/z$  414.2. Fragments labeled with an asterisk (\*) have lost a charge. The fragment ion  $b_9^{2+}\text{-P}$  shares the same mass as the signal derived from the precursor including a neutral loss of the phospho-moiety [M-P] and a subsequent loss of water (18/z).

The 2<sup>+</sup> and 3<sup>+</sup> charged precursors of peptide  $i_{115}\text{-S}_p\text{TFHAGQLR}$ , labeled with iTRAQ reagent 115, were used as an example for pS- and pT-peptides. The neutral loss of 98/z ([M-P]) represented the most intense peak in CID fragmentation mode (NCE range 25–85%) for both 2<sup>+</sup> and 3<sup>+</sup> charged precursors (Figure 2-5 (A, B)). For both precursor charge states SEQUESTs XCorr remained constant over the NCE range of 25–70% due to SEQUEST matched only a few fragment ion signals ( $b_2^+$  to  $b_5^+$ ;  $y_4^+$  to  $y_7^+$ ; Figure 2-2 (A)) alongside the neutral loss signals [ $b_9^+(-P)$  equals the precursor mass with subsequent loss of water], while for the 3<sup>+</sup> charged precursor, four y-ions and five b-ions could be matched (Figure 2-2 (B)). The high intensity of the neutral loss signal ([M-P]) and poor backbone fragmentation (< 20% compared to neutral loss signal) for CID fragmentation in the ion trap is illustrated in Figure 2-2 at an NCE of 35%. Typical for CID, the energy is distributed via the peptide backbone, leading to breaking the weakest bond (inducing the intense [M-P] signal) rather than peptide backbone fragmentation. To overcome this problem, multistage activation (MSA) was introduced (263).

## 2 Developing a Platform for relative Quantification using iTRAQ

In contrast to traditional MS<sup>3</sup>, in which the neutral loss fragment is isolated and used for further fragmentation, the MSA approach stores both, the fragment ions of the MS/MS scan and fragments generated by a further fragmentation (MS<sup>3</sup>) of the neutral loss ion. All ions are read out collectively in the ion trap, leading to spectra referred to as *pseudo* MS<sup>3</sup>, as they are a combination of both, MS/MS and MS<sup>3</sup> experiments. Using CID-MSA, the spectra from the 2<sup>+</sup> charged precursor were dominated by the neutral loss signal of [M-P] at low NCE values (Figure 2-5 (C)). Increasing the NCE further resulted in the neutral loss signal decreasing below detectable limits. Instead of the neutral loss signal dominating CID spectra, a further loss of 18/z, potentially by dehydration of the adjacent serine, from the neutral loss fragment [M-P] occurred and dominated spectra when NCE values of 35 to 50% were applied (Figure 2-5 (C)). For iTRAQ labeled peptides, a precursor fragment ion had been described previously, which loses iTRAQ and is charge reduced (63). Accordingly, precursors which had lost the H<sub>3</sub>PO<sub>4</sub> can lose the isobaric tag (144/z + H<sup>+</sup>) in addition, which was monitored at low collision energies ([M-P-iTRAQ]<sup>+</sup> and [M-P-iTRAQ-H<sub>2</sub>O]<sup>+</sup> in Figure 2-3 (A)).

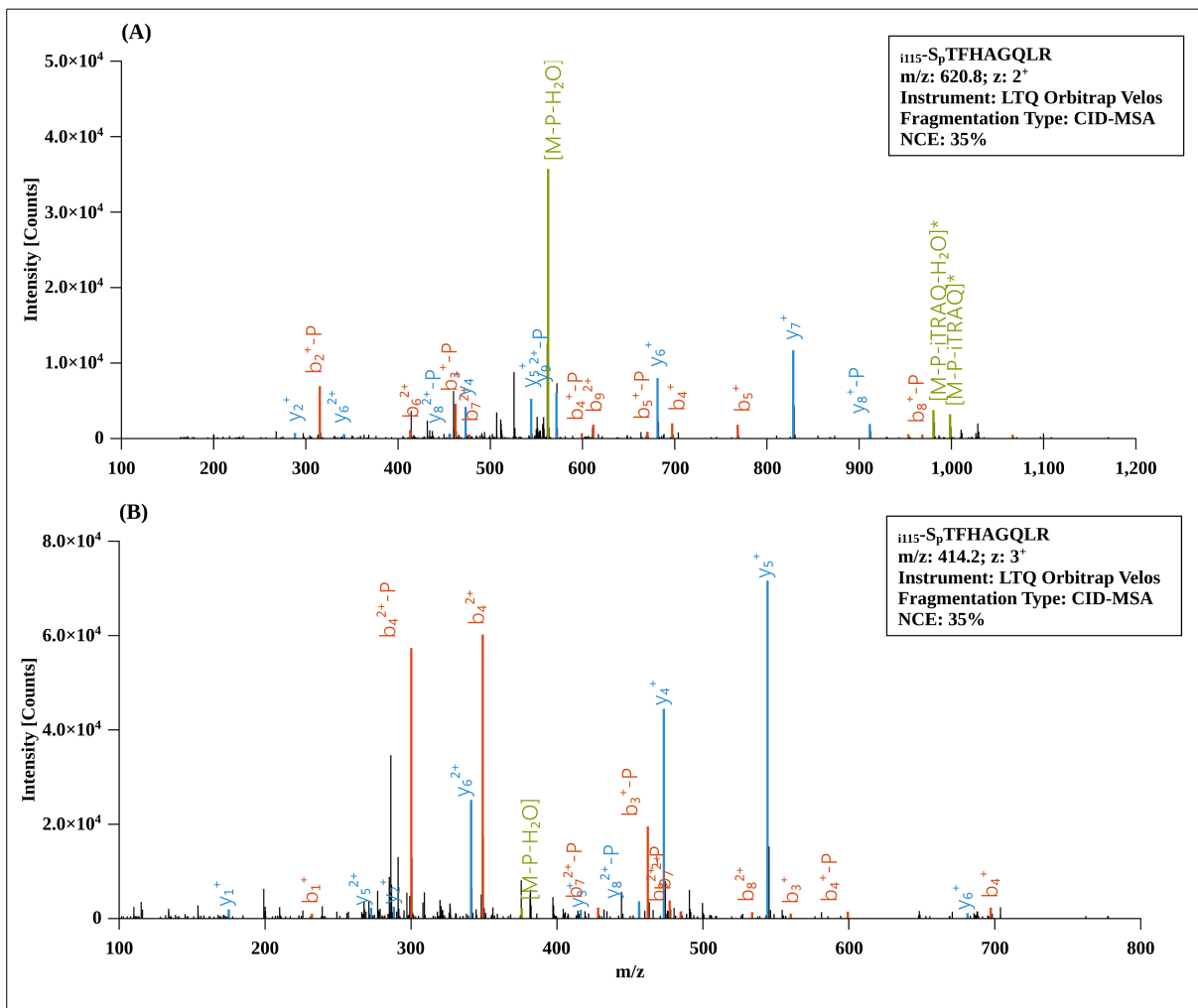


Figure 2-3: Offline MS/MS measurements in Orbitrap of the iTRAQ labeled phosphopeptide *i*<sub>115</sub>-S<sub>p</sub>TFHAGQLR. (A) CID-MSA, 2<sup>+</sup> charged precursor at m/z 620.8; (B) CID-MSA, 3<sup>+</sup> charged precursor at m/z 414.2. Fragments labeled with an asterisk (\*) have lost a charge.



## 2.1 iEM – iTRAQ Evaluation Method

For NCE values below 25% no identification results were obtained. NCE from 25 to 45% predominantly matched the  $y_4^+ - y_7^+$  and  $b_2^+ - P$  and  $b_3^+ - P$  ions at *sufficient* ion intensity (Figure 2-3), while at higher NCE values (>55%), backbone fragmentation increased and more b-series fragment ions appeared. The precursor fragment ions [M-P-iTRAQ] and [M-P-iTRAQ-H<sub>2</sub>O] were detectable above an NCE of 25% and increased with increasing NCE. Similar to CID, the XCorr for CID-MSA remained constant over the NCE range applied. For the 3<sup>+</sup> charged precursor, the neutral loss intensity distributions of [M-P] and [M-P<sub>80</sub>], are similar to that for the 2<sup>+</sup> charged precursor (Figure 2-5).

For HCD, the neutral loss signals of [M-P] and [M-P<sub>80</sub>] showed a maximum intensity at 35% NCE (Figure 2-5 (E)) while the curve shape is more similar to CID-MSA than to CID. For the 2<sup>+</sup> charged precursor after HCD excitation, fragment ion matching performed by SEQUEST showed a different behavior compared to CID and CID-MSA. For NCE values < 35% no fragment ions could be mapped. For the NCE values of 40 and 45%, the majority of the y-ion series was mapped, along with several from the b-ion series (Figure 2-4). Increasing the NCE to 50 and 55% a decreased number of ions was mapped. For NCE values >55%, mainly the  $b_1^+$  and  $y_1^+$  ions remained observable. For the 3<sup>+</sup> charged precursor, the assigned fragment ions for NCE values of 40 to 60% were identical. Increasing the NCE further led to peptide backbone fragmentation resulting in internal ions and immonium ions at the low m/z region. Comparison of the resulting XCorrs for fragments matched for the 2<sup>+</sup> and 3<sup>+</sup> charged precursors, revealed that the *bell shape like* curve for the 2<sup>+</sup> charged precursor is primarily due to the higher number of ions being mapped at NCE values 35 – 60%. In contrast, more fragment ions can still be matched at higher NCE values for 3<sup>+</sup> charged precursors.

Two pY-peptides were analyzed by offline-nano-ESI (Figure 2-6). Consistent with previously published findings (181), the peptide  $i_{115}\text{-V}_p\text{YELMR}$ , when subjected to CID fragmentation, generated an intense neutral loss signal of [M-P<sub>80</sub>]. This ion represents the most intense signal for NCE values higher than 25%. The intensity of the neutral loss signal of [M-P] is approximately 20% of the [M-P<sub>80</sub>] signal. However, contrary to the previous findings, the intensities of both the [M-P] and [M-P<sub>80</sub>] ion signals produced by the peptide  $i_{114}\text{-VIEDNE}_p\text{YTAR}$  were comparable which might be the result of an additional neutral loss of water. Using CID-MSA, the number of fragment ions matched was comparable to CID; however, the neutral loss signals of [M-P] and [M-P<sub>80</sub>] were reduced. For HCD fragmentation of the pY-peptides, the neutral loss signal of [M-P<sub>80</sub>] was detectable, and the XCorr values followed the same trend as with the pS- and pT-peptides, with the maxima centered at 45% NCE. The intensity of the important modification-specific and diagnostic pY immonium ion at m/z 216.04 (190) was also found to increase with increasing NCE. For these two peptides, a NCE of 35–40% was required to detect the diagnostic pY immonium ion in spectra.

## 2 Developing a Platform for relative Quantification using iTRAQ

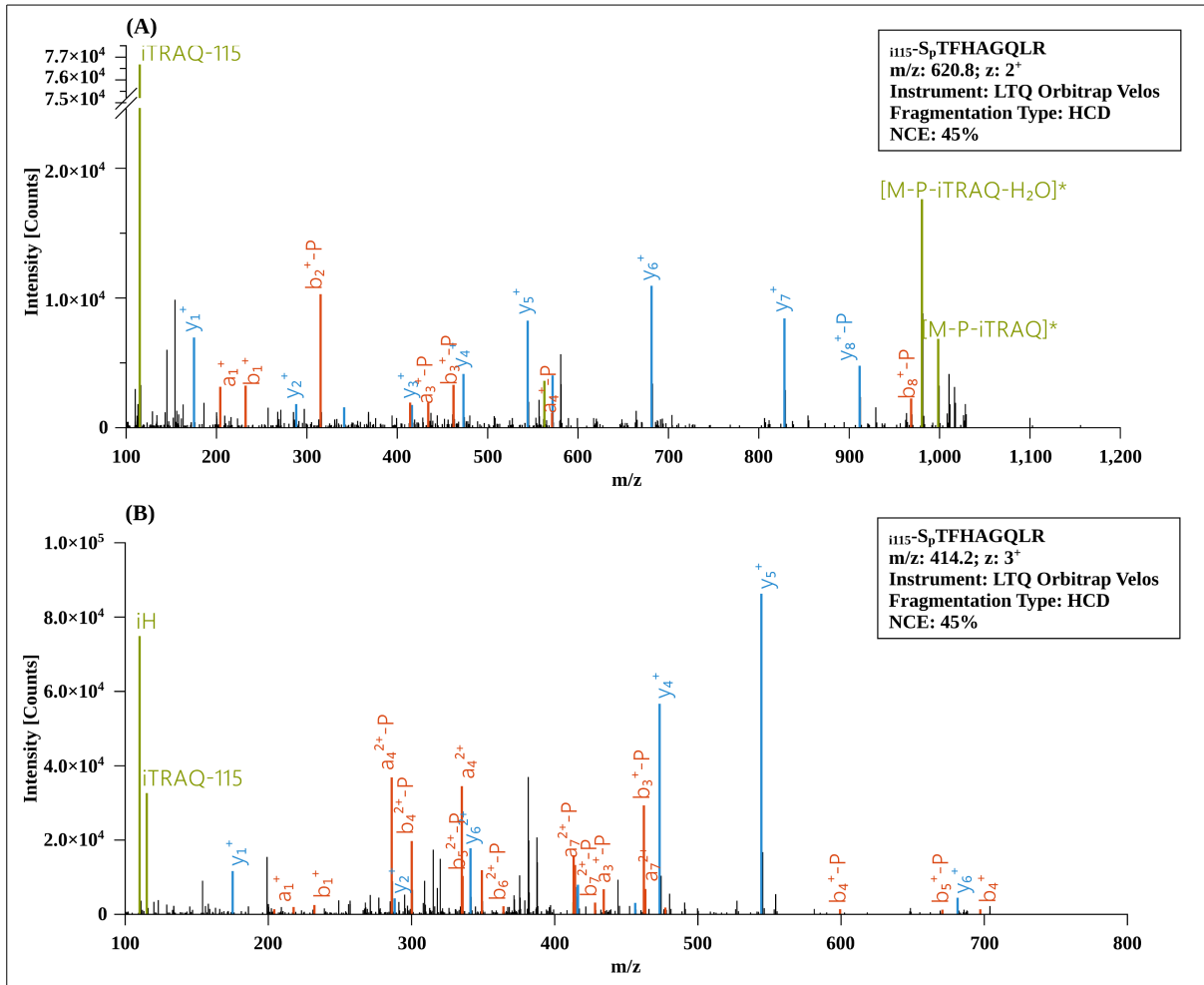


Figure 2-4: Offline MS/MS measurements in Orbitrap of the iTRAQ labeled phosphopeptide  $i_{115}$ -S<sub>p</sub>TFHAGQLR. (A) HCD, 2<sup>+</sup> charged precursor at m/z 620.8; (B) HCD, 3<sup>+</sup> charged precursor at m/z 414.2. Fragments labeled with an asterisk (\*) have lost a charge. The green labeled signal at m/z 110 (left to iTRAQ-115 (B)) belongs to the histidine immonium ion.

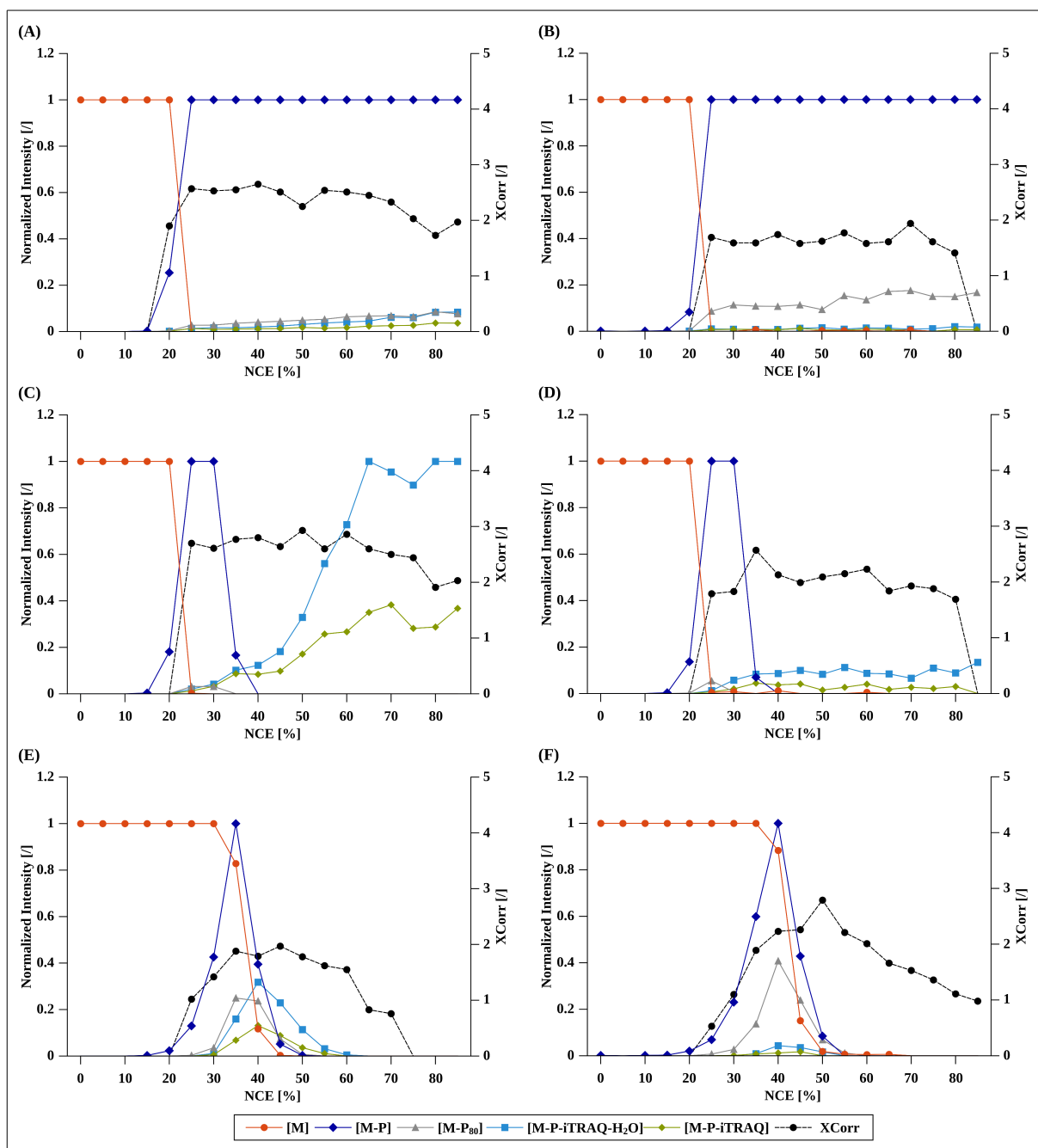


Figure 2-5: Dependence of ion intensities and XCorrs of labeled phosphopeptide  $i_{115}\text{-S}_p\text{TFHAGQLR}$  acquired in offline nano-ESI-Orbitrap MS on the NCE. (A) CID,  $2^+$  charged precursor at  $m/z$  620.8; (B) CID  $3^+$  charged precursor at  $m/z$  414.2; (C) CID-MSA,  $2^+$ ; (D) CID-MSA,  $3^+$ ; (E) HCD,  $2^+$ ; (F) HCD,  $3^+$ .

## 2 Developing a Platform for relative Quantification using iTRAQ

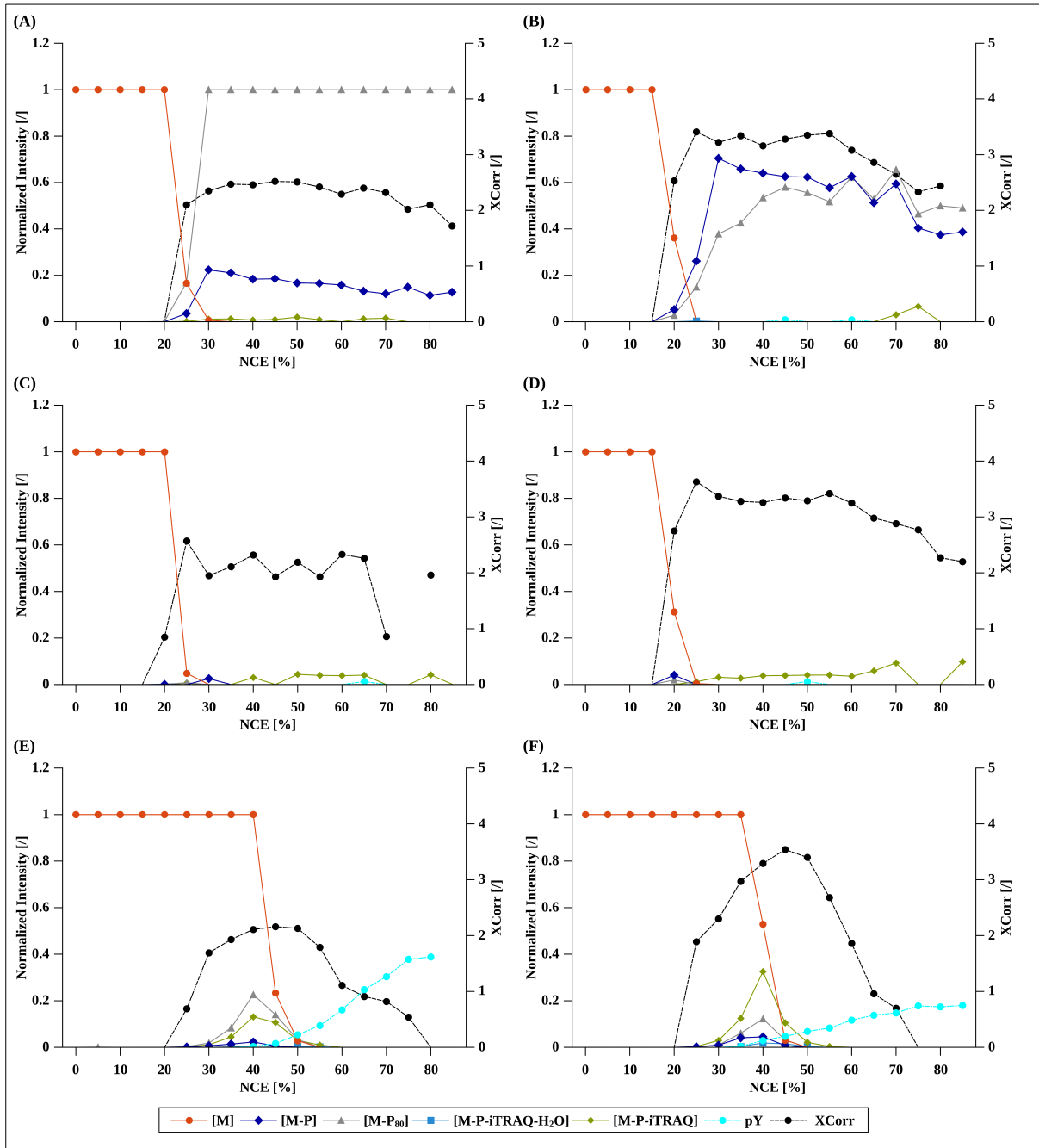


Figure 2-6: Offline MS/MS measurement of two iTRAQ labeled, 2<sup>+</sup> charged pY-peptides *i*<sub>115</sub>-V<sub>p</sub>YELMR (A, C, E), *m/z* 517.7; and *i*<sub>114</sub>-VIEDNE<sub>p</sub>YTAR (B, D, F) at *m/z* 717.3. (A, B): CID; (C, D): CID-MSA; (E, F): HCD.

### **Phosphorylation Site Mapping**

Database search algorithms such as SEQUEST, Mascot, X!Tandem, and OMSSA allow automatic evaluation for the quality of Peptide-Spectrum-Matches, enabling to process a large number of spectra rapidly. However, correct and confident identification of the phosphorylation amino acid, particularly assignment of closely spaced or adjacent phosphorylation sites, remains challenging. Furthermore, interference from noise and impurities during the precursor selection process can result in inaccurate assignments (280). The use of Mascot Delta Score (MD-score) has been proposed to evaluate the false-positive phosphorylation site localization results searched by MASCOT (276,281), while in conjunction with SEQUEST search results, a probability based score, termed ambiguity score (Ascore), has been introduced as a cutoff value to determine if a phosphorylation site has been correctly assigned (282). These scoring algorithms were not applied in this study, to be consistent with the results described above using SEQUEST XCorr.

For this purpose, data acquired from mix-2 (131 iTRAQ labeled phosphopeptides) were searched with Proteome Discoverer/SEQUEST. The rank 1 peptide hits with FDR < 1%, acquired either by HCD or CID-MSA-IT (acquired in the ion trap mass analyzer), were exported to Excel for examining the phosphorylation site(s) mapped by SEQUEST and summarized according to the peptide charge states and the fragmentation types in relation to the NCE applied (Figure 2-7). As expected by the XCorr distribution (Figure 2-5 (C)) for 2<sup>+</sup> charged precursors fragmented by CID-MSA, the number of identifications remained almost constant over a wide NCE range (25–70%), with a maximum at 45% NCE. This value also represented the optimal ratio for correctly vs incorrectly assigned phosphorylation sites (see Figure 2-7). In the NCE range of 25 – 70%, a larger number of 2<sup>+</sup> charged compared to 3<sup>+</sup> charged precursors were identified. This is possibly due to the charge distribution observed in these measurements (50% and 31% for 2<sup>+</sup> and 3<sup>+</sup> charged precursors, respectively). Furthermore, the number of phosphopeptides with incorrectly mapped phosphorylation sites is at least equal than the 2<sup>+</sup> charged precursors. For HCD, a higher dependency on the NCE applied can be observed for 2<sup>+</sup> and 3<sup>+</sup> charged precursors. For 2<sup>+</sup> charged precursors, the optimal NCE value lies between 35 and 45% NCE. For 3<sup>+</sup> charged precursors, the optimal values were shifted to 40 - 50% NCE. However, as observable by the XCorr behavior (Figure 2-5 (E, F)), the 2<sup>+</sup> charged precursors are more sensitive to the applied NCE values than the 3<sup>+</sup> charged precursors.

## 2 Developing a Platform for relative Quantification using iTRAQ

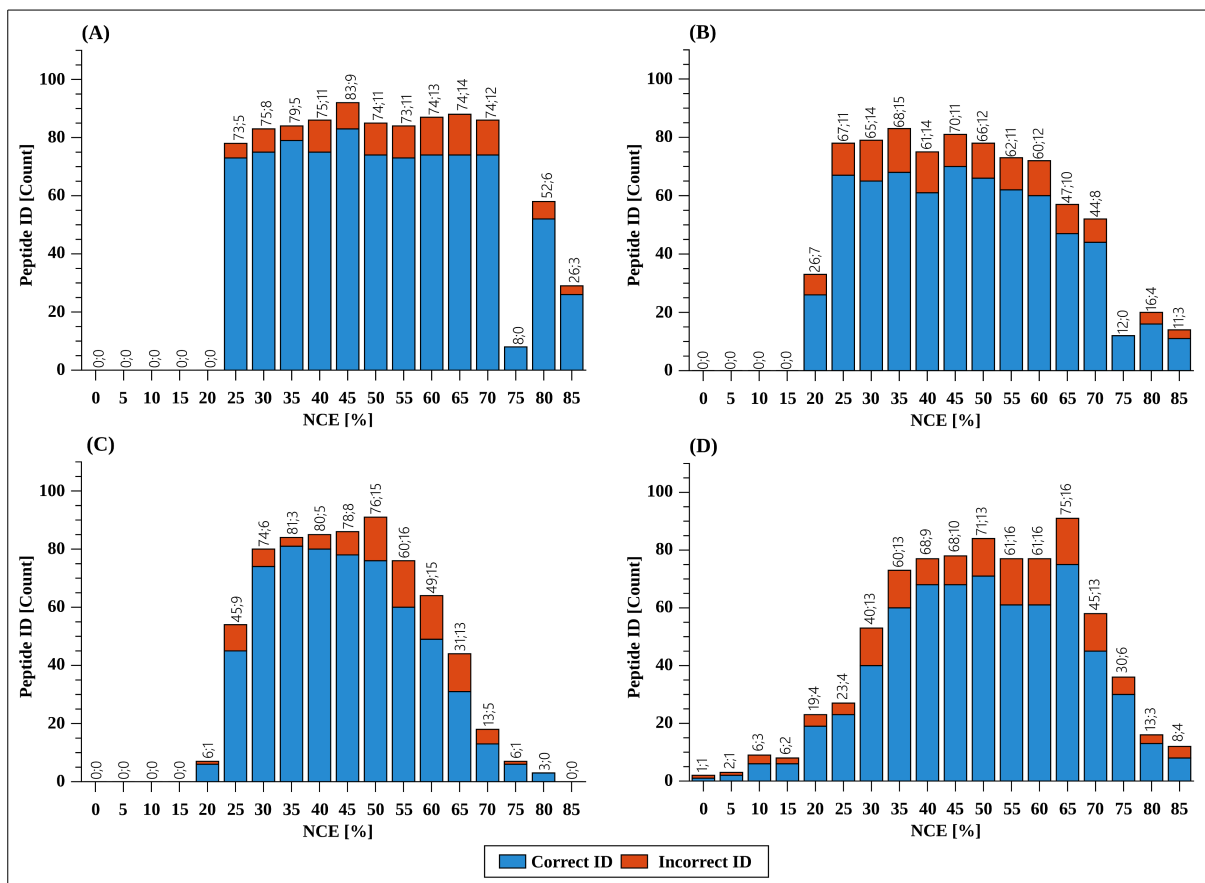


Figure 2-7: Comparison of the number of peptide identifications (ID) of mix-2 between correct (blue; lower number) and incorrect mapped phosphorylation sites (orange; upper number) at different NCE values. (A) CID-MSA, 2<sup>+</sup> charged precursors; (B) CID-MSA, 3<sup>+</sup>; (C) HCD, 2<sup>+</sup>; (D) HCD, 3<sup>+</sup>. The first number represents correct and the second number the incorrect identifications.

### Influence of NCE on Phosphopeptide Quantification

Protein quantification with iTRAQ labeling is based on the generation of specific reporter ions after MS/MS fragmentation. Furthermore, for proper quantification, sufficient ion statistics (i.e., peak intensities) of the reporter ions is a critical parameter. To obtain high signal intensities for reporter ions in the low m/z region, high NCE values can be applied. However, this is accompanied by a loss of signal intensities and sequence information in the middle to high m/z region (63). Furthermore, the acquisition time for HCD (up to 1 sec) is much higher than for CID-IT (ion trap, 0.1–0.2 sec) for the LTQ Orbitrap Velos (see Figure 1-12). Also, as shown above, the application of HCD at high NCE values negatively affects both peptide identification and phosphorylation site mapping.

A further aspect was to determine which NCE value should be applied for proper quantification of iTRAQ labeled phosphopeptides. For this purpose, four equal amounts of mix-1 were labeled individually with iTRAQ-4Plex reagents before combining and performing LC-ESI MS analysis using HCD fragmentation, and Orbitrap MS/MS detection. The iTRAQ reporter ion (114–117) raw intensities were exported from the .mgf file and further processed with the *in-house* scripts (VBA) to calculate the averaged normalized iTRAQ signal intensity,

## 2.1 iEM – iTRAQ Evaluation Method

the averaged relative iTRAQ signal intensity, the iTRAQ ratio (114 set as the denominator), and the standard deviation calculated for each data point. The peptide  $i_{114-117}\text{-S}_p\text{TFHAGQLR}$  (Figure 2-8) is used as an example for discussion.

For the  $2^+$  charged precursor, no iTRAQ reporter ion signals were observed at NCE below 30%. The averaged relative intensity of iTRAQ reporter ions rose sharply between NCE of 30 – 40%, and reached a maximum at NCE of  $45 \pm 5\%$ , decreasing gradually with higher NCE (Figure 2-8 (A)). The decrease of normalized iTRAQ reporter ion intensities at NCE above 55% was mainly due to the strong immonium ion of histidine ( $m/z$  at 110.072 Da) which dominated the spectrum at higher NCE. The relative intensity of iTRAQ reporter ions remained constant at NCE of 55 – 85% (Figure 2-9 (A)). Along with the iTRAQ reporter ions, the release of the intact protonated isobaric tag ( $m/z$  145.10) was monitored. As expected by the maxima at 40% shown for [M-P-iTRAQ] the non-dissociated iTRAQ also showed maximal intensities at the same NCE (Figure 2-8 (A)).

As the presence of the histidine immonium ion reduces the impact of iTRAQ reporter ions as most intense signals in spectra at higher NCE values, confirmation of the observable trends was assessed using two pY-peptides without histidine ( $i_{114-117}\text{-V}_p\text{YELMR}$  and  $i_{114-117}\text{-VIEDNE}_p\text{YTAR}$ ). Again, the normalized iTRAQ reporter ion intensity reached its maximum at 45% NCE (Figure 2-10).

For the  $3^+$  charged precursor (Figure 2-8 (B)), a markedly different curve shape was observable. The intensity did not increase in the same manner, instead the intensity (compared to base peak) increased linearly with NCE. The isobaric tag signal at  $m/z$  145.10 had a lower intensity in comparison to the  $2^+$  charged precursor. This might be due to the overall low intensity of the iTRAQ signal.

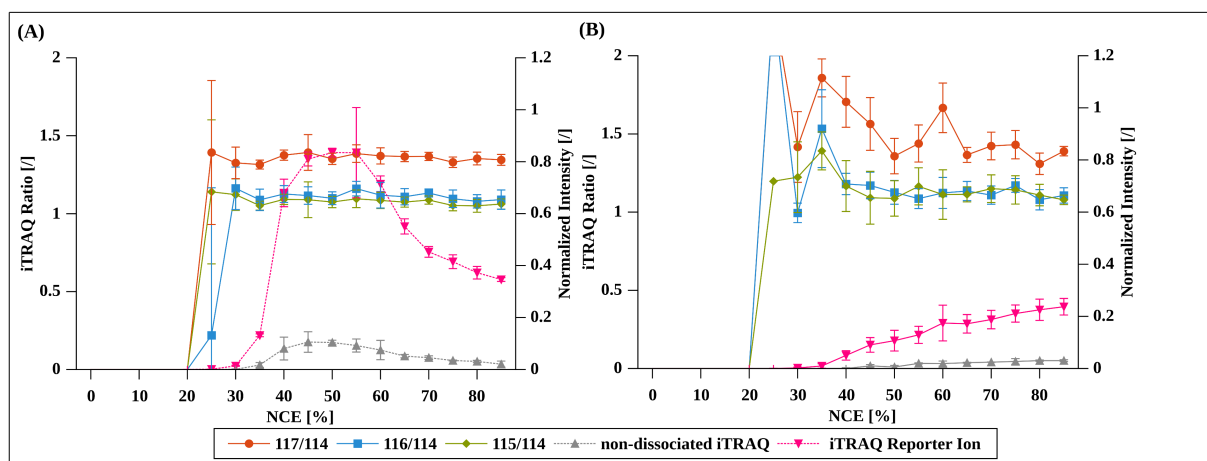


Figure 2-8: Quantitative analysis of the peptide  $i_{114-117}\text{-S}_p\text{TFHAGQLR}$  in dependence of NCE applied in HCD measurements; (A)  $2^+$  and (B)  $3^+$  precursor ion charge state.

## 2 Developing a Platform for relative Quantification using iTRAQ

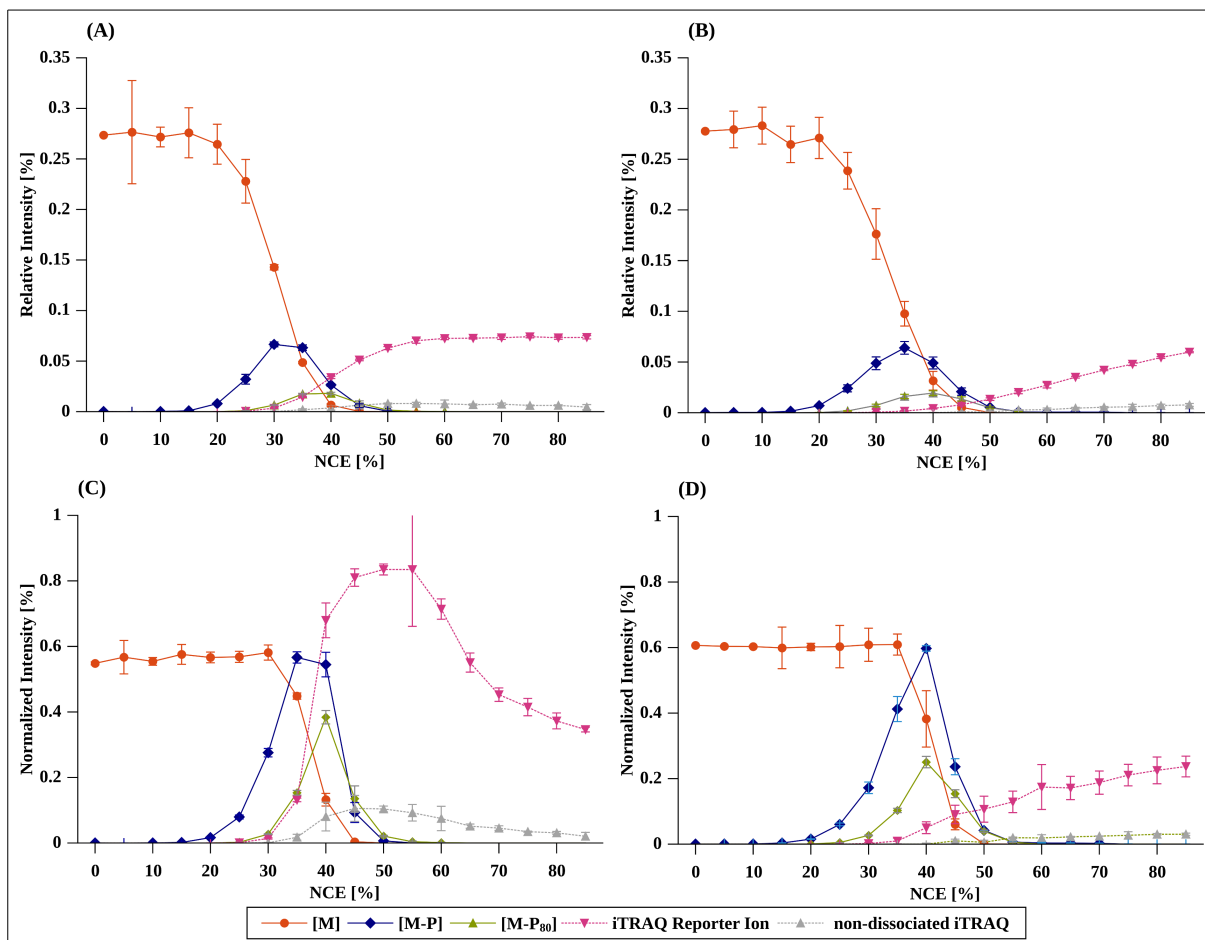


Figure 2-9: Online LC-ESI MS measurements of the iTRAQ labeled phosphopeptide  $i_{114-117}\text{-S}_p\text{TFHAGQLR}$  fragmented by HCD in dependence on the NCE. (A, B) relative intensities; (C, D) normalized intensities; for the  $2^+$  charged precursor  $m/z$  620.8 (A, C) and  $3^+$  charged precursor (B, D) at  $m/z$  414.2.

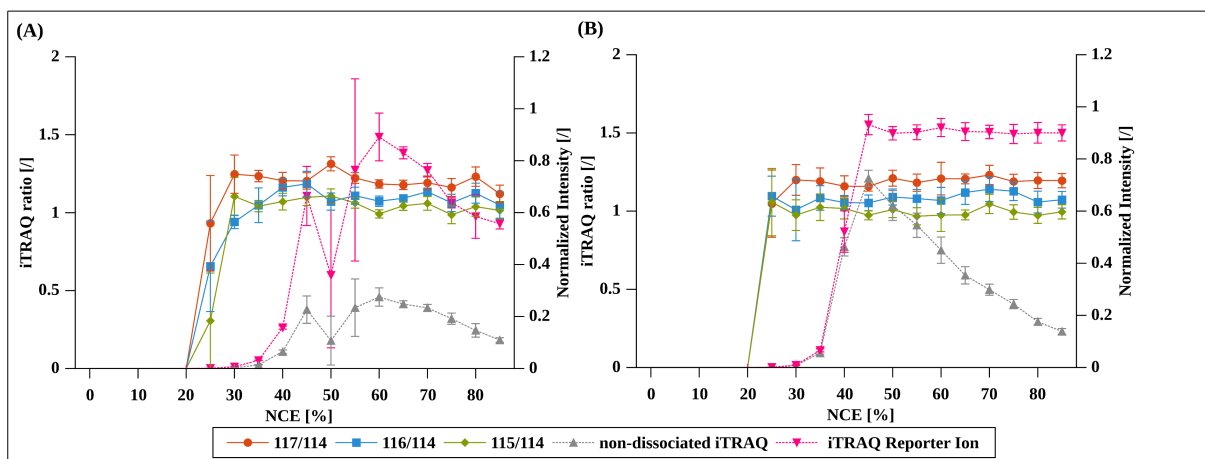


Figure 2-10: Comparison of normalized iTRAQ reporter ion intensities, the protonated non-dissociated isobaric tag (iTRAQ + balancer group) at  $m/z$  145.10 and the corresponding iTRAQ ratios (reporter ion 114 as denominator) depending on NCEs for HCD fragmentation for the  $2^+$  charged precursors of peptides (A)  $i_{114-117}\text{-V}_p\text{YELMR}$  and (B)  $i_{114-117}\text{-VIEDNE}_p\text{YTAR}$ .



## 2.1 iEM – iTRAQ Evaluation Method

Additionally, the generation of iTRAQ reporter ions from the pS-peptide  $i_{114-117}$ -GHL<sub>p</sub>SEGLVTK for both the 2<sup>+</sup> and 3<sup>+</sup> charged precursors are illustrated (Figure 2-11). Again, the normalized iTRAQ reporter ion intensity reached its maxima at an NCE of 45 – 50% for the 2<sup>+</sup> charged precursor and remained stable. However, for the 3<sup>+</sup> charged precursor, the histidine effect to replace iTRAQ reporter ion signals as most intense signals observed for the  $i_{114-117}$ -S<sub>p</sub>TFHAGQLR peptide was notably reduced. In agreement with the majority of other phosphopeptides, the iTRAQ reporter ions represented the most intense signal at NCE values greater than 60%. Further, the sensitivity and precision of iTRAQ ratios obtained at different NCE values was evaluated. The iTRAQ ratio for the 2<sup>+</sup> charged precursor remained constant and showed only a small standard deviation for NCE values >30%. In contrast, the iTRAQ ratio for the 3<sup>+</sup> charged precursor was more variable with a larger standard deviation (Figure 2-8 (A, B) or Figure 2-11 (E, F)). When the ratios of the other peptides analyzed were taken into account, an NCE value of 55% represents the minimum value required for accurate iTRAQ quantification of 3<sup>+</sup> charged phosphopeptides.

## 2 Developing a Platform for relative Quantification using iTRAQ

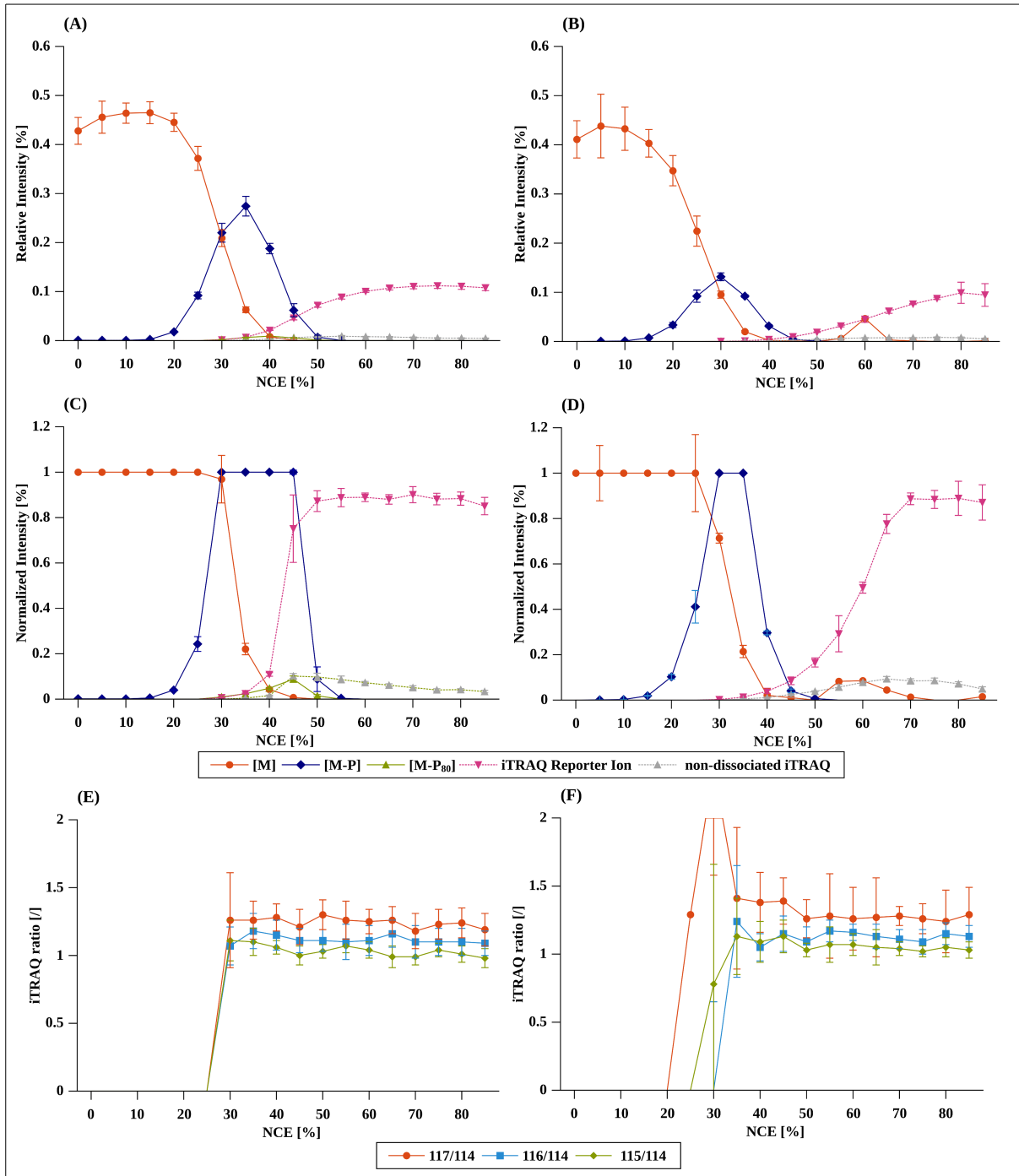


Figure 2-11: Online LC-ESI-MS/MS measurements of the iTRAQ labeled phosphopeptide  $i_{114-117}\text{-GHL}_p\text{SEGLVTK}$  fragmented by HCD. (A, B) relative intensities; (C, D) normalized intensities; (E, F) the corresponding iTRAQ ratios (reporter ion 114 as denominator), for the 2<sup>+</sup> charged precursor at m/z 704.9 (A, C, E) and the 3<sup>+</sup> charged precursor (B, D, F) at m/z 470.3.

### The Compromise Between Identification and Quantification of Phosphopeptides

As described above, the requirements for both identification and quantification follow inverse trends, being optimal at considerably different NCE values. To determine the optimal NCE for simultaneous identification and quantification via HCD while performing CID-MSA-IT for identification in parallel a complex peptide mixture (mix-2) containing 131 phosphopeptides was analyzed. The normalized signal intensities of the precursor and fragment ions described before were plotted for pS-, pT- and pY-peptides for both 2<sup>+</sup> and 3<sup>+</sup> charged states (Figure 2-13). The trends described for the single peptides were fully reflected in the average values obtained for this complex mixture. For the optimization of fragmentation for phosphopeptide identification and correct phosphorylation site mapping (Figure 2-7), CID-MSA-IT had only a low level of dependency on NCE. Additionally, pT-, pS-, and pY-residues were readily identifiable across a large range of NCE values (25 – 70%). However, for HCD, there is a much greater dependency on NCE for identification of phosphopeptides, and a comparatively narrow optimal range compared to CID. Therefore, the use of a higher NCE value of 45% to ensure optimal identification and phosphorylation site-mapping mapping in HCD can be recommended. In addition, the generation of the highly specific pY immonium ion signal at m/z 216.04 was monitored. Detection of this ion was only possible above a minimum NCE value of 45% after which it increased steadily with NCE (Figure 2-13 (E, F)).

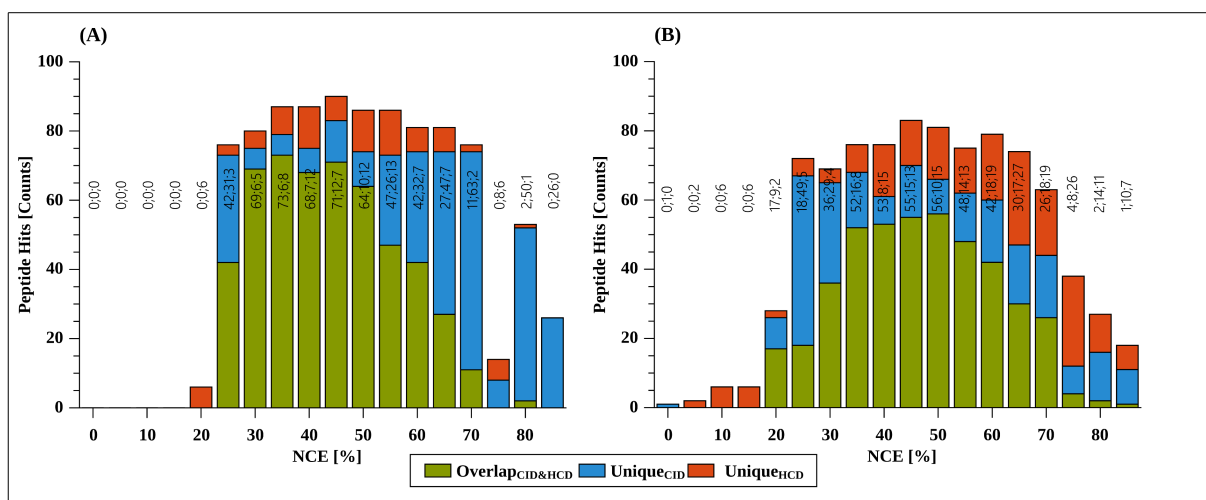


Figure 2-12: Comparison and overlap of the number of peptide identifications of mix-2 at different NCE. Green bars: peptide identified by both, CID-MSA-IT and HCD; blue bars: peptide only identified by CID; orange bars: peptide only identified by HCD. (A) 2<sup>+</sup> charged precursors; (B) 3<sup>+</sup> charged precursors.

Combined analysis of CID-MSA-IT and HCD spectra for identification and phosphorylation site mapping can considerably boost identification efficiency (Figure 2-12). Generally, CID in ion trap is a resonance based process, whereas HCD represents a beam-type CID event (283), which leads to different fragmentation patterns (Figure 2-2 to 2-4). For HCD, increasing NCE correlates to excessive fragmentation and a subsequent decrease in identification success (Figure 2-7). The acquisition of complementary fragmentation data can improve phosphopeptide identification. This becomes important if the applied NCE for HCD is high. This leads to the formation of intense iTRAQ reporter ions signals and to the formation of b<sub>1</sub><sup>+</sup> and y<sub>1</sub><sup>+</sup> ions, as an excess amount of

## 2 Developing a Platform for relative Quantification using iTRAQ

energy is introduced while peptide fragmentation. In such cases, CID-MSA-IT spectra ensure high quality identification, including phosphorylation site mapping. Quantification of isobaric labeled phosphopeptides by HCD also showed a dependency on the applied NCE. Furthermore, it was determined that the fragmentation efficiency was dependent on the charge state distribution of the precursor ions being analyzed. In line with this, recommendations for identification and accurate quantification of 2<sup>+</sup> charged precursors is a minimal NCE value of 45%, while for 3<sup>+</sup> charged phosphopeptides a higher NCE value of 60% is recommended.

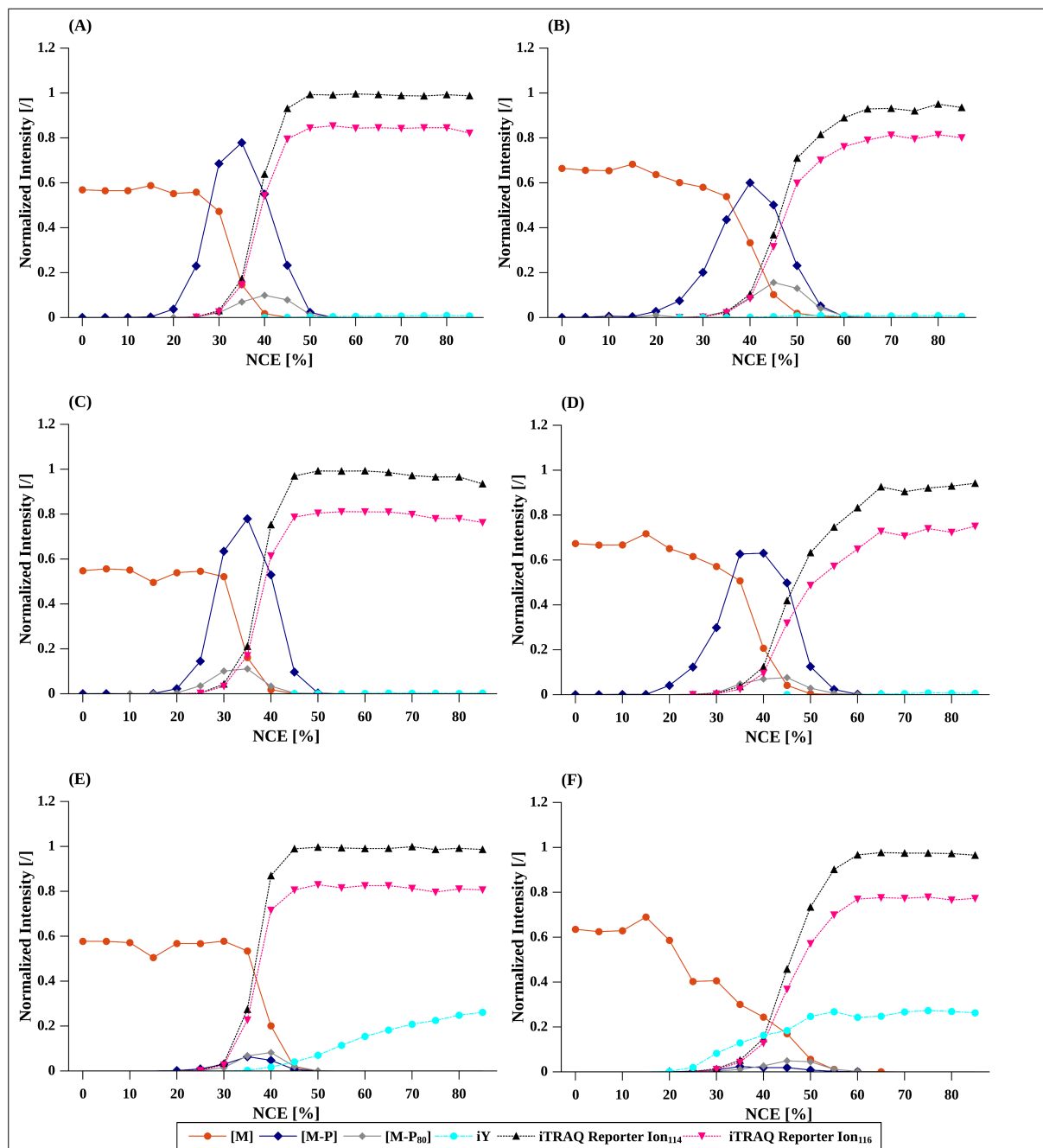


Figure 2-13: Comparison of online LC-ESI MS measurements of the iTRAQ labeled  $i_{114/i116}$ -phosphopeptide mix-2 fragmented by HCD at various NCE. Normalized intensities are shown, averaged over all identified peptides. (A): 51 pS-peptides, 2<sup>+</sup> charged; (B): 52 pS-peptides, 3<sup>+</sup> charged; (C): 14 pT-peptides, 2<sup>+</sup> charged; (D) 15 pT-peptides, 3<sup>+</sup> charged. (E): 30 pY-peptides, 2<sup>+</sup> charged; (F) 25 pY-peptides, 3<sup>+</sup> charged.

### 2.1.3.2 Ten Protein-mix

For simultaneous identification and quantification of complex samples, such as whole cell lysate, different requirements have to be fulfilled: for identification purposes, SEQUESTs XCorr and the number of identified peptides (unique and PSMs) are important criteria, whereas the formation of isobaric labeling derived reporter ions to guarantee accurate quantification is of highest demand for peptide and protein quantification.

The applied workflow to optimize tandem MS measurement parameters for non-phosphorylated peptides was similar to the one applied for phosphopeptides (chapter 2.1.3.1). The limited number of 260 tryptic peptides (*in-silico* digestion with: minimum peptide mass of 500 Da; no missed cleavages; no protein isoforms; no modifications) allowed to reproducibly select and fragment precursor ions of the ten protein-mix. This limited variations in precursor selection, which can be observed for more complex samples (81). Moreover, the precursor ion intensities were evenly distributed, and the effect of co-isolated precursors ions was minimized. This allowed to combine results of different LC-ESI MS experiments, each measured with a different NCE value for CID-IT and HCD fragmentation.

#### **Identification in Dependence of NCE**

In the first step of optimizing and analyzing the NCE values required for optimal peptide identification, the focus was set on SEQUESTs XCorr and the number of unique and totally identified (PSMs) peptides (Figure 2-14). For this purpose, the multistage feature of CID-IT was deactivated due to the low number of phosphopeptides, only originated from  $\alpha$ - and  $\beta$ -casein in the ten protein-mix. The NCE values were ranged from 0 to 85% for CID-IT and HCD, respectively. CID-IT spectra were recorded exclusively in the ion trap to benefit from its sensitivity and scan speed, whereas HCD spectra were recorded in the high resolution Orbitrap mass analyzer.

Upon CID-IT fragmentation, the 2<sup>+</sup> and 3<sup>+</sup> charged precursor ions were not affected by an increasing NCE from 30 – 60%. Both, the XCorr and the number of peptides (unique and PSMs) remained constant (Figure 2-14 (A, C)). A further increase of the NCE (> 60%) resulted in a fewer number of peptide identifications and a decreasing XCorr. 4<sup>+</sup> charged precursor, which were not analyzed in the phosphopeptide standard, were less affected by NCE variation compared to 2<sup>+</sup> charged precursor ions, although the trends observed for the 2<sup>+</sup> and 3<sup>+</sup> charge states were very similar (Figure 2-14 (E)).

For the peptide identification performed by SEQUEST of 2<sup>+</sup> charged precursors, HCD showed a different behavior compared to CID-IT (Figure 2-14 (B)). For NCE values < 25% only a limited number of peptides (< 100 PSMs) could be identified. For the NCE values of 30 – 50%, a constantly high number of peptides were identified. At 40% NCE, the highest number of peptides (unique and PSMs) were identified. Moreover, the XCorr was maximal at this NCE value.

Increasing the NCE to 55 and 60% decreased the ability to identify peptides, as 25% and 50% less peptides (PSMs) were identified, compared to the maximal number of identification at 40% NCE. This was a consequence of the XCorr, which decreased from a mean value of 2.7 at 40% NCE to 1.8 and 1.5 for 55% and 60% NCE, respectively. For NCE values  $\geq$  60%, the number of peptide identifications decreased further.

## 2 Developing a Platform for relative Quantification using iTRAQ

For the 3<sup>+</sup> charged precursors (Figure 2-14 (D)), the number of identified peptides for NCE values of 30 to 60% remained constant. However, the XCorr was affected by an elevated NCE value. The maximal XCorr of 3.0 was observed at 40% NCE, whereas the XCorr decreased almost linearly to 1.9 at an NCE value of 65%.

The number of identified peptides was barely affected by the NCE for 4<sup>+</sup> charged precursor ions. The maximal number of identifications (PSMs and unique) was observed at 50% NCE, whereas the maximal mean XCorr was calculated to be 3.5 at 45% NCE (Figure 2-14 (F)).

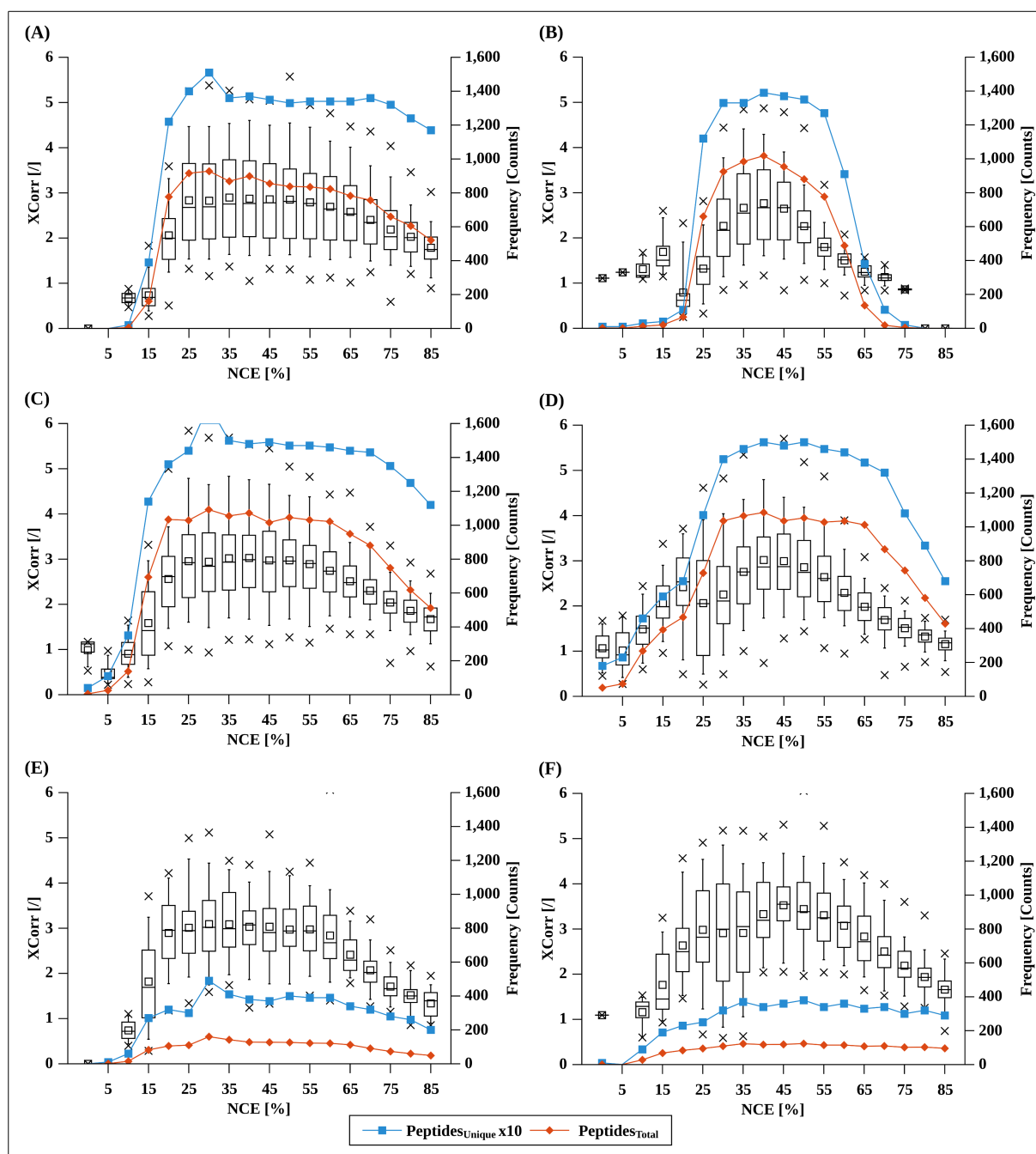


Figure 2-14: Comparison of SEQUESTs XCorr (Whiskers-Box-Plots) and the number of unique (x10) and total number of peptides identified via different NCEs applied for CID (A, C, E) and HCD (B, D, F) for 2<sup>+</sup> (A, B), 3<sup>+</sup> (C, D) and 4<sup>+</sup> (E, F) charged precursor ions.

### **Influence of NCE on the Formation of iTRAQ Reporter Ions**

Recently published methods for relative quantification using iTRAQ applied CID-IT spectra for identification and HCD spectra for quantification. This was a common approach for instruments such as the Orbitrap XL. To obtain a maximal number of peptide identifications, the low  $m/z$  region of HCD spectra was merged with the corresponding CID-IT spectra. These hybrid files were then applied to database searches (177). In other studies (211), an NCE value of 75% was used to guarantee intense reporter ions. However, these HCD spectra could only be utilized for quantification rather than both, identification and quantification simultaneously. This diminished the number of identified peptides after HCD fragmentation, as these spectra showed exclusively iTRAQ reporter ions signals and contained only a very few peptide backbone fragments (except such as  $y_1$  and  $b_1$ ) with low intensities.

Results from the phosphopeptide standard showed, that a minimal NCE of 45% and 55% is required for iTRAQ reporter ion formation for  $2^+$  and  $3^+$  charged precursor ions, respectively. However, as the number of precursor ions of the phosphopeptide standard was limited, the ten protein-mix sample was subjected to in-depth analysis. The focus was set on the intensity of the iTRAQ reporter ions, as stable iTRAQ ratios were observed upon applying 45% NCE. Therefore, the analysis of the iTRAQ ratios, such as  $iTRAQ_{114}/iTRAQ_{115}$  was omitted.

Along the normalized intensity (NI, Equation 2-1), also the relative intensity (RI, Equation 2-2) and the mean intensity was calculated for  $2^+$ ,  $3^+$  and  $4^+$  charged precursor ions upon HCD fragmentation (Figure 2-16).

For  $2^+$  charged precursor ions, barely no iTRAQ reporter ion signals were observed at NCE values below 35%. The averaged normalized intensity of iTRAQ reporter ions rose sharply between NCE of 35 and 45%, and remained almost constantly (Figure 2-16 (A)). The lower 5% value along with the averaged normalized intensity was mainly due to immonium ions such as histidine which dominate the spectra at higher NCE values, as it was shown for the phosphopeptides. For  $3^+$  charged precursor ions, a markedly different curve shape was observed (Figure 2-16 (B)). The normalized intensity did not increase sharply, instead the NI increased almost linearly with the NCE and remained constant at NCE values  $> 70\%$ . A minimum NCE value of 40% was required to detect iTRAQ reporter ions.

The curve shape of the  $4^+$  and  $3^+$  charged precursors were comparable, although NCE values  $\geq 75\%$  were required for maximal normalized intensities (Figure 2-16 (C)). An NCE of 45% was the minimum value to detect iTRAQ reporter ions, which was 10% and 5% higher than for  $2^+$  and  $3^+$  charged precursor ions. This might at least be partially an effect of the peptide lengths (Figure 2-15 (A)), which was significantly higher for  $3^+$  ( $\bar{\Delta}$  12.5 amino acids, AA) than for  $2^+$  charged precursors ( $\bar{\Delta}$  9.3 AA). Compared to doubly charged precursors, the number of amino acids per peptide almost doubled for  $4^+$  charged peptides ( $\bar{\Delta}$  17.5 AA).

For  $2^+$  charged precursor ions, the relative intensity vs. NCE plot can be described as a sigmoid curve (S-Function) and be grouped into two parts: a (roughly) linear one from 25 – 45% (with  $t < 2$ ; Equation 2-5) and a point of inflection at 50 – 55% (Figure 2-16 (D)). This curve shape was not observed for the  $3^+$  and  $4^+$  charged precursors (Figure 2-16 (E & F)). Instead, the curve shape was linear throughout the complete NCE range tested. This might be a consequence of the higher NCE value required for precursor charge states  $> 2^+$ .

## 2 Developing a Platform for relative Quantification using iTRAQ

$$\text{sig}(t) = \frac{1}{(1+e^{-t})}$$

Equation 2-5

As a higher variability of iTRAQ ratios was reported for lower (iTRAQ reporter ion) signals (284), the iTRAQ reporter ion data were further analyzed by their mean signal intensity (Figure 2-16 (G-I)): in Thermo Scientific's instrument software *tune page*, the target number of ions to be accumulated for a HCD scan is set to  $1 \times 10^5$  with a maximal injection time of 1 sec per default. Assuming, that one ion releases only one iTRAQ reporter ion, this also represents the maximal intensity (in counts) to be expected.

For  $2^+$  charged precursor ions, the averaged intensity increased from 30 – 55% NCE and reached its maximal value at  $6.9 \times 10^4$  (counts). The iTRAQ intensity decreased if NCE values  $>55\%$  were applied (Figure 2-15 (B) and Figure 2-16). This might be caused by consecutive fragmentation events of the iTRAQ reporter ions. For  $3^+$  charged precursors, the averaged reporter ion intensity increased linearly and its maximal value was reached at 80% NCE. Only one further NCE value was tested (85%), therefore, it cannot be excluded that the intensity might increase further with NCE. The curve shape of the  $4^+$  charged precursor was similar to the one for  $3^+$  charged precursor ions.

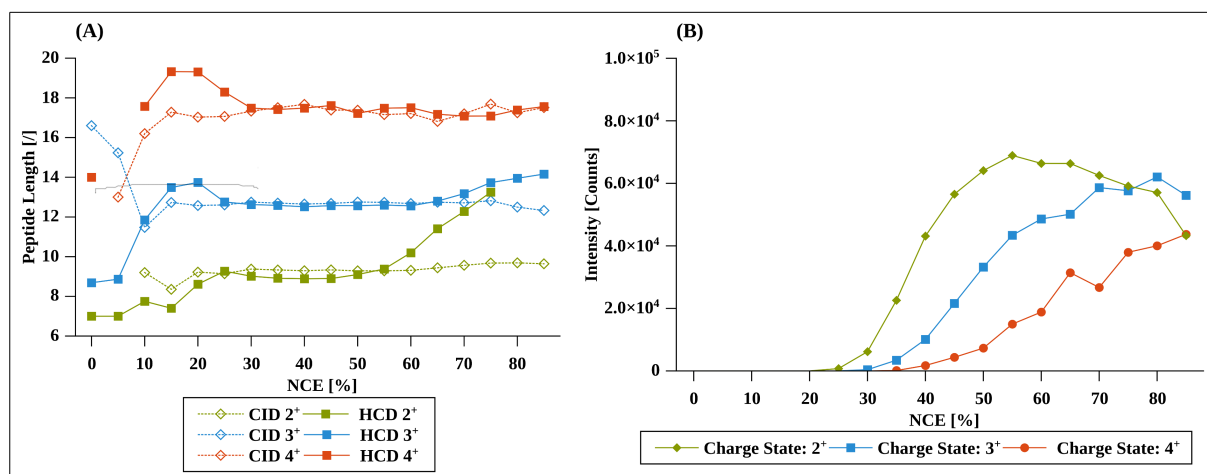


Figure 2-15: Averaged lengths of identified peptides for CID-IT and HCD in dependence of NCE applied for  $2^+$ ,  $3^+$  and  $4^+$  charged precursor ions (A) and the median raw intensity of iTRAQ reporter ion (B).



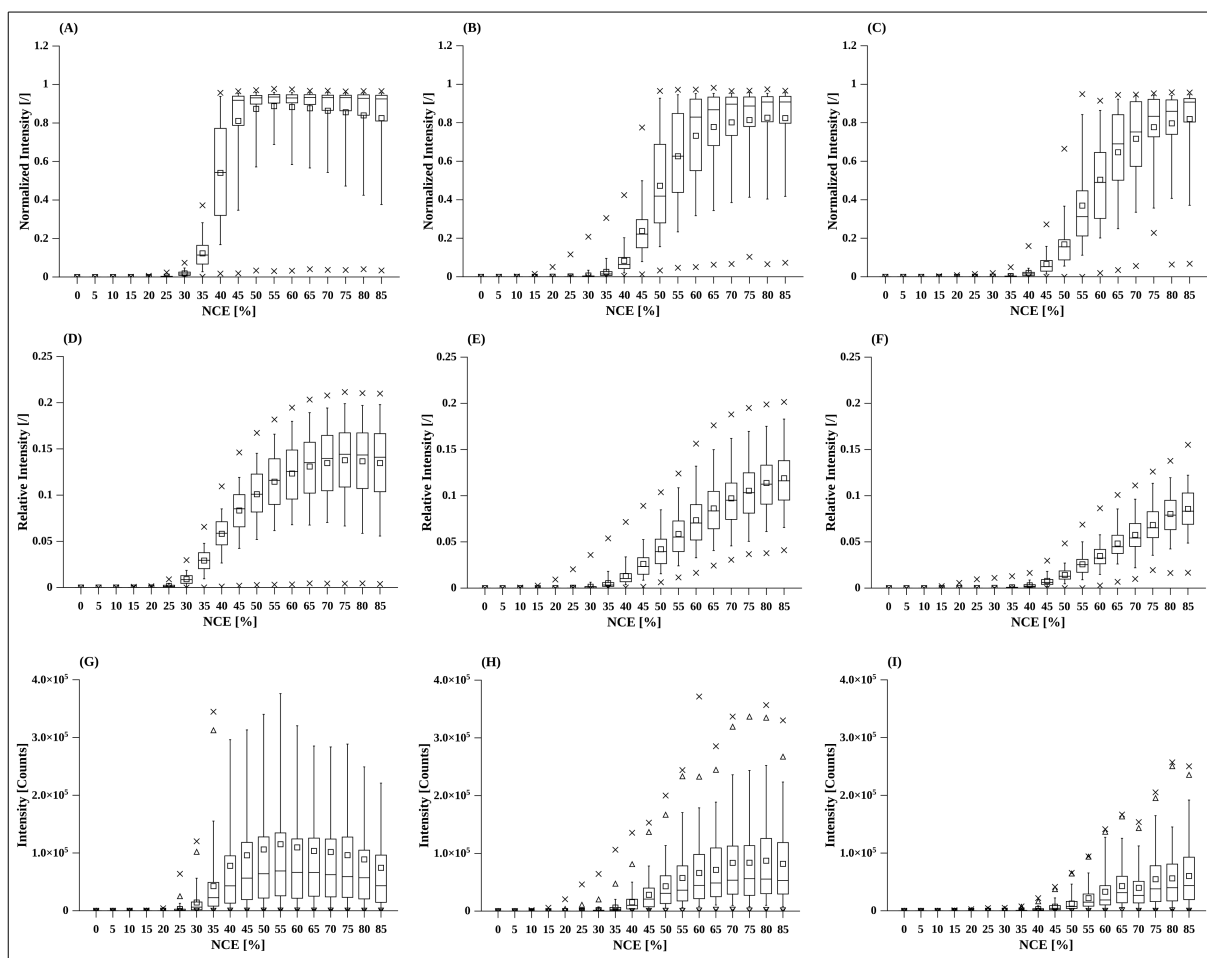


Figure 2-16: Whiskers-Box-Plots of iTRAQ reporter ion intensities after HCD fragmentation for the the protein-mix in dependence of NCE applied for  $2^+$  (A, D, G),  $3^+$  (B, E, H) and  $4^+$  (C, F, I) charged precursor ions. (A - C) Normalized Intensities; (D - F) Relative Intensities; (G - I) Intensities (raw values).

### **The Compromise Between Identification and Quantification**

The requirements for both identification and quantification follow inverse trends, as it was shown previously for the phosphopeptide standard. Although CID-MSA-IT was applied for the phosphopeptide samples, the results for non-phosphorylated peptides fragmented by CID-IT were similar. It was determined that CID-IT had only a low level of dependency on NCE and that peptides were readily identifiable across a large range of NCE values (25 – 60%).

However, for HCD, there was a much greater dependency on NCE for identification. Using an NCE of 40% ensured optimal identification results (Figure 2-14). At higher NCE values, the excessive fragmentation and the formation of iTRAQ reporter ions correlated with a subsequent decrease in identification success. In agreement with previous results of the phosphopeptide standard, an NCE value of 45% represented the optimal compromise to identify and quantify  $2^+$  charged precursor ions. Barely higher NCE values were required for  $3^+$  and  $4^+$  charged precursors with 60% and 65 – 70% NCE, respectively. This ensured to obtain a sufficient number of iTRAQ reporter ions signals, but yet a high number of tandem MS spectra were identifiable.

## 2 Developing a Platform for relative Quantification using iTRAQ

### 2.1.3.3 Biological Sample

The tryptic human macrophage digest had a higher complexity compared to the ten protein-mix. From previous experiments using a 2D-LC-ESI MS approach, the number of identifiable proteins from human macrophages was enumerated to be within the range of 1,000 to 1,500 (with FDR  $\leq$  1%; 2 unique peptides per protein).

Peptides derived from tryptic digest were iTRAQ-4Plex labeled and subjected to 1D-LC-ESI MS analysis on the LTQ Orbitrap Velos; this is similar to the ten protein-mix sample preparation. The CID-IT and HCD NCE values were sequentially increased from 0 to 85%. Peptides were separated using a linear three hour gradient; each NCE value was analyzed in duplicates to diminish technical variations.

### Pre-Processing of Tandem MS Data

To increase the number of precursor ions selected, the number of repeating scans of the same precursor ion (m/z) was set to three and one for the first and second replicate measurements, respectively. The mean number of MS1 scans was barely affected by the different strategies. On average, approximately 4,500 for the first and 5,000 for the second technical replicate were acquired (Figure 2-17). However, the number of MS2 spectra of the first technical replicate (25,000) was doubled compared to the second technical replicate (10,000 MS2 spectra). The lower number of acquired MS/MS spectra for the second technical replicate may be both beneficial (i.e., a higher chance to select low-abundant precursor ions) and detrimental (i.e., less MS2 spectra for a specific precursor ion for data merging processes). By combining both technical replicates, the input (in terms of tandem MS spectra) to extract peptide features as a function of NCE value for CID-IT and HCD was at its best.

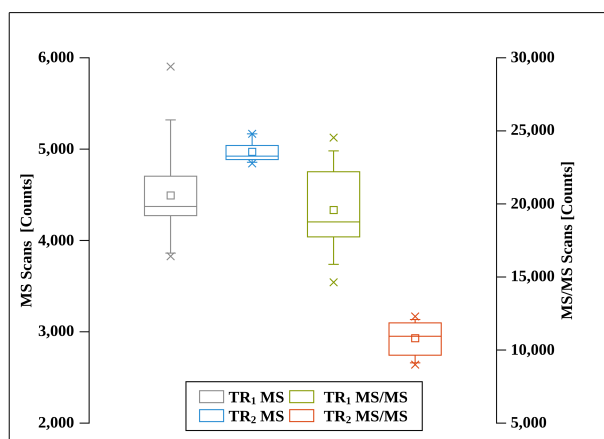


Figure 2-17: Whiskers-Box-Plot for both technical replicates TR1 (gray, green) and TR2 (blue, orange) showing the number of MS (gray and blue) and MS/MS scans (green and orange). The diversity of different MS/MS scans was related to the dynamic exclusion settings, as the number of repeating scans was lowered from 3 (TR<sub>1</sub>) to 1 (TR<sub>2</sub>) to increase the possibility for selecting low abundant precursor ions and precursors of different charge states, respectively.

### Analyzing the Retention Time Difference between Technical Replicates

In total, 36 LC-ESI MS runs were analyzed (2 TR x 18 LC MS runs). Although the LC gradient was not changed for the first and second technical replicate measurement, the retention time of peptides changed, as it is illustrated in Figure 2-18.

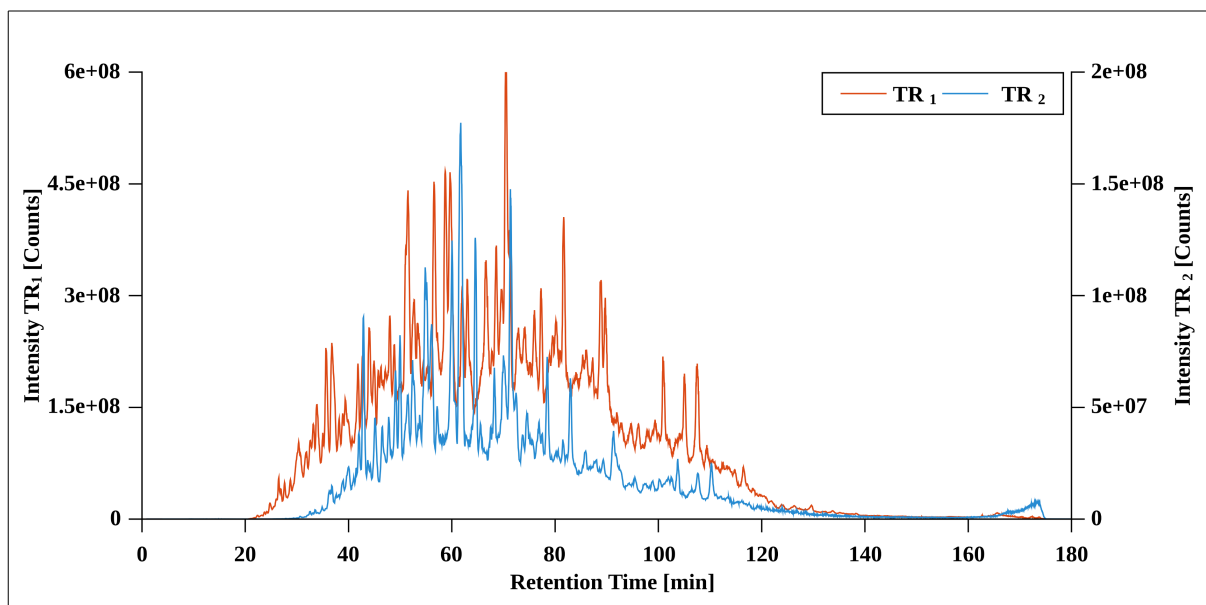


Figure 2-18: Total Ion Chromatogram of two technical replicate ( $TR_1$  (orange) and  $TR_2$  (blue)) demonstrating the effect of shifted retention times. The samples Normalized Collision Energy settings were: CID-IT 20%, HCD 50%.

The first peptides of the first technical replicate were detected at minute 20 (approximately) and at minute 30 for the second replicate. In Thermo Scientific's acquisition Software DCMS Link in combination to Chromeleon Xpress, the pressures for the loading- and micro pump were recorded in .raw files. Using these pressure profiles it was monitored, that after switching the 10-port valve from the loading- into the micro-pump flow, the LC column pressure for the micro-pump dropped sharply from approximately 170 to 120 bar. It was assumed, that this was the reason of the prolonged retention time. However, the column pressure drop only affected the retention time of peptides, but not their order of separation. Moreover, it was observed, that the effect of retention times difference between the first and second replicate ( $\Delta_{RT}$ ) was more significant for lower retention times ( $m/z$  560,  $\Delta_{RT} \sim 6$  min) than with prolonged retention times ( $m/z$  729,  $\Delta_{RT} \sim 2.3$  min) as illustrated in Figure 2-19.

## 2 Developing a Platform for relative Quantification using iTRAQ

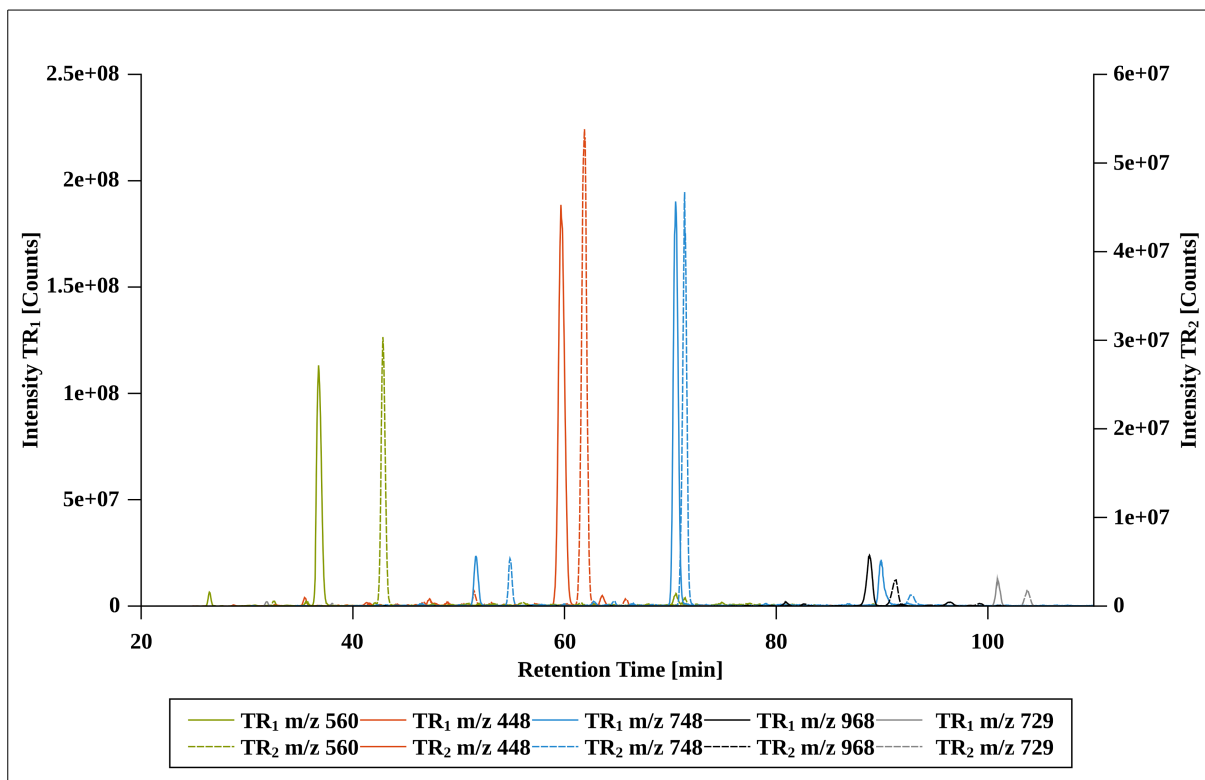


Figure 2-19: Extracted Ion Chromatograms (XICs) of five different precursor masses ( $m/z$  values) being roughly equally distributed via the LC separation from minute 40 to 100 to highlight the extend of retention time shifting.

### Retention Time Re-Calibration

The retention time can be used as a feature (in combination to precursors  $m/z$  and charge state) to re-map identify peptides at elevated NCE values if a series of replicates of the same sample is measured by LC-ESI MS. Frequently, this process is applied in label-free quantification experiments to align different samples; a process referred to as *re-match between runs* (e.g., in MaxQuant; (285)).

To finally use the retention time as a peptide re-mapping feature in combination to the precursor  $m/z$  value and charge state, all 36 LC-ESI MS were subjected to retention time re-calibration.

In the first step, a peptide master list was created in which all extracted LC-ESI MS run features were merged. This included information about the sequence, modification and charge state (static features). Additionally, the retention time and  $m/z$  value of peptides (dynamic features) were averaged via all 36 LC runs recorded to balance out the observed retention time shift (shown in Figure 2-18 and 2-19).

In the next step, each Peptide-Spectrum-Match (PSM) was compared to the master list. Both retention times (of a particular peptide species) from the master list and the LC-ESI MS run, were exported into a .txt file. This was done for each LC MS experiment separately. Additionally, the sequence, modification, charge states and  $m/z$  value of the peptide was exported.

The stepped NCE values applied for CID-IT and HCD (e.g., CID-IT 0% and HCE 30% within an LC-ESI MS run) allowed to receive a sufficient number of identifications for each .raw file. It is noteworthy, that the identifications obtained for CID-IT and HCD tandem MS spectra were merged for retention time re-calibration.

## 2.1 iEM – iTRAQ Evaluation Method

Finally, the lists containing the information about the retention times for the experimental and the master file were imported into QtiPlot. To obtain parameters to re-calculate retention times, a linear curve fitting approach was used. This is illustrated in Figure 2-20 for two technical duplicates.

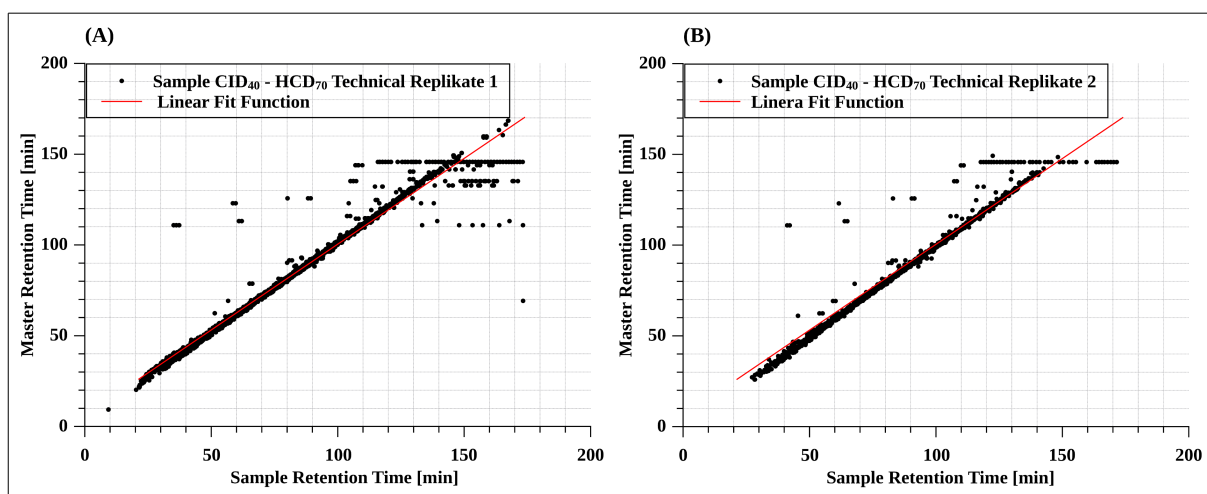


Figure 2-20: XY-scatter plot for sample CID 40 – HCD 70 for both the 1<sup>st</sup> and 2<sup>nd</sup> technical LC run measured. For sample normalization, a linear fit model was applied, as samples (=peptides) elution profile was shifted in total rather than changing the LC profile. The linear fit function was exported for each of the 36 runs and used for recalculation of retention times.

Using all 36 data sets separately in QtiPlot, the linear fit function for each LC- ESI MS run could be exported. Their corresponding curve fitting parameters are summarized in Table 2-4.

The retention time alignment plots (Figure 2-20) and the curve fitting parameters revealed that both, the retention time of the master file and each LC-ESI MS run were narrower for peptides eluting before minute 120 (approximately). At higher retention times (approx. 150 min), the number peptides that did not fit the linear curve fitting ascended. Potentially, these were peptides not sufficiently eluted from the trap-column in the previous LC-ESI MS run.

However, the linear curve fitting parameters (A & B in Table 2-4) were subsequently used to re-calculate each retention time point in the .mgf files. As the information about the fragmentation type is not encoded in .mgf files, this process was performed for CID-IT and HCD separately.

For the final database search, the retention time re-calibrated .mgf files were imported into the *Spectrum Files* node instead of the original .raw files (Figure 2-1). To obtain peptide list which were compatible to the previously compiled tools (e.g., chapter 2.1.3.2), peptide and protein features for CID-IT and HCD were exported into a single .txt file.

## 2 Developing a Platform for relative Quantification using iTRAQ

Table 2-4: The linear curve fitting ( $y = A*x + B$ ) parameters A, B and R<sup>2</sup> related to each LC run used for retention time recalculation in .mgf files.

Sample Information [AT & NCE]	TR <sub>1</sub>			TR <sub>2</sub>		
	A	B	R <sup>2</sup>	A	B	R <sup>2</sup>
CID 00% HCD 30%	0.93	7.02	0.96	0.99	-0.61	0.98
CID 05% HCD 35%	0.89	9.54	0.94	1	-0.77	0.98
CID 10% HCD 40%	1.04	-12.05	0.94	1.01	-2.13	0.96
CID 15% HCD 45%	1.01	-5.3	0.96	1.01	-1.51	0.95
CID 20% HCD 50%	0.88	8.54	0.92	1	-0.74	0.98
CID 25% HCD 55%	0.89	8.34	0.92	1.01	-0.89	0.97
CID 30% HCD 60%	0.93	6.83	0.95	1	-0.69	0.97
CID 35% HCD 65%	0.93	7.19	0.95	0.99	0.42	0.97
CID 40% HCD 70%	0.97	3.02	0.96	1	-0.49	0.97
CID 45% HCD 75%	0.95	5.9	0.96	1	-0.71	0.97
CID 50% HCD 80%	0.94	6.31	0.96	1	-0.38	0.96
CID 55% HCD 85%	0.96	5.69	0.96	1.01	-0.65	0.96
CID 60% HCD 00%	0.96	5.7	0.96	1.01	-1.51	0.96
CID 65% HCD 05%	0.97	5.16	0.96	1	-0.89	0.97
CID 70% HCD 10%	1.01	-4.29	0.97	1	-1.47	0.95
CID 75% HCD 15%	1.01	-3.46	0.96	1	-1.23	0.96
CID 80% HCD 20%	0.99	-1.14	0.97	1	-0.71	0.97
CID 85% HCD 25%	0.98	-0.12	0.96	0.99	-0.05	0.97

### Peptide Identification and Quantification in Dependence of NCE

To evaluate the performance of CID-IT and HCD for iTRAQ labeled peptides, a human macrophage digest was analyzed to evaluate the optimal conditions for identification and quantification, as it was done for the phosphopeptide standard and the ten protein-mix.

The number of peptides identified upon CID-IT and HCD fragmentation were plotted as a function of the NCE (Figure 2-21). As expected from previous results, the number of peptides identified (unique and totally) upon CID-IT excitation was steady at various NCE values (25 to 55%) for all charge states monitored (2<sup>+</sup> to 4<sup>+</sup>; Figure 2-21 (A, C, E)). However, the number of identification was always significantly lower (approximately 50%) than those identified by HCD (Figure 2-21 (B, D, F)).

For HCD, the number of peptides identified was highly dependent on the NCE value. The highest number of peptides (unique and PSMs) were identified in the NCE range of 40 ± 10%. However, the identification appeared to be more sensitive to NCE increases for 2<sup>+</sup> than for 3<sup>+</sup> and 4<sup>+</sup> charged precursors, as the extent of the *bell-curve shape* was most significant for 2<sup>+</sup> charged precursors. This curve shape was also observed for the remaining charge states.

## 2.1 iEM – iTRAQ Evaluation Method

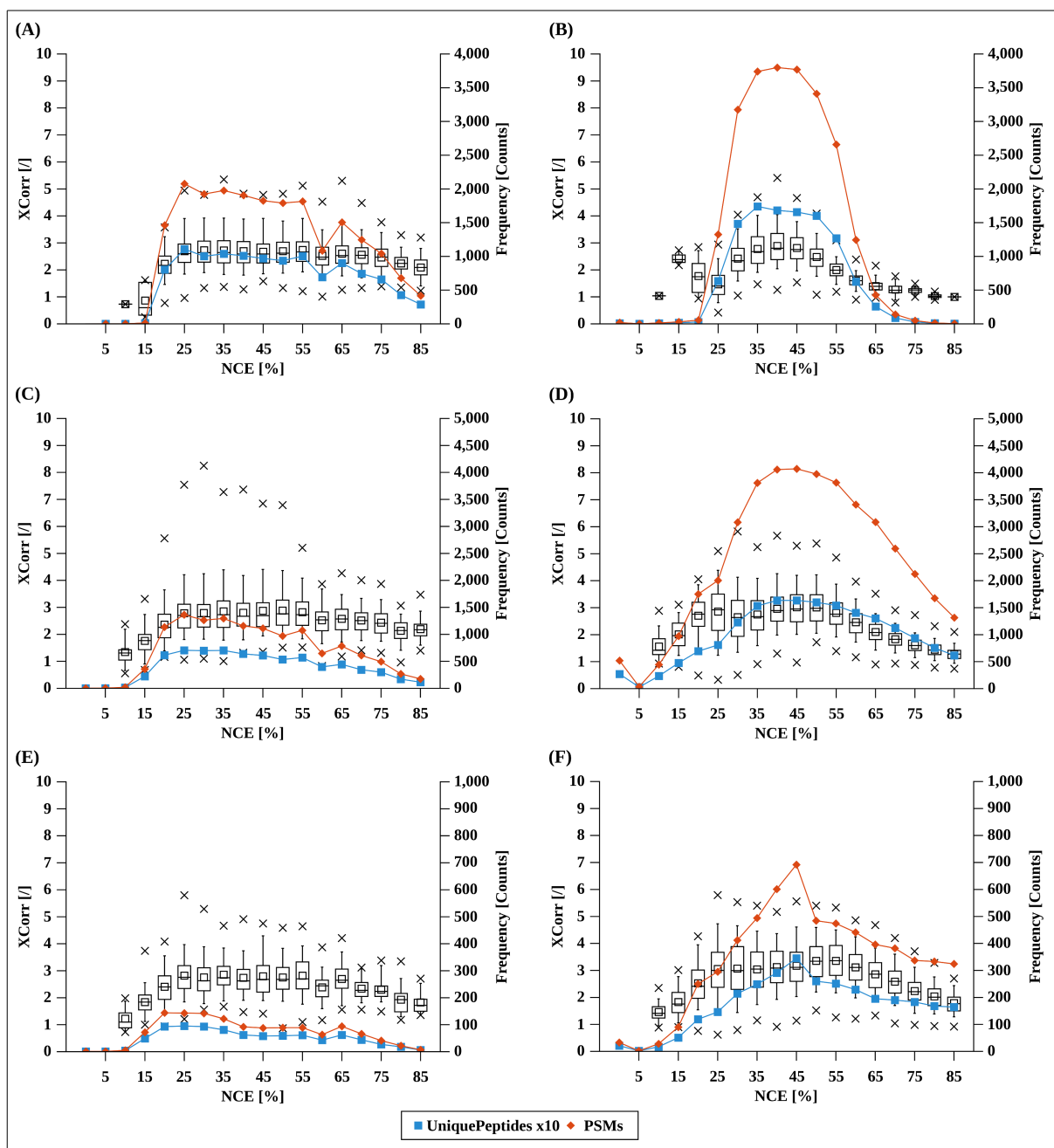


Figure 2-21: Comparison of SEQUESTs XCorr (Whiskers-Box-Plots) and the number of unique and total number of peptides identified via different Normalized Collision Energies applied for CID (A, C, E) and HCD (B, D, F) for 2<sup>+</sup> (A, B), 3<sup>+</sup> (C, D) and 4<sup>+</sup> (E, F) charged precursor ions.

For quantification, the normalized intensity, the relative intensity and the mean signal intensity were calculated for HCD as a function of the NCE value and the charge state of the precursor (Figure 2-22).

For 2<sup>+</sup> charged precursors, no iTRAQ reporter ion signals were observed at NCE values below 30%. The normalized intensity rose sharply between 35 and 45% and stayed constant (Figure 2-22 (A)). Contrary, the relative intensity reached its maximal value at elevated NCE values (> 65%; Figure 2-22 (D)). The mean intensity (Figure 2-22 (G)) decreased linearly after applying NCE values greater than 45 – 60%.

## 2 Developing a Platform for relative Quantification using iTRAQ

The curve shape of the normalized and relative intensity of the 3<sup>+</sup> and 4<sup>+</sup> charged precursor ions was remarkably different. Instead of the sharp rise, the intensity (relative and normalized) increased almost linearly with NCE (e.g., Figure 2-22 (E)). However, the maximal mean intensity was at 60% NCE (Figure 2-22 (H & I)), whereas the mean intensity at 45% is speculated to represent an outlier. In summary, similar optimal NCE values were explored for the peptide digest obtained from human macrophages to identify and quantify iTRAQ labeled peptides, compare to the previously analyzed iTRAQ labeled phosphopeptide standard and ten protein-mix.

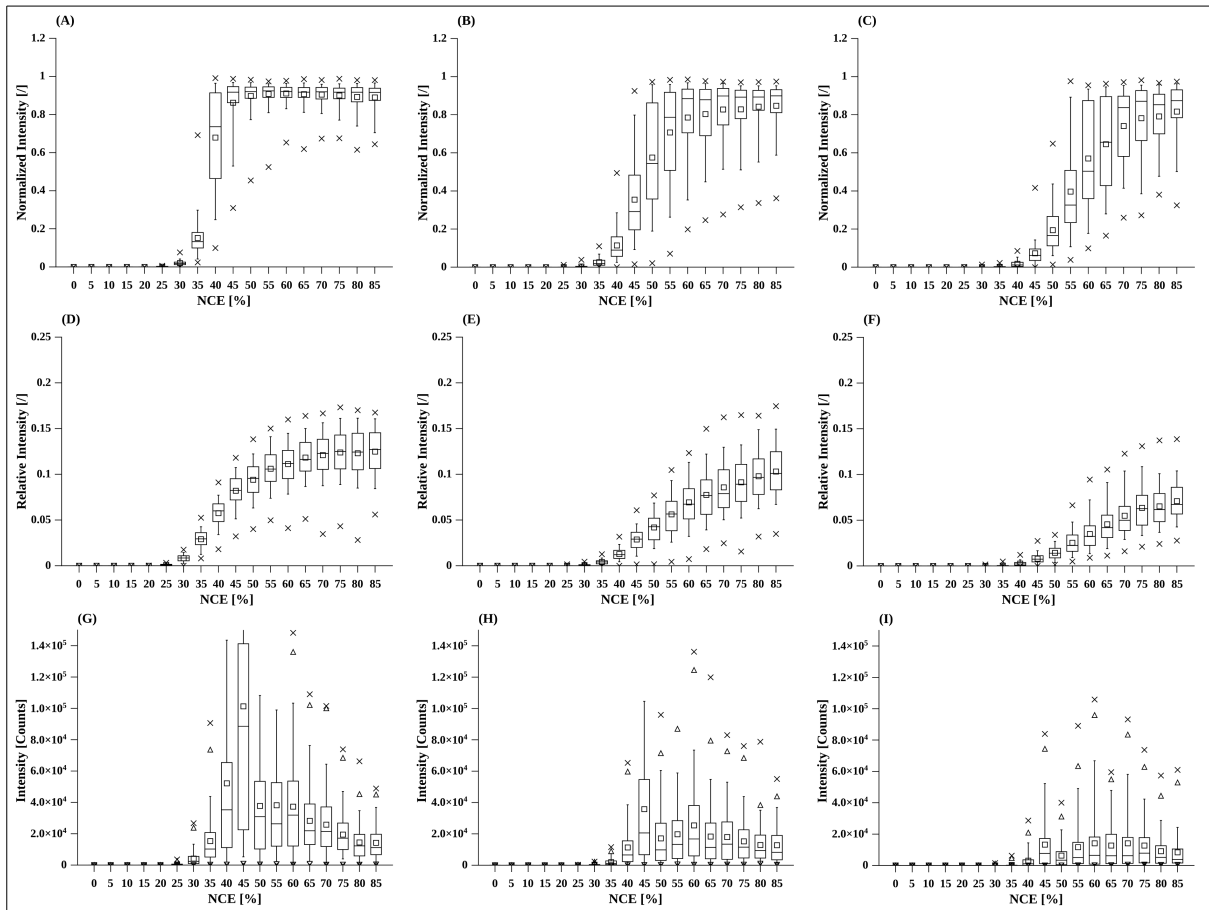


Figure 2-22: Whiskers-Box-Plots of iTRAQ reporter ion intensities after HCD fragmentation for the protein-mix in dependence of NCE applied for 2<sup>+</sup> (A, D, G), 3<sup>+</sup> (B, E, H) and 4<sup>+</sup> (C, F, I) charged precursor ions. (A - C) Normalized Intensities; (D - F) Relative Intensities; (G - I) Intensities (raw values).



### 2.1.4 Conclusion

The applied Normalized Collision Energy (NCE) is a parameter affecting spectra quality for identification and quantification, respectively. For identification, varying the NCE did not affect CID-IT and CID-MSA(-IT) spectra significantly if values between 25 – 55% were applied. However, for HCD optimal NCE values for 2<sup>+</sup>, 3<sup>+</sup> and 4<sup>+</sup> charged precursors were monitored: 2<sup>+</sup> charged precursors were highly affected by elevated NCE values, whereas the dependency was lower for precursors with charge states >2<sup>+</sup>. This was observed for both, non- and phosphorylated peptides. Generally, 4<sup>+</sup> charged peptides were tolerant to a wide range of applied NCE values. Although the NCE values mentioned here were obtained from a single LTQ Orbitrap Velos instrument, the values reported should be comparable to different instruments as the NCE value is re-adjusted during the instrument calibration and tuning procedure.

Low m/z fragment ions (such as iTRAQ reporter ions) cannot be stabilized in ion traps after CID-IT activation. To overcome this problem, HCD must therefore be applied. Generally, for reporter ion generation, MS/MS spectra with intense iTRAQ reporter ions are required. For 2<sup>+</sup> charged precursors, an NCE of 45% was evaluated to be sufficient to receive intense iTRAQ reporter ions and stable iTRAQ reporter ratios. This also represented the NCE value, at which iTRAQ reporter ions represent the base peak in MS/MS spectra. For 3<sup>+</sup> charged precursors, an elevated NCE was required (> 55%); which increased further (> 65%) for 4<sup>+</sup> charged precursor ions.

The acquisition of both, CID(-MSA)-IT and HCD in parallel represented an (almost) optimal trade-off for identification and quantification of iTRAQ labeled phosphorylated- and non-phosphorylated peptides, respectively. Although HCD MS/MS spectra contain both information (fragment and iTRAQ reporter ions), (phospho)peptide identification and correct phosphorylation site mapping (for phosphopeptides only) was still challenging if solely HCD was applied, as the number of ions with sequence information decreased with increasing NCE value. Generally, the process of phosphopeptide identification is enhanced by the application of both, CID(-MSA)-IT and HCD. While the MS full scan acquisition in the Orbitrap mass analyzer, precursor ions begin to be fragmented by CID-IT and are being analyzed in the ion trap in parallel to the completion of the MS full scan in the Orbitrap mass analyzer. This reduces the overall scan cycle compared to sequential MS full scan - CID/HCD measurement approaches (see Figure 1-12).

From the technical point of view, CID-IT takes advantage of sensitivity and speed, whereas HCD (recorded in Orbitrap mass analyzer) benefits from high mass accuracy. Consequently, complement data sets regarding measurement and fragmentation conditions are obtained by simultaneous application of both. For phosphopeptide analysis, CID-MSA-IT is superior to CID-IT as it creates sufficient b- and y-ion fragmentation. This minimizes neutral loss signals (e.g., [M-P]) but requires additional scan time. An additional benefit of CID(-MSA)-IT is that these spectra are more reproducible. Changing the NCE value did not effect the identification success as they do not show iTRAQ reporter ions. Consequently, CID data are more robust in terms of stable identifications, as the XCorr is directly correlated to the number of detected fragment ions observed. In contrast, HCD spectra benefit from the Orbitrap mass analyzer and the shorter but energetically more intense fragmentation processes, which leads to (at least) comparable identification results between CID-IT and HCD.

## 2 Developing a Platform for relative Quantification using iTRAQ

Moreover, spectra shape differs in terms of intensity distribution for b- and y-ions. In addition, a- and significant diagnostic (e.g., for pY-peptides at m/z 216.04) or other immonium ions can only be detected after HCD excitation and acquisition in the Orbitrap mass analyzer.

Interestingly, the number of identifications obtained by CID-IT dropped for the biological sample, where HCD clearly outperformed CID-IT. This might be an effect of the increased search space (human + *M. tuberculosis* vs. *E. coli* FASTA file) and the lower resolution of ion traps. However, the phosphopeptide and the ten protein-mix were measured 2012 and 2013, whereas the peptide digest from human macrophages was measured 2014. An instrument maintenance was performed prior sample measurement, which also included a clean of the ion trap. Possibly, particles (such as dust) might be present in the ion trap, which leads to de-charging of fragment ions. Consequently, less ions can be recorded in the ion trap which decreases the sensitivity of the ion trap, and could (at least partially) explain the drop in identification rate of CID-IT compared to HCD.

## 2.2 n-iTRAQ – Handling of Nominal Isobaric Labelings

### 2.2.1 Introduction

By variation of the Normalized Collision Energies applied for CID-IT, CID-MSA-IT and HCD, optimal fragmentation conditions for identification and quantification of iTRAQ-4Plex labeled peptides were established (see chapter 2.1). However, the isobaric labeling reagent iTRAQ-4Plex used in the previous section, is available in four different forms which generate the particular reporter ions observed at  $m/z$  114 to 117. Due to the composition of the reagents, encompassing different combinations of  $^{13}\text{C}$  and  $^{15}\text{N}$  isotopes, different modification masses are observed (see Figure 1-3): iTRAQ<sub>114</sub> has the highest (144.105918 Da) and iTRAQ<sub>115</sub> the lowest mass (114.099599 Da) of the four iTRAQ reagents; whereas iTRAQ<sub>116&117</sub> share the same number of carbon and nitrogen isotopes and, consequently have the same mass (144.102063 Da).

As a default setting for iTRAQ-4Plex experiments, database searches are performed with the modification mass of iTRAQ<sub>116&117</sub>, according to the Unimod recommendation (Accession # 214). During *in-silico* peptide index creation, this value is added to each static (N-terminal and K) and dynamic (Y) modification (see Figure 1-1). However, for peptides with intense iTRAQ<sub>114</sub> or iTRAQ<sub>115</sub> reporter ions, the *in-silico* calculated values have been demonstrated to differ from the experimental mass even (up to 3.8 ppm) for newly tuned and calibrated instruments (274). Peptides of severe differentially abundant proteins tend to contain intense iTRAQ reporter ion signals of one channel. Consequently, these peptides might be lost during database searches if the precursor mass tolerance window is set to narrow due to the only nominal isobaric iTRAQ-4Plex reagents.

High resolution mass spectrometers such as the LTQ Orbitrap Velos, can resolve ions from a resolution of 7,500 to 100,000 (Orbitrap mass analyzer). The highest resolution mode permits to differentiate ions at  $m/z$  400 with  $\Delta = 0.004$ . This is not sufficient to baseline separate two iTRAQ labeled precursor ion peaks (i.e., the same peptide with (i) iTRAQ<sub>114</sub> and (ii) iTRAQ<sub>115</sub> labeling), but it is sufficient to monitor how the precursor mass behaves at varying iTRAQ<sub>114</sub> to iTRAQ<sub>115</sub> ratios.

In order to demonstrate the potential errors introduced due to non isobaric iTRAQ reagent masses, ten samples with different iTRAQ reporter ion ratios are analyzed by LC-ESI MS and their *in-silico* calculated features systematically monitored: This is the (*in-silico* introduced) precursor mass shift at different iTRAQ<sub>114</sub> to iTRAQ<sub>115</sub> ratios, and the number of peptide identifications at different precursor mass tolerance windows set for database searches. The final goal is to present an alternative algorithm to improve the precursor ion mass calculation of iTRAQ-4Plex labeled peptides.

## 2 Developing a Platform for relative Quantification using iTRAQ

### 2.2.2 Material and Methods

#### 2.2.2.1 Chemicals

Chemicals used in this study are summarized in chapter 4.1.

#### 2.2.2.2 Sample Preparation

##### **Protein Digestion and iTRAQ-4Plex Labeling**

The iTRAQ labeled ten protein-mix (chapter 2.1.2.1) was used for sample preparation and LC-ESI MS analysis. The four iTRAQ labeled samples were combined in different ratios, resulting in a total of ten samples to simulate differentially abundant peptides, which are summarized in Table 2-5.

Table 2-5: Mixing schema of the ten protein-mix to obtain samples with (simulated) different peptide ratios. iTRAQ-114 to iTRAQ-117 represent the starting material after labeling, which were also measured.

Sample [Name]	iTRAQ ratio [I]	# MS/MS Spectra [Counts]
iTRAQ-114	1:0:0:0	2232.00 ± 119.65
iTRAQ-115	0:1:0:0	2789.00 ± 479.39
iTRAQ-116	0:0:1:0	2583.67 ± 379.76
iTRAQ-117	0:0:0:1	2620.67 ± 94.24
A <sub>(1:1:1:1)</sub>	1:1:1:1	2417.00 ± 56.04
B <sub>(4:4:1:1)</sub>	4:4:1:1	2760.67 ± 49.40
C <sub>(4:3:1:1)</sub>	4:3:1:1	2461.67 ± 127.58
D <sub>(3:4:1:1)</sub>	3:4:1:1	2346.67 ± 621.07
E <sub>(4:1:1:1)</sub>	4:1:1:1	2539.00 ± 68.51
F <sub>(1:4:1:1)</sub>	1:4:1:1	2586.33 ± 29.50

#### 2.2.2.3 Mass Spectrometry and Database Searches

##### **Instrument Setup**

For LC-ESI MS measurements, the LTQ Orbitrap Velos coupled to a Dionex U3000 HPLC system was employed (see chapter 2.1.2.2 and chapter 4.1 for additional details). After peptide loading and desalting (4 min; 30 µL/min), peptide separation (300 nL/min) was performed with a linear gradient from 5% to 40% B in 40 min, followed by a column wash step (total: 12 min) with a linear gradient from 40% B to 95% B. Inter run equilibration of the columns was set to 9 minutes. A Top 5 HCD method was written manually using vendors software: First, a full scan at a resolution of 100,000 with activated prescan was performed. For MS/MS, resolution of Orbitrap measurements was decreased to 7,500. The activation time was 0.1 ms, default charge was set to 2<sup>+</sup>, isolation width m/z 3. To obtain outbalanced HCD fragment ion spectra for simultaneous identification and quantification, the NCE was set to 47.5%. This value represents a compromise between 2<sup>+</sup> and 3<sup>+</sup> charged precursor ion fragmentation. Dynamic exclusion was enabled for all runs. The repeat count was set to 3 with a repeat duration of 20 sec and precursors, selected for MS/MS, were set for dynamic exclusion for 30 sec. Lock mass correction was enabled using the polydimethylcyclosiloxane signal m/z 445.120024.

### **Database Searches**

#### *Fixed Precursor Mass*

Recorded .raw files were processed using Proteome Discoverer (Version 1.4). Each of the ten samples (see Table 2-5) was searched in a MudPIT approach combining all three technical injections. The FASTA file was the same as applied before for the ten protein-mix (see chapter 2.1.2.3).

The workflow applied was build up by 9 different nodes (similar to Figure 2-1) with default settings if not otherwise specified: (0) Spectrum Files; (1) Spectrum Selector; (2) Scan Event Filter; (3) SEQUEST HT (see below); (4) Target Decoy PSM Validator (Target FDR (Strict): 0.01 and Target FDR (Relaxed): 0.05); (5) Reporter Ions Quantifier (iTRAQ-4Plex, respectively (Thermo Scientific Instruments; Integration Tolerance 20 ppm; Most Confident Centroid as Integration Method; Additional Settings: Show the Raw Quan Values; Apply Value Corrections, use All Peptides enabled and deactivated Experimental Bias).

The following settings were used for SEQUEST HT (3), enzyme name: trypsin (semi specific); max. missed cleavage sites: 2; min. peptide length: 6; max. peptide length: 144; precursor mass tolerance: 10 ppm; fragment mass tolerance: 0.02 Da; weight of all ions but b and y ions was 0, weight of b & y Ions: 1; dynamic modifications: phosphorylation (S, T, Y); oxidation (M), iTRAQ4plex (Y); static modifications: peptide N-Terminus: iTRAQ4plex (any N-Terminus); carbamidomethyl (C), iTRAQ4plex (K).

For evaluation of mass deviations introduced *in-silico*, Peptide-Spectrum-Matches obtained from Proteome Discoverer database search were filtered for high confidence (FDR 1%) and exported to .xlsx files.

#### *Variable Precursors Mass Tolerance*

The precursor mass tolerance window was varied from 0 (exactly: 0.01) to 20 ppm in 1 ppm steps for the workflow applied (see above). For analysis, only the four sample iTRAQ-114 to iTRAQ-117 combined in a MudPIT approach were used. All 12 runs (4 samples, 3 technical replicates) were combined into a single database search per each of the 21 ppm values set. Lists containing the Peptide-Spectrum-Matches were exported with high confidence only (1% FDR) and used for further data analysis.

### **Data Processing**

#### *Data Conversion*

For .raw-file conversion into a text base format (.mgf), Proteome Discoverer was utilized. The workflow used before was applied, but nodes (3) – (5) were replaced by a Spectrum Exporter Node.

#### *Data-Analysis (Fixed Precursor Mass Tolerance Window)*

Lists containing the Peptide-Spectrum-Matches were filtered for the *in-silico* calculated ppm value. Each entry was rounded to integer numbers (bin size 1 ppm) and the number of features within each bin was counted (Figure 2-23 and 2-24 (A)). Additionally, the PSM list of iTRAQ-114 was split into three different parts: the first contained peptides with 1x iTRAQ<sub>114</sub> modification, the second with 2x iTRAQ<sub>114</sub> modification and the third with

## 2 Developing a Platform for relative Quantification using iTRAQ

> 2x iTRAQ<sub>114</sub> modifications. Each of the lists was also used for feature calculation (Figure 2-24 (B)).

### Data-Analysis (Variable Precursor Mass Tolerance Window)

To demonstrate the dependency of PSMs to the precursor mass tolerance window set, each of the 21 PSM lists was searched separately (batch mode) with an *in-house* developed GAMBAS script. The number of PSMs for each sample (including the three technical replicates) and each precursor ppm value set was counted and exported into a tab-separated file. This list was imported into QtiPlot for data visualization (Figure 2-26).

### Algorithm to Re-Calculate the Precursor Mass (Fixed Precursor Mass Tolerance Window)

Three *in-house* developed VBA scripts were used to re-calculate the precursor mass (m/z value). Script (i) exported the information about iTRAQ reporter ions and header informations (retention time, precursor charge state, scan number, precursor m/z and raw file name) into a tab-separated .txt file (referred to as iTRAQ-list). For iTRAQ reporter ion signals, their intensity and the corresponding m/z values (withing 5 ppm) were exported. Additionally, the relative intensities of iTRAQ reporter ions were calculated (Equation 2-6). The iTRAQ<sub>relative intensity</sub> in combination to the known modification mass of iTRAQ 114-117 (see Figure 1-3; Equation 2-7) was used to calculate a specific iTRAQ modification mass for every single tandem MS spectrum (Equation 2-8). The amino acid masses (monoisotopic) used for precursor mass (m/z value) calculation are summarized in chapter 4.2.

$$\text{iTRAQ}_{\text{relative Intensity}}(x) = \frac{\text{iTRAQ-Intensity}(x)}{\sum_{x=114}^{x=117} \text{iTRAQ-Intensity}(x)} \quad \text{Equation 2-6}$$

$$\text{iTRAQ}_{\text{part}}(x) = \text{iTRAQ}_{\text{relative Intensity}}(x) \times \text{iTRAQ}_{\text{mz}}(x) \quad \text{Equation 2-7}$$

$$\text{iTRAQ}_{\text{corrected}} = \sum_{x=114}^{x=117} \text{iTRAQ}_{\text{relative Intensity}}(x) \quad \text{Equation 2-8}$$

Script (ii) imported the information after database search (PSM list) and the previously created iTRAQ-list [see (i)]. Firstly, features of the identified peptides were loaded (sequence, modification, raw file, retention time, precursor charge state, tandem MS scan number, mass accuracy, precursor mass (m/z)) and the number of modifications (iTRAQ, oxidation and carbamidomethyl). Afterwards, the iTRAQ-list (i) was searched for the corresponding *spectrum hit*. A hit was true if both (PSM and entry within the iTRAQ-list) share the same RT, charge, precursor mass, scan number and raw file name.

The previously calculated iTRAQ<sub>corrected</sub> value from the iTRAQ-list (i) was then used to re-calculated the peptides precursor mass and the corresponding mass deviation (in ppm). To test the applied workflow, the calculation was repeated replacing iTRAQ<sub>corrected</sub> with its classical value (iTRAQ<sub>116&117</sub>:144.120026 Da).

Script (iii) was used for data visualization. The mass deviation values (for each, the original ppm value obtained by Proteome Discoverer and the re-calculated) were rounded to integer values and the number of hits within one bin (one bin = 1 ppm) was counted.

## 2.2.3 Results and Discussion

### 2.2.3.1 Monitoring ppm Deviations Introduced by Different iTRAQ Ratios

To analyze differences (introduced by varying iTRAQ ratios) for *in-silico* calculated precursor masses, ten samples were analyzed by MS. Six of these samples showed different iTRAQ ratios. The latter four (iTRAQ-114 to iTRAQ-117) represented the maximal permissible value, as they only contained one iTRAQ labeling (see Table 2-5 for details).

Sample A showed iTRAQ ratios of 1:1:1:1 (iTRAQ<sub>114</sub> : iTRAQ<sub>115</sub> : iTRAQ<sub>116</sub> : iTRAQ<sub>117</sub>). After performing a database search in MudPIT approach 2,016 PSMs were exported at 1% FDR (see Figure 2-23) and the average mass accuracy was calculated to be 0.5 ppm via all PSMs. Grouping the PSMs by creating bins of 1 ppm size by rounding to integer values, most of the peptides were identified in the  $\pm 0$  ppm (37%) and +1 ppm (38%) bin, respectively. This represented approximately 75% of all identified peptides. One tenth of the peptide identification was found in the -1 ppm bin. This mass deviation distribution and accuracy was expected due to the application of lock mass correction for MS and MS/MS analysis, as shown recently (85). For this test sample A<sub>(1:1:1:1)</sub>, no negative effect was observed using the iTRAQ<sub>116&117</sub> mass (144.102063 Da) for database searches.

Increasing the amount of both, iTRAQ<sub>114</sub> and iTRAQ<sub>115</sub> (sample B<sub>(4:4:1:1)</sub>; Figure 2-23 (B)) in relation to iTRAQ<sub>116</sub> and iTRAQ<sub>117</sub>, a similar distribution to sample A<sub>(1:1:1:1)</sub> was observed. However, the percentage of PSMs within the bins  $\pm 0$  ppm and +1 ppm was lowered to 50%. A further of 10% was identified within the bins -1 and +2 ppm, respectively. The averaged mass deviation was still very accurate at + 0.7 ppm.

For the sample C<sub>(4:3:1:1)</sub> a shift to slightly higher mass deviations (+ 1.7 ppm) was observable (Figure 2-23 (A)). Beside this, also the distribution in the bin 0 to +2 ppm changed. Only 14% of PSMs were identified in the  $\pm 0$  ppm, whereas approximately 30% were identified within the +1 ppm bin. A further of 20% was found with a mass accuracy of 2 ppm. The shift to higher masses deviations for this sample C<sub>(4:3:1:1)</sub> was introduced as iTRAQ<sub>114</sub> has a higher mass and intensity than iTRAQ<sub>115</sub>. Moreover, the outbalancing effect, as shown for sample A<sub>(1:1:1:1)</sub> and B<sub>(4:4:1:1)</sub> of iTRAQ<sub>114</sub> and iTRAQ<sub>115</sub>, was not sufficient, resulting in the observed mass shift. For sample D<sub>(4:3:1:1)</sub> this was less prominent and can be explained by the smaller mass deviation of iTRAQ<sub>115</sub> (-0.002464 Da) compared to iTRAQ<sub>116&117</sub>, whereas iTRAQ<sub>114</sub> differs by +0.003855 Da (Figure 2-23 (B)).

Simulating four times higher abundant proteins (sample E<sub>(4:1:1:1)</sub>; Figure 2-23 (A)), the mass discriminating effect increased significantly. The average deviation was calculated to be 2.8 ppm, although lock mass correction was enabled. 50% of PSMs were grouped to the +2 and +3 ppm bin, and a further of 12% and 15% to the bins +1 and +4 ppm, respectively. Again, the mass shift effect was less prominent for samples with dominating iTRAQ<sub>115</sub> (sample F<sub>(1:4:1:1)</sub>; Figure 2-23 (B)). Approx. 66% of all PSMs were identified within the bins -1 and -2 ppm.

## 2 Developing a Platform for relative Quantification using iTRAQ

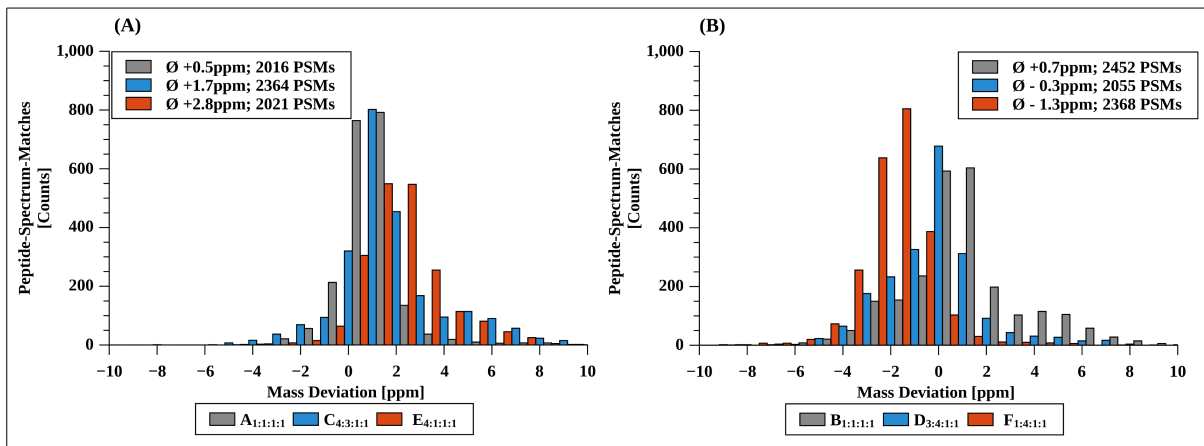


Figure 2-23: The frequency of peptides identified within the bins of -10 to +10 ppm for the samples with increasing content of iTRAQ<sub>114</sub> labeling (A, C & E) (left side (A)) and with increasing content of iTRAQ<sub>115</sub> (B, D, & F) (right side (B)).

Assuming, a protein is highly abundant in only one of the four biological entities, the mass deviation effect should increase if these samples were labeled with iTRAQ<sub>114</sub> or iTRAQ<sub>115</sub>. This was simulated by analyzing the four iTRAQ labelings separately (samples: iTRAQ-114, iTRAQ-115, iTRAQ-116 and iTRAQ-117 (Figure 2-24 (A)). Again, for all measurements, the lock mass correction was enabled to ensure comparable results. For the samples iTRAQ-116 and iTRAQ-117, the averaged mass deviation was 0.1 and 0.0 ppm, demonstrating perfectly the advantageous application of lock mass for internal re-calibration. For both samples, more than 55% of PSMs were grouped into the  $\pm 0$  ppm bin. Enlarging the monitored mass deviation window to -1 to +1, more than 92% of the PSMs were matched.

For the sample iTRAQ-115 and iTRAQ-114, the average mass shift was -2.8 and +4.3 ppm, respectively. Again, the effect was smaller for iTRAQ<sub>115</sub> labeled samples, as it was shown before. Moreover, the PSM per bin distribution was significantly different compared to the samples labeled with iTRAQ<sub>116</sub> and iTRAQ<sub>117</sub>. For the iTRAQ-114 sample, at least 10% of the PSMs were found in each of five bins (+2 to +6 ppm). Again, this was less intense for the sample iTRAQ-115 (-1 to -5 ppm).

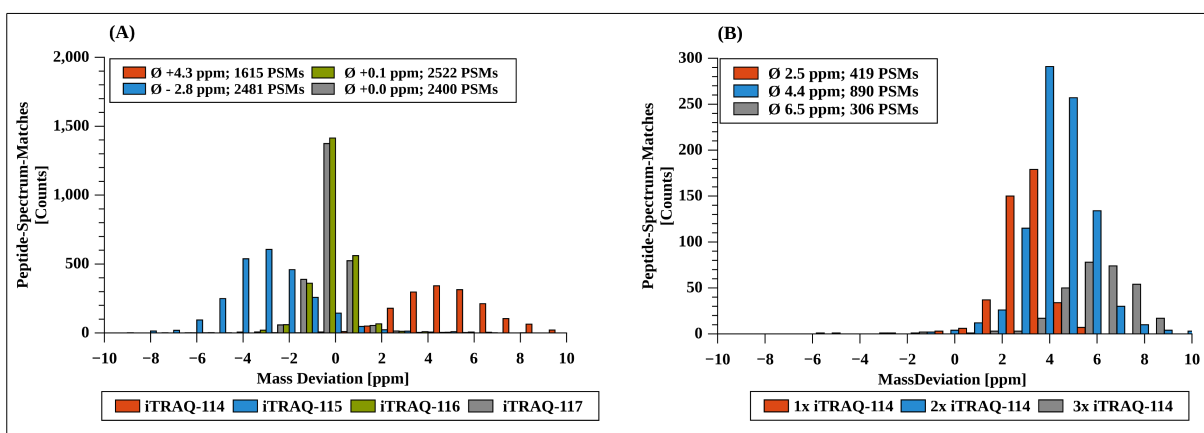


Figure 2-24: The frequency of peptides identified within the bins of -10 to +10 ppm for the samples iTRAQ-114 to iTRAQ-117 (A) and in-depths analyzing the number of iTRAQ labelings attached for sample iTRAQ-114 (B).



## 2.2 n-iTRAQ – Handling of Nominal Isobaric Labelings

By clustering the identified peptides into groups with one (N-terminal only), two (N-terminal + C-terminal K) or more (>2, N-terminal + C-terminal K + tyrosine or a missed cleavage) iTRAQ labelings attached, a correlation between the number of iTRAQ labelings and the mass deviation was discovered (Figure 2-24 (B)). For singly iTRAQ<sub>114</sub> labeled peptides, the average mass deviation shifted to +2.5 ppm, whereas it increased to +4.4 ppm for doubly labeled peptides. If a peptide species contained more than two iTRAQ<sub>114</sub> labelings, the mass deviation was calculated to be +6.5 ppm (in average).

In summary, the application of iTRAQ<sub>116&117</sub> as input-values to calculate the theoretical precursor masses and mass deviation values, respectively, did not alter the theoretically calculated precursor masses (m/z) if the iTRAQ reporter ions were equivalent distributed (sample A to D). However, for peptides with significantly differing iTRAQ ratios (sample E, F and iTRAQ-114, iTRAQ-115, respectively) the application of iTRAQ<sub>116&117</sub> as input-values for mass calculation effected the PSMs results. In consequence, these relevant peptides might not be considered as peptide candidates for database searches if the precursor mass tolerance window is set to  $\pm 5$  ppm.

To demonstrate the relation of iTRAQ ratios to the mass accuracy distribution (ppm), Whiskers-Box-Plots were created for all ten samples analyzed (Figure 2-25). The maximal (x), the 1%; 99% cut-off (triangle) and 5%; 95% (end of vertical lines) values were included to extend the significance. The maximal values were frequently close to  $\pm 10$  ppm, as expected for a database search with a precursor mass tolerance window set to  $\pm 10$  ppm. However, the 5%; 95% values shifted step less to higher (or lower) mass deviations with increasing iTRAQ<sub>114</sub> (iTRAQ<sub>115</sub>) intensities. Additionally, the distribution accumulated with increasing number of iTRAQ labelings being attached to a peptide. For iTRAQ-116 and iTRAQ-117, the distribution (25, 50 & 75%) were close to 0 ppm, whereas iTRAQ-114 mass accuracy was broader distributed (as explained before). Peptides with severe differences in iTRAQ ratios represent those being (potentially) biologically relevant, which might be prone by an *in-silico* introduced imprecision during precursor mass calculation.

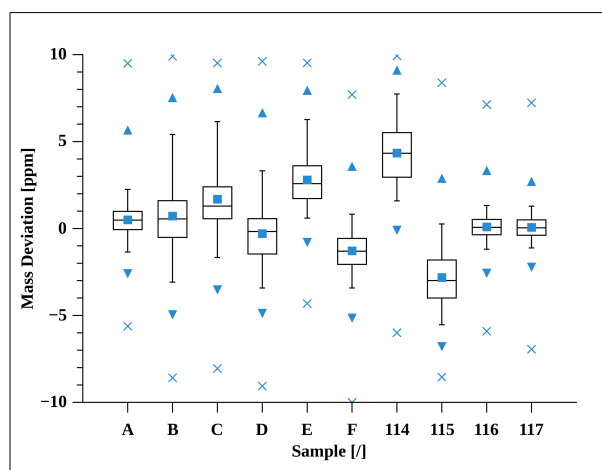


Figure 2-25: Whiskers-Box-Plots for the ten samples analyzed regarding the *in-silico* introduced mass shift. Rectangular: median value; triangle: 1%, 99%; x: maximal values. Box-Plots represent values of 25, 50 and 75%.

## 2 Developing a Platform for relative Quantification using iTRAQ

### 2.2.3.2 Variations of Precursor Mass Accuracy Window

As outlined in the previous section, the theoretically calculated peptide mass and its mass deviation were a function of the iTRAQ reporter ion intensities observed in MS/MS spectra. To evaluate if this (*in-silico* introduced) mass shift had an influence on the number of peptide identified, a series of 21 database searches were performed. For each, the precursor mass tolerance window was increased by 1 ppm from 0 ppm (actually 0.01 ppm as the minimum value available in Proteome Discoverer 1.4) to 20 ppm. Finally, the number of PSMs within one sample (e.g., iTRAQ-114) and one precursor mass tolerance value set (e.g., 5 ppm), respectively, were exported. For visualization, the PSMs of the four samples (iTRAQ-114 to iTRAQ-117) were plotted in relation to the precursor mass tolerance window (Figure 2-26).

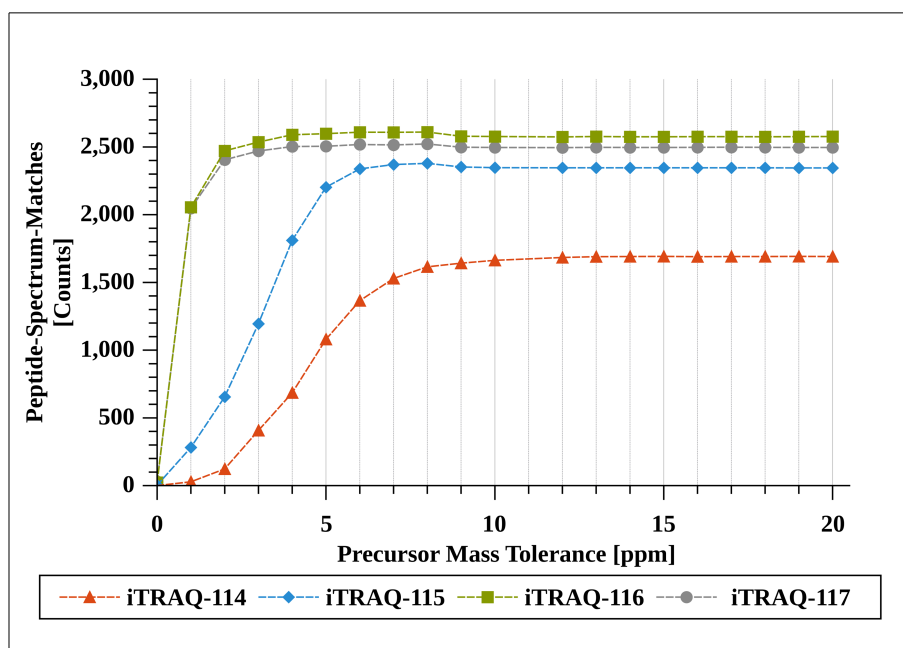


Figure 2-26: Number of identified peptides for the samples iTRAQ-114 to iTRAQ-117 with a low (0.01 ppm) to a high (20 ppm) predefined precursor mass tolerance.

iTRAQ<sub>116&117</sub> are the default iTRAQ modification masses used for database search. This relation was also obvious for the number of Peptide-Spectrum-Matches. For both, iTRAQ-116 and iTRAQ-117, a low precursor mass tolerance window of 2 ppm was required to receive stable Peptide-Spectrum-Matches (change < 10%). This was in good agreement with previous results, as the averaged mass deviation value (with a fixed precursor mass tolerance of 10 ppm) was close to  $\pm 0$  ppm.

The mass difference between iTRAQ<sub>115</sub> and iTRAQ<sub>116</sub> ( $\Delta=0.002464$  Da) is smaller compared to the one for iTRAQ<sub>114</sub> and iTRAQ<sub>116</sub> ( $\Delta=0.003855$  Da). This was also reflected by the number of PSMs: for iTRAQ-115, the number of PSMs was stable (change < 10%) if a precursor mass tolerance window of 5 – 6 ppm was defined in Proteome Discoverer. In contrast, a value of approx. 7 ppm was required for the sample iTRAQ-114.

Summarizing, the minimum precursor mass tolerance values set for database searches reported here for iTRAQ-114 to iTRAQ-117 reflected the maximal values, which should be considered to be expected for high precision MS data (e.g., by the application of lock mass correction or offline re-calibration). For outbalanced or

## 2.2 n-iTRAQ – Handling of Nominal Isobaric Labelings

ratios of 1.3:1 (iTRAQ<sub>114</sub> and iTRAQ<sub>115</sub>), the mass shift effect was not significant. For highly differentially abundant proteins (such as recombinant ones), a higher precursor mass tolerance window might be required. First, the introduced *in-silico* mass shifts due to different iTRAQ modification masses and, second, by a general (*experimental*) calibration shift due to climatic and electronic changes of the instrument.

### 2.2.3.3 Post Database Search Re-Calculation of Precursor Masses

In general, proteins included in FASTA files are *in-silico* digested to receive a list of peptide candidate for Peptide-Spectrum-Matching processes. To each peptide within the list, the corresponding masses of static and dynamic modifications are added and the final peptide masses calculated. For isobaric iTRAQ-4Plex labeling experiments, the masses of iTRAQ<sub>116&117</sub> are used for peptides mass calculation. This agrees with the UniMod recommendation (Accession # 214) and is acceptable for similar iTRAQ reporter ion intensities. However, this procedure is not optimal for highly up- or down-regulated proteins. In the following section, an upstream workflow to database searches is described which eliminates the observed effect. Principally, this procedure is not limited to an upstream approach but requires a more flexible FASTA file index creation (see discussion).

#### Extracting the Required Information for Precursor Mass Re-Calculation

During precursor selection for MS/MS fragmentation, information about the precursor (e.g., charge state and m/z value) were directly available in the scan header. Similar, the iTRAQ reporter ion intensities were present in MS/MS spectra (e.g., .mgf files) after precursor fragmentation using suitable collision energies and fragmentation techniques (see chapter 2.1). After having performed a database search and peptide validation, information about the precursor amino acids composition and its modifications were immediately available. All these features were required to re-calculate the precursor mass more accurately.

#### Re-Calculating the Precursors Masses

The complete workflow to re-calculate the iTRAQ modification and precursor mass is illustrated in Figure 2-27. As an example, the tryptic peptide TLNFNAGEPELLMLANWRPAQPLK, labeled on the N-terminus and the C-terminal lysine, was chosen. In addition to FDR based peptide validation, the peptide was annotated *in-silico* (filter: 20 ppm mass accuracy, intensity cut-off filter: 5%; see Figure 2-27 (A)) to verify the correct peptide spectrum match.

The first step in re-calculating the precursor mass (m/z value) was to sum up the four iTRAQ intensities (iTRAQ<sub>total intensity</sub>). For the example peptide, this equaled an intensity of iTRAQ<sub>total intensity</sub> = 26,525.68 (compare Figure 2-27 (B)). Each of the four iTRAQ reporter ion intensities were then set into relation to this value, resulting in values referred to as iTRAQ<sub>relative intensity(x)</sub> with x = 114 to 117. The corresponding percentile distribution is illustrated by a pie-chart (Figure 2-27 (C)). As shown by zoom-in for the lower m/z area (Figure 2-27 (B)), iTRAQ<sub>114</sub> dominated the remaining iTRAQ reporter ions channels with a very high relative intensity of iTRAQ<sub>relative intensity 114</sub> = 64.52%.

To correct the iTRAQ modification mass, classically utilized (144.102063 Da) for the observed distribution (here: with over dominating iTRAQ<sub>114</sub>), each of the relative intensities were multiplied by its known and correct

## 2 Developing a Platform for relative Quantification using iTRAQ

modification mass. For iTRAQ<sub>114</sub>, the relative intensity (0.6452) was multiplied by 144.105918 Da resulting in a value of iTRAQ<sub>part\_114</sub> = 92.970245 Da (see box in Figure 2-27 (D)). Finally, all four product values were summed up (see histogram in Figure 2-27 (D)). This resulted in the new iTRAQ modification mass (iTRAQ<sub>corrected</sub> = 144.104275 Da), which was highly specific for this MS/MS spectra. This value was utilized instead of the standard iTRAQ mass (144.102063 Da) to re-calculated the precursors mass and m/z value, respectively. For the Peptide-Spectrum-Match shown here, the newly calculated iTRAQ modification mass (iTRAQ<sub>corrected</sub>) differed by 0.002212 Da to the classically applied iTRAQ mass (iTRAQ<sub>116&117</sub>). This difference, which was *solely* introduced during precursor mass calculation and did *not* represent an experimental measurement error, caused a mass shift to 2.1 ppm. For the iTRAQ intensity corrected modification mass (iTRAQ<sub>corrected</sub>), the mass deviation was calculated to be 0.7 ppm, which was significantly more accurate.

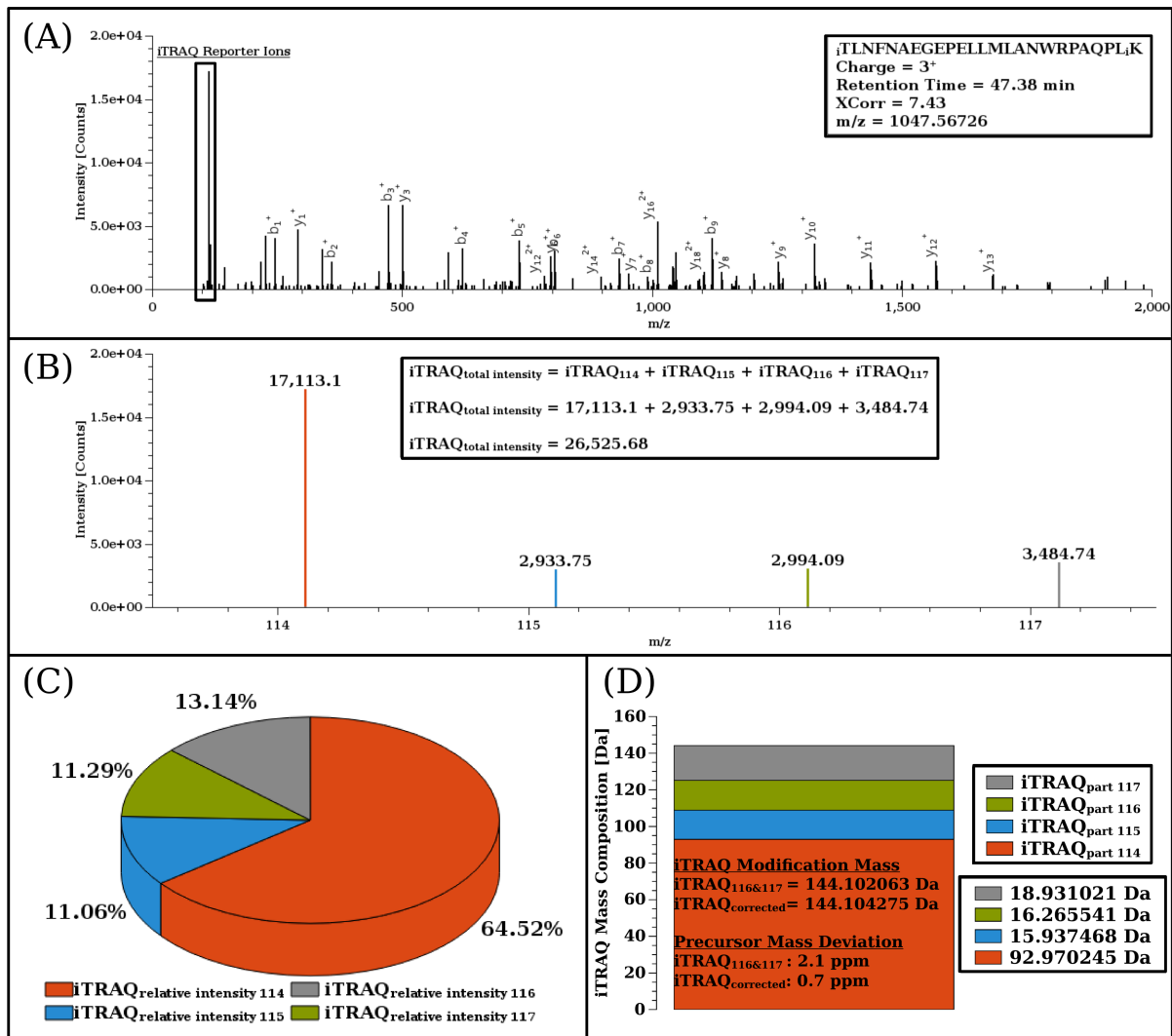


Figure 2-27: Workflow applied in this study to re-calculate the peptides precursor mass for every spectrum separately. (A) represents the acquired and annotated HCD MS/MS spectra; (B) a zoom-in of the lower m/z region with the four iTRAQ reporter ions and their raw intensities; (C) pie-chart demonstrating the impact of each iTRAQ reporter ion to the sum of all four; (D) the specific iTRAQ<sub>part</sub> values and their final, newly calculated iTRAQ mass.

### The Application of Highly Specific iTRAQ Modification Masses

The workflow to re-calculate the precursor mass ( $m/z$  value) described above was additionally applied to the four iTRAQ channels measured separately (iTRAQ-114 to iTRAQ-117) and the six different iTRAQ ratio samples (samples  $A_{1:1:1:1}$  to  $F_{1:4:1:1}$ ). The developed algorithm was not only calculating the precursor mass ( $m/z$ ) based on the iTRAQ reporter ions observed, but also used the classical approach (iTRAQ<sub>116&117</sub> masses) to validate the correctness of the calculations compared to Proteome Discoverer. Here, the differences introduced were smaller than 0.05 ppm and were related to the combinational effect of different decimal places used during calculation and the formation of 1 ppm bins for data visualization.

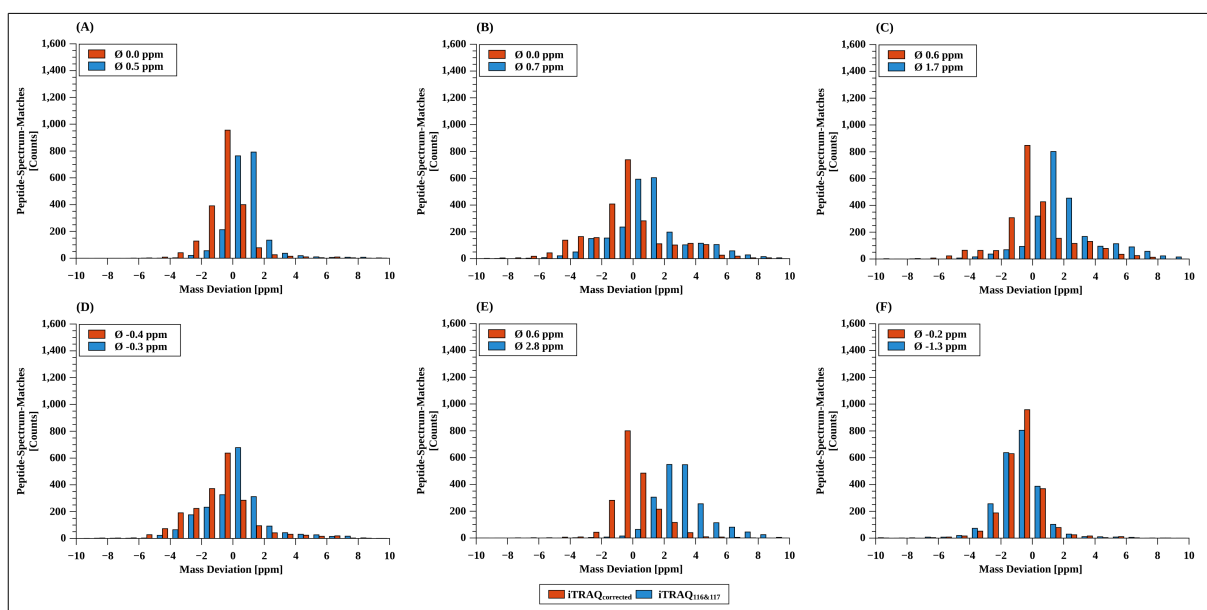


Figure 2-28: The frequency of peptides identified within the bins of -10 to +10 ppm for the samples  $A_{(1:1:1:1)}$  to  $F_{(4:1:1:1)}$  were plotted with (iTRAQ<sub>corrected</sub>, red color) and without (iTRAQ<sub>116&117</sub>, blue color) utilization of the re-calculation algorithm.

For sample  $A_{1:1:1:1}$ , a high mass accuracy was observed (0.5 ppm) even without iTRAQ modification mass correction, as it was expected due to the outbalancing effect of iTRAQ<sub>114</sub> and iTRAQ<sub>115</sub>. However, the little shift to positive values can be explained by the higher mass difference of iTRAQ<sub>114</sub> compared to iTRAQ<sub>115</sub>. This effect was not observed after the iTRAQ reporter ion intensities were taken into account. Here, the median mass accuracy was calculated to be 0.0 ppm (exactly -0.01 ppm). Additionally, the frequency of PSMs within the  $\pm 0$  ppm bin increased and the number of PSMs was distributed equally in the  $\pm 1$  ppm bins (Figure 2-28 (A)). For the sample  $B_{(1:1:1:1)}$ , a similar effect was observed (Figure 2-28 (B)). In Sample  $C_{(4:3:1:1)}$ , the effect of dominating iTRAQ<sub>114</sub> was more significant with +1.7 ppm. By using the re-calculation algorithm, the mass accuracy improved by 1.1 ppm to a final (averaged) value of 0.6 ppm. Again, a higher number of PSMs were found within the  $\pm 0$  ppm bin (Figure 2-28 (C)). Only for sample  $D_{(3:4:1:1)}$  was a slightly negative effect observed after re-calculation (before: -0.3 ppm and after re-calibration: -0.4 ppm). Currently, the cause of this effect is unknown. For the *highly differentially regulated* samples  $E_{(4:1:1:1)}$  and  $F_{(1:4:1:1)}$ , the effect of improved precursor mass calculation was significant. For sample E, the mass deviation decreased from 2.8 to 0.6 ppm and from -1.3 to

## 2 Developing a Platform for relative Quantification using iTRAQ

-0.2 ppm for sample F, respectively.

The improved mass accuracy described so far represented values, which might be observed in biological samples. To demonstrate the maximal improvement by applying the re-calculation algorithm, samples iTRAQ-114 to iTRAQ-117 were analyzed accordingly (Figure 2-29 (A - D)).

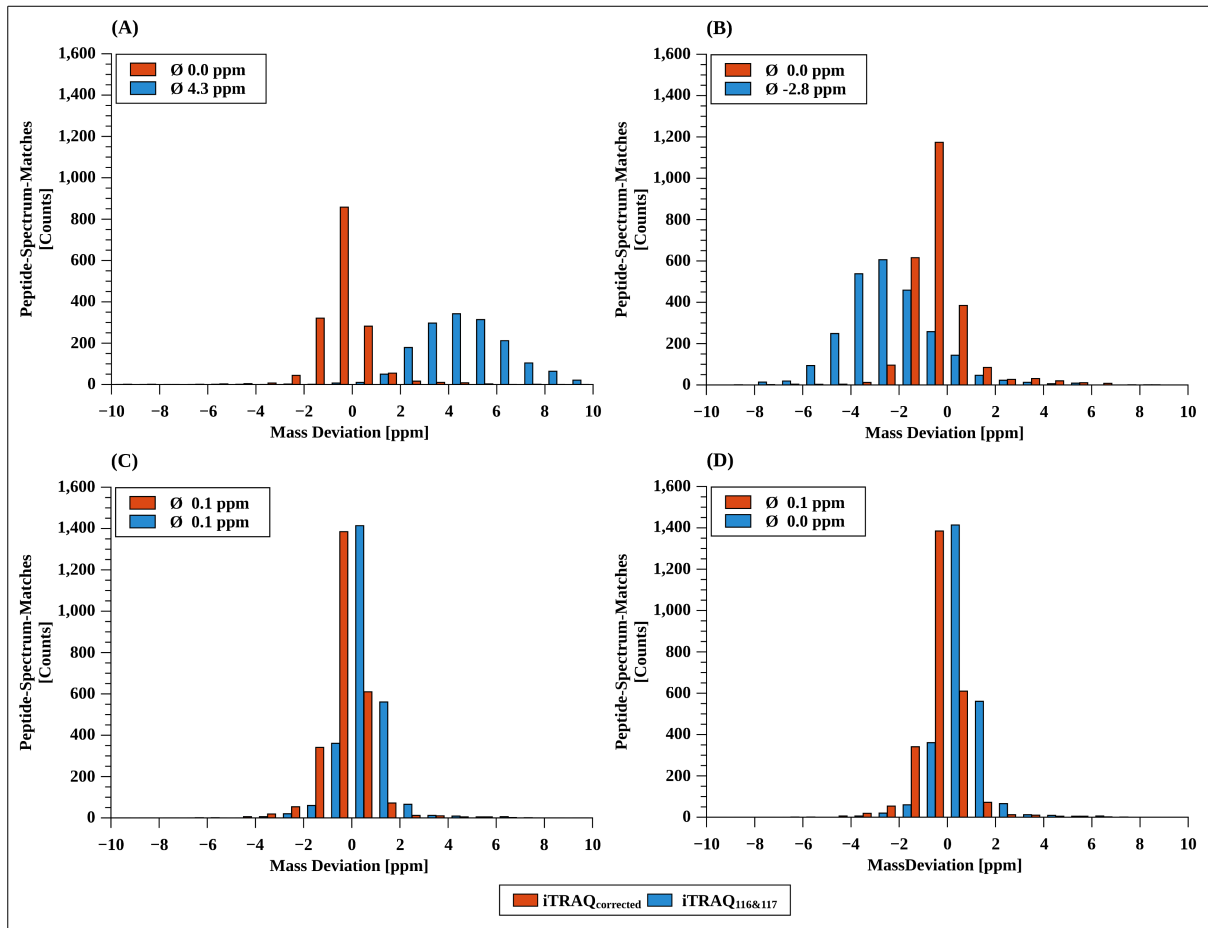


Figure 2-29: The frequency of peptides identified within the bins of -10 to +10 ppm for the samples iTRAQ-114 to iTRAQ-117 (A - D) were plotted with (*iTRAQ<sub>corrected</sub>*, red) and without (*iTRAQ<sub>116&117</sub>*, blue) utilization of the re-calculation algorithm.

As expected, the differences for iTRAQ-116 and iTRAQ-117 were negligible (mass accuracy greater than 0.1 ppm; (Figure 2-29 (C & D))). For iTRAQ-115, the mass accuracy increased from -2.8 ppm to 0.0 ppm. This represented a shift of nearly 3 ppm, which was also demonstrated by its frequency histogram (Figure 2-29 (B)). Before re-calculation, the largest mass deviation (+4.3 ppm) was observed for the sample iTRAQ-114. Taking the iTRAQ intensity into account, the mass accuracy significantly improved (0.0 ppm, in average). Additionally, the effect of higher mass deviations with a higher number of iTRAQ labelings attached was compensated. This was demonstrated by the high number of PSMs within the  $\pm 0$  ppm (53%), -1 ppm (20%) and +1 ppm (17%) bin, respectively.

The distribution and statistical relevance of the identified peptides was plotted via a Whiskers-Box-Plot (Figure 2-30), as it was done for the original data (Figure 2-25). Before re-calibration, the data with high

## 2.2 n-iTRAQ – Handling of Nominal Isobaric Labelings

iTRAQ<sub>114</sub> and iTRAQ<sub>115</sub> intensities shifted to higher and lower mass deviations, respectively. Additionally, a broader distribution was observed. However, after re-calibration, the positive effect was not only recognize able by the average mass deviation (in ppm), but also by smaller 5%; 95% and 25%; 75% values (*triangle and vertical line endings*).

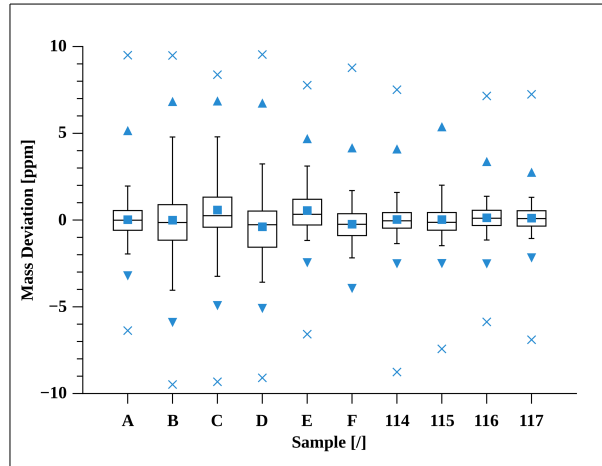


Figure 2-30: Whiskers-Box-Plots for the ten samples analyzed regarding the in-silico introduced mass shift. Rectangular: median value; triangle: 1%, 99%; x: maximal values. Box-Plots represent values of 25, 50 and 75%.

## 2 Developing a Platform for relative Quantification using iTRAQ

### 2.2.4 Conclusion

Relative quantification of proteins is a key feature to identify differentially regulated proteins of two or more biological entities. Using isobaric tags for relative and absolute quantitation (iTRAQ) differences in protein abundance are monitored via different iTRAQ reporter ion intensities upon precursor fragmentation. Consequently, the iTRAQ reporter ion distribution can be interpreted as a collection of differentially labeled precursor ions, each with a specific mass which depends on the iTRAQ-4Plex reagent being used.

This mass shift is induced by different iTRAQ reagents masses and is outbalanced if equally distributed iTRAQ reporter ion intensities are present in MS/MS spectra. Generally, the intensity of iTRAQ reporter ions in tandem MS spectra is linearly correlated to the corresponding precursor ion composition (see Figure 1-4). However, by increasing the amount of iTRAQ<sub>114</sub> and iTRAQ<sub>115</sub> labeled precursor ions for fragmentation, respectively, a mass shift (ppm) is introduced compared to the *in-silico* calculated peptide masses, which utilize iTRAQ<sub>116&117</sub> as the default input value for precursor mass calculation. This effect was most prominent in the analyzed data set if precursor ions accumulated for MS/MS fragmentation were dominated by iTRAQ<sub>114</sub> and iTRAQ<sub>115</sub>, respectively. As iTRAQ reagents are coupled predominantly to primary amines (N-terminal & K) and to a smaller extent to tyrosines, the mass shifting effect is also a function of the number of labelings attached.

This difference in iTRAQ reagent masses also affected database search results, as a minimum precursor mass tolerance window of 5 – 7 ppm was required for exclusively iTRAQ<sub>114</sub> and iTRAQ<sub>115</sub> labeled peptides. This is close to the *gold-standard* applied nowadays in proteomics experiments (10 ppm), reducing the allowed experimental mass shift to 3 ppm instead of 10 ppm. To compensate for this error introduced by *in-silico* calculation of the precursor mass, an algorithm was developed which used the iTRAQ reporter ion intensities to re-calculated the precursor mass accordingly. With its application, the mass accuracy of samples dominated by iTRAQ<sub>114</sub> could significantly be improved from +4.3 to 0 ppm.



## 2.3 Y-iTRAQ – Accurate Identification of iTRAQ Tyrosine Labeled Peptides

### 2.3.1 Introduction

The increasing mass accuracy and resolution of mass spectrometers has shown, that the non-equivalent masses of the four iTRAQ-4Plex reagents affect database search results. Suitable, high resolution MS instruments have also allowed to detect and discover a variety of diagnostic immonium ions, such as for pY-peptides or specific fragmentation pathways (190,286). However, a high number of MS/MS spectra still remains unidentified (275). Only 5 - 20% of low resolution ion trap and 15 - 70% quadrupole TOF tandem MS spectra are typically identified and validated peptide hits (287–290). In addition, it has been reported that the number of identified peptides decreases after isobaric labeling. Here, the number of identified peptides decreases while the mass of isobaric tags increases, i.e., iTRAQ-4Plex yields higher identification rates than TMT-6Plex or iTRAQ-8Plex do (63,272).

In order to establish a platform for relative quantification of (phosphorylated) peptides using isobaric tags, in-depth analysis was performed to further validate and enhance peptide identifications of iTRAQ labeled peptides. Modification of tyrosine, a side-reaction during iTRAQ labeling, is commonly neglected in database search approaches. However, the Paragon algorithm of the software suite ProteinPilot allows for this modification. In contrast to conventional search engines in which discrete settings are required, peptide features such as cleavage events, amino acid polymorphism or modifications are modeled with probability values by the Paragon algorithm (275). Although the probability of the iTRAQ labeling is set considerably to lysine and N-termini (*fixed groups*; Paragon probability: 0.993 and 0.932, respectively), the (relative) high abundance of iTRAQ labeled tyrosine, serine and threonine has been demonstrated recently (291) and is also regarded by the Paragon algorithm (*variable groups*; Paragon probability: 0.036, 0.011, 0.0027, respectively). This is a rather high frequency compared to biological relevant (dynamic) modifications, such as phosphorylation events (pS, pT, pY with Paragon probabilities of 0.002, 0.007 and 0.00001, respectively), which are often set as variable modification in database searches although no phosphopeptide enrichment and emphasis analysis was performed.

MS/MS spectra of peptides containing tyrosine exhibit an immonium ion at  $m/z$  136.07 (292). Further, pY-peptides frequently deliver a diagnostic ion at  $m/z$  216.04 (190). Both immonium ions differ only by the mass of a phosphate group (79.97 Da; Unimod Accession # 21). As both, the iTRAQ reagent and the phospho-moiety bind to the phenolic hydroxyl group of tyrosine, an iY-iTRAQ-4Plex immonium ion at  $m/z$  280.17 (136.07 Da +144.10 Da) is expected. The aim of this section is to demonstrate the occurrence of this ion under different fragmentation conditions and to analyze their fragmentation pattern and influence on the retention time of reversed phase chromatography.

## 2 Developing a Platform for relative Quantification using iTRAQ

### 2.3.2 Material and Methods

#### 2.3.2.1 Chemicals

Chemicals used in this study are summarized in chapter 4.1.

#### 2.3.2.2 Sample Preparation and iTRAQ Labeling

##### **Protein Digestion and iTRAQ-4Plex Labeling**

The same ten protein-mix used to optimize iTRAQ labeled peptide fragmentation conditions (chapter 2.1) was used in this section. In addition, a single protein ( $\alpha$ -lactalbumin; 100  $\mu$ g starting material) digestion and iTRAQ-4Plex labeling was applied using the same workflow as for the ten protein-mix.

##### **Phosphoprotein Digestion and iTRAQ-8Plex Labeling**

The protocol applied for SDS-PAGE, in-gel digestion and subsequent iTRAQ-8Plex labeling is described in chapter 2.4.2.2.

##### **Dephosphorylation and iTRAQ-4Plex Labeling of a Phosphopeptide Standard**

Phosphopeptides were iTRAQ-4Plex labeled as described previously (chapter 2.1.2.1). Divergent to the previous protocol: for dephosphorylation, the sample (131 phosphopeptides; 10  $\mu$ L each) was split in two aliquots and both (D for dephosphorylation, P<sub>1&2</sub> for phosphorylation) were dried by vacuum evaporation. Afterwards, sample D was resolubilized in 40  $\mu$ L alkaline phosphatase buffer (1:10 diluted; supplied by Roche Diagnostics GmbH) and 20 U of alkaline phosphatase were added. The reaction mixture was incubated at 37°C for 3 h with gentle shaking. Heat inactivation of the alkaline phosphatase was performed by incubation of the samples at 95°C for 10 minutes. Afterwards, sample D was evaporated to dryness. Both samples (D & P) were resolubilized in 35  $\mu$ L buffer A (25% ethanol, 50% iTRAQ dissolution buffer (both provided by AB SCIEX), 23.75% H<sub>2</sub>O and 1.25% EDTA solution (10mM)). 40  $\mu$ L of iTRAQ reporter ion solution were added to each sample accordingly: iTRAQ-4plex-114 P<sub>1</sub>; iTRAQ-4plex-116 P<sub>2</sub>; iTRAQ-4plex-115 sample D. After labeling, each sample was evaporated to dryness and resolubilized for 10 minutes in 20  $\mu$ L of 3% ACN with 0.1% FA. Finally, samples were diluted to a total of 30  $\mu$ L. For LC-ESI MS measurements, 2  $\mu$ L of sample were injected per run (only LTQ Orbitrap Velos).

### 2.3.2.3 Mass Spectrometry, Database Searches and Data Analysis

#### LC-ESI MS on a LTQ Orbitrap Velos Equipped with ETD Support

##### *Instrumental Setup*

The same measurement parameters as described previously (phosphopeptide mix-2 and ten protein-mix, chapter 2.1) were applied. In addition to CID-IT and HCD, also PQD was used for fragmentation and analysis in the ion trap mass analyzer (iTRAQ-4Plex labeled ten protein-mix). Again, 18 raw files with different collision energies were recorded (0 – 85% NCE; only PQD). For peptide separation, the LC-ESI MS setup was unchanged except specific PQD settings: activation Q 0.9; activation time 0.1.

For ETD analysis, a TOP 5 *Nth order double play* method was applied to analyze the singly digested protein  $\alpha$ -lactalbumin. Samples were loaded with 30  $\mu$ L/min and separated using 300 nL/min. For peptide separation, a shorter gradient was applied: 5 – 10% B within 1 min, 10 – 15% B within 1 min, 15 – 40% B within 9 min, following column washing (40 – 95% B within 1 min kept constant for 1 min, followed column recalibration at 5% B for 15 min). MS full scans were recorded in the Orbitrap mass analyzer with a resolution of 30,000. Subsequently, the five most intense precursor ions were selected (isolation window of 3 Da) and fragmented with ETD using a minimal signal required of 500 (*counts*), a default charge of 2<sup>+</sup>, an activation Q and time of 0.25 and 160 ms, respectively. First mass was set to m/z 100. Lock mass correction was enabled with m/z 445.120024. Precursor charge states < 2<sup>+</sup> were rejected. Dynamic exclusion list options were: repeat count 1, repeat duration 30, exclusion list size 500, exclusion duration 30 sec, exclusion mass width relative to low/high  $\pm$  10 ppm.

##### *Database Searches*

Recorded .raw files (iTRAQ-4Plex only) were searched using Proteome Discoverer (1.4) against the bF-FASTA (see chapter 2.1.2.3). For analyzing the phosphopeptide standard, the same FASTA file was used but without the *E. coli* proteins. For the iTRAQ-8Plex samples, the same bF-FASTA file was used.

The workflow applied was build up by 11 different nodes with default settings if not otherwise specified: (0) Spectrum Files; (1) Spectrum Selector, (2) MS2-Spectrum Processor, (3) & (7) Scan Event Filter to split into CID and HCD fragmentation; (4) & (8) SEQUEST HT (see ten protein-mix in chapter 2.1.2.3); (5) phosphoRS 3.0; (9) Reporter Ions Quantifier (iTRAQ-4Plex, respectively (Thermo Scientific Instruments; Integration Tolerance 20 ppm; Most Confident Centroid as Integration Method; Additional Settings: Show the Raw Quan Values; Apply Value Corrections, use All Peptides enabled and deactivated Experimental Bias); (10) Event Detector; (11) Precursor Ions Area Detector and (12) Percolator.

For PQD and ETD data, only the Scan Event Filter was changed to PQD (ion trap) and ETD (Orbitrap), respectively. Settings for the mass analyzers remained unchanged.

## 2 Developing a Platform for relative Quantification using iTRAQ

### Data Analysis

The following analysis was performed for the iTRAQ-4Plex labeled ten protein-mix and iTRAQ-8Plex labeled phosphoprotein sample acquired on the LTQ Orbitrap Velos. Peptide-Spectrum-Matches (PSMs) were filtered for high confidence (FDR = 1%) and exported to .txt files.

Firstly, exported PSMs were filtered for iTRAQ-4/8Plex modifications on tyrosine, respectively. An *in-house* written *Command Line Interface* (CLI) script (using GAMBAS) was used to count the number of identified peptides of the Y-iTRAQ modified peptide and its unmodified counterpart in all recorded 2x 18 (iTRAQ-4Plex; ten protein-mix) and 9 (iTRAQ-8Plex; phosphoprotein sample) LC-ESI MS runs, respectively. Simultaneously, the median of the retention time of each peptide (Y-iTRAQ and its unmodified counterpart, respectively) was calculated and used to analyze potential retention time shifts after iTRAQ labeling (Table 2-6 & 2-7). To compare precursors intensities of Y-iTRAQ labeled peptides with their unmodified form, extracted ion chromatograms (XICs) for each precursor were calculated and exported using XCalibur. Data were imported into QtiPlot for visualization (Figure 2-40).

For automated fragment ion matching of MS/MS spectra, an *in-house* written CLI script (GAMBAS) was used. Here, the precursor mass and the corresponding b- and y-fragment ions (minimum fragment ion charge state = 1<sup>+</sup>; max.: precursor charge state) were imported and automatically annotated using a  $\pm 20$  ppm fragment ion window (Orbitrap mass analyzer). Only m/z signals having a relative intensity of 5 (HCD data) or 1% (ETD data) compared to the base peak were used for annotation. For CID and PQD data (ion trap), a cut off filter at 1% of the most intense peak was applied. The mass accuracy was 0.2 Da at m/z 400, equaling a mass accuracy windows of 500 ppm. The generated tab-separated lists were finally imported into QtiPlot for visualization.

For analyzing the relative intensities (intensity of ion of interest versus the most intense signal in spectra) of specific peptide signals, the same framework utilized previously (chapter 2.1) was used in this section with minor modifications. Peptide features of HCD data (retention time (RT), charge state (z) and m/z value) for both, the modified (Y-iTRAQ) and unmodified form, were used to identify the corresponding spectra with the following settings: retention time (RT) window of 5 min; mass deviation of maximal 20 ppm. Here, the intensity of the precursor [M], iTRAQ reporter ions, undissociated iTRAQ (iTRAQ + Balancer Group), tyrosine immonium ion (iY), tyrosine immonium ion with intact iTRAQ (iY-iTRAQ) and the most intense signal in spectra (base peak) were exported for each spectra, m/z value, charge state and NCE, respectively. Finally, features were grouped (with: same m/v value (20 ppm), charge state, RT (5 min) and NCE) for median and standard deviation calculation. The four iTRAQ reporter ion signals were grouped and averaged. All monitored ions were set into relation to the base peak. Finally, data were plotted versus the applied NCE value (Figure 2-38).

## 2.3 Y-iTRAQ – Accurate Identification of iTRAQ Tyrosine Labeled Peptides

### LC-ESI MS on a Synapt G2s

#### *Instrumental Setup*

Nano-HPLC tandem mass spectrometric experiments were performed on a Dionex U3000 (Dionex, Idstein, Germany) coupled to a Synapt G2s (Waters, Milford, MA). Solvents and flow rates for HPLC remained unchanged compared to the previously described setup of the LTQ Orbitrap Velos. iTRAQ-4Plex labeled peptides originated from  $\alpha$ -lactalbumin digestion were separated using the following gradient: 5% B to 10% B within 1 minute, linear increase from 10% B to 50% B within 40 min, followed by a column wash step (5 min) with a linear gradient from 50% B to 95% B. Inter run equilibration of the analytical and pre-column was set to 11 minutes.

Synapt G2Ss source temperature was set to 80°C, and the sampling cone and source offset voltage to 40 and 80 V, respectively. Full MS scans were recorded for 1 sec in resolution mode using continuum data format. Lowest and highest masses were set to  $m/z$  300 and 2,000. For MS/MS, a maximal scan time of 0.5 sec was allowed. The lower mass range was enlarged to  $m/z$  50, keeping the highest mass constant. No lockmass was used for the analysis. Trap collision energy was increased in 5 eV increments, starting with 0 eV ending with 100 eV, resulting in a total of 21 LC-ESI MS runs.

#### *Data Conversion and Database Search*

Recorded .raw folders were converted to .mgf file using ProteoWizard (Version 3.0.5759; (293)). Files were re-written using an *in-house* build script (GAMBAS) to be compatible for database search using SearchGUI (Version 1.16.2; (294)) and PeptideShaker (Version 0.25.2; (295)). The database search was only performed to ensure comparable retention times of the peptide species via the 21 LC-ESI MS runs. The final peptide identification was performed manually. Search settings were: 10 protein-mix FASTA file containing the used proteins; enzyme specificity was set to trypsin; fixed modification: carbamidomethyl (C), iTRAQ-4Plex (N-Terminal & K); variable modifications: oxidation (M), iTRAQ(Y); max. missed cleavage: 2; precursor mass tolerance: 200 ppm; fragment mass tolerance: 0.5 Da; fragment ion type: b and y ions; min and max precursor charge state: 2<sup>+</sup> and 5<sup>+</sup>, respectively. As done for LTQ Orbitrap Velos data, the peptide of interest (VGINYWLAHK) was manually validated by fragment ion annotation (see LTQ Orbitrap Velos data analysis section) and LC retention time within the 21 raw files recorded. For visualization purpose, spectra with a collision energy of 45 eV were chosen (100 ppm, cut-off filter: 2.5% of most intense ion). The analysis of the CE versus relative intensity plots (Figure 2-38) were done as described before for LTQ Orbitrap Velos HCD data. Solely, the mass accuracy window was enlarged to 100 ppm.

## 2 Developing a Platform for relative Quantification using iTRAQ

### LC-MALDI MS on an AB SCIEX MALDI TOF/TOF 5800

#### *Instrumental Setup*

Prior to MS analysis with a MALDI TOF/TOF 5800 mass spectrometer (AB Sciex, Darmstadt, Germany), two technical replicates were separated on a Dionex U3000 nano-LC system (Dionex, Idstein, Germany) and subsequently spotted on a MALDI target (AB Sciex) using a Probot microfraction collector (Dionex). One  $\mu\text{L}$  of iTRAQ-4Plex labeled peptides originated from  $\alpha$ -lactalbumin digestion were injected for each technical replicate. Columns, solvents and flow rates applied for peptide separation were the same as described above for LTQ Orbitrap Velos setup with the exception that eluates A and B (micro-pump) contained TFA instead of FA. After desalting the samples 4 min on the trap column, the following gradient was applied: 5 to 10% B in 1 min, 10 to 50% B in 55 min, 50 to 95% B in 10 min, 95% B for 5 min, 95 to 5% B in 0.1 min, 5% B for 19.9 min. The eluate from HPLC separation was automatically mixed in a ratio of 1:3 (3 mg/mL in 70% ACN, 0.1% TFA with CHCA matrix, 14 nM Glu-fibrinopeptide B as internal calibrant) and spotted onto the target from 10 to 84 min (15 s intervals per spot).

Default calibration with the 4700 Proteomics Analyzer Standards Kit was performed in MS and MS/MS mode prior to MS measurements. In addition, MS spectra were internally calibrated to Glu-fibrinopeptide B. MS spectra of the fractions 25 to 65 minutes were acquired in a mass range of 700 to 4000 m/z with 1000 total shots per spectrum. Precursors were selected by the 4000 Series Explorer Software (AB SCIEX) using the following criteria: minimum S/N ratio: 25; minimal fraction width: 1; maximal precursors per spot: 10. MS/MS spectra of the strongest precursors were acquired first at either medium or high gas pressure with 3000 total shots per precursor.

#### *Data Conversion*

Spectra were exported to .mgf files using the TS2Mascot software (version 1.0.0, Matrix Science, London, UK). Settings were: export of peaks from 60 Da to 35 Da below the precursor mass; minimum of S/N ratio: 0; monoisotopic peaks only. Subsequently, database searches were performed using Proteome Discoverer to validate peptide identifications in addition to manual spectra validation by fragment annotation (see above).

#### *Database Searches*

Proteome Discoverer workflow was build up by 5 different nodes with default settings if not otherwise specified: (0) Spectrum Files; (1) Spectrum Selector (min precursor mass: 700 Da, max precursor mass: 4000 Da), (2) Scan Event Filter (Mass Analyzer: TOFMS); (3) Mascot (Version: 2.2.07, further details see below) and Target Decoy PSM Validator. A Mascot database search versus a 10 protein-mix FASTA file containing cRAP like proteins (see chapter 4.3) with a maximum of 2 missed cleavage sites allowed was performed. Instrument was set to MALDI-TOF-TOF with a precursor mass tolerance of 100 ppm and a fragment mass tolerance of 0.3 Da. Other settings: fixed modifications: carbamidomethyl (C), iTRAQ-4Plex (N-terminal & K); dynamic modifications: iTRAQ (Y), phosphorylation (S,T,Y), oxidation (M). For spectra annotation, a ppm window of 100 and an intensity cut-off of 2.5% of most intense signal was set.

## 2.3.3 Results

### 2.3.3.1 Detection of iY-iTRAQ-4Plex Specific Immonium Ions

iTRAQ-4/8Plex reagents are coupled via NHS chemistry onto primary amines of peptides (N-termini and lysine side chains). However, to a smaller extent (< 3%; (60)) the reagents can also react with tyrosine's hydroxyl-group (Figure 2-31).

Using higher collision induced dissociation (HCD) on the Orbitrap Velos and beam-type CID on TOF instruments, the generation of the prominent tyrosine immonium ion (iY) at  $m/z$  136.07 is a known fact. After phosphorylation (and its increase in mass (+ 79.96 Da)) a characteristic immonium ion (pY,  $m/z$  216.04) is observable for phosphotyrosine (see Figure 2-31 and chapter 2.1.3.1). Consequently, for Y-iTRAQ modified peptides, which are also bound via the hydroxyl-group to tyrosine, a specific immonium ion should be detectable after HCD fragmentation at  $m/z$  280.17.

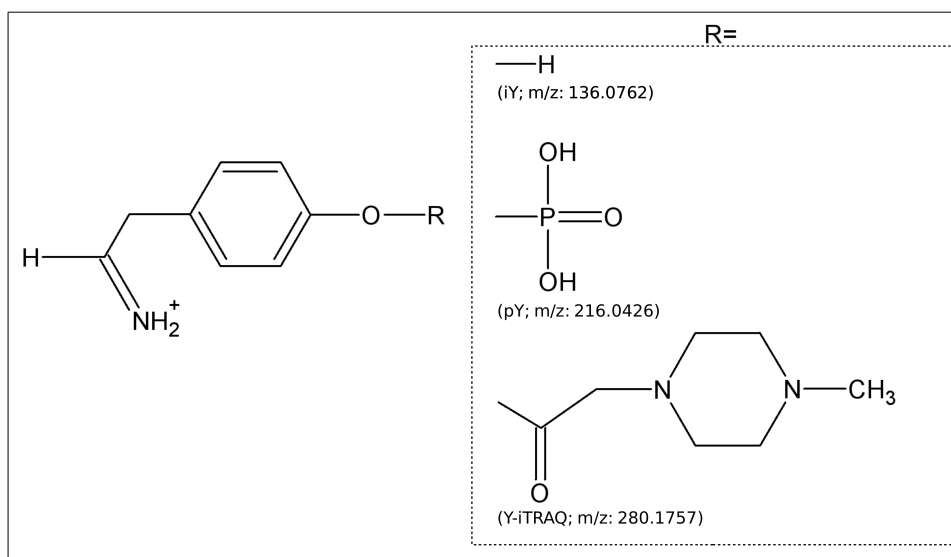


Figure 2-31: Structure and  $m/z$  value of the immonium ions generated after MS/MS fragmentation for the native (-H), phosphorylated ( $H_3PO_4$ ) and iTRAQ labeled tyrosine.

In the following, iTRAQ-4Plex labeled tandem MS data upon HCD fragmentation on the LTQ Orbitrap Velos will be described. Within the ten protein-mix analyzed, the peptide VGINYWLAHK (Figure 2-32) was confidently identified in two forms: (i) one was iTRAQ labeled at the N-terminus and at the  $\epsilon$ -amino group of lysine ( $m/z$ : 744.93,  $z=2^+$ ; referred to as unmodified,  ${}_i$ VGINYWLAH $_i$ K), whereas the second form (ii) showed an additional iTRAQ modification on position Y<sub>5</sub> ( $m/z$  816.98;  $z = 2^+$ ;  ${}_i$ VGIN $_i$ YWLAH $_i$ K)). The annotated spectra of the unmodified iTRAQ tyrosine peptide (Figure 2-32 (A)) showed an outbalanced b- and y-ion series with intense iTRAQ reporter ions. The b-ion series could be identified from  $b_1^+ - b_8^+$ ; except  $b_6^+$  and  $b_9^+$ , which did not pass the filter set to 20 ppm and the relative intensity cut-off set to 5%, respectively. Similar results were observed for the y-ion series. All fragment ions were matched from  $y_1^+$  to  $y_7^+$ .  $y_8^+$  and  $y_9^+$  did not pass the mass accuracy filter set. The  $y_{10}^+$  fragment ion was only monitored as non dissociated precursor ion  $[M+2H]^{2+}$ . The peak labeled with an asterisk (\*) represented the charge reduced intact precursor with a loss of iTRAQ.

## 2 Developing a Platform for relative Quantification using iTRAQ

Undissociated but singly charged iTRAQ ( $m/z$  145.10) was also detected with a relative intensity of 50% (not marked in (Figure 2-32)). The iY immonium ion at  $m/z$  136.07 was observed with a relative intensity of 1.3%.

For the Y-iTRAQ modified form (Figure 2-32 (B)), a similar fragment ion pattern was observed. All but four fragment ions ( $b_9^+$ ,  $b_{10}^+$ ,  $y_8^+$  and  $y_{10}^+$ ) could be matched. After labeling tyrosine with iTRAQ, a shift of +144.10 Da for the  $b_5^+$  ( $m/z$  691.38  $\rightarrow$  835.48) and  $y_6^+$  ( $m/z$  961.53  $\rightarrow$  1105.63) masses (and following ones) was observed. The manual annotation confirmed, in addition to the precursor mass and identification by SEQUEST, the correct peptide identification in both forms. The proposed Y-iTRAQ immonium ion (iY-iTRAQ) at  $m/z$  280.17 was detected for the Y-iTRAQ modified form (Figure 2-32 (B)) but not for the unmodified one; its relative intensity was 3.3% (compared to the most intense signal in the spectrum).

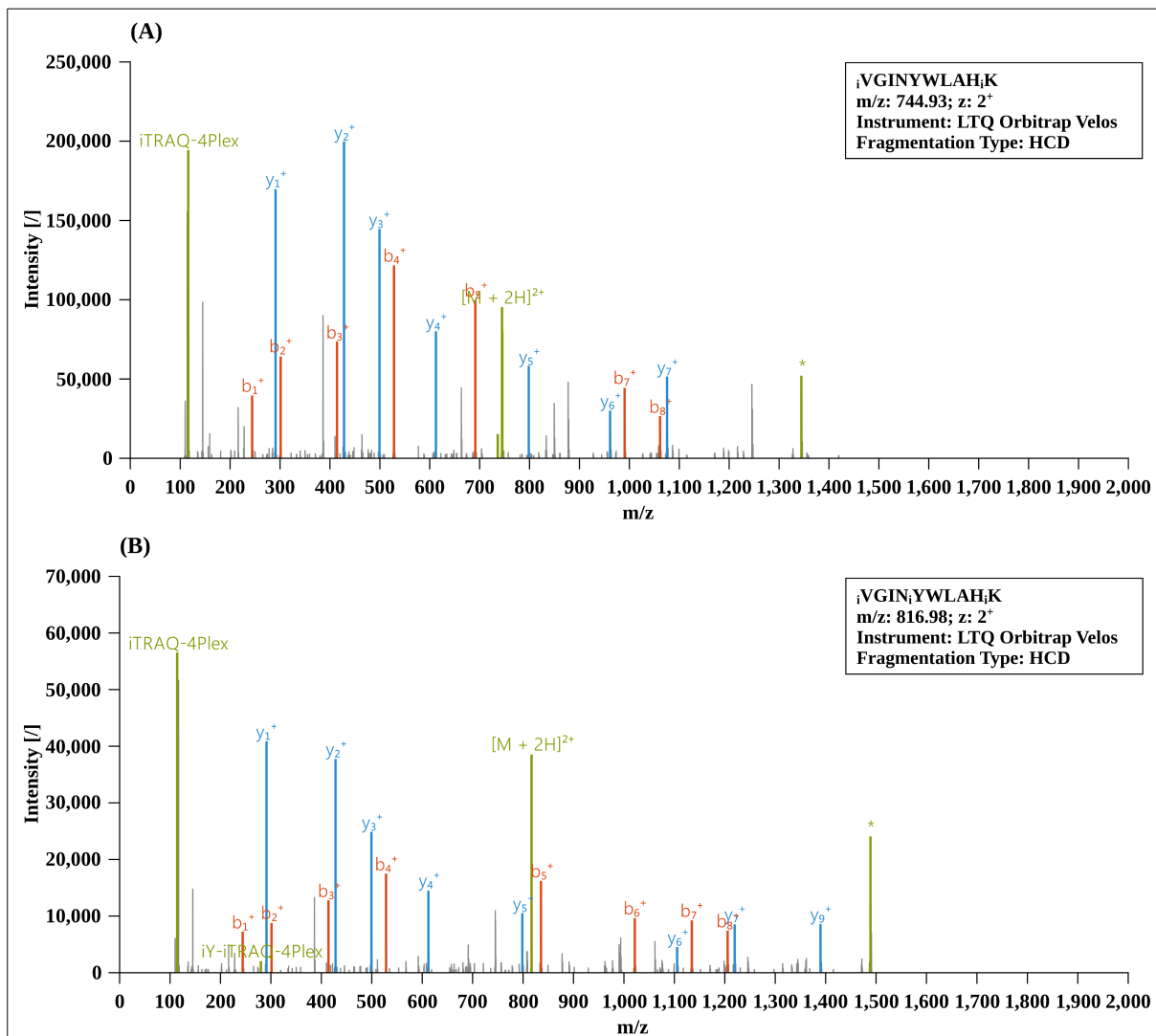


Figure 2-32: Annotated MS/MS spectra after HCD fragmentation (LTQ Orbitrap Velos) shows the b- and y-ions matched for the peptide VGINYWLAHK in its typical iTRAQ labeled form  $i$ VGINYWLAH;K (A) and with an additional iTRAQ labeling at tyrosine  $i$ VGIN $i$ YWLAHiK (B). The additional iTRAQ modification is observable by a mass increase of +144.10 Da for the  $b_5^+$  and  $y_6^+$  (and following) ions. At  $m/z$  280.17, the Y-iTRAQ(-4Plex) immonium ion is labeled and only present in the modified form (B). The peaks labeled with an asterisk (\*) belong to the intact precursor ion, which has lost one (charged) iTRAQ labeling (+ balancer group).



### 2.3 Y-iTRAQ – Accurate Identification of iTRAQ Tyrosine Labeled Peptides

Using Q-TOF and TOF/TOF instruments, which also generate beam-type CID spectra, the same peptide (VGINYWLAHK) was analyzed in its unmodified and Y-iTRAQ labeled form on the Synapt G2s (LC-ESI MS) and the AB Sciex 5800 MALDI TOF/TOF (LC-MALDI MS).

For the 2<sup>+</sup> charged precursor (Synapt G2S, Figure 2-33), the a- ( $a_3^+ - a_6^+$ ), b- ( $b_1^+ - b_6^+$ ) and y-ion ( $y_1^+ - y_5^+$ ) series were observed for the unmodified form. Most of the immonium ions were present but not fulfilling the mass accuracy criteria or intensity cut-off filter set. The coverage of a-, b- and y-ion series of the Y-iTRAQ modified form were similar ( $a_3^+ - a_6^+$ ;  $b_1^+ - b_8^+$ ;  $y_1^+ - y_8^+$ ), also including the Y-iTRAQ modified  $b_5^+$ - and  $y_6^+$ -ions. Using a collision energy of 45eV, the iY-iTRAQ immonium ion (m/z 280.17) intensity was (approximately) half the intensity of the iTRAQ reporter ions. Consequently, this diagnostic ion was one of the most intense signals in the spectra.

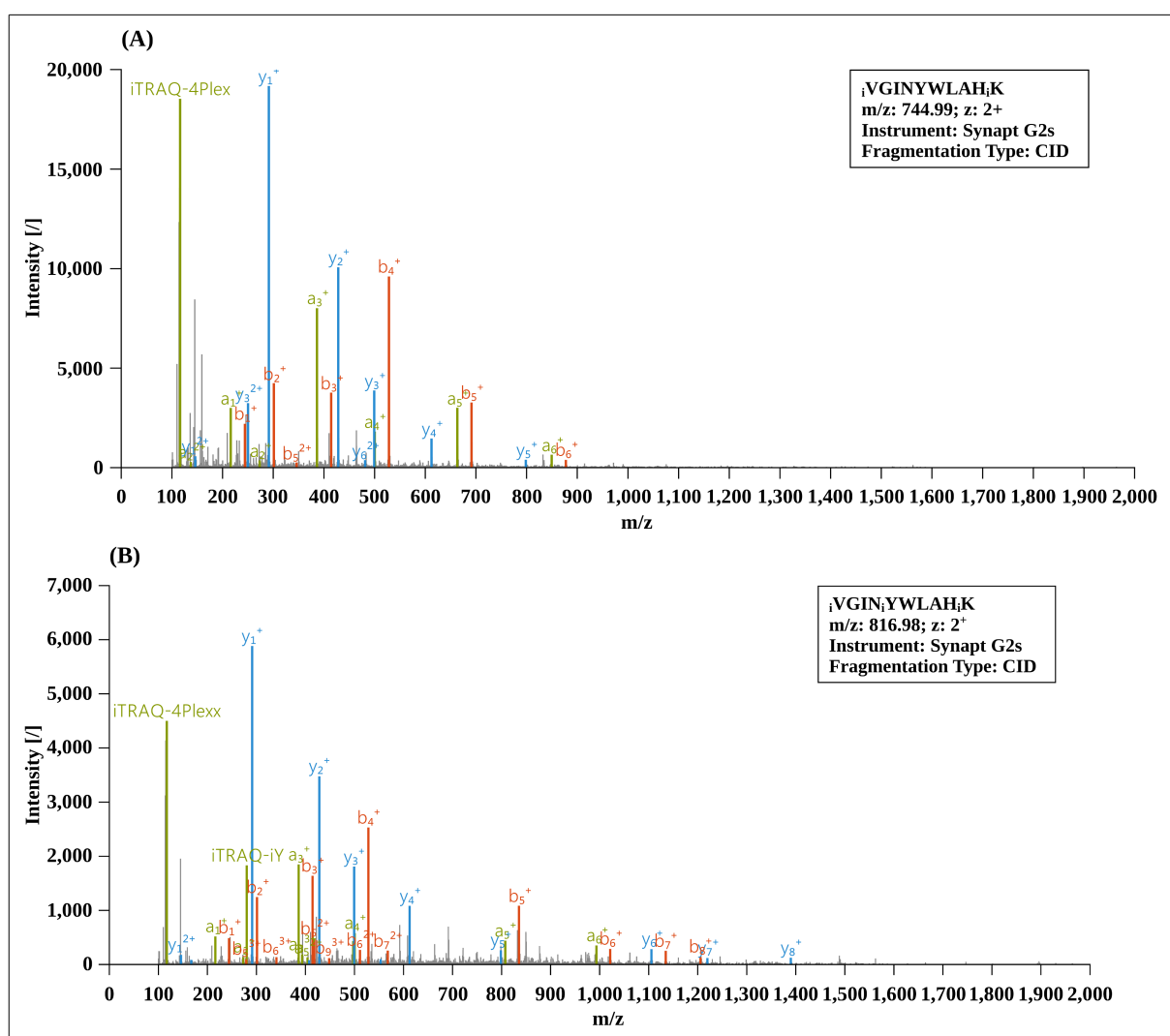


Figure 2-33: Annotated MS/MS spectra after CID fragmentation (Synapt G2s, ESI mode) shows the b- and y-ions matched for the peptide VGINYWLAHK in its common iTRAQ labeled form  $i$ VGINYWLAHK (A) and with an additional iTRAQ labeling on the tyrosine  $i$ VGIN $i$ YWLAHiK (B). The additional iTRAQ modification is observable by a mass increase of +144.10 Da for the  $b_5^+$  and  $y_6^+$  (and following) ions. For the latter, the Y-iTRAQ immonium is present at m/z 280.17, whereas the tyrosine immonium ion (m/z 136.07) shows a similar

## 2 Developing a Platform for relative Quantification using iTRAQ

intensity for the doubly iTRAQ derivatized form (A).

Switching to LC-MALDI MS (AB SCIEX 5800 MALDI TOF/TOF), in which predominantly  $1^+$  charged precursors are formed during ionization, an almost complete sequence coverage was observed (Figure 2-34). For both peptide species, the  $b_1^+ - b_9^+$  and  $y_1^+ - y_9^+$  ion series were identified. Additionally, the complete a-ion series was matched but  $a_4^+$  for the unmodified and  $a_8^+$  and  $a_9^+$  for the Y-iTRAQ modified form (not annotated in spectra). Taking advantage of TOF instruments to detect low mass ions, all but two (iG & iI) immonium ions were present (not annotated in spectra). Importantly, the iY-iTRAQ immonium ion at  $m/z$  280.17 was present in the modified, but missing in the unmodified form. The intensity of this ion was approximately 15% and 10% compared to the iTRAQ reporter ions for high and medium gas pressure (Figure 2-34 (B, D)).

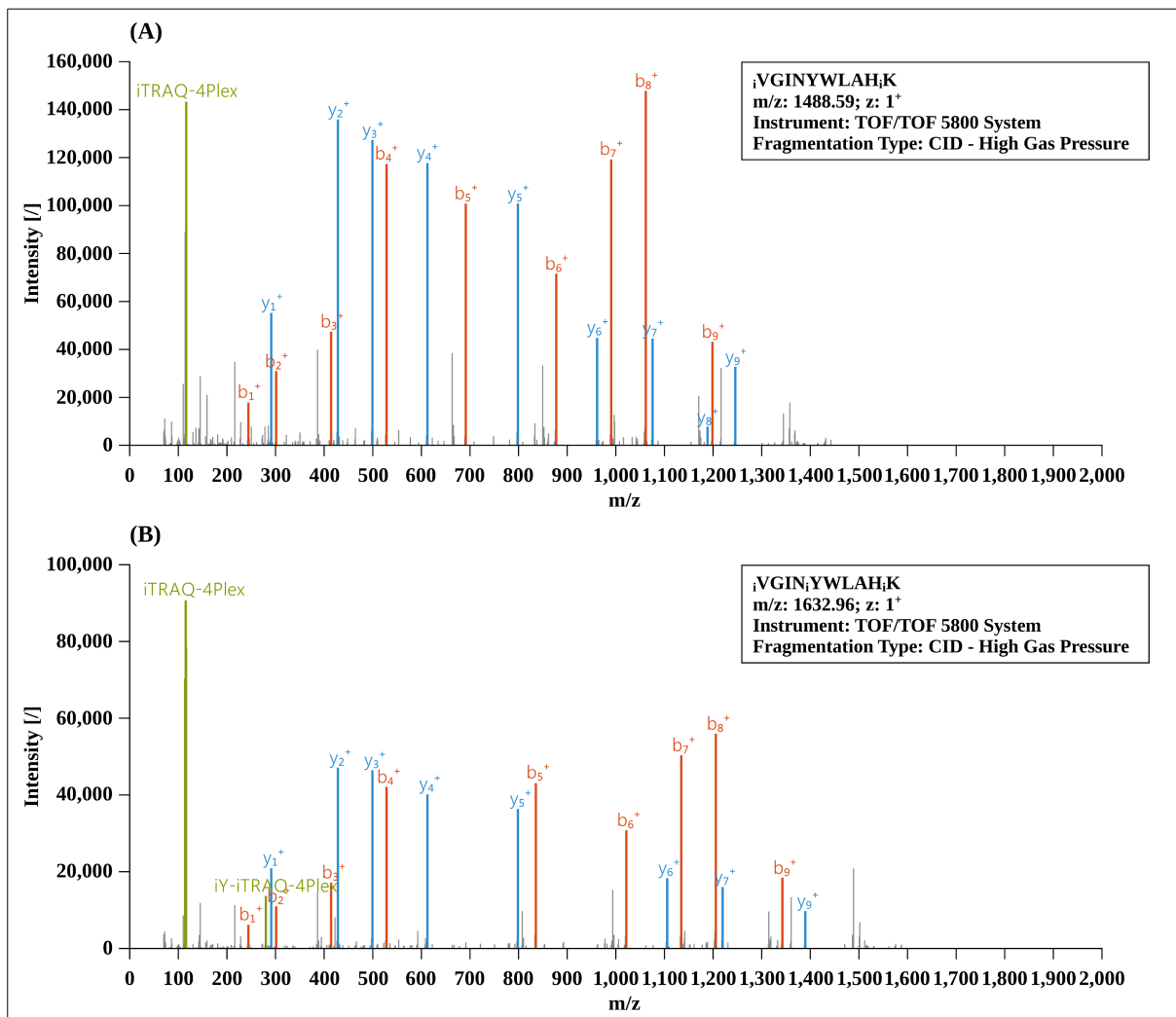


Figure 2-34: Annotated MS/MS spectra after CID fragmentation (AB SCIEX 5800 MALDI TOF/TOF) with higher (A, B) and medium gas pressure (C, D) for the peptide VGINYWLAHK in its native iTRAQ labeled form iVGINYWLAH;K (A,C) and with the additional iTRAQ labeling on tyrosine iVGIN;YWLAH;K (B, D). The Y-iTRAQ immonium at  $m/z$  280.17 is only present in the Y-iTRAQ modified form and more prominent after fragmentation with higher gas pressure (B). Matrix used for MALDI measurements was CHCA.

### 2.3 Y-iTRAQ – Accurate Identification of iTRAQ Tyrosine Labeled Peptides

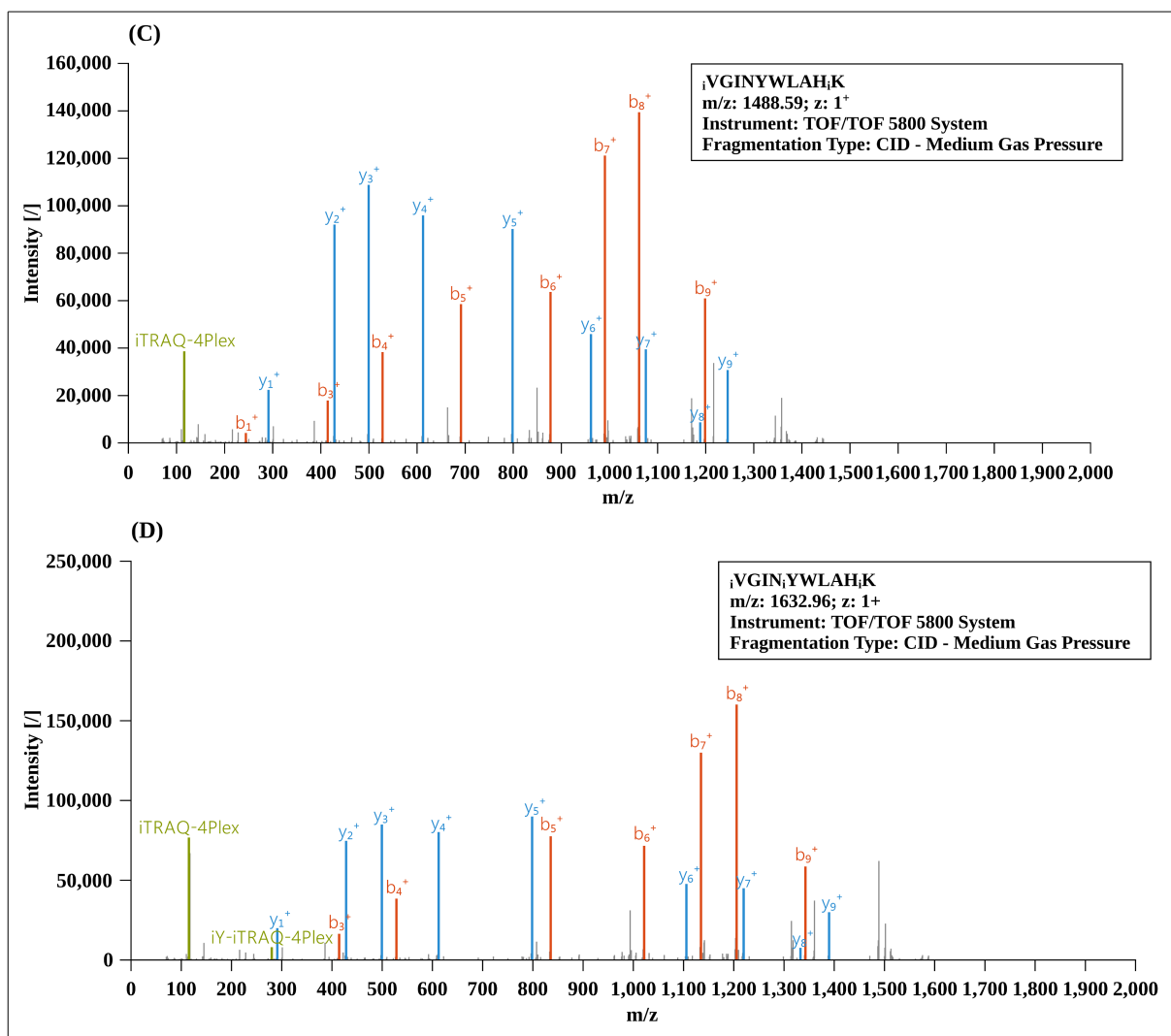


Figure 2-34 (continued): Annotated MS/MS spectra after CID fragmentation (AB SCIEX 5800 MALDI TOF/TOF) with higher (A, B) and medium gas pressure (C, D) for the peptide VGINYWLAHK in its native iTRAQ labeled form  $i$ VGINYWLAH<sub>3</sub>K (A,C) and with the additional iTRAQ labeling on tyrosine  $i$ VGIN<sub>i</sub>YWLAH<sub>3</sub>K (B, D). The Y-iTRAQ immonium at m/z 280.17 is only present in the Y-iTRAQ modified form and more prominent after fragmentation with higher gas pressure (B). Matrix used for MALDI measurements was CHCA.

## 2 Developing a Platform for relative Quantification using iTRAQ

Low  $m/z$  ions (such as iTRAQ reporter ions) are difficult to detected in ion trap mass analyzers due to their instabilities while trapping (208); which is referred to as  $1/3^{\text{rd}}$  cut-off. Therefore, the lowest  $m/z$  value monitored for the peptide VGINYWLAHK was  $m/z$  210 using CID-IT fragmentation on the LTQ Orbitrap Velos. The ion series coverage was comparable for both (Figure 2-35), the unmodified and Y-iTRAQ modified form ( $b_1^+ - b_9^+$ ,  $y_1^+ - y_9^+$  but  $b_1^+$  for the modified form), validating the correct peptide identification. In contrast to quadrupole like fragmentation techniques, the iY-iTRAQ immonium ion was not detected (CID-IT).

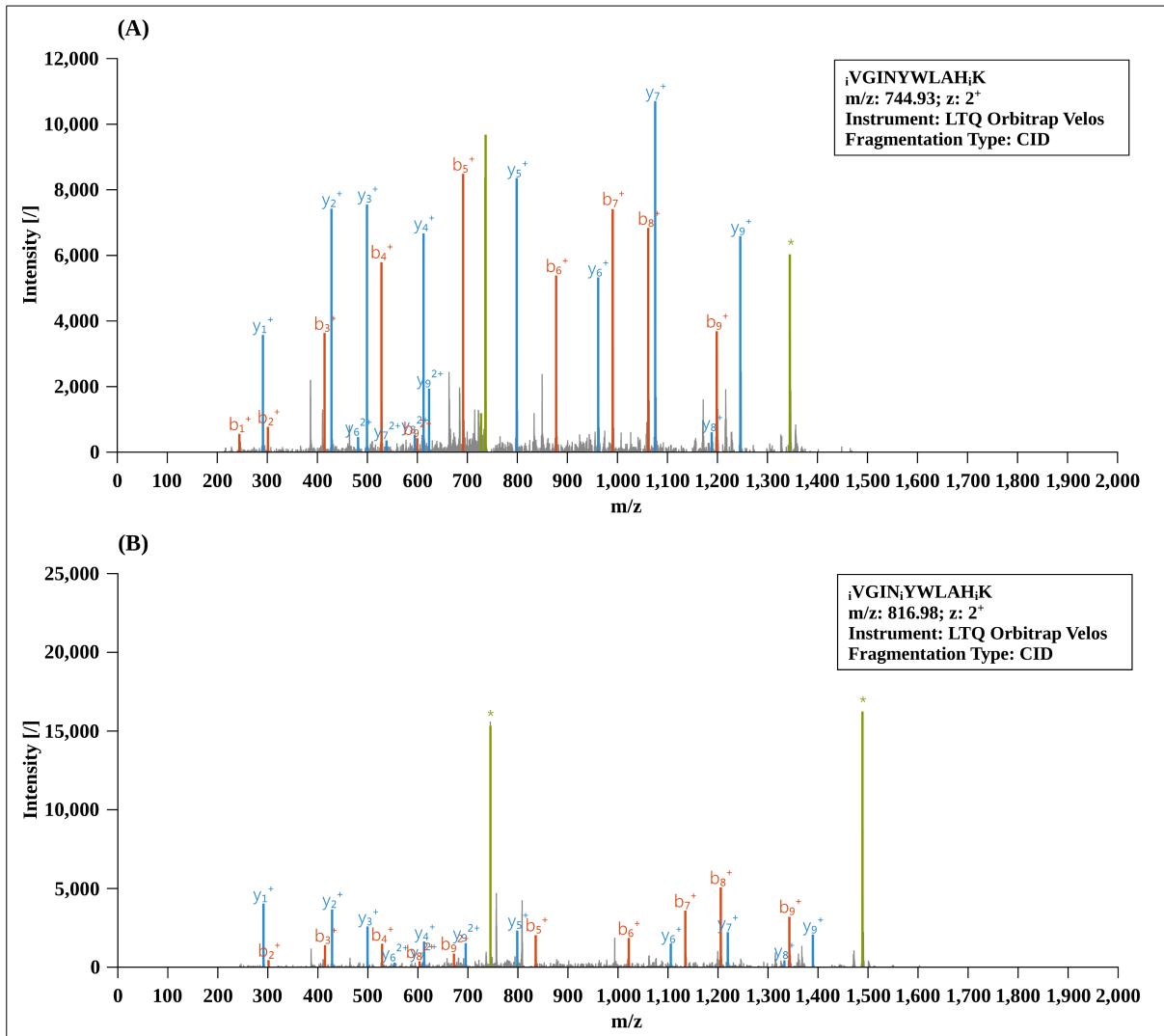


Figure 2-35: Annotated MS/MS spectra after CID fragmentation (Orbitrap Velos, recorded in ion trap) shows the b- and y-ions matched for the Peptide VGINYWLAHK in its typical iTRAQ labeled form  $\gamma$ VGINYWLAH $\gamma$ K (A) and with an additional iTRAQ labeling on the tyrosine  $\gamma$ VGIN $\gamma$ YWLAH $\gamma$ K (B). The additional iTRAQ modification is observable by a mass increase of +144.10 Da for the b<sub>5</sub><sup>+</sup> and y<sub>6</sub><sup>+</sup> (and following) ions.

### 2.3 Y-iTRAQ – Accurate Identification of iTRAQ Tyrosine Labeled Peptides

PQD is a fragmentation technique introduced to overcome the problem to detect low  $m/z$  ions in ion traps. It was also used to analyze the presence of the iY-iTRAQ immonium ion. Similar to CID in ion trap fragmentation, an almost complete ion series was detectable ( $b_2^+ - b_9^+$ ,  $y_1^+ - y_9^+$  but  $b_1^+$  &  $b_2^+$  for the modified form; Figure 2-36). However, the iY-iTRAQ immonium was not detected. Because neither the iY-iTRAQ immonium ion nor iTRAQ reporter ions were detected, the fragmentation settings for PQD may need to be further optimized.

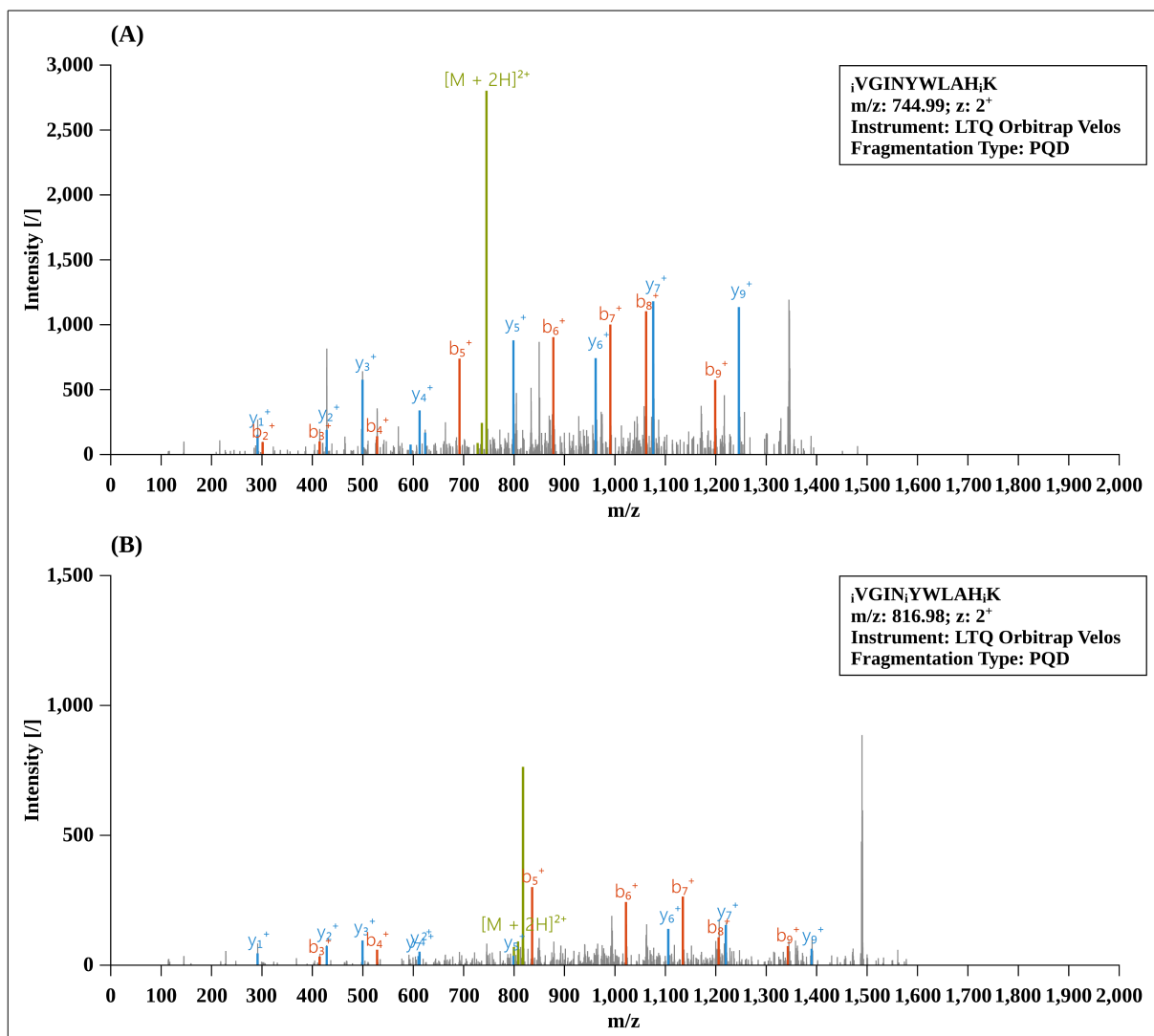


Figure 2-36: Annotated MS/MS spectra after PQD fragmentation (LTQ Orbitrap Velos) shows the  $b$  and  $y$ -ions matched for the Peptide VGINYWLAHK in its common iTRAQ labeled form  $i$ VGINYWLAH,K (A) and with an additional iTRAQ labeling on the tyrosine  $i$ VGIN,YWLAH,K (B).

## 2 Developing a Platform for relative Quantification using iTRAQ

ETD is a fragmentation technique generating c- and z-ions and therefore complements collision induced fragmentation techniques. ETD was also used to fragment the peptide VGINYWLAHK in both modifications forms (Figure 2-37). Using a relative long activation time (160 ms) to fragment the 3<sup>+</sup> charged precursor ion, an almost complete ion series was observed for the unmodified peptide (c<sub>1</sub><sup>+</sup> – c<sub>8</sub><sup>+</sup>, z<sub>1</sub><sup>+</sup> – z<sub>9</sub><sup>+</sup>, (Figure 2-37 (A)). For the Y-iTRAQ modified form (Figure 2-37 (B)), the ions c<sub>2</sub><sup>+</sup> – c<sub>4</sub><sup>+</sup>, c<sub>6</sub><sup>+</sup> and z<sub>1</sub><sup>+</sup> – z<sub>9</sub><sup>+</sup> were matched. Both immonium ions (iY at m/z 136.07 and iY-iTRAQ at m/z 280.17) were not detectable, similar to CID and PQD.

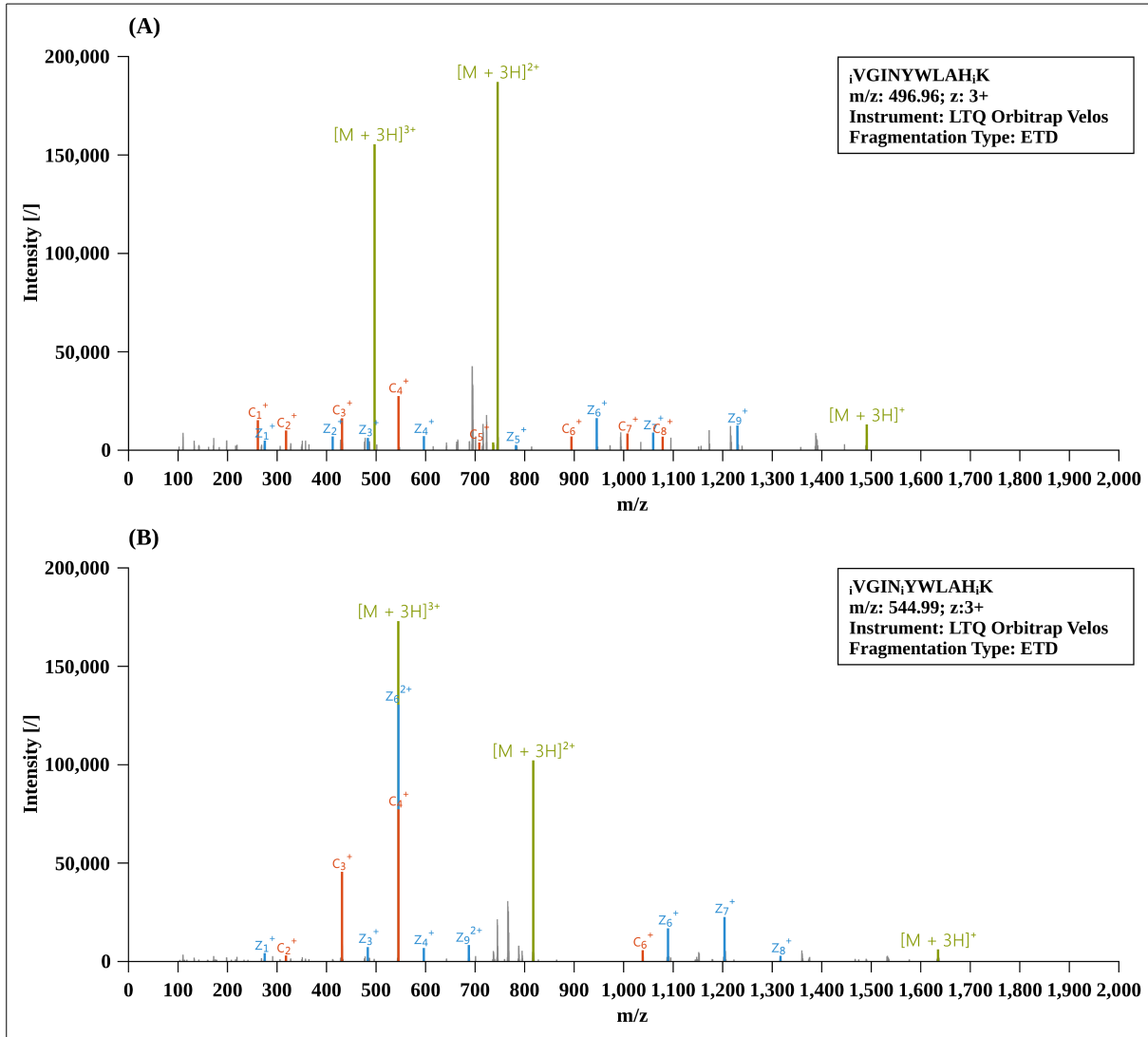


Figure 2-37: The annotated MS/MS spectra after ETD fragmentation (Orbitrap Velos, spectra recorded in Orbitrap mass analyzer) for the peptide VGINYWLAHK in its native iTRAQ labeled form *i*VGINYWLAH<sub>i</sub>K (A) and with the additional iTRAQ labeling on tyrosine *i*VGIN<sub>i</sub>YWLAH<sub>i</sub>K (B).

### 2.3.3.2 Y-iTRAQ Immonium Ion Formation in Dependence of Collision Energy

As shown previously, the intensity of pY immonium ion from pY-peptides increased with applied collision energy (see chapter 2.1.3.1). In contrast, the Y-iTRAQ-4Plex immonium ions carry a group (iTRAQ + balancer group) which is designed to dissociate readily. Therefore, the Y-iTRAQ-4Plex modified peptides were further analyzed to evaluate how the precursor [M], iTRAQ reporter ions [iTRAQ], non-dissociated iTRAQ (= iTRAQ + balancer group, [non-dissociated iTRAQ]) and the tyrosine related immonium ions [iY and iY-iTRAQ-4Plex] were present in spectra at different collision energies. For analysis, the 2<sup>+</sup> charged precursors of the peptide VGINYWLAHK was chosen and subjected to fragmentation by LC-ESI MS on the LTQ Orbitrap Velos (HCD fragmentation) and Synapt G2S (CID, trap fragmentation).

In agreement with previous results for HCD fragmentation (Orbitrap Velos; see chapter 2.1.3.1), the precursor ion represented the most intense ion for 2<sup>+</sup> charged precursors in spectra up to an NCE of 35% (Figure 2-38 (A & B)). With a further increase of 10%, the precursor was no longer detectable. The iTRAQ reporter ion signal reached highest signal intensity when NCE energies ≥ 40% were applied. Undissociated iTRAQ represented a singly charged ion liberated from precursors without further dissociation and showed its maximal relative intensity at 40% NCE. A further increase in NCE led to a drop in intensity, possibly caused by a further fragmentation into reporter ions.

For the unmodified peptide, the tyrosine immonium ion (iY) at m/z 136.07 was detectable after applying an NCE of 40% and its intensity increased linearly with NCE (Figure 2-38 (A), blue line). In contrast, the iY-iTRAQ-4Plex immonium ion (iY-iTRAQ) behaved very differently (Figure 2-38 (A), green line): for NCEs between 35 – 40%, this ion was detectable and its intensity increased up to 60% NCE. At higher NCEs, its intensity dropped, but still had an intensity of approximately 20% compared to iTRAQ reporter ions at 85% NCE.

For the Synapt G2s, which was operated in trap CID mode (from 0 to 100 eV in 5eV steps per LC run), similar trends were observable. For the unmodified form (Figure 2-38 (C)), the tyrosine immonium ion (iY) increased in intensity with increasing CE as observed for HCD fragmentation. The undissociated iTRAQ reporter ion (iTRAQ + balancer group) was decreasing with increasing CE values after exceeding 40 eV. However, the most prominent difference was observed for the iY-iTRAQ immonium ion compared to HCD data (Figure 2-38 (D)): for CE values of 60 – 65 eV, the ion intensity was comparable with iTRAQ reporter ions. A further increase in CE resulted in a lower relative intensity of iY-iTRAQ as seen for HCD fragmentation.

## 2 Developing a Platform for relative Quantification using iTRAQ

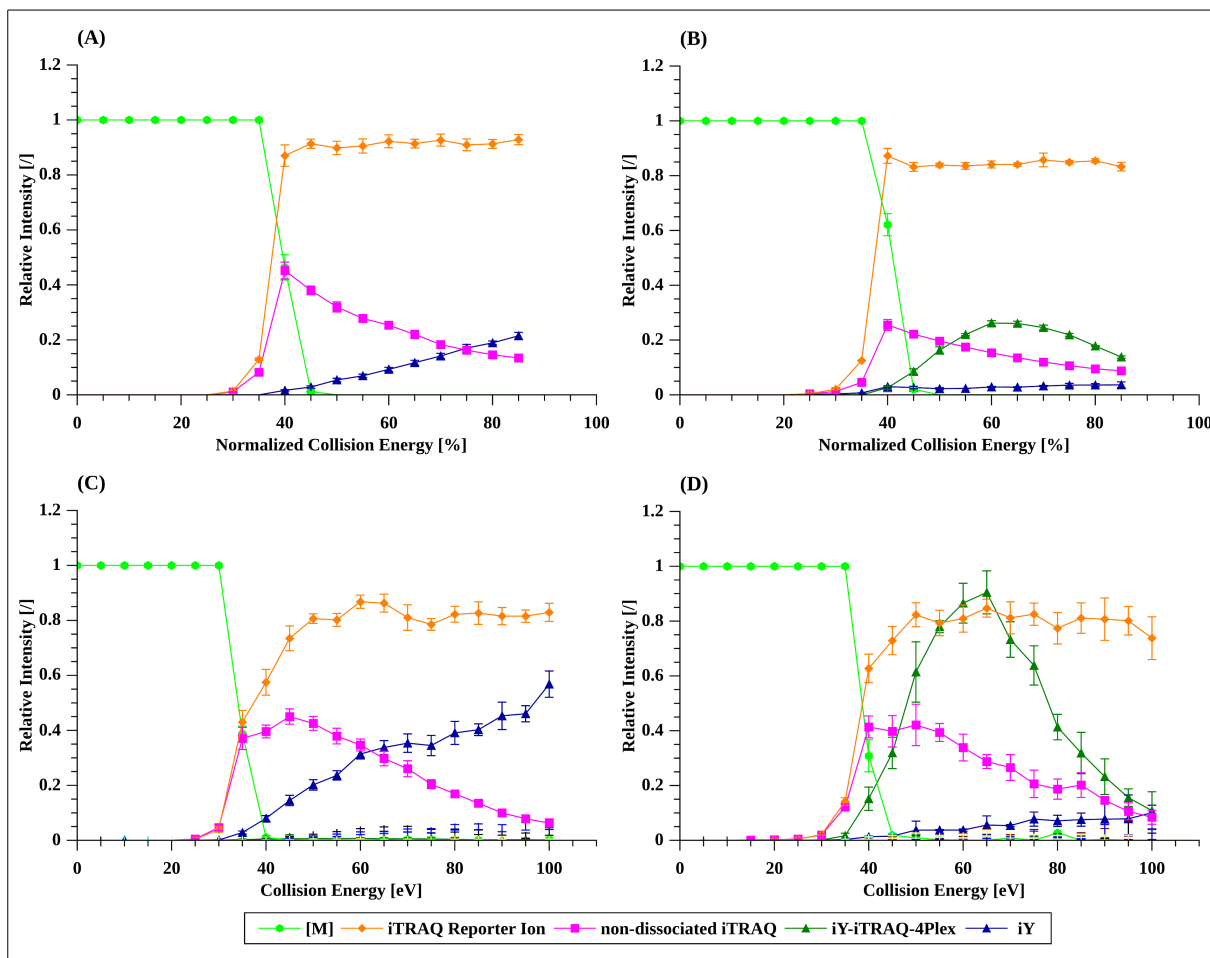


Figure 2-38: The relative intensity (see Material and Methods for description) was calculated for the precursor [M], the non-dissociated iTRAQ reporter ions (iTRAQ + balancer group), iTRAQ reporter ions iTRAQ<sub>114</sub>-iTRAQ<sub>117</sub>, tyrosine iTRAQ-4Plex immonium ion (iY-iTRAQ-4Plex) and the tyrosine immonium ion (iY) at different collision energies (NCEs & CEs) for the peptide VGINYWLAHK. Ion properties were calculated for the common iTRAQ modified (<sub>i</sub>VGINYWLAH<sub>i</sub>K, (A, C)) and the Y-iTRAQ modified form (<sub>i</sub>VGIN<sub>i</sub>YWLAH<sub>i</sub>K, (B, D)) for 2<sup>+</sup> charged precursors. HCD data (LTQ Orbitrap Velos; A,B) and CID data (Synapt G2S; C, D) were compared to demonstrate similar fragmentation pattern.

### 2.3.3.3 Detection of iY-iTRAQ-8Plex specific Immonium Ions

After demonstrating the presence of an iY-iTRAQ immonium ion for iTRAQ-4Plex labeled peptides, a dataset of iTRAQ-8Plex labeled phosphoproteins was investigated to evaluate the presence of an iY-iTRAQ-8Plex immonium ion after beam-type CID fragmentation, although the structure of iTRAQ-8Plex has not published.

The peptide LFRQFYQLD was used as an example, which was identified in both forms (with and without iTRAQ modification, Figure 2-39). After HCD fragmentation of the 2<sup>+</sup> charge precursor, the matched number of b- and y-fragment ions by SEQUEST was significantly higher for the unmodified peptide (10/18; <sub>i</sub>LFRQFYQLD) rather than the modified (<sub>i</sub>LFRQF<sub>i</sub>YQLD) form (7/18). Not all of the fragment ions matched by SEQUEST exceeded the intensity cut-off value of 1% (compared the most intense signal in spectra) used to mark

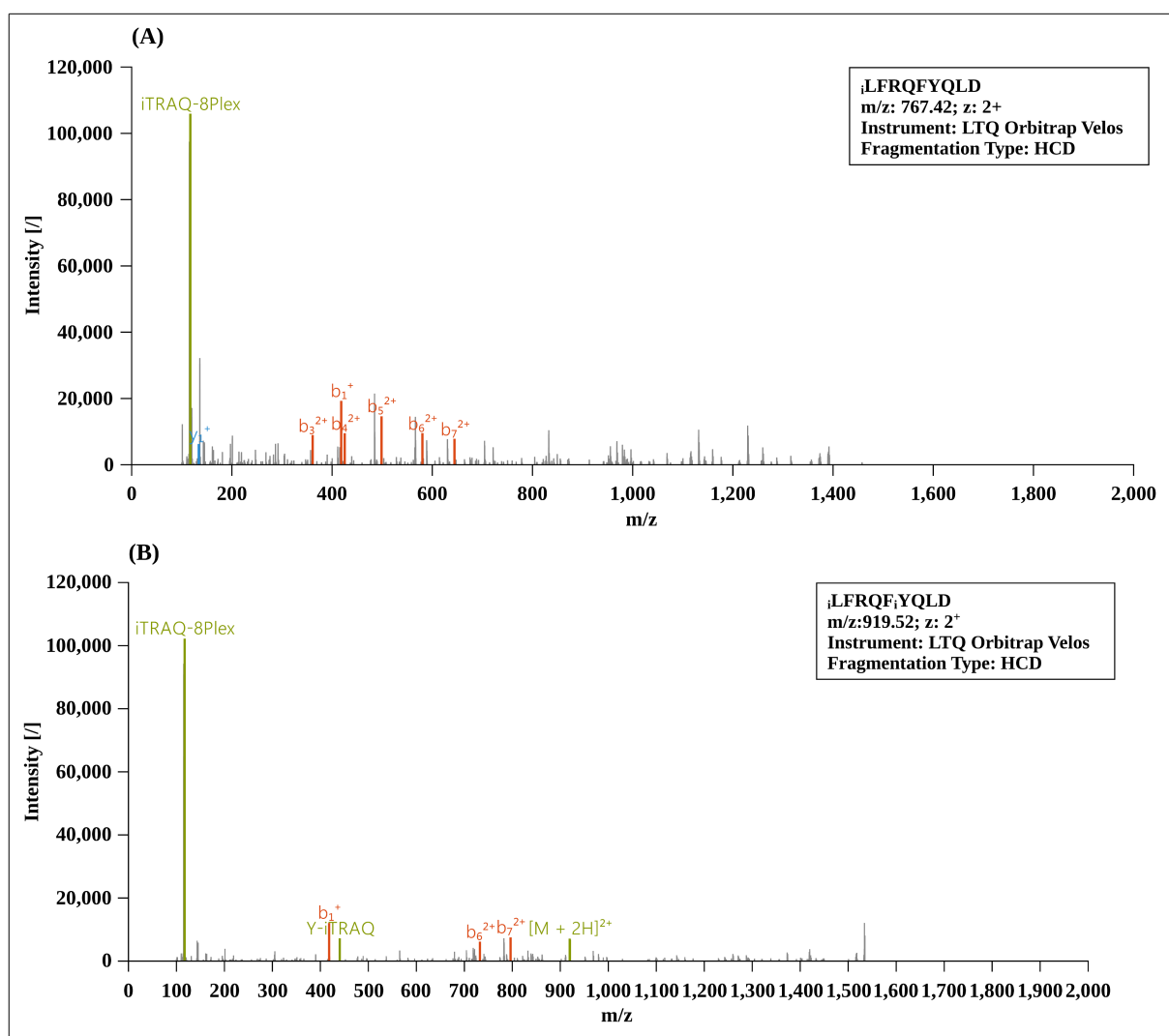


### 2.3 Y-iTRAQ – Accurate Identification of iTRAQ Tyrosine Labeled Peptides

fragments ions in tandem MS spectra in Figure 2-39. Within the peptide, tyrosine is located at the sixth position. Both  $b_6^{2+}$  ions were detected at  $m/z$  580.33 and 732.43 for the unmodified and modified form, respectively. This clearly showed the correct identification of this peptide in addition to validate the peptide by database search.

For the modified form ( $i$ LFRQF<sub>i</sub>YQLD, Figure 2-39 (B)), a signal at  $m/z$  440.28 was detected with a high intensity similar to the remaining b- and y-ions. This signal matched to the theoretically calculated  $i$ Y-iTRAQ-8Plex immonium ion. Supportingly, this ion was not detectable in the unmodified form.

Additionally, an iTRAQ-8Plex modified tryptic peptide (EPMIGVNQELAYFYPELFR, Figure 2-39 (C)) was also analyzed. For the unmodified form ( $i$ EPMIGVNQELAYFYPELFR), 17 of 38 b- and y-fragment ions could be matched by SEQUEST. For the Y-iTRAQ-8Plex labeled form ( $i$ EPMIGVNQELA<sub>i</sub>YFYPELFR), slightly more fragment ions could be matched (21/38). The  $i$ Y-iTRAQ-8Plex specific immonium ion at  $m/z$  440.28 was again only present in the Y-iTRAQ modified peptide (Figure 2-39 (D)). This indicates its uniqueness in terms of detection.



## 2 Developing a Platform for relative Quantification using iTRAQ

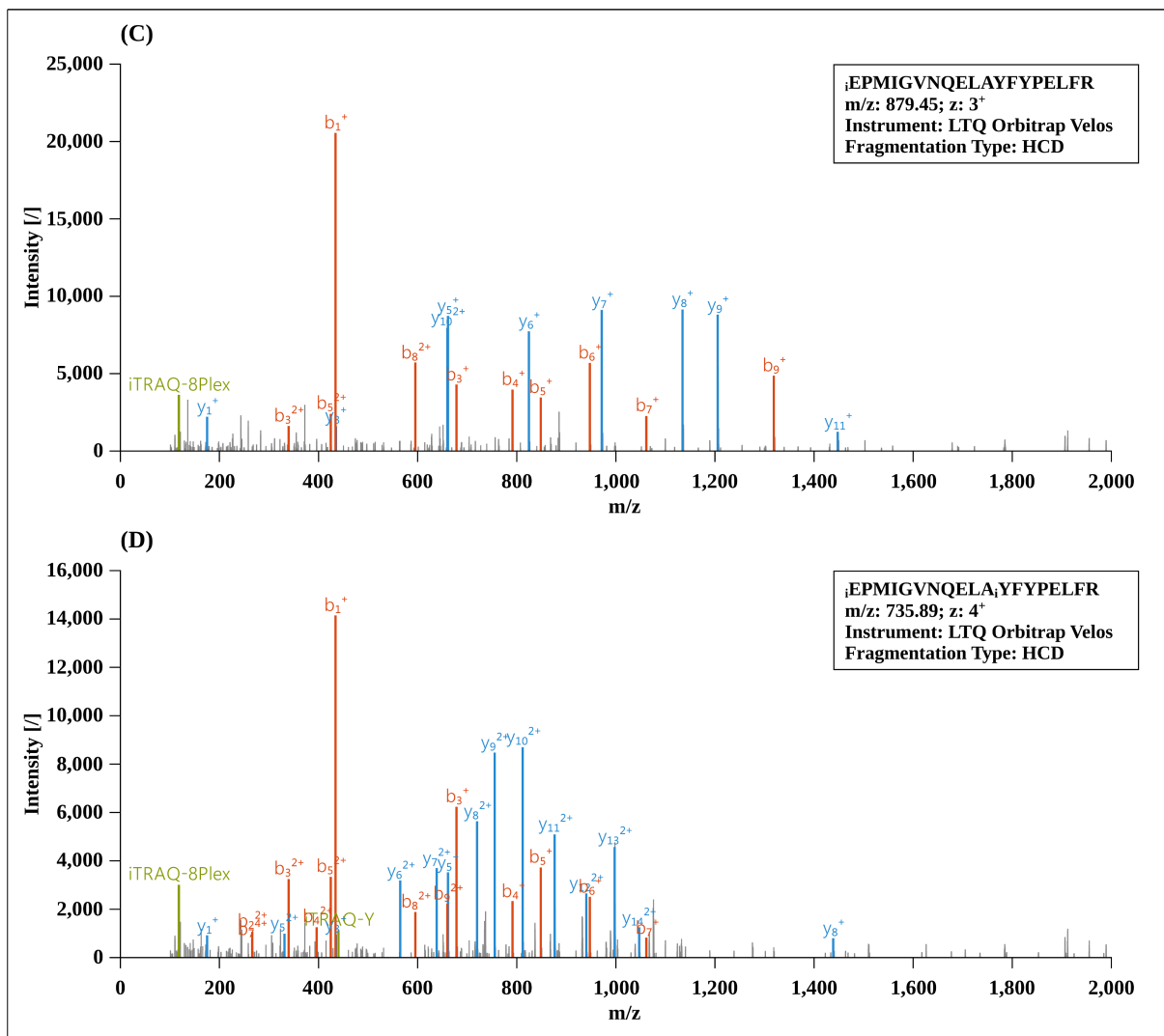


Figure 2-39: Annotated MS/MS spectra after HCD fragmentation (LTQ Orbitrap Velos) shows the b- and y-ions matched for the *Staphylococcus aureus* V-8 Protease (GluC) derived peptide LFRQFYQLD (A, B) and the tryptic peptide EPMIGVNQELAYFYPELFR in its common iTRAQ labeled form *i*LFRQFYQLD (A), *i*EPMIGVNQELAYFYPELFR (C) and with an additional iTRAQ labeling on the tyrosine, *i*LFRQFYQLD (B), *i*EPMIGVNQELAYFYPELFR (D). The additional iTRAQ modification mass shift is obvious for the b<sub>6</sub><sup>+</sup> (and following) ions. For the latter, the *i*Y-iTRAQ immonium is present at m/z 440.28, which equals the mass of tyrosine immonium ion (m/z 136.07) plus an iTRAQ-8Plex labeling (+304.20 Da for iTRAQ<sub>115, 118, 119 & 121</sub> and +304.21 Da for iTRAQ<sub>113, 114, 116 & 117</sub>; according to UniMod Accession numbers 730 and 731).

### 2.3.3.4 Influence of iTRAQ tyrosine Labeling on the LC Retention Time

After demonstrating the uniqueness of iY-iTRAQ-4/8Plex immonium ions for iTRAQ tyrosine labeled peptides, the influence of this additional iTRAQ modification was further investigated. Not only did the peptide mass increase by 144.0995 Da (iTRAQ-4Plex) and 304.2053 Da (iTRAQ-8Plex), respectively, but also the physiochemical properties of a peptide changed. Therefore, the retention time difference for unmodified and iTRAQ tyrosine labeled peptides were compared to elucidate if one of the species eluted earlier by reversed phase LC separation. Such information can be used to validate potential Y-iTRAQ modified peptides in large scale iTRAQ proteomics studies.

For the peptide VGINYWLAHK, the XICs of the unmodified (red) and Y-iTRAQ-4Plex labeled form (blue) were extracted and plotted (Figure 2-39). Two pieces of information can be seen in this figure: First, the XIC of the modified form ( $8.51 \times 10^5$ ) was significantly lower compared to the unmodified form ( $1.16 \times 10^7$ ). However, the modified form exhibits 7% of the unmodified maximal XIC intensity. Second, the Y-iTRAQ-4Plex modified form eluted approximately 4 minutes earlier. To verify if this was reproducible for other peptides showing both modification forms, the analysis was repeated for the peptide HIIVACEGNPYVPVHFDAVS (unmodified: green and Y-iTRAQ-4Plex: pink) and NDTGSTDYGILQINSR (unmodified: orange Y-iTRAQ-4Plex: dark yellow). Again, the retention time shift was varying between three to four minutes. Here, the relative intensity of the Y-iTRAQ-4Plex modified forms were 10 and 14%, respectively. This high XIC ratio was unexpected due to the described number of side-reactions being smaller than 3%, but might be caused by the excess of iTRAQ reagent used for labeling. A further set of 36 peptides was analyzed based on their retention time difference between the unmodified and Y-iTRAQ-4Plex modified form (Table 2-6). Only for the peptide VAGTWYSLAMAASDISLLDAQSAPLR, a higher retention time was observed for the Y-iTRAQ-4Plex modified peptide. The overall retention time difference was about 3.3 minutes.

A further analysis of the iTRAQ-8Plex labeled peptides showed similar results (Table 2-7). In total, 79 peptides could be identified in both forms (unmodified and Y-iTRAQ-8Plex). 75 iTRAQ-8Plex modified peptides, equaling 95% of the peptides, elute earlier than their unmodified counterpart. Only four iTRAQ-8Plex modified peptides did not follow this trend (marked bold in Table 2-7). Three of the four peptides shared the same peptide structure (T)VYQH(Q)KA, which might be an indicator that this is a peptide specific effect. Moreover, these peptides eluted earliest (at approx. 22 min) compared to the remaining peptides listed in Table 2-7. For the fourth peptide (QYLYQGPIVLNPWDQV), the higher retention time of the Y-iTRAQ modified peptide cannot be explained. In average, the difference in retention time between unmodified and Y-iTRAQ-8Plex increased to 4.15 minutes. This might be caused by slight variations during LC separation or, more likely, the overall bigger structure of iTRAQ-8Plex compared to iTRAQ-4Plex.

## 2 Developing a Platform for relative Quantification using iTRAQ

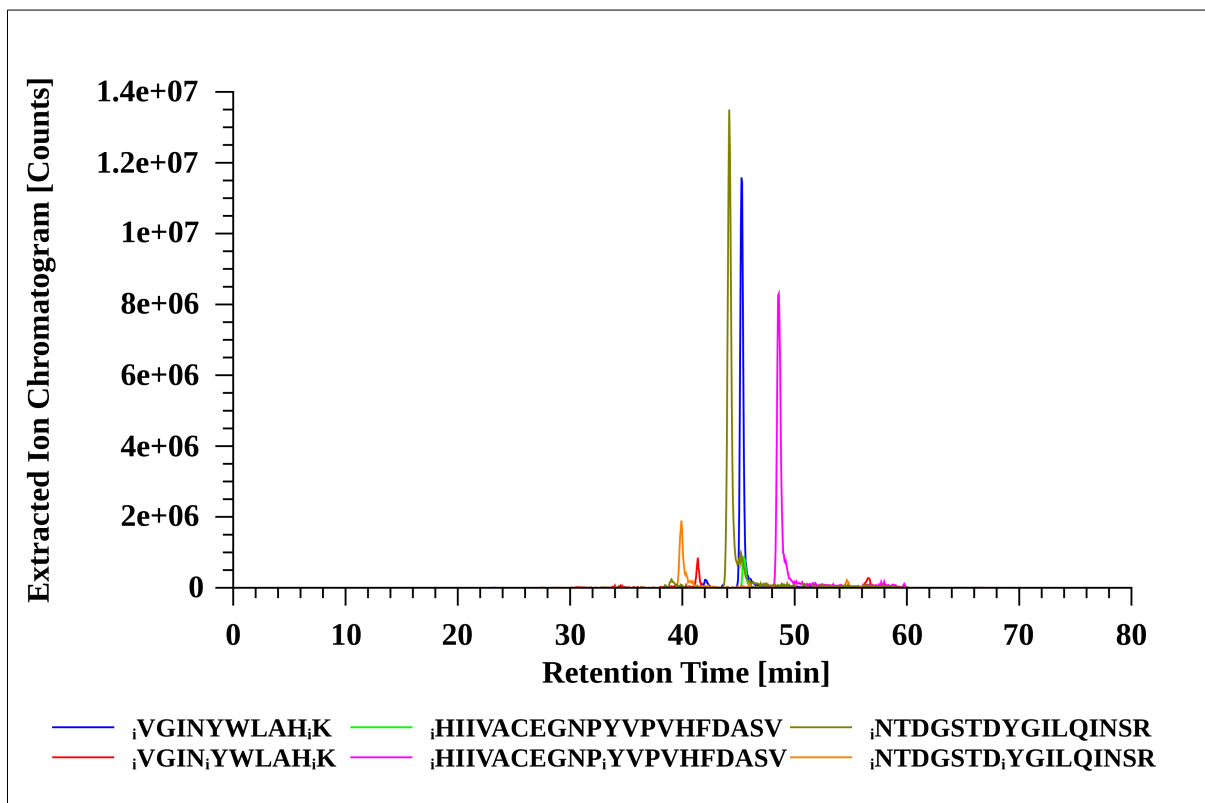


Figure 2-40: Overlay of Extracted Ion Chromatograms (XICs) for three tryptic peptides (iTRAQ-4Plex labeled) being modified on the  $\epsilon$ -amino and N-terminal amino group (upper row, unmodified) or carrying an additional modification on tyrosine (lower row, Y-modified).

Table 2-6: Median retention times [min] and the number of features (counts) were calculated for the iTRAQ-4Plex labeled ten protein-mix. Based on the total number of Peptide-Spectrum-Matches of 2x 18 LC runs, the retention time differences for tyrosine iTRAQ-4Plex (Y-iTRAQ) labeled and in its native form (unmodified, without iTRAQ modification on tyrosine). Except one peptide (VAGTWYSLAMAASDISLLDAQSAPLR), the retention time for the Y-iTRAQ labeled peptide species was usually lower compared to the native form.

Peptide Sequence	Modification	Median RT [min]		Number of Features	
		Y-iTRAQ	native	Y-iTRAQ	native
-	(including the Y-iTRAQ)				
TVYQH <sub>i</sub> QK	N-Term(iTRAQ4plex); Y3(iTRAQ4plex); K7(iTRAQ4plex)	23.74	24.92	2	128
RHGLDN <sub>i</sub> YR	N-Term(iTRAQ4plex); Y7(iTRAQ4plex)	25.59	25.84	53	166
HGLDN <sub>i</sub> YR	N-Term(iTRAQ4plex); Y6(iTRAQ4plex)	26.38	27.16	107	316
KYIPG <sub>i</sub> TK	N-Term(iTRAQ4plex); K1(iTRAQ4plex); Y2(iTRAQ4plex); K7(iTRAQ4plex)	28.73	30.64	50	229
AVPYP <sub>i</sub> QR	N-Term(iTRAQ4plex); Y4(iTRAQ4plex)	29.35	31.83	90	144
YPNCAY <sub>i</sub> K	N-Term(iTRAQ4plex); Y1(iTRAQ4plex); C4(Carbamidomethyl); K7(iTRAQ4plex)	30.00	32.68	105	219
KPVEEYANCHLAR	N-Term(iTRAQ4plex); K1(iTRAQ4plex); Y6(iTRAQ4plex); C9(Carbamidomethyl)	31.46	33.39	131	193
KTGQAPGFTYTDANK	N-Term(iTRAQ4plex); K1(iTRAQ4plex); Y10(iTRAQ4plex); K15(iTRAQ4plex)	32.05	34.07	40	664
QHMS <sub>i</sub> SSTAASSNYCNQMMK	N-Term(iTRAQ4plex); Y15(iTRAQ4plex); C16(Carbamidomethyl); K21(iTRAQ4plex)	32.19	34.32	159	100
DGPLTGT <sub>i</sub> YR	N-Term(iTRAQ4plex); Y8(iTRAQ4plex)	31.26	34.50	76	139
YIPG <sub>i</sub> TK	N-Term(iTRAQ4plex); Y1(iTRAQ4plex); K6(iTRAQ4plex)	32.08	35.50	44	372
KPVDEYKDCHLAQVPSHTVVAR	K1(iTRAQ4plex); N-Term(iTRAQ4plex); Y6(iTRAQ4plex); K7(iTRAQ4plex); C9(Carbamidomethyl)	34.40	35.71	2	63

## 2.3 Y-iTRAQ – Accurate Identification of iTRAQ Tyrosine Labeled Peptides

Table 2-6 (continued): Median retention times [min] and the number of features (counts) were calculated for the iTRAQ-4Plex labeled ten protein-mix. Based on the total number of Peptide-Spectrum-Matches of 2x 18 LC runs, the retention time differences for tyrosine iTRAQ-4Plex (Y-iTRAQ) labeled and in its native form (unmodified, without iTRAQ modification on tyrosine). Except one peptide (VAGTWYSLAMAASDISLLDAQSAPLR), the retention time for the Y-iTRAQ labeled peptide species was usually lower compared to the native form.

Peptide Sequence	Modification (including the Y-iTRAQ)	Median RT [min]		Number of Features	
		Y-iTRAQ	native	Y-iTRAQ	native
-					
TGQAPGFTYTDANK	N-Term(iTRAQ4plex); Y9(iTRAQ4plex); K14(iTRAQ4plex)	33.61	36.68	95	423
YLGEEYVK	N-Term(iTRAQ4plex); Y1(iTRAQ4plex); K8(iTRAQ4plex)	36.38	40.19	83	190
KTEREDLIAYLK	N-Term(iTRAQ4plex); K1(iTRAQ4plex); Y10(iTRAQ4plex); K12(iTRAQ4plex)	38.20	41.31	3	413
YAAELHLVHWNTK	N-Term(iTRAQ4plex); Y1(iTRAQ4plex); K13(iTRAQ4plex)	40.15	43.15	47	255
ADRDQYELLCLDNTR	N-Term(iTRAQ4plex); Y6(iTRAQ4plex); C10(Carbamidomethyl)	39.70	43.25	15	282
NTDGSTDYGLQINSR	N-Term(iTRAQ4plex); Y8(iTRAQ4plex)	40.04	44.56	182	609
VLVLDTDYK	N-Term(iTRAQ4plex); Y8(iTRAQ4plex); K9(iTRAQ4plex)	39.75	44.72	89	249
TEREDLIAYLK	N-Term(iTRAQ4plex); Y9(iTRAQ4plex); K11(iTRAQ4plex)	41.62	44.81	11	286
EETLMEYLENPKK	N-Term(iTRAQ4plex); Y7(iTRAQ4plex); K12(iTRAQ4plex); K13(iTRAQ4plex)	43.38	47.40	7	416
GYSLGNVWVCAAK	N-Term(iTRAQ4plex); Y2(iTRAQ4plex); C9(Carbamidomethyl); K12(iTRAQ4plex)	42.26	47.60	103	728
HIIVACEGNPYVPVHFDASV	N-Term(iTRAQ4plex); C6(Carbamidomethyl); Y11(iTRAQ4plex)	45.51	48.77	189	424
EDPQTFYAVAVVK	N-Term(iTRAQ4plex); Y7(iTRAQ4plex); K14(iTRAQ4plex)	45.58	50.82	41	335
MYLGYEYVTAIR	N-Term(iTRAQ4plex); Y2(iTRAQ4plex)	45.87	51.14	167	272
EDLIAYLK	N-Term(iTRAQ4plex); Y6(iTRAQ4plex); K8(iTRAQ4plex)	44.29	51.23	66	1441
EETLMEYLENPK	N-Term(iTRAQ4plex); Y7(iTRAQ4plex); K12(iTRAQ4plex)	46.56	51.85	98	490
VGINYWLAHK	N-Term(iTRAQ4plex); Y5(iTRAQ4plex); K10(iTRAQ4plex)	41.37	51.95	303	3047
AVVQDPALKPLALVYGEATSR	N-Term(iTRAQ4plex); K9(iTRAQ4plex); Y15(iTRAQ4plex)	48.28	52.68	38	610
GITWKEETLMEYLENPK	N-Term(iTRAQ4plex); K5(iTRAQ4plex); Y12(iTRAQ4plex); K17(iTRAQ4plex)	50.03	53.34	13	305
VYVEELKPTPEGDLEILLQK	N-Term(iTRAQ4plex); Y2(iTRAQ4plex); K7(iTRAQ4plex); K20(iTRAQ4plex)	50.27	53.74	15	279
TAGWNIPMGLLYNK	N-Term(iTRAQ4plex); Y12(iTRAQ4plex); K14(iTRAQ4plex)	49.40	54.32	65	257
YLLFCMENSAPLQSLACQCLV R	Y1(iTRAQ4plex); N-Term(iTRAQ4plex); C5(Carbamidomethyl); C18(Carbamidomethyl); C20(Carbamidomethyl)	51.13	55.52	10	280
YGDFGTAAQQPDGLAVVGVFLK	N-Term(iTRAQ4plex); Y1(iTRAQ4plex); K22(iTRAQ4plex)	53.56	58.05	33	421
VAGTWYSLAMAASDISLLDAQS APLR	N-Term(iTRAQ4plex); Y6(iTRAQ4plex)	59.43	58.29	135	24
AIAANEADAVTLDAGLVYDAYL APNNLKPVVAEFGYSK	N-Term(iTRAQ4plex); Y21(iTRAQ4plex); K28(iTRAQ4plex); K38(iTRAQ4plex)	58.10	59.66	6	75
YLEFISDAIIHVLHSAK	N-Term(iTRAQ4plex); Y1(iTRAQ4plex); K16(iTRAQ4plex)	58.52	59.80	18	70
SAGWNIPIGLLYCDLPEPR	N-Term(iTRAQ4plex); Y12(iTRAQ4plex); C13(Carbamidomethyl)	54.59	59.81	98	574
DMPIQAFLLYQEPVLGPVR	N-Term(iTRAQ4plex); Y10(iTRAQ4plex)	55.34	60.22	52	521

## 2 Developing a Platform for relative Quantification using iTRAQ

Table 2-7: Median retention times [min] and the number of features (counts) were calculated for the iTRAQ-8Plex labeled 10 protein-mix. Based on the total number of Peptide-Spectrum-Matches of 9 LC runs, the retention time differences for tyrosine iTRAQ-8Plex (Y-iTRAQ) labeled and in its native form (without iTRAQ modification on tyrosine). Except four peptides (bold), the retention time for the Y-iTRAQ labeled peptide species was usually smaller compared to the native form.

Peptide Sequence	Modification	Median RT [min]		Number of Features	
		Y-iTRAQ	native	Y-iTRAQ	native
-	(Including Y-iTRAQ)				
<b>TVYQHQQ</b>	N-Term(iTRAQ8plex); Y3(iTRAQ8plex); K7(iTRAQ8plex)	22.27	22.21	1	75
<b>VYQHQQ</b>	N-Term(iTRAQ8plex); Y2(iTRAQ8plex)	22.67	20.29	21	4
<b>VYQHQKA</b>	N-Term(iTRAQ8plex); Y2(iTRAQ8plex); K6(iTRAQ8plex)	23.16	21.97	15	37
AVPYPQR	N-Term(iTRAQ8plex); Y4(iTRAQ8plex)	25.96	26.20	7	65
ITVDDKHYQK	N-Term(iTRAQ8plex); K6(iTRAQ8plex); Y8(iTRAQ8plex); K10(iTRAQ8plex)	26.25	27.07	1	20
QLDAYPS	N-Term(iTRAQ8plex); Y5(iTRAQ8plex)	29.59	30.70	12	2
VPYPQRDMPIQA	N-Term(iTRAQ8plex); Y3(iTRAQ8plex); M8(Oxidation)	29.61	31.09	11	14
QYTDAPSF	N-Term(iTRAQ8plex); Y2(iTRAQ8plex)	30.56	32.78	7	34
KAVPYPQRDMPIQAF	N-Term(iTRAQ8plex); K1(iTRAQ8plex); Y5(iTRAQ8plex); M10(Oxidation)	31.33	32.96	15	116
RYLGYLE	N-Term(iTRAQ8plex); Y5(iTRAQ8plex)	32.95	35.11	1	2
VPYPQRDMPI	N-Term(iTRAQ8plex); Y3(iTRAQ8plex)	33.10	35.25	2	34
RQFYQLDAYPS	N-Term(iTRAQ8plex); Y9(iTRAQ8plex)	33.15	35.39	2	21
SRYPYGLNYY	N-Term(iTRAQ8plex); Y3(iTRAQ8plex)	33.48	40.99	2	6
YKVPQLEIVPNSAEER	N-Term(iTRAQ8plex); Y1(iTRAQ8plex); K2(iTRAQ8plex)	33.57	35.36	3	90
KAVPYPQRDMPIQAF	N-Term(iTRAQ8plex); K1(iTRAQ8plex); Y5(iTRAQ8plex)	33.59	35.46	19	93
DVPSERYLGYLE	N-Term(iTRAQ8plex); Y7(iTRAQ8plex); Y10(iTRAQ8plex)	33.67	35.26	2	94
QKFPQYLQY	N-Term(iTRAQ8plex); K2(iTRAQ8plex); Y9(iTRAQ8plex)	33.96	35.69	1	26
YYVPLGTQY	N-Term(iTRAQ8plex); Y2(iTRAQ8plex); Y9(iTRAQ8plex)	34.32	36.56	3	50
YVPLGTQY	N-Term(iTRAQ8plex); Y8(iTRAQ8plex)	34.37	37.60	43	41
YKVPQLEIVPNSAEER	N-Term(iTRAQ8plex); Y1(iTRAQ8plex); K2(iTRAQ8plex); S12(Phospho)	34.49	43.48	46	255
FYQLDAYPS	N-Term(iTRAQ8plex); Y2(iTRAQ8plex); Y7(iTRAQ8plex)	34.87	36.99	1	22
KAVPYPQRDMPIQAF	N-Term(iTRAQ8plex); K1(iTRAQ8plex); Y5(iTRAQ8plex); M10(Oxidation)	34.93	36.79	2	58
ALPQYL	N-Term(iTRAQ8plex); Y5(iTRAQ8plex)	35.14	39.69	10	9
DVPSERYLGYLE	N-Term(iTRAQ8plex); Y10(iTRAQ8plex)	35.26	38.56	96	151
QLDAYPSGAW	N-Term(iTRAQ8plex); Y5(iTRAQ8plex)	35.33	39.70	26	76
QFYQLDAYPS	N-Term(iTRAQ8plex); Y8(iTRAQ8plex)	35.98	39.84	3	14
FALPQYLK	N-Term(iTRAQ8plex); Y6(iTRAQ8plex); K8(iTRAQ8plex)	36.02	38.99	2	72
ALNEINQFYQK	N-Term(iTRAQ8plex); Y9(iTRAQ8plex); K11(iTRAQ8plex)	36.09	38.41	3	185
YPPGPIPN	N-Term(iTRAQ8plex); Y1(iTRAQ8plex)	36.34	40.73	3	16
YQKFPQYLQY	N-Term(iTRAQ8plex); K3(iTRAQ8plex); Y10(iTRAQ8plex)	36.37	38.51	3	64
YYVPLGTQY	N-Term(iTRAQ8plex); Y9(iTRAQ8plex)	36.58	40.57	50	31
LLYQEPVLGPVR	N-Term(iTRAQ8plex); Y3(iTRAQ8plex)	36.62	40.06	2	93
NQFYQKFPQYL	N-Term(iTRAQ8plex); K6(iTRAQ8plex); Y10(iTRAQ8plex)	36.73	39.80	3	16
YPELF	N-Term(iTRAQ8plex); Y1(iTRAQ8plex)	37.08	41.87	2	1
FYQLDAYPS	N-Term(iTRAQ8plex); Y7(iTRAQ8plex)	37.36	42.29	39	66
LAYFYPE	N-Term(iTRAQ8plex); Y3(iTRAQ8plex); Y5(iTRAQ8plex)	37.52	40.38	2	29
YPPGPI	N-Term(iTRAQ8plex); Y1(iTRAQ8plex)	37.53	42.76	1	11
QLDAYPSGAWY	N-Term(iTRAQ8plex); Y11(iTRAQ8plex)	37.63	41.86	80	42

## 2.3 Y-iTRAQ – Accurate Identification of iTRAQ Tyrosine Labeled Peptides

Table 2-7 (continue): Median retention times [min] and the number of features (counts) were calculated for the iTRAQ-8Plex labeled 10 protein-mix. Based on the total number of Peptide-Spectrum-Matches of 9 LC runs, the retention time differences for tyrosine iTRAQ-8Plex (Y-iTRAQ) labeled and in its native form (without iTRAQ modification on tyrosine). Except four peptides (bold), the retention time for the Y-iTRAQ labeled peptide species was usually smaller compared to the native form.

Peptide Sequence	Modification	Median RT [min]		Number of Features	
		Y-iTRAQ	native	Y-iTRAQ	native
-	(Including Y-iTRAQ)				
LYQEPVLGPV	N-Term(iTRAQ8plex); Y2(iTRAQ8plex)	37.78	42.66	1	25
VLSRYPSYGLNYY	N-Term(iTRAQ8plex); Y13(iTRAQ8plex)	37.84	47.27	2	31
VYFPFGPIPN	N-Term(iTRAQ8plex); Y2(iTRAQ8plex)	37.99	42.94	4	47
YFYPELFRQ	N-Term(iTRAQ8plex); Y3(iTRAQ8plex)	38.38	42.64	30	125
MPPFKYPVEPFTE	N-Term(iTRAQ8plex); K5(iTRAQ8plex); Y6(iTRAQ8plex)	38.50	41.87	3	58
VYFPFGPI	N-Term(iTRAQ8plex); Y2(iTRAQ8plex)	38.76	44.78	21	44
YFYPELF	N-Term(iTRAQ8plex); Y1(iTRAQ8plex); Y3(iTRAQ8plex)	39.08	42.81	1	20
LFQFYQLD	N-Term(iTRAQ8plex); Y6(iTRAQ8plex)	39.28	43.44	75	166
QYTDAPSFSDIPNPI	N-Term(iTRAQ8plex); Y2(iTRAQ8plex)	39.56	43.77	29	209
QSLVYFPFGPIPN	N-Term(iTRAQ8plex); Y5(iTRAQ8plex)	39.65	44.75	43	143
WYYVPLGT	N-Term(iTRAQ8plex); Y2(iTRAQ8plex)	39.69	45.83	1	1
VYFPFGPIPNSLPQ	N-Term(iTRAQ8plex); Y2(iTRAQ8plex)	39.79	45.27	2	62
FLLYQEPV	N-Term(iTRAQ8plex); Y4(iTRAQ8plex)	39.79	45.68	35	63
SLVYFPFGPIPN	N-Term(iTRAQ8plex); Y4(iTRAQ8plex)	39.86	45.17	9	40
GYLEQLL	N-Term(iTRAQ8plex); Y2(iTRAQ8plex)	39.91	45.22	6	54
YIPIQYVLSR	N-Term(iTRAQ8plex); Y6(iTRAQ8plex)	39.96	43.34	1	40
LVYFPFGPIPN	N-Term(iTRAQ8plex); Y3(iTRAQ8plex)	40.24	46.44	2	23
LAYFYPE	N-Term(iTRAQ8plex); Y5(iTRAQ8plex)	40.34	45.60	49	66
INNQLFPYPY	N-Term(iTRAQ8plex); Y8(iTRAQ8plex)	40.67	44.85	6	42
QSLVYFPFGPI	N-Term(iTRAQ8plex); Y5(iTRAQ8plex)	40.84	46.75	48	75
IHPFAQTQSLVYFPFGPIPN	N-Term(iTRAQ8plex); Y12(iTRAQ8plex)	40.98	44.63	44	63
GAWYYVPL	N-Term(iTRAQ8plex); Y5(iTRAQ8plex)	41.13	47.12	2	13
AQTQSLVYFPFGPIPNSLPQ	N-Term(iTRAQ8plex); Y8(iTRAQ8plex)	41.39	46.37	10	39
SLVYFPFGPI	N-Term(iTRAQ8plex); Y4(iTRAQ8plex)	41.47	47.14	2	44
FYPELF	N-Term(iTRAQ8plex); Y2(iTRAQ8plex)	41.52	47.62	13	9
SLVYFPFGPIPNSLPQ	N-Term(iTRAQ8plex); Y4(iTRAQ8plex)	42.29	47.48	2	45
YFYPELF	N-Term(iTRAQ8plex); Y3(iTRAQ8plex)	42.85	49.81	10	30
AYFYPELF	N-Term(iTRAQ8plex); Y4(iTRAQ8plex)	42.92	49.13	16	32
DMPIQAFLLYQEPVLGPVR	N-Term(iTRAQ8plex); M2(Oxidation); Y10(iTRAQ8plex)	43.36	47.30	22	117
LYQGPIVLNPW	N-Term(iTRAQ8plex); Y2(iTRAQ8plex)	43.83	49.21	15	77
FLLYQEPVLGPV	N-Term(iTRAQ8plex); Y4(iTRAQ8plex)	43.92	50.56	43	257
LYQEPVLGPVRGPFPIIV	N-Term(iTRAQ8plex); Y2(iTRAQ8plex)	44.29	48.37	3	77
DMPIQAFLLYQEPVLGPVR	N-Term(iTRAQ8plex); Y10(iTRAQ8plex)	45.02	49.34	10	87
EPMIGVNQELAYFYPELFR	N-Term(iTRAQ8plex); M3(Oxidation); Y14(iTRAQ8plex)	45.24	49.22	41	57
LLYQEPVLGPVRGPFPIIV	N-Term(iTRAQ8plex); Y3(iTRAQ8plex)	45.99	50.63	8	79
AYPSGAWYYVPLGTQYTDAP SFSDIPNPIGSENSE	N-Term(iTRAQ8plex); Y16(iTRAQ8plex)	47.27	51.85	2	22
LGYLEQLL	N-Term(iTRAQ8plex); Y3(iTRAQ8plex)	47.57	54.27	35	81
AYPSGAWYYVPLGTQYTDAP SFSDIPNPIGSE	N-Term(iTRAQ8plex); Y16(iTRAQ8plex)	47.88	52.72	2	3
YLGYLEQLLR	N-Term(iTRAQ8plex); Y1(iTRAQ8plex)	49.49	51.06	4	74
CYQSYSTEMSITDCRETG	N-Term(iTRAQ8plex); C1(Carbamidomethyl); Y5(iTRAQ8plex); S6(Phospho); T7(Phospho); M8(Oxidation); S9(Phospho); C13(Carbamidomethyl)	50.91	52.10	11	8
<b>QYLYQGPIVLNPWDQV</b>	N-Term(iTRAQ8plex); Y2(iTRAQ8plex)	52.78	50.31	2	102

## 2 Developing a Platform for relative Quantification using iTRAQ

### 2.3.3.5 N-terminal Tyrosine as a Special Variant of Y-iTRAQ Modifications

In an *in-silico* trypsin digested human proteome (reference proteome; no isoforms; no miss cleavages; minimal mass: 400 Da; maximal mass: 10,000 Da; downloaded May 2014), approximately 3.5% of all peptides contained an N-terminal tyrosine. Although this is a rather small number, the fragmentation behavior and the generation of specific immonium ions of N-terminal iTRAQ labeled tyrosine peptides were analyzed. Furthermore, its N-Terminal phosphorylated form was analyzed accordingly.

The peptide YSLTVAVK was chosen as a model peptide, which was identified in three forms: (i) with iTRAQ labeling at the N-Terminus and the C-terminal lysine ( ${}_i\text{YSLTVAV}_i\text{K}$ ; unmodified), (ii) with an additional phosphorylation at the N-Terminus ( ${}_{i,p}\text{YSLTVAV}_i\text{K}$ ; phosphorylated) or (iii) with with an additional iTRAQ labeling at the N-terminus ( ${}_{i,i}\text{YSLTVAV}_i\text{K}$ ; Y-iTRAQ modified).

For the unmodified peptide (i), an almost complete b-ion series ( $b_1 - b_6$ ) was observed, although  $b_5$  and  $b_6$  did not pass the mass accuracy filter limited to 20 ppm (Figure 2-41 (A)). Additionally, the y-ions  $y_1 - y_4$  were detected, which proved the correct peptide identification. The Y-iTRAQ specific immonium ion at  $m/z$  280.17 described previously is identical in mass and structure to the  $a_1^+$ -ion for N-terminal tyrosine iTRAQ-4Plex labeled peptides (*unmodified*). As commonly observed for HCD fragmentation (189), the  $a_1^+$  ion was observed (relative intensity of 2.5%). Its corresponding  $b_1^+$ -ion was detected with a relative intensity of 12.8%, which was significantly higher compared to the  $a_1^+$ -ion. This was expected, as the frequency and abundances of a-ions are often lower than those of corresponding  $b_1^+$ -ions (296). This higher intensity of the  $b_1^+$ -ion helps to differentiate between an iY-iTRAQ immonium ion generated from peptides like  ${}_i\text{VGIN}_i\text{YWLAH}_i\text{K}$  (iY-iTRAQ immonium ion), and N-terminal peptides carrying one iTRAQ modification ( ${}_i\text{YSLTVAV}_i\text{K}$  with  $a_1^+$ - and  $b_1^+$ -ion).

Further, the phosphorylated form (ii) of this peptide was analyzed ( ${}_{i,p}\text{YSLTVAV}_i\text{K}$ , Figure 2-41 (B)). Again, the same b- and y-ions were detected and passing the filter settings of 20 ppm ( $b_1 - b_4$  &  $y_1 - y_4$ ). Interestingly, the corresponding a-ion series ( $a_1 - a_6$ ) was more intense (and passing the filter settings), indicating that the phosphorylation of tyrosine changes the fragmentation behavior of this peptide. Again, the  $a_1^+$ - and  $b_1^+$ -ions were detected with a relative intensity of 2.3% and 12.9%, respectively. The intensity was similar to the non phosphorylated peptide ( ${}_i\text{YSLTVAV}_i\text{K}$ ; unmodified).

A special modification represented the doubly iTRAQ-4Plex labeling of the N-terminal tyrosine within the peptide  ${}_{i,i}\text{YSLTVAV}_i\text{K}$  (Figure 2-41 (C)). Here, iTRAQ is once coupled via the primary amino group and once via the hydroxyl group to tyrosine. The b-ions ( $b_1 - b_6$ ) were detected, whereas the y-ion series was similar to the two other peptide forms ( $y_1 - y_5$ ). Remarkably, the corresponding a-ion series was also present ( $a_1 - a_5$ ). Here, the doubly iTRAQ-4Plex labeled  $a_1$ -ion was not detected in its singly charged form, but with a relative intensity of 12.8% in its doubly charged form ( $a_1^{2+}$ ). This higher intensity might be caused by the iTRAQ-4Plex labelings, which have a higher proton affinity (approximately 50 kJ/mol higher for tertiary amines than for primary amines (297)). A similar effect was observed for the remaining b-ions. However, the y-ion series was present in its singly charged form, which can be explained by the attachment of only one (C-terminal lysine) instead of two (N-terminal and hydroxyl-group of tyrosine) iTRAQ-4Plex labelings.



### 2.3 Y-iTRAQ – Accurate Identification of iTRAQ Tyrosine Labeled Peptides

For the Y-iTRAQ modified peptide (iii), the relative intensity of the  $a_1^{2+}$ -ion was higher (12.8%) compared to the corresponding  $b_1^+$  and  $b_1^{2+}$  ion (2.5% and 7.2%). Surprisingly, the iY-iTRAQ immonium ion at  $m/z$  280.17 was also detected with a relative intensity of 7%. At this point, it cannot be confirmed whether this ion was caused by N-terminal or side chain dissociation of the iTRAQ-4Plex group or a combination of both.

However, the three peptide modification forms (i – iii) and their fragmentation behavior was further analyzed under elevated NCE conditions (NCE = 75%) to verify the presence of the iY-, pY- and iY-iTRAQ immonium ions (Figure 2-42). Unexpected, the pY- and iY-iTRAQ-4Plex immonium ions represented the most intense sequence specific signals in spectra; only the non-sequence informative iTRAQ reporter ions were of higher relative intensity. This supported the assumption that the N-terminal iTRAQ dissociated further rather than the side chain iTRAQ group of tyrosine.

## 2 Developing a Platform for relative Quantification using iTRAQ

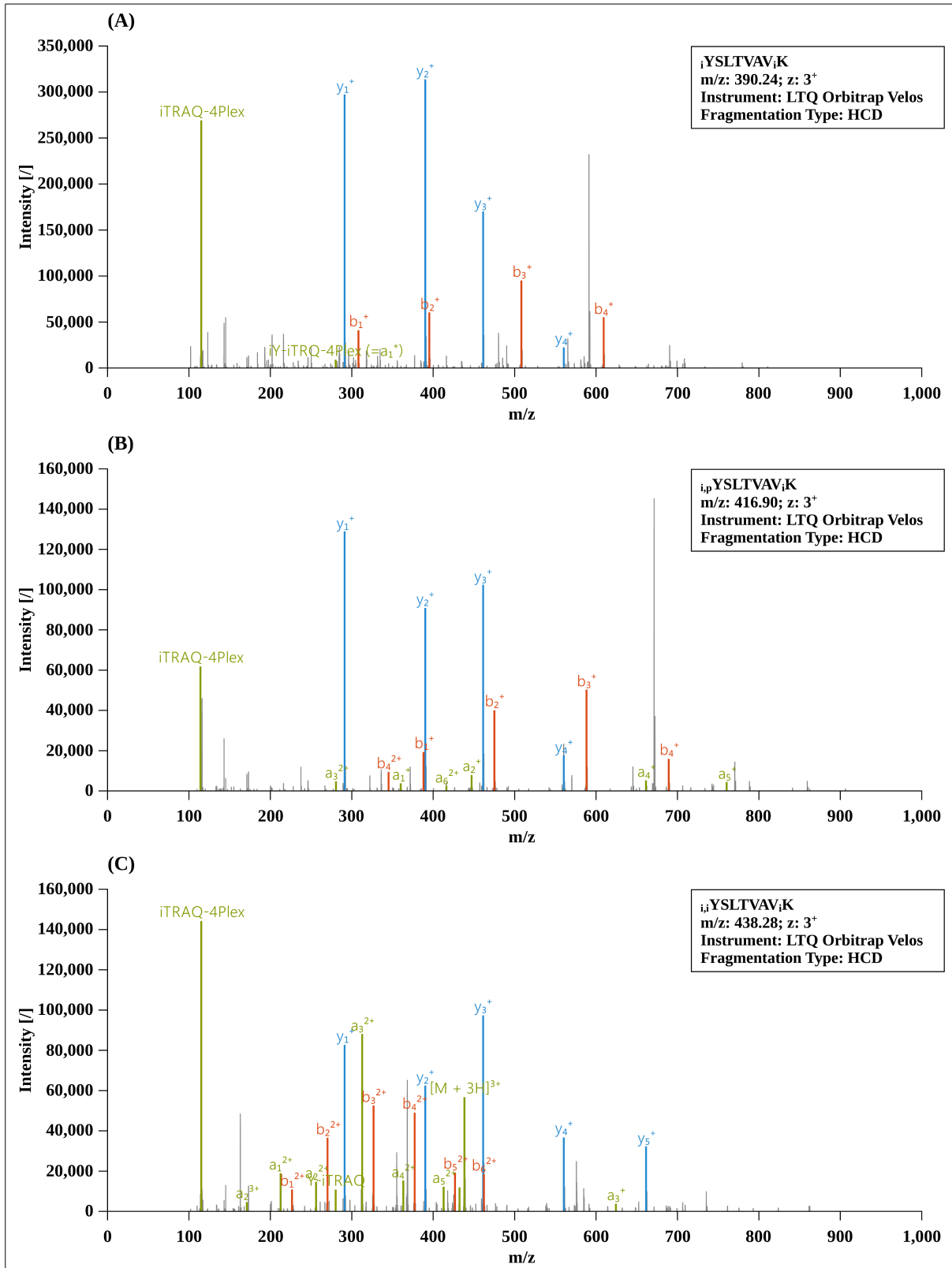


Figure 2-41: The iTRAQ-4Plex labeled peptide YSLTVAK was identified to be present in three forms:  $i$ YSLTVAV;K (A), with an additional phosphorylation at the N-terminus  $i_p$ YSLTVAV;K (B) and in a N-terminal doubly iTRAQ labeled form  $i_i$ YSLTVAV;K (C). For  $i$ YSLTVAV;K (A), the Y-iTRAQ specific immonium ion at m/z 280.17 is equaling the  $a_1^+$  ion. After phosphorylation (B), the pY-iTRAQ immonium ion (=  $a_1^+$  ion) was detected

### 2.3 Y-iTRAQ – Accurate Identification of iTRAQ Tyrosine Labeled Peptides

at  $m/z$  360.14. Replacing the phospho-moiety with an iTRAQ group, the Y-2x(iTRAQ) immonium ion was detected at  $m/z$  424.27. All MS/MS spectra were acquired after HCD fragmentation in the Orbitrap mass analyzer. For the latter peptide ( $i_i$ YSLTVAV $_i$ K), the additional iTRAQ modification had an influence on the b- and a- ion series, which were predominantly present in the doubly charged form.

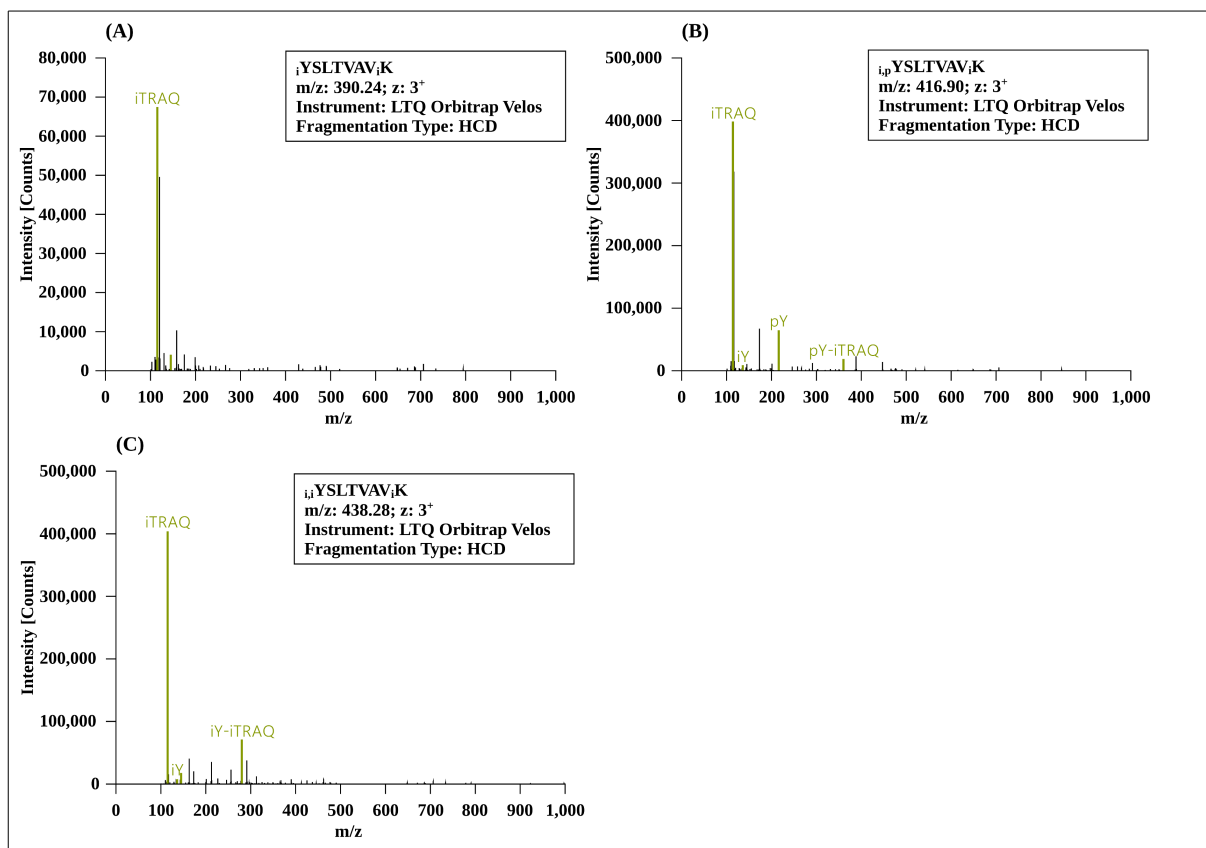


Figure 2-42: Peptide fragmentation of the iTRAQ-4Plex labeled peptide YSLTVAVK was shown before (Figure 2-41) for the three observed modification forms:  $i_i$ YSLTVAV $_i$ K (A), with an additional phosphorylation at the N-terminus  $i_p$ YSLTVAV $_i$ K (B) and in a N-terminal doubly iTRAQ labeled form  $i_i$ YSLTVAV $_i$ K (C). To demonstrate the impact/presence of the observed specific immonium ions, the MS/MS spectra after HCD fragmentation at NCE = 75% are shown. For  $i_i$ YSLTVAVK, the Y-iTRAQ immonium ion was not detectable. For the phosphorylated form, the tyrosine- (iY,  $m/z$  136.07), the phosphotyrosine- (pY,  $m/z$  216.04) and the phosphotyrosine-iTRAQ immonium ion (pY-iTRAQ,  $m/z$  360.14) were detectable. For the doubly iTRAQ labeled N-terminal tyrosine ( $i_i$ YSLTVAV $_i$ K), the tyrosine- (iY,  $m/z$  136.07) and Y-iTRAQ immonium (Y-iTRAQ,  $m/z$  280.17) were present. The doubly iTRAQ labeled immonium ion (Y-2x(iTRAQ),  $m/z$  424.27) was not detectable.

## 2 Developing a Platform for relative Quantification using iTRAQ

### 2.3.4 Conclusion

Only 5 - 20% of low resolution ion trap and 15 - 70% quadrupole TOF tandem MS spectra are typically identified and validated peptide hits (287–290). The causes are various: (i) only 6-10 modifications are usually considered to limit the search space during database search, which in combination to (ii) unexpected cleavage sites, (iii) modifications or (iv) amino acid substitutions causes most of the spectra to be unassigned (275).

For iTRAQ labeling experiments, an unwanted and often neglected modification represents the binding of iTRAQ to the hydroxyl-group of tyrosine. However, as this modification is known to occur (60,291), it is mandatory to include this dynamic modification in computational proteomics workflows to identify and validate Peptide-Spectrum-Matches by database search algorithms.

To further validate this modification manually or by computational proteomics pipelines, a diagnostic iY-iTRAQ ions can assist this process. For both iTRAQ-4Plex and iTRAQ-8Plex, specific signals at  $m/z$  280.17 and 440.28, respectively, could be unambiguously presented. These immonium ions were (potentially) generated by double side cleavage (C- and N-terminal) which is common for immonium ions. Although a variety of different fragmentation techniques were applied, these ions were only detected after beam-type CID fragmentation (e.g., HCD on the Orbitrap Velos). Its relative intensity in MS/MS spectra was demonstrated to be a function of the applied collision energy, and exhibits a maximal intensity at (slightly) elevated collision energies. It is noteworthy that this ion was also detected at higher NCE values, increasing its diagnostic ion significance, as other sequence specific fragment ions were depleted at these energy regimes. Further, the Y-iTRAQ modification of tyrosine changed the retention time compared to their unmodified counterparts upon peptide separation by reversed phase chromatography.

Summarizing, Y-iTRAQ features (iY-iTRAQ immonium ion in combination with the shorter retention time by reversed phase chromatography) will help to increase the number of identifiable spectra in quantitative proteomics approaches. It is important to include all sequence specific signals (so-called tags) of tandem MS spectra, as this information is useful in validating peptide identifications. This is especially true for the amino acid tyrosine (and other amino acids such as phenylalanine, valine, isoleucine, leucine, tryptophan, proline and histidine) which was demonstrated previously to be a good predictor of peptide composition (298). These ions cannot be detected in ion trap devices, however, the incorporation of immonium ions into search engines could improve the application of beam-type CID fragmentation, as a higher number of signals of MS/MS spectra are used which will increase the homology between the experimental and theoretically calculated spectrum.

## 2.4 MECi – Targeted Phosphoprotein Quantification

### 2.4.1 Introduction

For quantitative phosphoproteomics studies using iTRAQ labeling, optimal MS instrumentation conditions for adequate fragment ion (identification) and reporter ion generation (quantification) have been established (chapter 2.1). In addition, the effect of non accurate *in-silico* precursor mass ( $m/z$  value) calculation for highly differentially abundant peptides could be compensated by an *in-house* developed algorithm. To further increase the number of Peptide-Spectrum-Matches after iTRAQ labeling, a diagnostic ion for iTRAQ-4Plex and iTRAQ-8Plex labeled tyrosine peptides has been identified, which can be used as additional features to validate peptide identifications.

The complement the platform to relatively quantify phosphopeptides by isobaric labeling, the establishment of a new procedure to improve the identification and relative quantification of phosphoproteins by increasing the number of quantifiable (phospho)peptides is the aim of this section.

An important aspect of phosphoprotein analysis is to localize the phosphorylation site exactly, which is despite improvements (e.g., Phospho RS (299), Mascot MD-Score (276), MaxQuant PTM score (300), NCE optimization (chapter 2.1)) still challenging (301). To enlarge the number of phosphorylated peptides to be selected for fragmentation and subsequent peptide identification (or to localize the phosphorylation site), the application of a multi-protease approach has been shown to be superior to the single protease digestion (302) either in parallel or subsequently use (303–307).

Several workflows have been published to quantify changes of phosphorylated protein abundances using isobaric labeling strategies (277,308,309). However, to increase the number of phosphorylated peptides that can be used for identification and determination of the exact phosphorylation site, a novel multi-protease digestion procedure (two specific and two unspecific proteases) in combination to iTRAQ labeling will be established. This workflow aims to increase the number of quantifiable (phospho)peptides to accurately relatively quantify the difference in phosphoproteins. An important aspect of this workflow is to ensure the compatibility of isobaric labeling with SDS-PAGE and phosphopeptide enrichment strategies using  $\text{TiO}_2$ .

To demonstrate the influence and the advantage of a multi-protease digest method with subsequent iTRAQ labeling, various identification features (such as the sequence coverage) will be compared to the results of single protease digestion. Further, the compatibility of the multi-protease approach with relative phosphorylation site quantification will be shown.

## 2 Developing a Platform for relative Quantification using iTRAQ

### 2.4.2 Material and Methods

The complete workflow of the newly developed approach covering all the steps except MS analysis is summarized schematically in Figure 2-43.

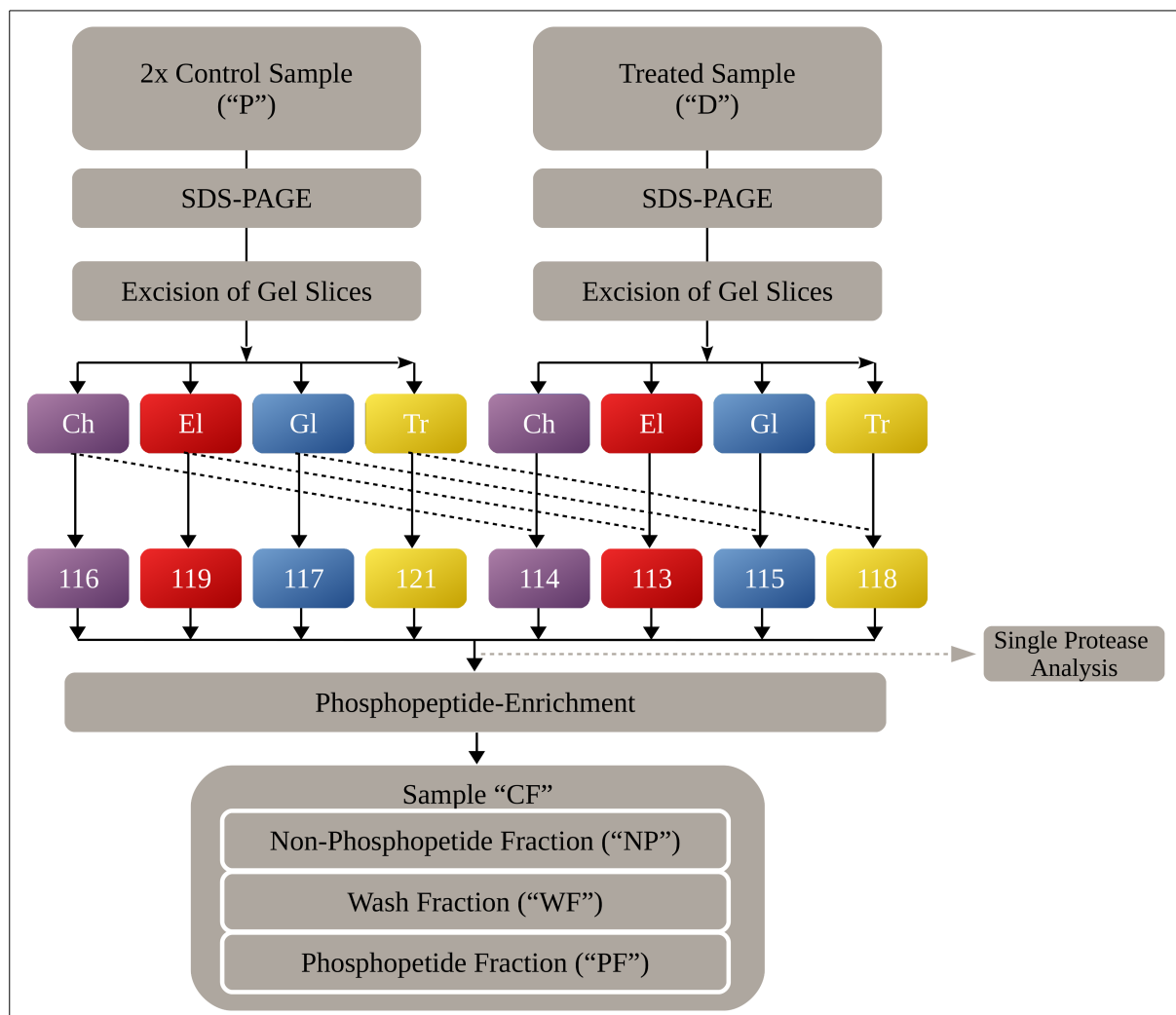


Figure 2-43: Schematic diagram of the multi-protease workflow for relative quantification using iTRAQ-8Plex. The control sample ("P" phosphorylated) was directly used for iTRAQ labeling, whereas the treated sample ("D" dephosphorylated) was a 1:1 mixture of the phosphorylated and dephosphorylated sample. This simulated two biologically differently phosphorylated samples. Afterwards, both samples are digested separately using four enzymes (chymotrypsin, elastase, *Staphylococcus aureus* V-8 Protease (GluC) and trypsin). Each digest is labeled with a unique iTRAQ-8Plex channel. The pooled sample is then enriched for phosphopeptides in a batch process using  $\text{TiO}_2$  (ratio 1:20). Samples labeled with gray dashes ("Single Protease Analysis") are only used for method development and not part of the workflow for analysis. All samples were subsequently analyzed by LC-ESI MS.

### 2.4.2.1 Chemicals

Chemicals used in this study are summarized in chapter 4.1.

### 2.4.2.2 Sample Preparation and Analytical Workflow

#### Preparation of SDS-Gel

The SDS-gels were prepared using standard techniques (310). Briefly, the resolving gel (12%) was prepared by adding 3.4 mL ddH<sub>2</sub>O, 4ml 30% acrylamind/bisacrylamid, 2.5 mL gel buffer (1.5 M Tris-HCl, pH 8.8) and 0.1 mL 10% SDS. Immediately prior to casting, 50  $\mu$ L 10% APS (ammonium persulfate) and 5  $\mu$ L TEMED were added and swirled gently to initiate polymerization. After 2h of polymerization, the stacking gel (8%) was prepared accordingly: 4.7% ddH<sub>2</sub>O, 2.7 ml 30% acrylamind/bisacrylamid, 2.5 mL gel buffer (0.5 M Tris-HCl, pH 6.8) and 0.1 mL 10% SDS. Prior to pouring, 50  $\mu$ L 10% APS and 10  $\mu$ L TEMED were added. To ensure complete polymerization, the gels were left overnight at 4°C, prior to sample loading.

#### Preparation of Casein-Sample

To simulate a control and treated sample, the proteins  $\alpha$ -S1-,  $\alpha$ -S2- and  $\beta$ -casein were treated with alkaline phosphatase (AP). Ten  $\mu$ L of  $\beta$ -casein and 10  $\mu$ L of  $\alpha$ -casein solution (c = 10  $\mu$ g/ $\mu$ L in H<sub>2</sub>O) were combined. The sample was buffered with 40  $\mu$ L of AP-buffer (diluted 1:10 with ddH<sub>2</sub>O; AP-buffer was supplied by Roche Diagnostics GmbH, Mannheim, Germany) and split into two aliquots: 6  $\mu$ L of the AP solution (containing the alkaline phosphatase) was added to one sample, referred to as <sup>1</sup>D(ephosphorylated), while 6  $\mu$ L of ddH<sub>2</sub>O was added to the control sample or <sup>1</sup>P(hosphorylated). Both samples were incubated at 30°C overnight. Samples were dried using vacuum centrifugation and reconstituted in 50 $\mu$ L TEAB solution (50 mM) and 50  $\mu$ L sample buffer (0.5 M Tris-HCL, 1mM EDTA, pH 8.5). Samples were denatured at 85°C for 5 min and 5  $\mu$ L loaded onto each SDS gel band.

#### SDS-Gel Separation

SDS-Gels were run using 80 Volts for 30 min, for sample concentration, followed by an increase to 110 Volts for protein separation. SDS-gels were fixed for 2 h (50% methanol, 2% phosphoric acid (PA)) and stained for 10 min (34% methanol, 2% PA, 17% ammonium sulphate (AS), 0.066% Coomassie Brilliant Blue G-250). SDS-gel were destained overnight with 5% ethanol and 10% acetic acid.

#### Gel-Slice Exaction and Protein Digestion

Gel bands were excised and then simultaneously destained and washed (3 times) with a 30% ACN solution (100  $\mu$ L) followed by an additional wash with 50% ACN (100 $\mu$ L). Prior disulfide-bond reducing, samples were dried under vacuum centrifugation for 30 min at 30°C. Reduction was performed using 80  $\mu$ L TCEP (6mg/10mL) at 50°C for 30 min. Samples were stored at 4°C for 1-2 h prior to alkylation. Alkylation was performed for 25 min in dark at room temperature using IAA (50  $\mu$ L IAA solution with 2.5 mg IAA/1mL) and 50  $\mu$ L of TEAB buffer (50 mM). Residual IAA was quenched using a solution of 50  $\mu$ L TCEP and 200  $\mu$ L TEAB

## 2 Developing a Platform for relative Quantification using iTRAQ

buffer for 15 min (room temperature), followed by an additional washing step with 200  $\mu$ L TEAB buffer (15 min, room temperature). Gel pieces were shrunk for 15 min at room temperature (100  $\mu$ L; 50% ACN, 50% TEAB) and dried under vacuum centrifugation for 30 min.

For protein digestion, chymotrypsin, elastase, Staphylococcus aureus V-8 Protease (*GluC*) and trypsin were used. With the exception of trypsin, a protease:protein ratio of 1:10 (w/w) was applied. For trypsin a ratio of 1:20 was used. Each protease stock solution was diluted with TEAB buffer and 20  $\mu$ L of protease solution was added to the dehydrated gel pieces. Gels were allowed to swell for 15 min at room temperature. Afterwards, 100  $\mu$ L of TEAB buffer was added and the samples were left to digest overnight at 37°C.

For extraction of peptides, 50  $\mu$ L of TEAB added. After 30 min at room temperature, the supernatant was transferred to a new Eppendorf tube and 200  $\mu$ L extraction buffer (60% v/v ACN, 0.5% v/v TFA) was added to the gel pieces. Sample extraction was performed for 30 min, followed by a second extraction with 100% ACN (30 min). The pooled samples were dried under vacuum centrifugation and resuspended in 16  $\mu$ L of loading buffer A (3% ACN, 0.1% TFA). One  $\mu$ L of this sample was removed and used for MS analysis to ensure complete digestion (*not* included in Figure 2-43). To simulate differentially regulated phosphorylation degrees, the samples originated from control were used without any further processing for iTRAQ labeling. The dephosphorylated sample instead, was a 1:1 combination of both, the control and treated sample, as outlined in Figure 2-43 by gray dotted line.

### **iTRAQ Labeling**

Samples were dried under vacuum centrifugation and re-hydrated using 20  $\mu$ L of iTRAQ-8Plex labeling buffer (60% isopropanol, 40% iTRAQ-Dissolution-Buffer). iTRAQ reporter ion solutions were prepared according to the manufacture's instructions. For labeling, samples were incubated with a 2 fold molar excess of iTRAQ reagent (10  $\mu$ L) for 2 h in the dark at room temperature. The labeling schema is outlined in Table 2-8 and Figure 2-43. To ensure complete labeling and to hydrolyze unreacted iTRAQ reagents, samples were stored at -20°C prior to vacuum centrifugation. Samples were rehydrated with 15  $\mu$ L of loading buffer A and 1  $\mu$ L was used for iTRAQ labeling screening.

*Table 2-8: Each of the two samples (control (P) and treated sample (D)) were digested with one of the four proteases applied and labeled with one of the eight iTRAQ-8Plex reagents, respectively. By shuffling of the iTRAQ reporter ion channels, samples can also be analyzed by HCD (as performed in this study) and ETD, which gives complementary fragment ions (c- and z-ions).*

<b>iTRAQ Reporter Ion</b>	<b>Enzym</b>	<b>Sample</b>	<b>HCD Reporter Ion</b>	<b>ETD Reporter Ion</b>
113	elastase (El)	D	113	101
114	chemotrypsin (Ch)	D	114	101
115	GluC (Gl)	D	115	102
116	chemotrypsin (Ch)	P	116	104
117	GluC (Gl)	P	117	104
118	trypsin (Tr)	D	118	106
119	elastase (El)	P	119	106
121	trypsin (Tr)	P	121	108



### **Phosphopeptide Enrichment with TiO<sub>2</sub> resin**

Phosphopeptide enrichment was performed as described previously (311). Briefly, TiO<sub>2</sub> was resuspended in loading buffer I (80% ACN, 5% TFA, 1M glycolic acid (GA)) to a TiO<sub>2</sub> concentration of 1 – 10 g/L. After vortexing, TiO<sub>2</sub>, equivalent to 20 times the weight of starting protein material, was transferred to a new Eppendorf tube and the volume adjusted to 100 µL using loading buffer I. After 15 min of shaking, TiO<sub>2</sub> was pelleted by centrifugation and the supernatant discarded.

Samples were re-suspended in loading buffer I (20 µL) and transferred to the TiO<sub>2</sub> beads, vortexed and shaken at high rpm (1300) for 15 min. TiO<sub>2</sub> was pelleted and the supernatant removed (fraction: non-phosphopeptide). Three washing steps (100 µL each) with washing buffer I (80% ACN, 1% TFA), washing buffer II (20% ACN, 0.2% TFA) and water were performed for 15 min and the supernatant from all wash fractions combined (fraction: wash). Prior to phosphopeptide elution, the TiO<sub>2</sub> beads were dried under vacuum centrifugation (30 min, room temperature) and eluted using 100 µL of elution buffer (1% NH<sub>4</sub>OH, approx. pH 10). After 15 min of incubation, the sample was centrifuged and the supernatant transferred to a new Eppendorf tube (fraction: phosphopeptide).

### **2.4.2.3 Mass Spectrometry**

#### **LC-ESI MS**

The same LTQ Orbitrap Velos / HPLC setup as described before (chapter 2.1) was utilized for HPLC-ESI MS experiments. Generally, 9 µL of sample (12 µL for LC-ESI MS experiments with imported precursor exclusion lists) was injected and desalted for 4 min (30 µL/min), prior to peptide separation (300 nL/min) using a linear increasing gradient of solvent B from 5% to 55% within 51 min. A column washing step was added by increasing the solvent B to 95% within 5 min and kept constant for an additional 5 min. Afterwards, the column was re-equilibrated for 9 min using 5% solvent B. Between each sample injection, a wash run with the same method was applied.

XCalibur was used to write a tandem MS/MS method utilizing both, CID-IT and HCD fragmentation (see Figure 1-12). The MS method was split into MS full scan (Resolution 60,000; m/z 300 – 2000) with prescan activated. In parallel, the 5 most intense precursor ions were selected for CID-IT fragmentation (35% NCE) and recorded in ion trap. For HCD, the same precursor ions were selected and fragmented with an NCE of 45%. Fragment ions were recorded with a resolution of 7,500 in Orbitrap mass analyzer (m/z selection window 3; first m/z value: 100). Only precursors with a charge state > 1<sup>+</sup> and an intensity > 500 counts were selected for fragmentation. Repeat count was set to 3 within a repeat duration of 20 sec. After fragmentation, each precursor was dynamically excluded for 30 sec.

The exclusion lists were generated after performing the database searches (see below) of the 3 phosphopeptide enrichment runs (non- phosphopeptide-, wash-, and phosphopeptide-fraction) with all PSMs passing the set filters (high confidence, 1<sup>st</sup> Rank only). Export options used: (i) export to m/z values; (ii) retention time windows width (min): 5; (iii) lower retention time window (min): 5; upper retention time windows (min): 75; (iv) mass precision (decimals) 5; (v) max concurrent entries: maximum.

## 2 Developing a Platform for relative Quantification using iTRAQ

### Database Searches

For data analysis, Proteome Discoverer 1.4 with the search algorithm SEQUEST HT (supplied with PD 1.4.0.288) was used. The workflow was built up as outlined in Figure 2-1 with default settings unless otherwise specified: (0) Spectrum Files; (1) Spectrum Selector; (2) MS2-Spectrum Processor; (3) & (7) Scan Event Filter to split into CID and HCD fragmentation; (4) & (8) SEQUEST HT (see below); (5) phosphoRS 3.0; (9) Reporter Ions Quantifier: iTRAQ-8Plex (Thermo Scientific Instruments; Integration Tolerance 20 ppm; Most Confident Centroid as Integration Method; Additional Settings: Show the Raw Quan Values; Apply Value Corrections, use All Peptides enabled and deactivated Experimental Bias); (10) Event Detector; (11) Precursor Ions Area Detector and (12) Percolator or Target Decoy PSM Validator.

For SEQUEST HT database searches, two different FASTA files were used: (sF) with 403 entries (10 protein-mix, 180 phosphopeptide sequences (Intavis standard) and keratin related proteins (see chapter 2.1.2.3)) and (bF) containing the proteins within the sF-FASTA file and in addition, the reference proteome of *E. coli*, resulting in a total number of 4526 proteins. *No-enzyme* specificity was used with a minimum and maximum peptide length of 4 and 144, respectively. Precursor mass tolerance was set to 10 ppm, 0.3 Da for CID and 0.02 Da for HCD fragmentation. The maximum number of dynamic and equal modifications allowed per peptide were 12 and 8, respectively. Phosphorylation (S, T & Y) and oxidation (M) were set as variable modifications. Static modifications included iTRAQ-8Plex (peptide N-terminus and K), and carbamidomethylation (C). Results were filtered with medium confidence (5% FDR) and high confidence (1% FDR) for identification and quantification, respectively.

For evaluation of the computational workflow, four different settings were tested: Target Decoy PSM Validation (supplied with Proteome Discoverer 1.4) with sF- and bF-FASTA and Percolator (supplied with Proteome Discoverer 1.4 ) with sF- and bF-FASTA file, respectively. For final data analysis, the workflow containing Percolator and bF-FASTA file was chosen (referred to as reference workflow).

### Data-Analysis

#### *Optimization of the Computational Proteomics Workflow for No-Enzyme Database Searches*

Grouped and ungrouped peptides (5% FDR) for the four different enzymes (see section above) were exported into different .xls files. Information regarding protein sequence coverage [%], number of unique peptides and Peptide-Spectrum-Matches (PSMs) were extracted from protein tables without any additional post-processing (Figure 2-44). For the comparison of PSMs identified by CID and HCD (Figure 2-45), for both phosphorylated and non-phosphorylated peptides, *in-house* developed VB(A) scripts were used to split the data. Three biological replicates were used to outbalance variations within samples.

$$Ratio_{HCD, PSM} = \frac{HCD}{(CID + HCD)} \quad \text{Equation 2-9}$$

### *Identification: Determining the Benefit of the Multi-Protease Approach*

For each of the iTRAQ labeled protease digests (chymotrypsin (Ch), elastase (El), GluC (Gl), trypsin (Tr)) the sequence coverage [%], number of unique peptides and the Peptide-Spectrum-Matches (PSMs) were exported to .xls files (see section above) and compared (Figure 2-46).

The *in-silico* combined samples from the *Single Protease Analysis* (Figure 2-43) were referred to as CETG. Technical replicates from the same sample were analyzed in a MudPIT approach.

Following TiO<sub>2</sub> enrichment, two technical replicates and an exclusion list LC-ESI MS run of the phosphopeptide-, wash- and non-phosphopeptide fraction were searched in a MudPIT approach (referred to as CF). All database searches used the reference workflow.

A heat map algorithm using RGB style for visualization of amino acid occurrence was programmed (blue for low (0%), green (50%) and red for high values (100%)) and used to visualize the different frequency of amino acids in the experiments. For calculation of the corresponding RGB color, the lowest and highest number of amino acid occurrence was searched within all input files (Figure 2-47 to 2-49). The function to calculate the RGB color is shown in chapter 4.1.

### *Phosphopeptide Identification: Application of Multiple Proteases to Increase the Number of Unique Phosphopeptides*

The exported peptide group lists for identification of the sample CF (.xlsx files, see above) were used to compare the four different proteases in terms of their number of unique phosphopeptide sequences identified. To further classify the peptides according to one of the proteases used for digestion, the C-terminal amino acid was used. Chymotrypsin and GluC share two cleavage sites (D, E). Elastase and chymotrypsin share leucine cleavage specificity. These were classified as Ch-Gl and Ch-El, respectively (Figure 2-51). The total number of C-terminal amino acids not being derived by one of the protease were grouped under the term *unspecific* (un). Histidine (un-H) and tyrosine (un-T) were highly present within the group *unspecific C-terminal amino acids* and shown additionally (Figure 2-51).

To compare the number of identified unique phosphopeptides (A, C, E) and the number of Peptide-Spectrum-Matches of phosphorylated peptides (B, D, F) for the four singly used proteases, their *in-silico* combined sample (CETG) and the enriched fraction CF were plotted (Figure 2-50).

### *iTRAQ Quantification*

For the quantitative phosphopeptide analysis, only PSMs with high confidence were used to exclude potentially wrong annotations. Due to the *no-enzyme* database search, every possible peptide sequence of the supplied FASTA file (falling within the specified precursor mass tolerance window) can be matched to the fragmented precursor and be used for subsequent fragment ion matching and scoring. Although peptides used for analysis needed to pass FDR < 1%, only peptides having the correct C-terminus (e.g. R or K for trypsin) and high intensive reporter ion channels (118 (sample D) and 121 (sample P) for trypsin) in comparison to the residual iTRAQ reporter ion channels were used (minimum value  $iS > 0.4$ ;  $iS$  calculation see Equation 2-10).

## 2 Developing a Platform for relative Quantification using iTRAQ

$$iS = \frac{i\text{TRAQ Reporter Ion } x_1 + i\text{TRAQ Reporter Ion } x_2}{\sum i\text{TRAQ Reporter Ion}} \quad \text{Equation 2-10}$$

Afterwards, based on the sequence and its singly charged precursor mass (mass deviation < 30 ppm), the median and standard deviation of the iTRAQ *w*-Ratio (see Equation 2-11) of the reporter ions were calculated.

$$w\text{-Ratio} = \frac{i\text{TRAQ Reporter Ion } x_1}{i\text{TRAQ Reporter Ion } x_1 + i\text{TRAQ Reporter Ion } x_2} \quad \text{Equation 2-11}$$

### 2.4.3 Results

The aim of this study was to create a workflow using a multi-protease approach in combination with iTRAQ-8Plex labeling to comprehensively identify, assign and quantify protein phosphorylation between two biological entities. The workflow was designed to be fully compatible with SDS-PAGE, iTRAQ labeling and phosphopeptide enrichment. For digestion, two unspecific (chymotrypsin, elastase) and two highly specific (trypsin, GluC) proteases were chosen to generate a variety of unique and overlapping peptide sequences. An *in-silico* digestion (mmass, Version 5.5.0) of  $\alpha$ -S1-,  $\alpha$ -S2-,  $\beta$ - and  $\kappa$ -casein was performed using elastase (V, I, A, L, S, G), chymotrypsin (Y, W, F, L), GluC (only E, TEAB buffer (312)) and trypsin (R, K) with no missed cleavages allowed. The minimal peptide length was four amino acids, other peptides were filtered out. Average peptide lengths of 6, 12, 14 and 16 amino acids were obtained for elastase, chymotrypsin, GluC and trypsin, respectively. This indicated a good distribution of unique peptide sequence lengths for the four proteases used.

#### 2.4.3.1 Percolator and FDR Analysis to Identify iTRAQ-8Plex Labeled Peptides

The quantitative multi-protease approach with iTRAQ-8Plex labeling for targeted phosphopeptide analysis requires either (i) a database search without enzyme specificity or (ii) a custom digested protein FASTA file combined at the peptide level with preselected missed cleavage specifications. For the latter, identified peptides have to be mapped back to their protein origin either manually or with additional scripting tools. In contrast, the unspecific (*no-enzyme*) database search can be implemented in Proteome Discoverer directly. For *no-enzyme* searches, the entire protein is indexed which increases the number of potential peptide candidates used for the database search compared to a targeted approach. In targeted approaches, only peptides with predefined C-terminal residues are used (e.g., R and K for trypsin). Therefore, a higher number of peptide decoy sequences is available for *no-enzyme* database searches and can be used as for Peptide-Spectrum-Match analysis and validation. Thus, (phospho)peptides passing the FDR approach were of higher quality compared to those from a smaller decoy database. For the aforementioned reasons an unspecific database search was used for peptide identification applying the multi-protease approach.

An important aspect of this workflow was to analyze the differences between two FASTA files: (i) a FASTA file containing 403 protein entries (10 protein-mix plus impurities listed in chapter 4.1) and (ii) the same FASTA with the addition of an *E. coli* reference proteome (see chapter 2.1.2.3 for details) resulting in a total of 4,526 protein sequences. In addition, two peptide validation strategies commonly used for peptide identification were evaluated: the FDR approach (Peptide Validator) was compared to the machine learning code Percolator.

Independent of the validation method and FASTA file utilized during database search, an almost complete sequence coverage was observed for the three target proteins:  $\alpha$ -S1-,  $\alpha$ -S2 and  $\beta$ -casein (Figure 2-44 (A)). Additionally,  $\kappa$ -casein was also identified. However, this protein is a known contamination for  $\alpha$ - and  $\beta$ -casein protein standards. For both FASTA files an average increase of ~10% in unique peptides was observed for all three biological replicates if the Peptide Validator (FDR approach) is replaced by the semi-supervised machine learning code Percolator (Figure 2-44 (B)). This effect was slightly more pronounced for the bF-FASTA file. This was expected due to the higher number of potential peptide sequences falling within the  $\pm 10$  ppm precursor

## 2 Developing a Platform for relative Quantification using iTRAQ

mass tolerance window set in Proteome Discoverer. About 66% of the unique peptides identified (same sequence & monoisotopic precursor charge state  $[M+H]^+$ ) were identified by both, the FDR approach and Percolator (bF-FASTA file). Approximately 10% and 23% peptides were uniquely identified with the FDR or the Percolator approach, respectively. This is a huge difference considering, that the only change in the workflow was the peptide validation method. The benefit of using Percolator was even more visible at the PSMs level (Figure 2-44 (C)). An increase of 41% and 55% for the number of PSMs was observed for the small FASTA (sF) and big FASTA file (bF), respectively. For all protein features monitored (coverage, unique peptides and PSMs), the standard deviation for  $\kappa$ -casein was highest, followed by  $\alpha$ -S2-casein. The differences were introduced while sample generation (e.g., excision of the SDS-gel), as the protein area differed significantly for  $\alpha$ -S1- and  $\kappa$ -casein but were matchable for  $\alpha$ -S1- and  $\beta$ -casein (Figure 2-44 (D)).

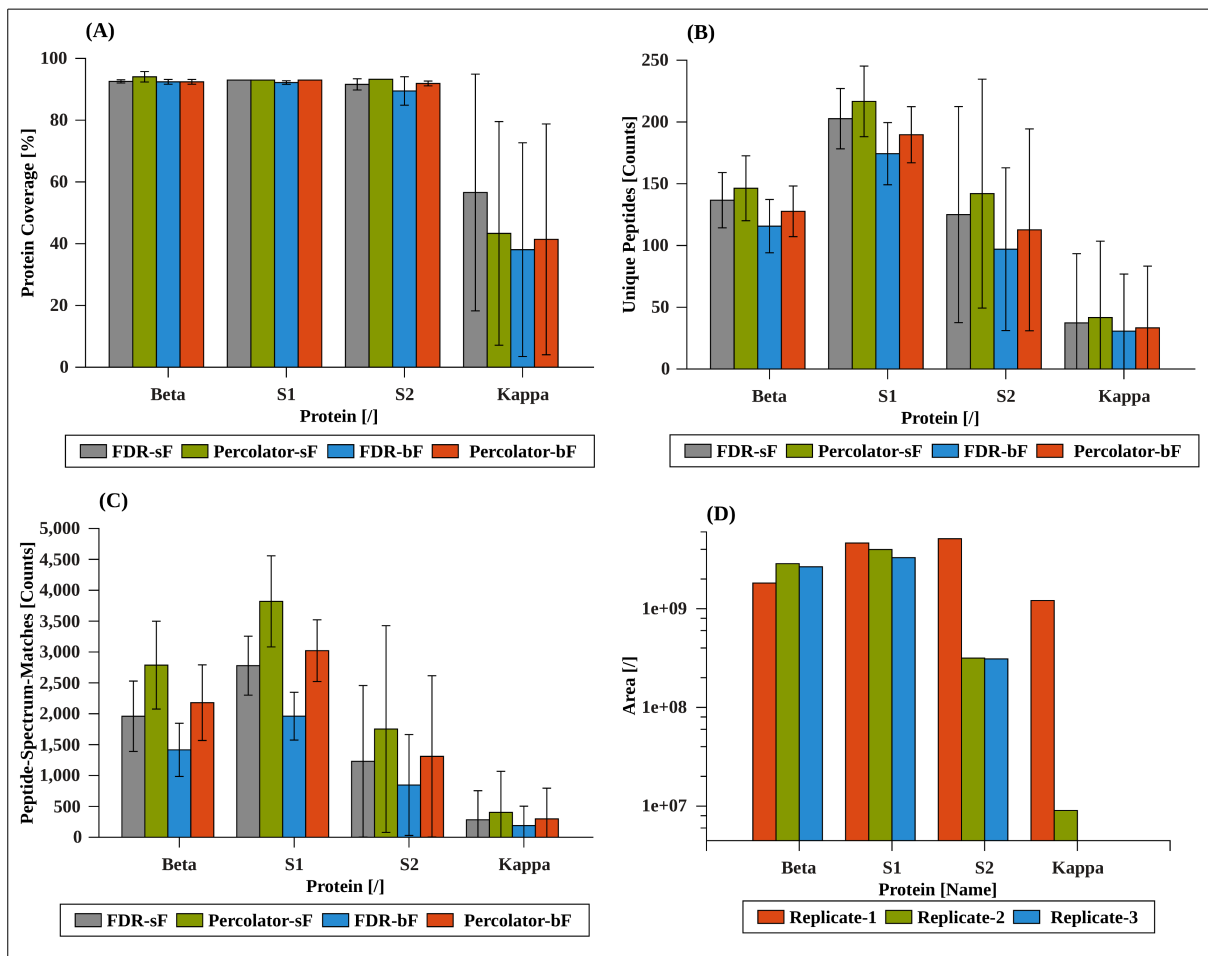


Figure 2-44: For the two workflows using a small FASTA file (sF) and a 10x bigger FASTA file (bF), respectively, the protein sequence coverage with a classical FDR approach and the improvement after percolating the results are demonstrated for the protein coverage (A), the number of unique peptides (B), the PSMs (C) and the protein area (D). For data generation the casein sample after iTRAQ-8Plex labeling and enrichment for phosphorylated peptides was used (sample CF, see workflow). These cover the non-phosphorylated- (NP), wash- (WF) and phosphorylated fraction (PF) and were searched in MudPIT approach.

## 2.4 MECi – Targeted Phosphoprotein Quantification

For data acquisition, a dual CID/HCD fragmentation method was applied to generate a dataset of CID-IT (resonance) and beam-type CID spectra. As different energy regimes are applied and the time of energy transfer to peptide species differs between the two fragmentation types, the spectra shape can be significantly different (283,313), as it was shown before for pS-peptides (Figure 2-2 to Figure 2-4).

The Peptide-Spectrum-Matches were further split into CID-IT and HCD fragmentation and plotted (Figure 2-45 (A)) for the FDR and Percolator approach. For HCD spectra, the increase in identifications was ~20% and ~26% for the sF- and bF-FASTA file, respectively, after percolating (compared to FDR). In contrast, the number of identification increased significantly by 96% and 169% (sF- and bF-FASTA file, respectively) after percolating the CID-IT data.

The significant increase in identifications for CID-IT after percolating was further compared to HCD. Generally, HCD identified more peptides, independently on the applied peptide validation method. However, with the classical FDR approach, the  $\text{Ratio}_{\text{HCD,PSM}}$  was 73% and 80% (sF- and bF-FASTA file), demonstrating the clear benefit of beam-type fragmentation followed by analysis in the high resolution Orbitrap mass analyzer. After percolation, the ratio dropped for both, the sF- and bF-FASTA file ( $\text{Ratio}_{\text{HCD,PSM}} = 62\%$  and  $65\%$ ).

The dataset was further split into phosphorylated and non-phosphorylated peptides to determine the influence of Percolator and FDR analysis. After percolating, the increase in phosphorylated and non-phosphorylated peptides identified was ~7% and ~31% for HCD and ~75% and ~200% for CID-IT, respectively (data not shown). For both fragmentation techniques combined (CID + HCD), the increase in Peptide-Spectrum-Matches was higher for non-phosphorylated peptides (Figure 2-45 (B)). This might be due to (i) the lower overall number of (experimental and theoretical possible) phosphorylated peptides, (ii) the lower fragmentation efficiency of phosphopeptides using CID-IT (intense neutral loss signal) or (iii) the Percolator algorithm is more efficient for peptides with only fixed modifications forms. Generally, accurate peptide identification based on fragment ion matching approaches were more sensitive for CID-IT in terms of search space increase due its lower mass accuracy (analyzed in ion trap mass analyzer) compared to HCD, which can only be analyzed in the Orbitrap mass analyzer.

To summarize, Percolator performed significantly better than a classical FDR analysis. This was independent of the fragmentation or peptide type. Interestingly, percolating on an *E. coli* sized database performed similar to a FASTA file with approximately one tenth of the proteins, using the FDR approach. As a consequence of these findings, all subsequent database searches were performed using the bF-FASTA file with Percolator.

## 2 Developing a Platform for relative Quantification using iTRAQ

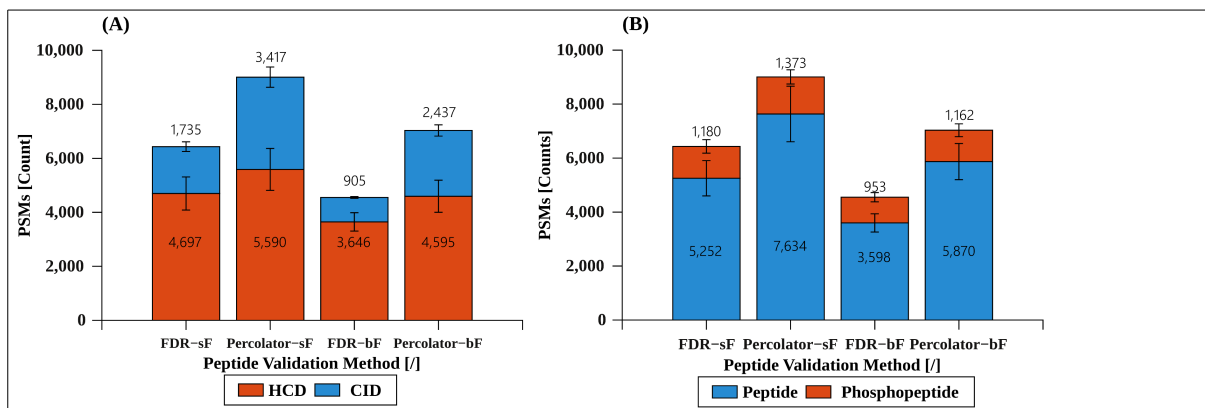


Figure 2-45: For the two workflows using a small FASTA file (sF) and a 10x bigger FASTA file (bF), respectively, (A) PSMs were split into CID-IT and HCD fragmentation and (B) for non- and phosphorylated peptides. For data generation the casein sample after iTRAQ-8Plex labeling and enrichment for phosphorylated peptides was used (sample CF, see workflow). These cover the non-phosphorylated- (NP), wash- (WF) and phosphorylated Fraction (PF) and were searched in MudPIT approach.

### 2.4.3.2 The Advantage of a Multi-Protease Approach

To determine the benefit of the multi-protease versus a single protease approach, the sequence coverage [%], the number of unique peptides and PSMs for all data-sets were compared (Figure 2-46). Trypsin (Tr) achieved a protein coverage ranging from ~ 63 to 78% (except  $\kappa$ -casein; Figure 2-46 (A)). The relative unspecific protease elastase (El) performed best with protein coverages of ~ 73 to 89%. The *in-silico* combination of the four proteases (CETG) and the sample CF (covering the non-phosphorylated, wash and phosphopeptide enriched fractions) both resulted in protein coverages of ~ 92% for the three major proteins ( $\alpha$ -S1-,  $\alpha$ -S2- &  $\beta$ -casein).

Protein sequences were imported from UniProt without any further slicing. Therefore, the first ~15 amino acids (~7% of the protein sequence) represented signal peptides not translated to the full length protein. Consequently, the 92% (*in-silico* CETG and CF sample) represented the maximal to be expected sequence coverage. These data showed, that by using multiple proteases the protein sequence coverage increased. It is noteworthy, that all fractions after phosphopeptide enrichment were analyzed by mass spectrometry.

To further analyze the identification results, the number of unique peptides for each protease, their *in-silico* combination (CETG) and for the sample CF were plotted (Figure 2-46 (B)). On average (three biological replica analyzed), only 91 unique casein-related peptides were identified with trypsin. This represented the lowest number of unique peptides identified from the individual proteases (GluC: 104, chymotrypsin: 162, elastase: 246).

The *in-silico* combination (CETG) and the multi-protease approach (CF) clearly outperform the single protease approach with a total of 550 and 463 unique peptides, respectively. Alongside to the number of unique peptides, similar results were obtained for the PSMs. Due to the higher number of .raw files used for database search, the number of PSMs were biased towards the multi-protease approach (CETG and CF).



## 2.4 MECi – Targeted Phosphoprotein Quantification

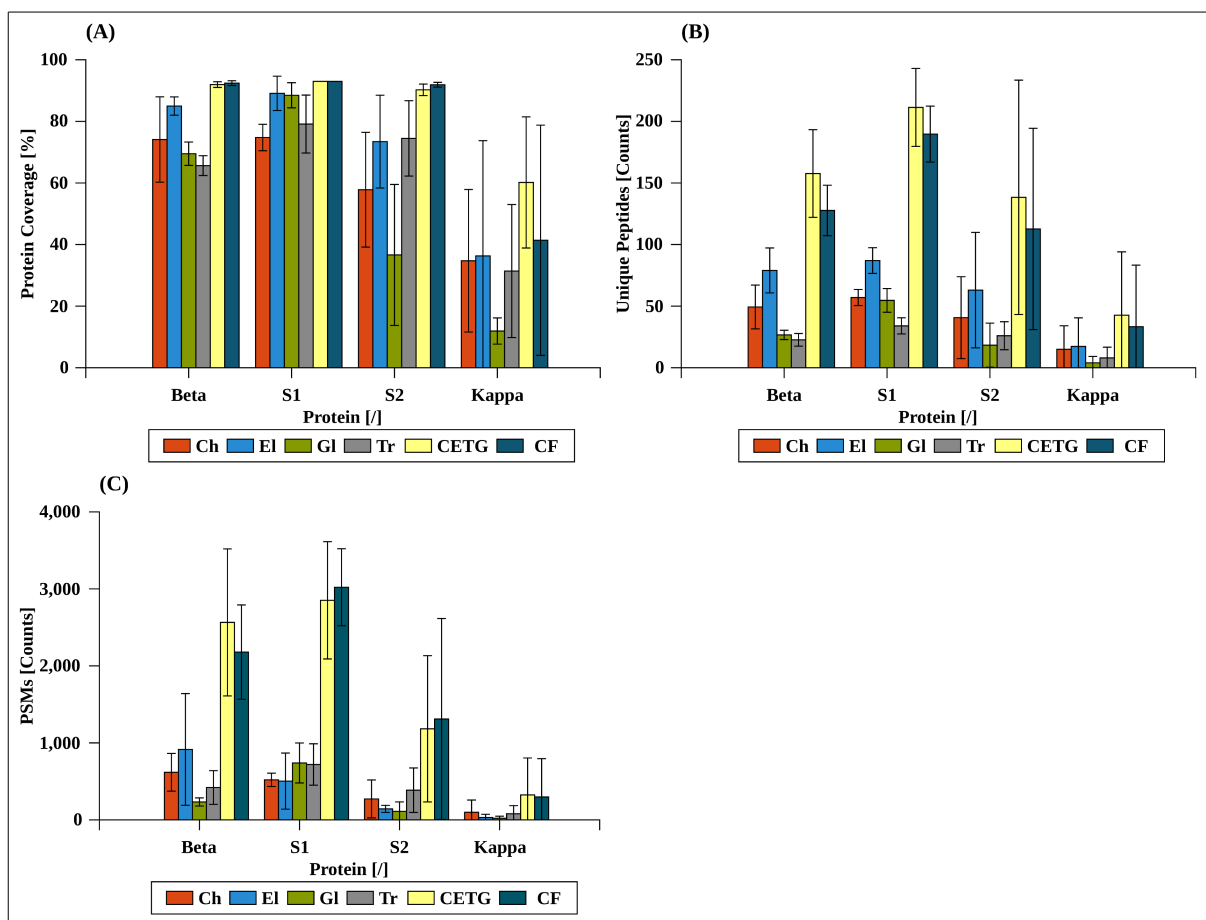


Figure 2-46: To visualize the differences in protein coverage (A), the number of unique peptides (B) and the number of PSMs, these values were plotted for the four singly used proteases (chymotrypsin (Ch), elastase (El), GluC (Gl) and trypsin (Tr); each sample D + P in MudPIT approach), for the *in-silico* combination of the four proteases and the phosphopeptide enrichment fractions (CF).

To determine which protein sequences were more likely to be identified (and not only the number and sequence of unique peptides) the occurrence of each amino acid by *unique peptides* was compared for each individual protease, the *in-silico* combination (CETG) and for the multi-protease approach (CF) after TiO<sub>2</sub> enrichment, respectively. Here, the Heat-Map minimal and maximal values were extracted from all data-sets (referred to as *global*, Figure 2-47 (A) to 2-49 (A)). Whereas for the second data analysis, the minimal and maximal values were extracted within each data-set (e.g., trypsin for  $\alpha$ -S1-casein), referred to as *local* (Figure 2-47 (B) to 2-49 (B)).

For  $\alpha$ -S1- and  $\alpha$ -S2-casein, the number of amino acids extracted from identified unique peptides was rather low for the four singly applied protease digests (Ch, El, Gl & Tr), being indicated by a mainly blue to slightly green color (up to 30%, Figure 2-47 (A)) for the global analysis approach. For  $\beta$ -casein, a higher number of amino acids using the proteases elastase and chymotrypsin were present.

Comparing the *in-silico* combination (CETG) and the enriched sample, the benefits were obvious by the increased frequency of the amino acids. This trend was similar for all three casein-proteins. With this visualization method, the increase in phosphopeptide identifications after TiO<sub>2</sub> enrichment is clear, for example,

## 2 Developing a Platform for relative Quantification using iTRAQ

the phosphorylated amino acids pS<sub>56</sub>, pS<sub>61</sub>,pS<sub>63</sub> and pS<sub>130</sub> showed a higher frequency after enrichment compared to the *in-silico* combined sample (CETG).

Each data-set was also subjected to Heat-Map generation using the local maximal values (Figure 2-47 (B) to 2-49 (B)). This permits to observe specific advantages for each of the proteases used for protein digestion and subsequent analysis of the number of unique peptides.

For  $\alpha$ -S1-casein, the amino acids within the sequence IGVNQEL (position #150) were frequently identified using chymotrypsin. For the remaining proteases, different sequences from  $\alpha$ -S1-casein were predominantly identified after database search such as, SMKEGIHAQQKEPM (position #139) for elastase, DQAME for GluC (position #66) and DIGSESTEDQAMEDIK (position # 58) for trypsin.

The sequence (..R)L(KKY..) (#116) of  $\alpha$ -S1-casein was difficult to be identified independently of the protease used. This was also reflected for the *in-silico* combined sample (CETG) and phosphopeptide enriched fraction (CF) and might give a hint towards inadequate ionization and fragmentation or amino acid polymorphisms, which has to be analyzed subsequent by X!Tandem, *de-novo* and manual spectra interpretation.

For  $\beta$ -casein, similar effect were regarding the protein coverages were observed (Figure 2-48). Trypsin identified hardly sequences between the amino acids #128 to #190. Here, the obtained tryptic peptides contained more than 25 amino acids. Peptide fragmentation and identification is usually optimal for peptides with a sequence length of 5 to 25 AA (314). Consequently, the protein section #128 – #190 was hampered to be identified. However, these protein sections were frequently identified after protein digestion using elastase and chymotrypsin. Another interesting protein section is located at position #101 (PFLQPEVVM). This section was identified frequently (red colored) by chymotrypsin, elastase and GluC (red colored) but less by trypsin (only green colored). By applying the multi-protease approach (CETG & CF), almost all protein regions could be identified. Solely the region of INKKI (position #41) was not identified in any sample.

The results obtained by  $\alpha$ -S2-casein digestion were different among the four proteases. For every protease, different protein sequences were identified. For example, the triply-phosphorylated peptide (with pS<sub>71-73</sub>) was predominantly identified using GluC and elastase, whereas the rate of identifying this triply-phosphorylated peptide was much lower for trypsin.

## 2.4 MECi – Targeted Phosphoprotein Quantification

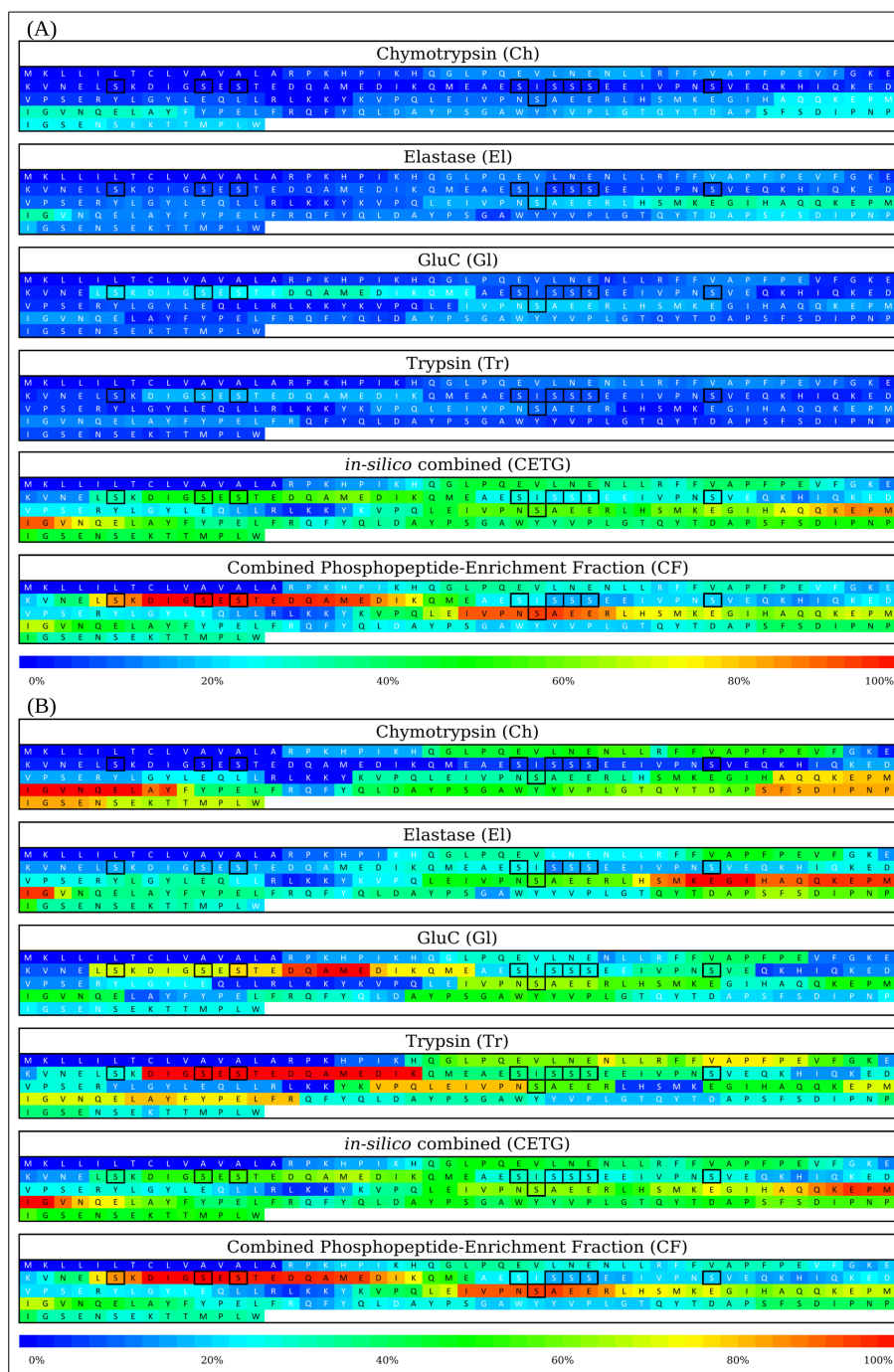


Figure 2-47: Heat-Map in RGB style to visualize the frequency of each amino acid found within all unique peptide groups identified by Proteome Discoverer for  $\alpha$ -S1-casein using the reference workflow. The information about the four singly used proteases chymotrypsin, elastase, GluC and trypsin, the *in-silico* combined sample (CETG) are compared to the enriched sample (CF). The enriched fraction covers the samples non-phosphorylated- (NP), wash- (WF) and phosphopeptide-fraction (PF). The same data sets (unique peptides) were used. However, for (A) the highest number of amino acid occurrence was identified within each experiment, whereas for (B), this was performed over all experiments. Especially for the phosphorylated amino acids pS<sub>56</sub>, pS<sub>61</sub>, pS<sub>63</sub> and pS<sub>130</sub> the positive effect of phosphopeptide enrichment using TiO<sub>2</sub> is clearly demonstrate able.

## 2 Developing a Platform for relative Quantification using iTRAQ

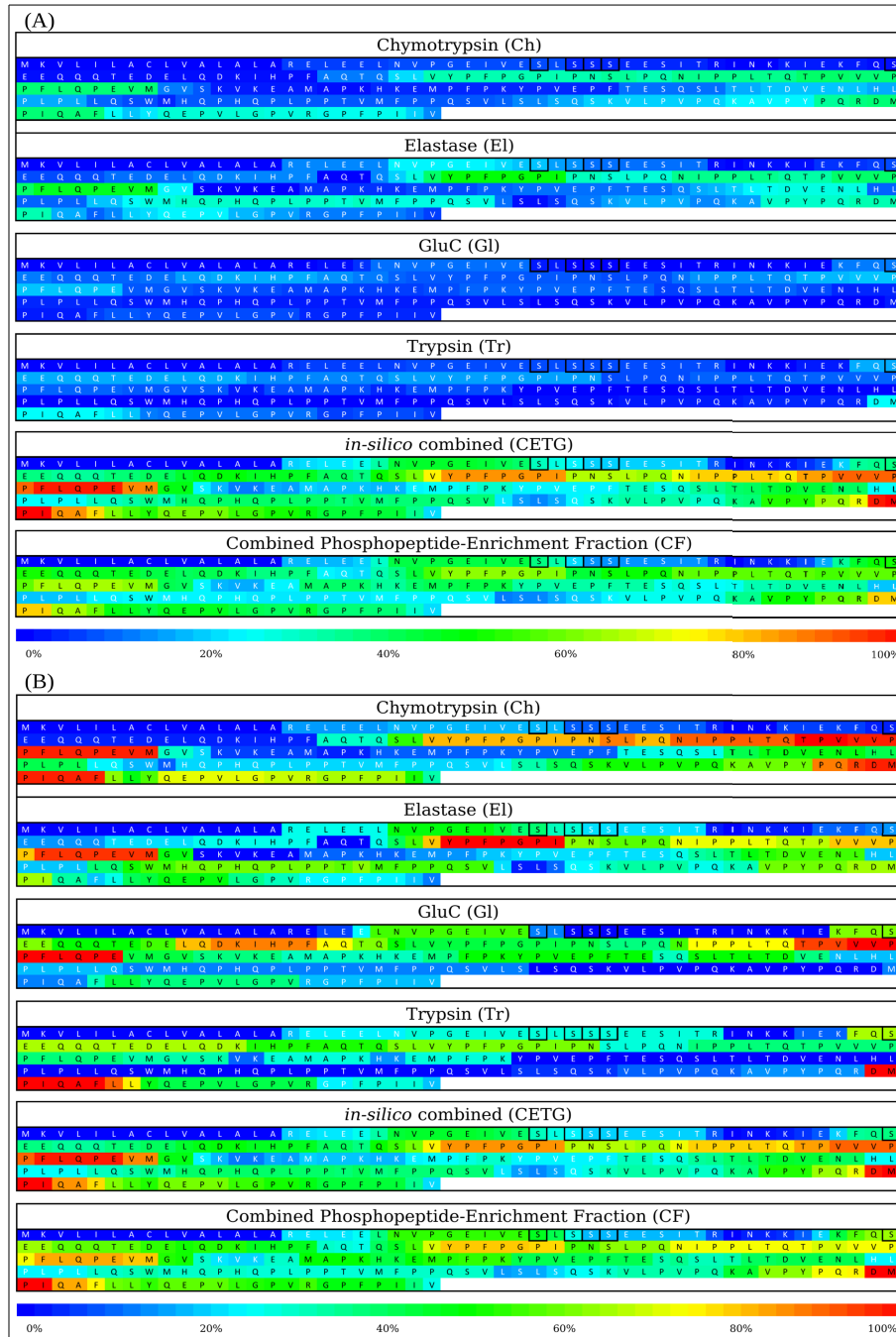


Figure 2-48: Heat-Map in RGB style to visualize the frequency of each amino acid found within all unique peptide groups identified by Proteome Discoverer for  $\beta$ -casein using the reference workflow. Here, the number of PSMs within a peptide groups was not taken into account. The information about the four singly used proteases chymotrypsin (Ch), elastase (El), GluC (Gl) and trypsin (Tr), the combination of all four (ME) as well as the *in-silico* combined sample (CETG) are compared to the enriched sample (CF). The enriched fraction covers the samples non-phosphorylated- (NP), wash- (WF) and phosphopeptide-fraction (PF). The same data sets (unique peptides) were used.

## 2.4 MECi – Targeted Phosphoprotein Quantification

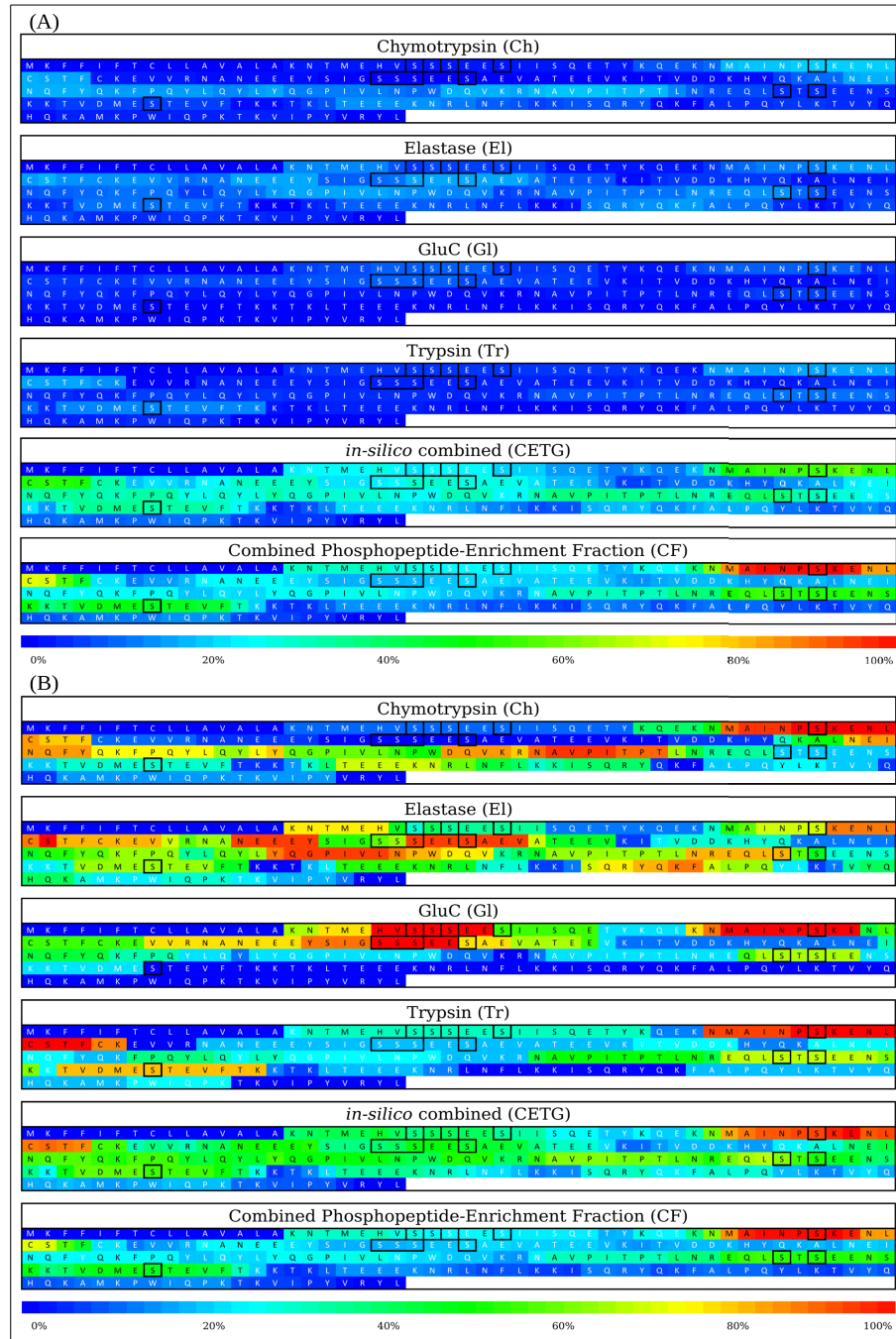


Figure 2-49: Heat-Map in RGB style to visualize the frequency of each amino acid found within all unique peptide groups identified by Proteome Discoverer for  $\alpha$ -S2-casein using the reference workflow. Here, the number of PSMs within a peptide groups was not taken into account. The information about the four singly used proteases chymotrypsin (Ch), elastase (El), GluC (Gl) and trypsin (Tr), the combination of all four (ME) as well as the *in-silico* combined sample (CETG) are compared to the enriched sample (CF). The enriched fraction covers the samples non-phosphorylated- (NP), wash- (WF) and phosphopeptide-fraction (PF). The same data sets (unique peptides) were used.

## 2 Developing a Platform for relative Quantification using iTRAQ

### 2.4.3.3 Enrichment of iTRAQ-8Plex labeled Phosphopeptides

Enrichment techniques are a prerequisite for the analysis of phosphorylated peptides in large-scale phosphoproteomics studies, as they are frequently of low stoichiometry (122). In this study, TiO<sub>2</sub> was chosen for enrichment as it is robust regarding buffer and salts (125). Three fractions were collected and analyzed: non-phosphorylated-, wash- and the phosphopeptide enriched fraction. In pre-experiments using  $\beta$ -casein, the enrichment efficiency and selectivity was tested using different ratios of TiO<sub>2</sub>: starting from the widely applied 1:5 or 1:6 (protein:TiO<sub>2</sub>) ratio for in-solution full proteome studies (272), increasing amounts of TiO<sub>2</sub> (1:10, 1:15 and 1:20) were used for workflow optimization. Finally, a protein to TiO<sub>2</sub> ratio of 1:20 was found to be optimal for enrichment. At lower ratios, the phosphopeptide pS<sub>50</sub> was also identified in the non-phosphorylated and washing fractions. This phenomenon can be caused by the high amount of phosphorylated peptides or by impurities introduced by SDS-PAGE separation performed at protein level.

In order to demonstrate the benefits of using a multi-protease approach (Figure 2-50), the number of unique (A, C, E) phosphopeptides and the total number of Peptide-Spectrum-Matches (PSMs) of phosphorylated peptides (B, D, F) were analyzed for the four proteases alone, their *in-silico* combination (CETG) and the phosphopeptide enriched fraction (CF). Although the sample complexity is rather low compared to a whole cell lysate, not all phosphorylated amino acids could be identified with the four proteases in the single protease approach. By combining the four proteases *in-silico*, not surprisingly, both, the number of unique and PSMs increased. However, the best result was obtained after TiO<sub>2</sub> enrichment, which demonstrated the excellent compatibility of the chemicals used with the optimized TiO<sub>2</sub> enrichment procedure. In particular for  $\alpha$ -S2-casein, the number of identified peptides (both unique and PSMs) was significantly higher after enrichment compared to the *in-silico* combined sample (Figure 2-50; (E & F)).

## 2.4 MECi – Targeted Phosphoprotein Quantification

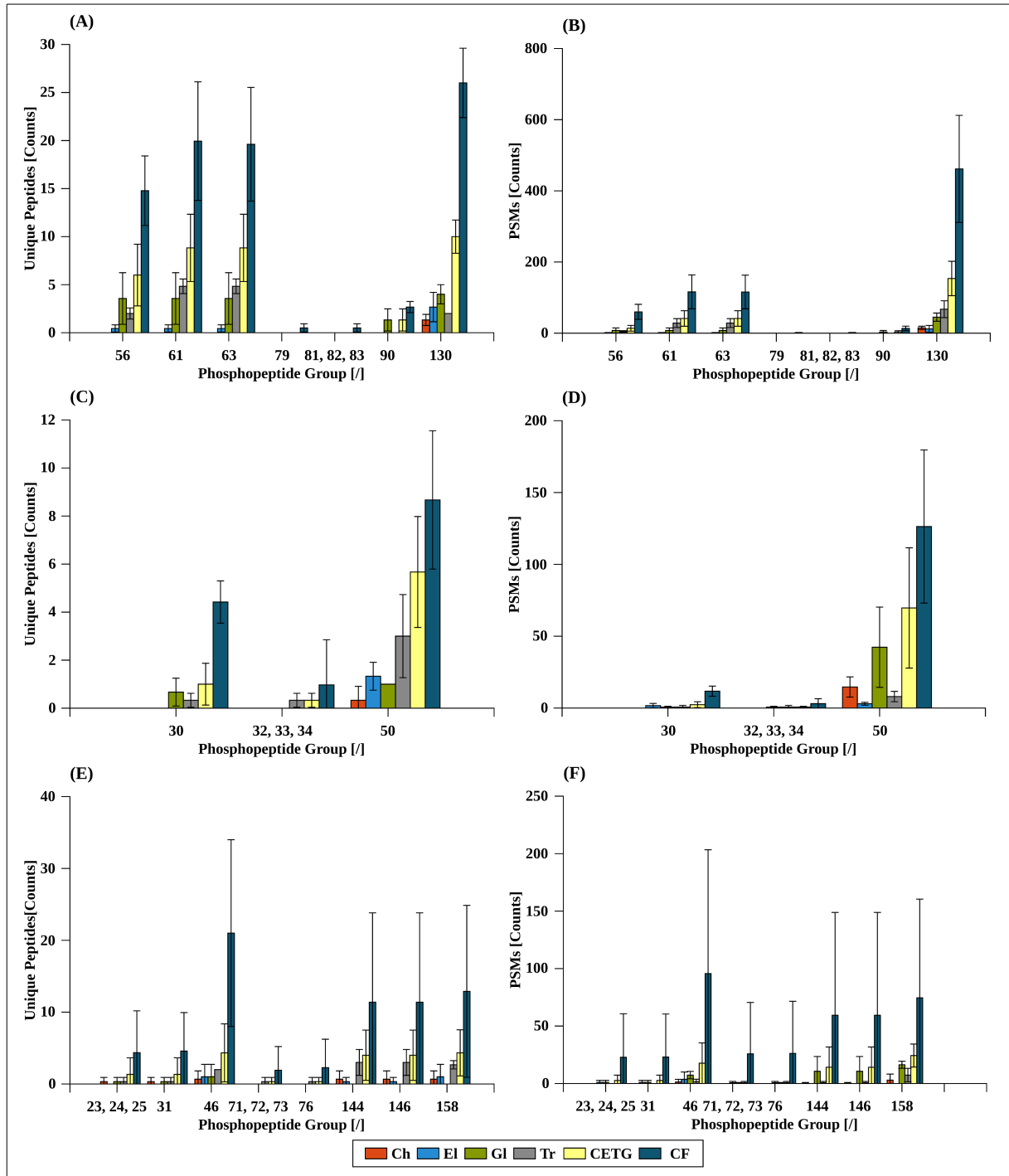


Figure 2-50: Number of unique identified phosphopeptide sequences (A, C, E) and their total Peptide-Spectra-Matches (PSMs; (B, D, F)) for  $\alpha$ -S1-casein (A, B),  $\beta$ -casein (C, D) and  $\alpha$ -S2-casein (E, F) identified by Proteome Discoverer 1.4 using the reference workflow (see Material and Methods for further details). The column diagrams contain the information about the four singly used proteases chymotrypsin (Ch), elastase (El), GluC (Gl) and trypsin (Tr). Additionally, the in-silico combined sample (CETG) is compared to the enriched sample (CF). The enriched fraction covers the samples non-phosphorylated- (NP), wash- (WF) and phosphopeptide-fraction (PF).

## 2 Developing a Platform for relative Quantification using iTRAQ

### 2.4.3.4 Phosphoprotein Quantification using iTRAQ-4x2Plex

The multi-protease approach has been shown to yield an increased number of phosphorylated peptides in terms of PSMs and unique peptide sequences (302). However, the previously published methods (302,315) lack either (i) a quantitative dimension or (ii) the multi-protease approach. The experimental design (Figure 2-43) described in the present work filled this gap by using a quantitative (iTRAQ-8Plex labeling) multi-protease approach which allowed to perform relative quantification. Due to the unique iTRAQ reporter ion labeling, principally this can be achieved using both HCD and ETD fragmentation (Table 2-8).

#### **Grouping of Quantifiable Peptides**

To quantify phosphorylated peptides, it was not possible to simply group unique peptide sequences according to the C-terminal amino acids (Figure 2-51). This is due to the overlapping cleavage site specificities of proteases: for chymotrypsin two overlapping amino acids (D, E) with GluC and one overlapping amino acid (L) with elastase are known. These were classified as Ch-Gl and Ch-El, respectively. For singly & doubly phosphorylated peptide sequences (pS<sub>56</sub>, pS<sub>61</sub>, pS<sub>63</sub> and pS<sub>130</sub> of  $\alpha$ -S1-casein) the number of unique peptides identified was similar between the four proteases (protease groups, respectively). However, for peptide sequences in which multiple phosphorylation sites (e.g., pS<sub>81-83</sub> of  $\alpha$ -S1-casein) were present, the detection and identification was more difficult. This can be seen for  $\alpha$ -S1-caseins phosphorylated amino acids pS<sub>79</sub> and pS<sub>81-83</sub> for which considerably fewer unique peptides were identified.

Beside the specific C-terminal amino acids, a high number of unspecific cleavage events were identified, such as C-terminal threonine and histidine (Figure 2-51). From pre-experiments performed with a similar mix of  $\beta$ -,  $\alpha$ -S1- and  $\alpha$ -S2-casein, these were shown to mainly originate from elastase (T) and chymotrypsin (H) cleavage. However, using the iTRAQ ratios (IS ratio; Equation 2-10) peptides could be easily be re-mapped to the specific protease from which they were generated (for direct comparison).

The peptide  ${}_{i}\text{LEIVPN}_{p}\text{SAEERL}$  was chosen as an example which, from the C-terminal amino acid, can not be unambiguously assigned to either chymotrypsin or elastase due to overlapping cleavage specificities. I/L cannot be differentiated by LE-CID and are often set equally for database searches. However, based on the intense iTRAQ<sub>113</sub> and iTRAQ<sub>119</sub> signals, it was unambiguously that this peptide was generated by elastase rather than chymotrypsin (Figure 2-52).

Information regarding the iTRAQ reporter ions was used in combination with the peptides C-terminal amino acid to remove potential false positive identifications. *In-house* developed VB(A) scripts were used to group peptides according to the protease, which they were formed by based on the C-terminal amino acid. Moreover, the iTRAQ reporter ions intensities were used to calculate the median and standard deviations for the w-Ratio (Equation 2-11). The corresponding peptide w-ratios are summarized in chapter 4.4.



## 2.4 MECi – Targeted Phosphoprotein Quantification

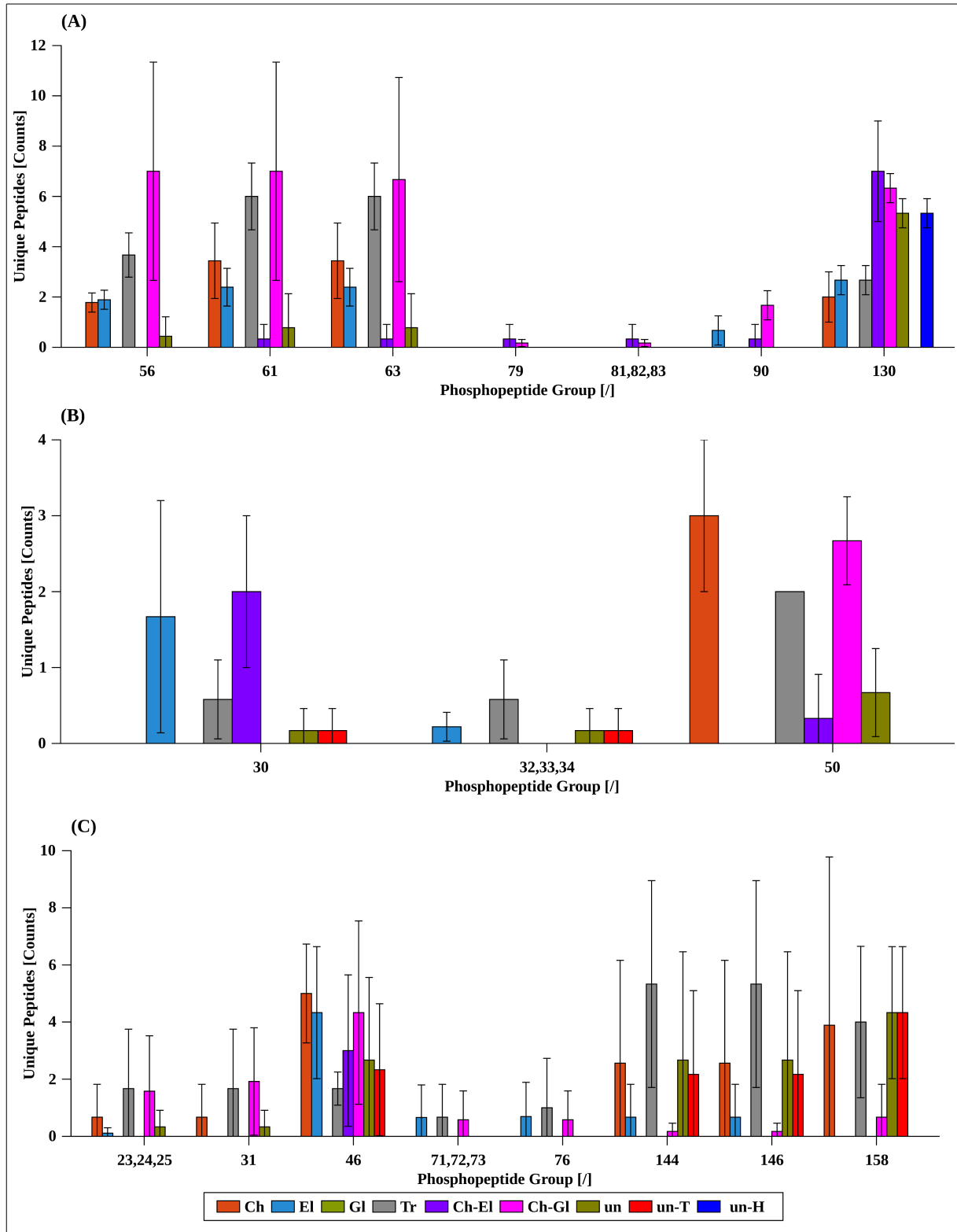


Figure 2-51: The identified phosphorylated peptides within the protein  $\beta$ -casein (A) and  $\alpha$ -S2-casein (B) were split into protease groups according to the C-terminal amino acid. Chymotrypsin and GluC (Ch-Gl) share two P1' and chymotrypsin and elastase (Ch-El) one P1' cleavage sites, respectively, making them indistinguishable without further bioinformatics tools (147). For the remaining cleavage sites (chymotrypsin, Ch; elastase, El; GluC, Gl; trypsin, Tr) only unique C-terminal cleavage sites are known. Other C-terminal amino acids were also

## 2 Developing a Platform for relative Quantification using iTRAQ

summed up and shown (unspecific, un). The amino acids threonine (un-T) and histidine (un-H) were highly present and shown additionally. For the peptide group 130, all unspecific amino acids were also a part of the Histidine group (un-H).

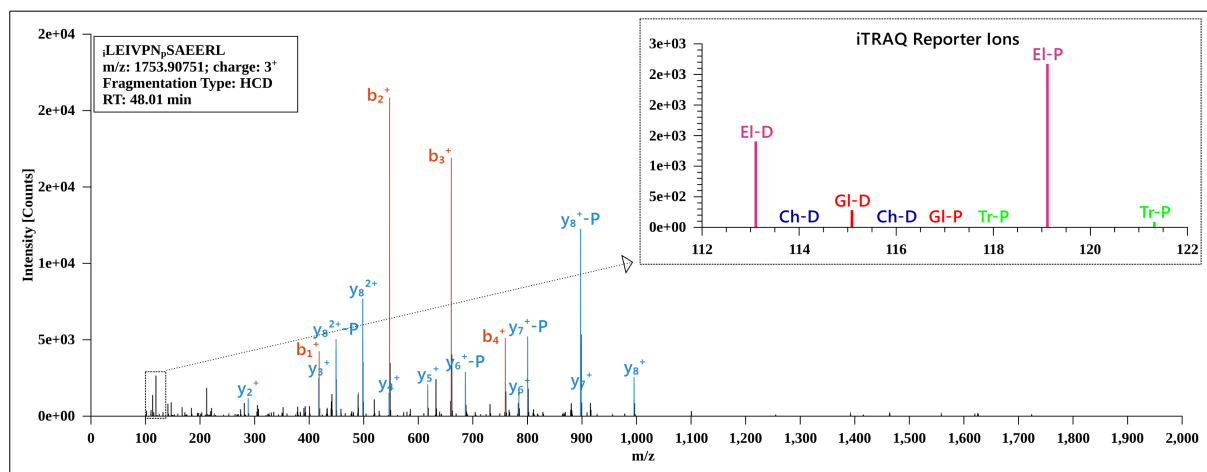


Figure 2-52: The annotated HCD MS/MS spectrum (filter: max fragment ion deviation:  $\pm 20$ ppm, signal-intensity  $> 5\%$  of most intense signal in spectra) for the iTRAQ-8Plex labeled peptide *i*LEIVPN<sub>p</sub>SAEERL showing an almost complete sequence coverage (b- and y-ion series combined). Additionally, fragment ions formed by neutral loss events similar to [M-P], were observed (e.g., y<sub>6</sub><sup>+</sup>-P). Although this peptide can be generated by elastase or chymotrypsin cleavage (I & L can be set to equal in database searches, e.g., MaxQuant (300)) the iTRAQ reporter ion intensities (without impurity correction) clearly shows its origin by elastase digestion (intense iTRAQ<sub>113</sub> and iTRAQ<sub>119</sub> signal).

### Implementation of (Phospho)Peptide Group Quantification

Classically for iTRAQ related proteomics studies, the iTRAQ ratio is generated by dividing one iTRAQ reporter ion signal intensity (or area) by another (e.g., iTRAQ<sub>115</sub>/iTRAQ<sub>114</sub>). In this study, an alternative calculation was used (w-Ratio) to overcome the problem of division by 0, which was expected to occur for dephosphorylated peptides. The w-Ratio scale of 0 to 1 improved the data visualization process to compare differentially abundant peptides. In the following, the numerator was chosen to be the iTRAQ reporter ion of the dephosphorylated sample D for all calculations of the w-Ratio. With the experimental sample set up, dephosphorylated peptides should show w-Ratios towards 1, as these non-phosphorylated peptides should be only present in low abundance (or not at all) in the natively phosphorylated sample. In contrast, the phosphorylated counterparts should have a w-Ratio of 0.3 (equals iTRAQ ratio of 0.5), because the sample D is a 1:1 composition of dephosphorylated sample and the natively phosphorylated casein sample.

The chosen standard proteins  $\alpha$ -S1-,  $\alpha$ -S2- and  $\beta$ -casein were used to validate the quantitative multi-protease workflow. The number of phosphorylation sites and its degree is high ( $\alpha$ -S1-casein = 8-9,  $\alpha$ -S2-casein = 10-13 and  $\beta$ -casein = 4-5 number of phosphorylation sites). Rather than being distributed evenly across the entire protein, many of the phosphoserine residues exist in clusters, e.g., pS<sub>81-83</sub> in  $\alpha$ -S1-casein and pS<sub>32-34</sub> in  $\beta$ -casein. Therefore, caseins are the ideal model substances to evaluate whether the developed workflow was capable of identifying both singly- and multiply-phosphorylated peptides.

## 2.4 MECi – Targeted Phosphoprotein Quantification

For quantification, only high confident peptides (FDR < 1%) were used which is a common approach for quantitative proteomics studies (113). For demonstration purposes, all quantified tryptic peptides for  $\alpha$ -S1-casein (Figure 2-53) were extracted from the data-set to present differences for de- & phosphorylated peptides (*w*-Ratio of 1 and 0.3, respectively) and peptide without differential expression (*w*-Ratio = 0.5). Peptide groups were clustered according to their iTRAQ-8Plex reporter ions so that the results achieved by each protease could be compared for the three casein proteins (Figure 2-54 to 2-58).

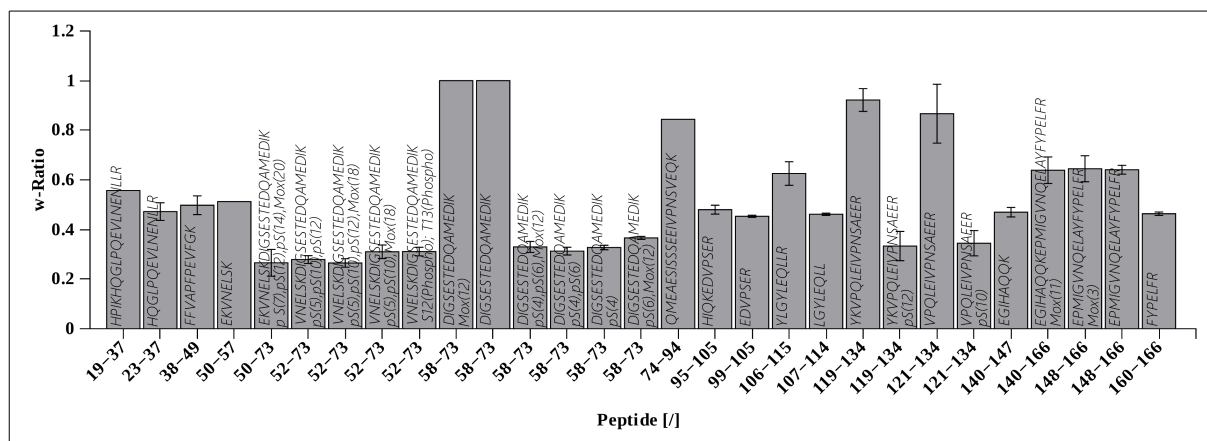


Figure 2-53: All tryptic peptides identified and quantified for  $\alpha$ -S1-casein. After re-grouping of unique peptides the median and the standard deviation for each peptide group was calculated. For dephosphorylated peptides, a *w*-Ratio towards 1 is expected, whereas for the phosphorylated counterpart a *w*-Ratio of 0.3 should be observed.

### Quantifying (Phospho)Peptide Groups obtained by a Multi-Protease Digest

#### *Quantification Results for $\alpha$ -S1-Casein*

$\alpha$ -S1-casein contains phosphorylated serines at positions 56, 61, 63, 79, 81-83, 90 and 130 many of which were quantified by tryptic peptides (Figure 2-54 (D)). The triply phosphorylated peptide (Peptide<sub>52-73</sub>), containing one tryptic missed cleavage, and the doubly phosphorylated form (without a missed cleavage; Peptide<sub>58-73</sub>) were quantified with *w*-Ratios of 0.27 and 0.31, respectively. Interestingly, both peptides were also identified with a missing phosphorylation site exhibiting slightly higher *w*-Ratios than their triply and doubly phosphorylated counterparts (e.g., peptide #4 and #5 in Figure 2-54 (D)). Similar results were obtained with GluC and elastase, i.e., various degrees of phosphorylation were observed for phosphopeptides containing residues pS<sub>56</sub>, pS<sub>61</sub> and pS<sub>63</sub> (see below). Both  $\alpha$ -S1- and  $\alpha$ -S2-casein contain various number of phosphorylated serines ( $\alpha$ -S1 = 8-9,  $\alpha$ -S2 = 10-13) and thus, the various degrees of peptide phosphorylation for these proteins may be a result of phosphoheterogeneity, which is characteristic for casein proteins (316), or be a result of the alkaline phosphatase treatment.

For the tryptic peptide<sub>58-73</sub>, both forms (de- and phosphorylated) were present. The non-phosphorylated peptide (#7 and #8 in Figure 2-54 (D)) showed a *w*-Ratio of 1, whereas their phosphorylated counterparts (#9 – #12) showed the expected *w*-Ratio of 0.3. The results found for this peptide demonstrated the advantages of analyzing all fractions of the phosphopeptide enrichment. Both the phosphorylated and non-phosphorylated forms can be identified giving complementary quantitative information. For the phosphorylation site at pS<sub>130</sub>,

## 2 Developing a Platform for relative Quantification using iTRAQ

both the dephosphorylated and phosphorylated forms were identified for each peptide with w-Ratios of 0.9 and 0.3 for dephosphorylated and the phosphorylated peptides, respectively. In total for trypsin, 11 phosphopeptides and 6 dephosphorylated counterparts were identified for  $\alpha$ -S1-casein. However, while many of the phosphorylation sites on  $\alpha$ -S1-casein were identified with trypsin, phosphoserine residues at positions pS<sub>79</sub>, pS<sub>81-83</sub> and pS<sub>90</sub> were not. The remaining peptides, not being part of a potential phosphorylation site, were also plotted together with the de- and phosphorylated to demonstrate that these scatter around a w-Ratio of 0.5 (Figure 2-53).

Interestingly, for GluC (Figure 2-54 (C)), only singly and doubly phosphorylated peptides were identified. This might be caused by the acidic C-Terminal amino acid (E) in addition to the negatively charged phospho-moiety lowering the ionization efficiency. The peptide<sub>55-65</sub> and peptide<sub>55-76</sub> were identified in both the doubly phosphorylated and completely dephosphorylated forms, with expected w-Ratios. The peptide<sub>77-92</sub> covered in total five known phosphorylation sites, but only the completely dephosphorylated form was identified (w-Ratio = 0.9). Although the phosphorylated form was not identified from GluC digests, a potential posttranslational modification can be assumed due to the high iTRAQ w-Ratio, which is in contrast to the remaining peptides (see Table 4-4 for further details). This demonstrated again the benefit of analyzing all fractions following TiO<sub>2</sub> enrichment. The phosphorylation site at position pS<sub>90</sub>, was covered by both forms with expected w-Ratios (Peptide<sub>86-92</sub>). The phosphorylation site pS<sub>130</sub> (Peptide<sub>126-132/133/140</sub>) was identified with various sequence lengths (with or without missed cleavages) and with and without oxidation of methionine, all of which were quantified correctly.

For the elastase digest (Figure 2-54 (B)), all phosphorylation sites (source: UniProt, last modified June 11, 2014, version 118) within the protein  $\alpha$ -S1-casein were quantified. This was not observed for peptides obtained from the GluC and trypsin digest. This again clearly demonstrated the benefit of a quantitative multi-protease approach. Similarly to GluC and trypsin, the peptide<sub>53-72/75</sub> covering three phosphorylation sites (pS<sub>56</sub>, pS<sub>61</sub>, pS<sub>63</sub>) was identified as a triply and doubly phosphorylated peptide with elastase. Surprisingly, no completely dephosphorylated counterpart was identified. The peptide<sub>76-107</sub> covered five phosphorylation sites, and this peptide was also identified with the correct w-Ratio. The Peptide<sub>81-89</sub>, covering four potential phosphopeptides positions (pS<sub>81</sub>, pS<sub>82</sub>, pS<sub>83</sub> & pS<sub>90</sub>) was identified in its dephosphorylated form with the correct w-Ratio of 0.9. As demonstrated before for GluC, the phosphorylation site pS<sub>130</sub> was identified in a various peptide sequence lengths (missed cleavages) and dynamic modifications (peptides #8 – #16 in Figure 2-54 (C)).

In terms of quantifying  $\alpha$ -S1-casein phosphorylation sites, identification and quantification performed the worst for the chymotrypsin digest (Figure 2-54 (A)). The Peptide<sub>77-85</sub> showed w-Ratios of  $\sim 0.4$  which did not match to those observed with the other proteases. For the phosphorylation site pS<sub>130</sub>, the expected w-Ratios were observed although, one of the peptides (Peptide<sub>126-140</sub>) had a too low w-Ratio.

## 2.4 MECi – Targeted Phosphoprotein Quantification

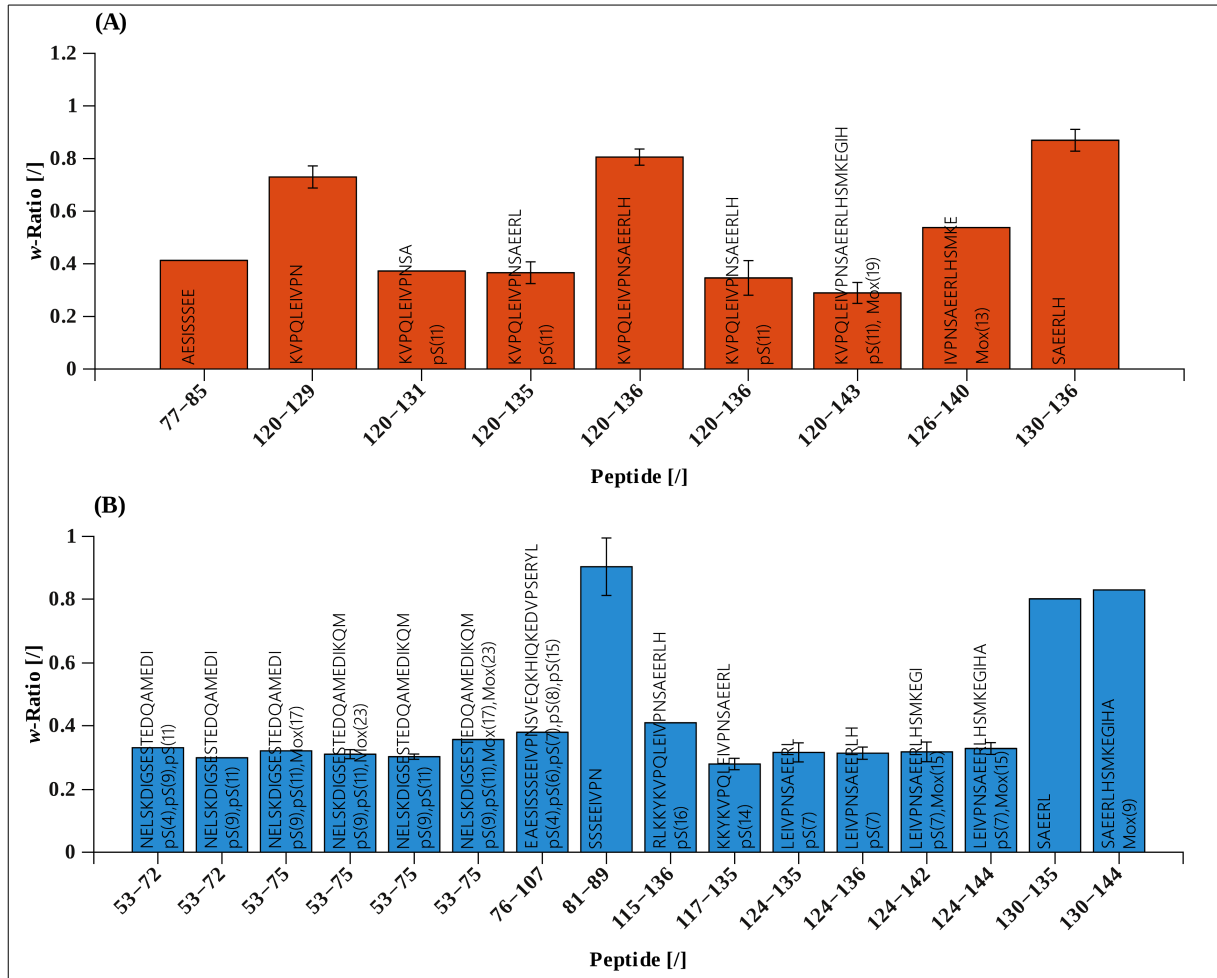


Figure 2-54:  $\alpha$ -S1-casein peptides covering a potential phosphorylation site were extracted from the identified PSM list and further split into the related proteases according to their C-terminus and the iTRAQ reporter ion signals for easier visualization (chymotrypsin in yellow, elastase in cyan, GluC in olive and trypsin in gray). Finally, after re-grouping of unique peptides the median and the standard deviation for each peptide group was calculated. For dephosphorylated peptides, a w-Ratio towards one is expected, whereas for the phosphorylated counterpart a w-Ratio of 0.3 should be observed.

## 2 Developing a Platform for relative Quantification using iTRAQ

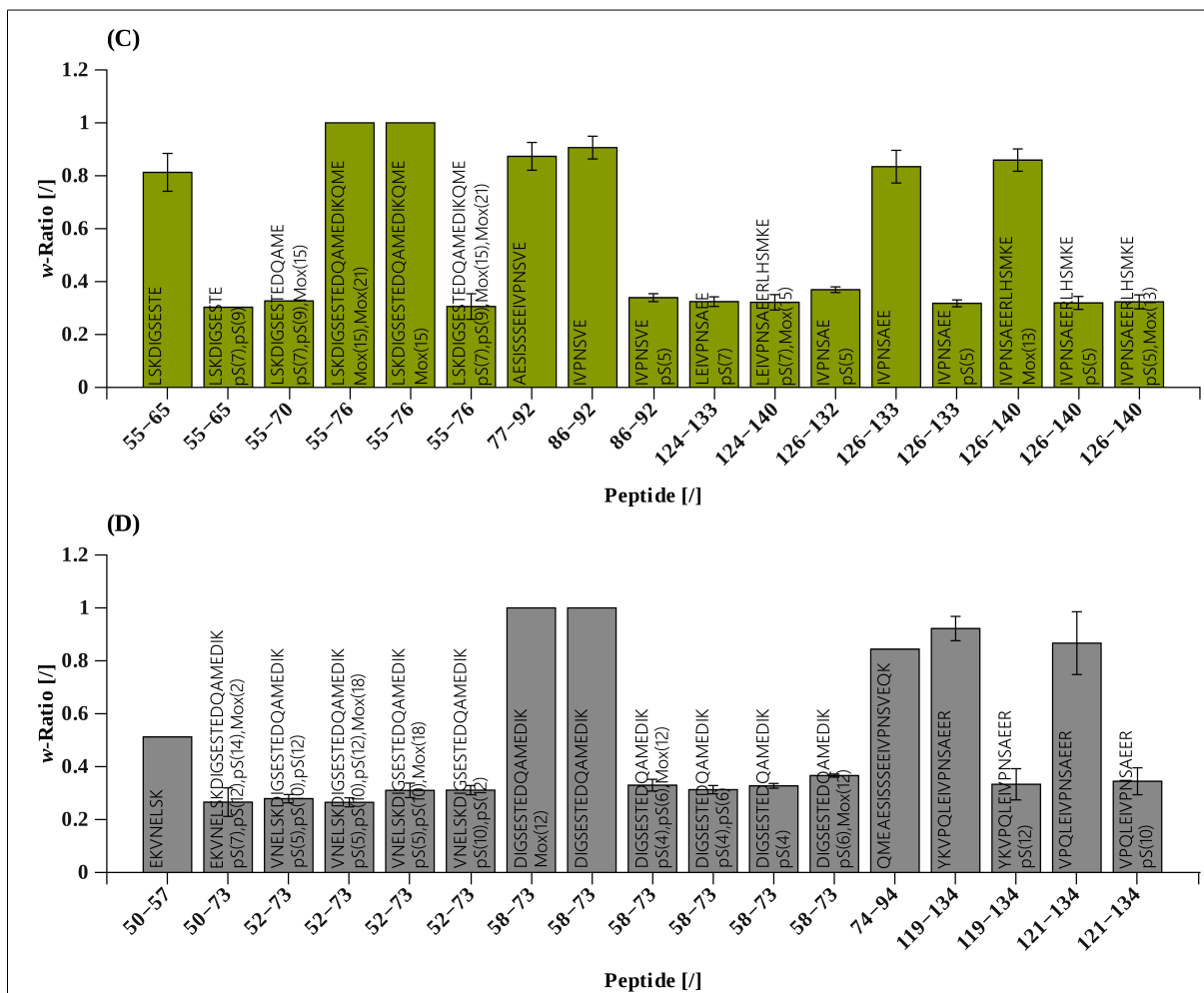


Figure 2-54 (continued):  $\alpha$ -S1-casein peptides covering a potential phosphorylation site were extracted from the identified PSM list and further split into the related proteases according to their C-terminus and the iTRAQ reporter ion signals for easier visualization (chymotrypsin in yellow, elastase in cyan, GluC in olive and trypsin in gray). Finally, after re-grouping of unique peptides the median and the standard deviation for each peptide group was calculated. For dephosphorylated peptides, a w-Ratio towards one is expected, whereas for the phosphorylated counterpart a w-Ratio of 0.3 should be observed.

### Quantification Results for $\beta$ -Casein

$\beta$ -casein contains two isolated (pS<sub>30</sub> and pS<sub>50</sub>) and three clustered (pS<sub>32,33,34</sub>) phosphorylation sites. Due to the lower number of phosphorylation sites, all peptides identified with the four proteases were plotted on one graph (Figure 2-55). Using the quantitative multi-protease approach, all phosphorylation sites could be identified and correctly quantified (except the chymotryptic Peptide<sub>44-67</sub>). For tryptic  $\beta$ -casein digestion, only two peptide sequences (without missed cleavages or dynamic modifications) were expected due to the clustering of the phosphorylation sites (pS<sub>30-34</sub>). Experimentally, these two peptide sequences were identified using trypsin (Peptide<sub>16-40</sub> and Peptide<sub>48-63</sub>). Additionally, one tryptic phosphorylated peptide (Peptide<sub>45-63</sub>) with one missed cleavage site was identified and quantified correctly (w-Ratio = 0.3). Elastase was cleaving after residue L<sub>21</sub> and L<sub>31</sub> and producing a shorter phosphopeptide (Peptide<sub>22-31</sub>, w-Ratio of 0.4 and 0.9 for the phosphorylated and

## 2.4 MECi – Targeted Phosphoprotein Quantification

dephosphorylated form, respectively). The phosphorylation site pS<sub>50</sub> was quantified correctly using GluC derived peptides (w-Ratio of 0.3 and 0.9 for the phosphorylated and dephosphorylated form, respectively), although three phosphorylated peptides with one to three missed cleavages were identified. However, the triply phosphorylated peptide was not identified after elastase or GluC digestion.

For chymotrypsin, only the phosphorylation pS<sub>50</sub> was identified. Two of the corresponding phosphorylated peptides (Peptide<sub>42-67</sub> and Peptide<sub>46-67</sub>) were quantified with the expected w-Ratios of 0.3. Surprisingly, the chymotryptic peptide KIEKFQSEEQQTEDELQDKIHFP (Peptide<sub>44-67</sub>) showed a w-ratio of approx. 1 for both, the phosphorylated and dephosphorylated peptide (Peptide<sub>49-67</sub>). This was unexpected due to the excellent performance of the remaining three proteases. By closer inspection of the raw-data, this phenomenon could (partially) be explained. The phosphopeptide was identified twice as 5<sup>+</sup> and three times as 4<sup>+</sup> charged precursor. All of these share the same retention time window (within 30 sec due to dynamic exclusion) and precursor mass deviation (3 ppm). However, the iTRAQ ratios for the 5<sup>+</sup> charged precursor followed the expected trends (w-ratio = 0.3), whereas those of the 4<sup>+</sup> charged precursor ion did not meet the expectations (w-ratio = 1). Here, only the iTRAQ reporter ion at m/z 114 was detected in tandem MS spectra. Currently, this observation cannot be explained.

With the multi-protease approach, in total 16 peptides were quantified for β-casein; with averaged w-Ratios of 0.32 and 0.89 for the phosphorylated and dephosphorylated peptides, respectively.

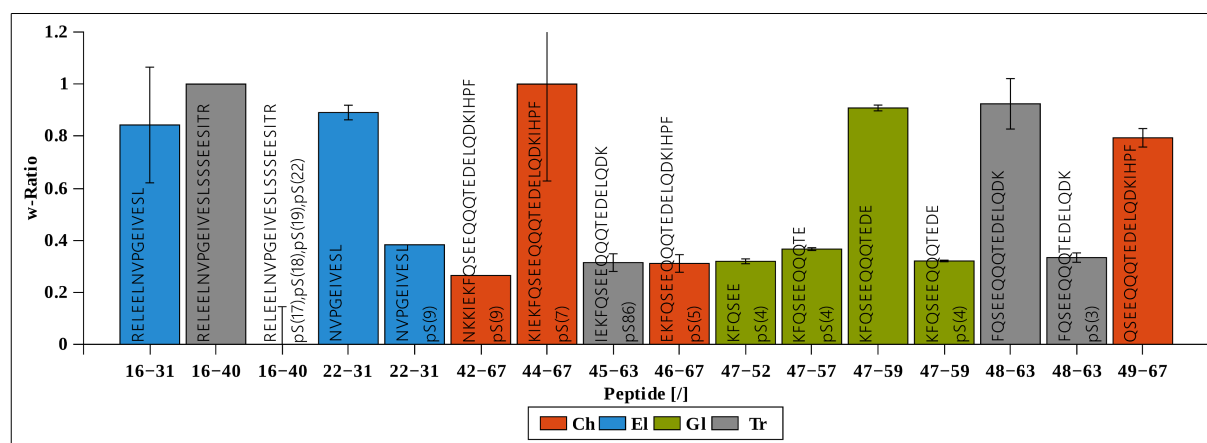


Figure 2-55: β-casein peptides covering a potential phosphorylation site were extracted from the identified PSM list and further split into the related proteases according to their C-terminus and the iTRAQ reporter ion signals for easier visualization (chymotrypsin in yellow, elastase in cyan, GluC in olive and trypsin in gray). Finally, after re-grouping of unique peptides the median and the standard deviation for each peptide group was calculated. For dephosphorylated peptides, a w-Ratio towards 1 is expected, whereas for the phosphorylated counterpart a w-Ratio of 0.3 should be observed.

### Quantification Results for α-S2-Casein

All phosphoserine residues of α-S2-casein were identified and quantified (Figure 2-56 to 2-58). Interestingly, the phosphoserine at pS<sub>46</sub> showed no quantitative difference (w-Ratio = 0.5) for both the phosphorylated or dephosphorylated forms. For this two phosphoserine residue, peptides from each protease

## 2 Developing a Platform for relative Quantification using iTRAQ

exhibited a w-Ratio of ~ 0.5. The different degree of phosphorylation was stimulated by alkaline phosphatase treatment. This phosphoserine residues may not be accessible to the alkaline phosphatase. Samples were incubated under relatively mild conditions: 30°C overnight at pH 8.0 (Tris-HCl). Elevated temperatures or denaturing conditions, e.g., low amounts of detergents, may have resulted in a more complete dephosphorylation of  $\alpha$ -S2-casein. The remaining phosphorylated peptides, however, behaved as expected with an average w-Ratio of 0.3 for the remaining phosphopeptides.

In summary, the phosphorylation sites from all the caseins were accurately quantified with the new developed workflow. With quantitative information from four proteases, the multi-protease iTRAQ-4x2Plex labeling strategy could be used not only to identify global changes in protein phosphorylation but also accurately identify up- and down-regulation of individual phosphorylation sites, as shown for  $\alpha$ -S2-casein at pS<sub>31</sub> and pS<sub>46</sub>.

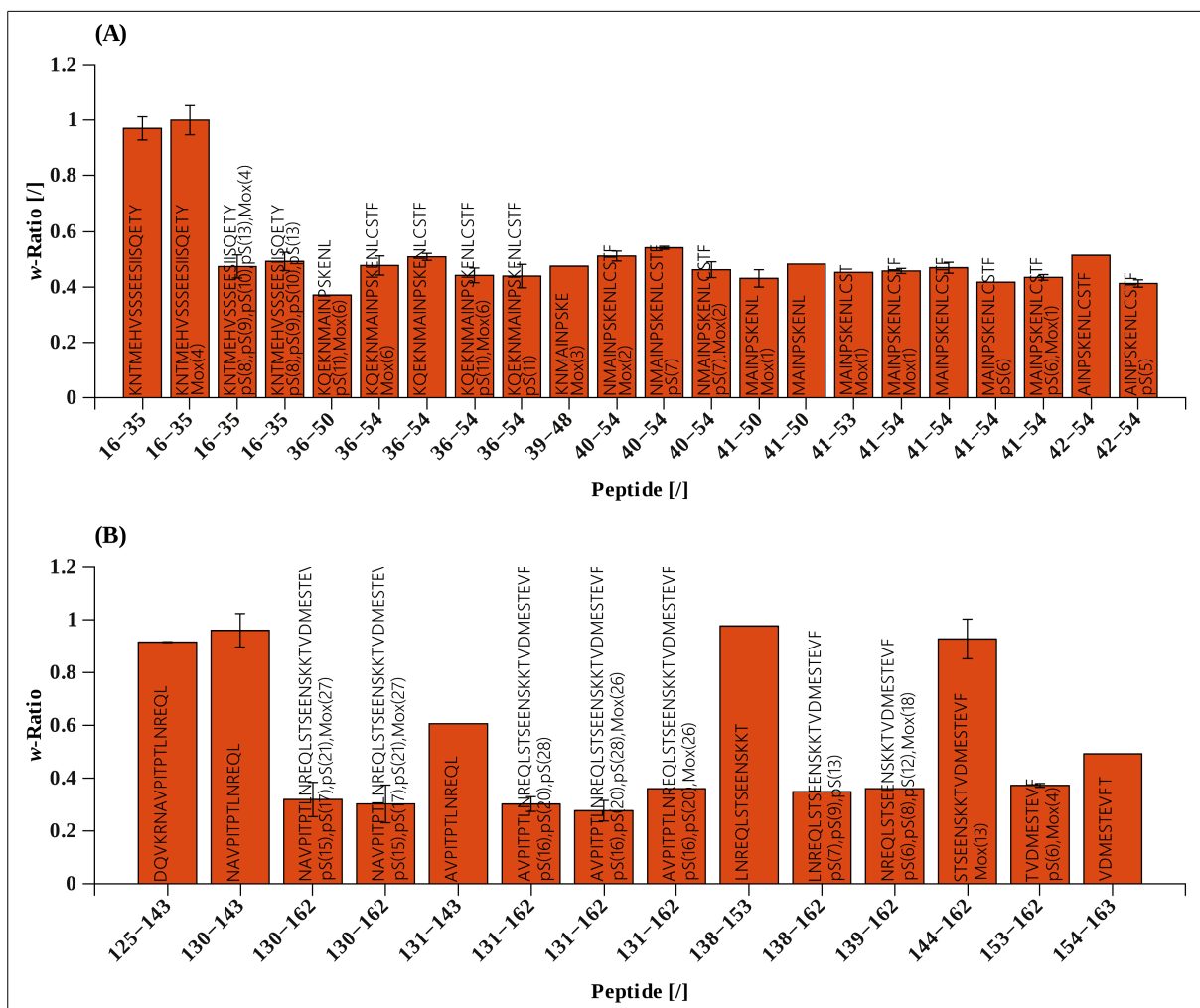


Figure 2-56:  $\alpha$ -S2-casein chymotryptic peptides covering a potential phosphorylation site were extracted from the identified PSM list and further split into the related proteases according to their C-terminus and the iTRAQ reporter ion signals. Finally, after re-grouping of unique peptides the median and the standard deviation for each peptide group was calculated. For dephosphorylated peptides, a w-Ratio towards 1 is expected, whereas for the phosphorylated counterpart a w-Ratio of 0.3 should be observed.



## 2.4 MECi – Targeted Phosphoprotein Quantification

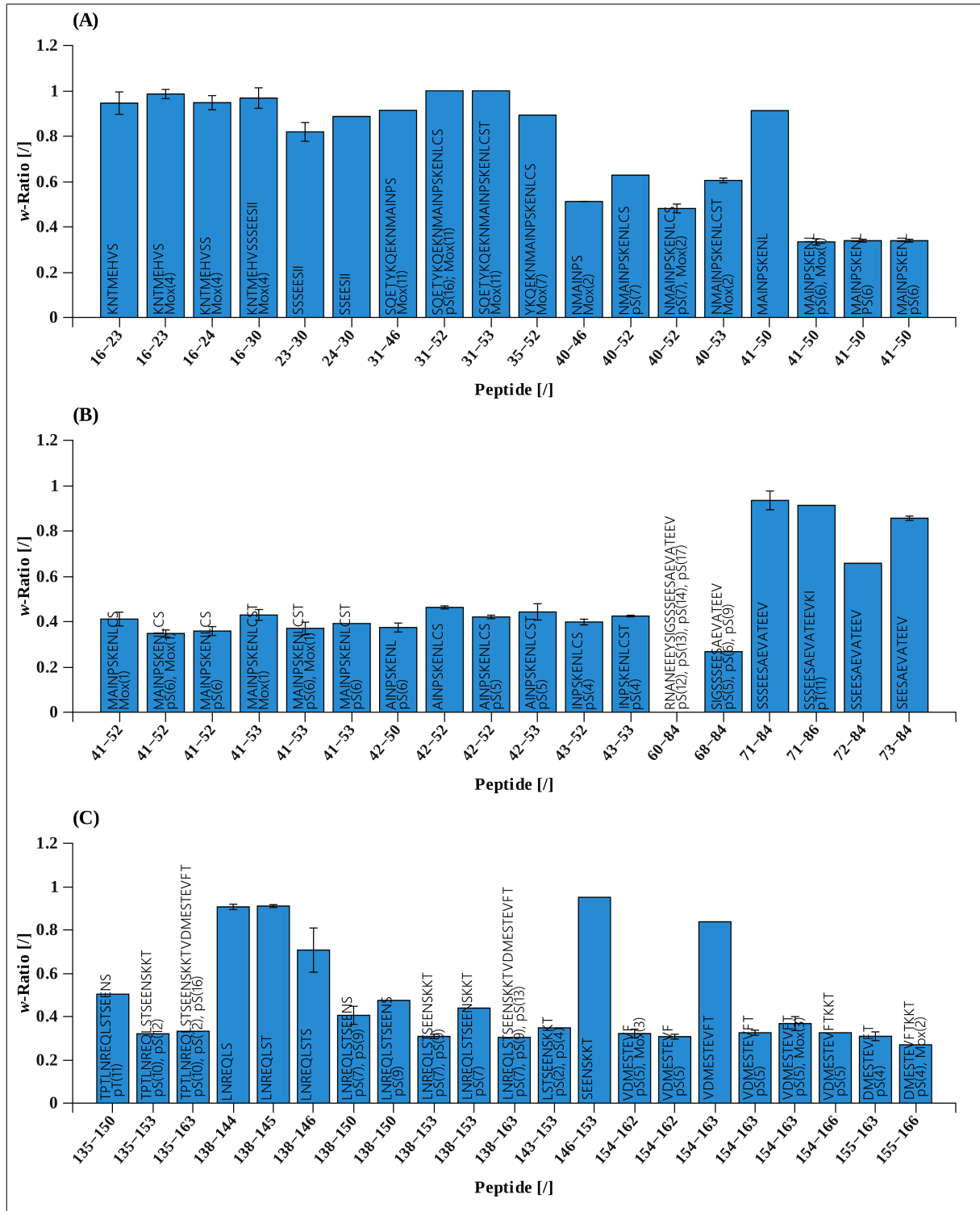


Figure 2-57:  $\alpha$ -S2-casein elastase derived peptides covering a potential phosphorylation site were extracted from the identified PSM list and further split into the related proteases according to their C-terminus and the iTRAQ reporter ion signals. Finally, after re-grouping of unique peptides the median and the standard deviation for each peptide group was calculated. For dephosphorylated peptides, a w-Ratio towards 1 is expected, whereas for the phosphorylated counterpart a w-Ratio of 0.3 should be observed.

## 2 Developing a Platform for relative Quantification using iTRAQ

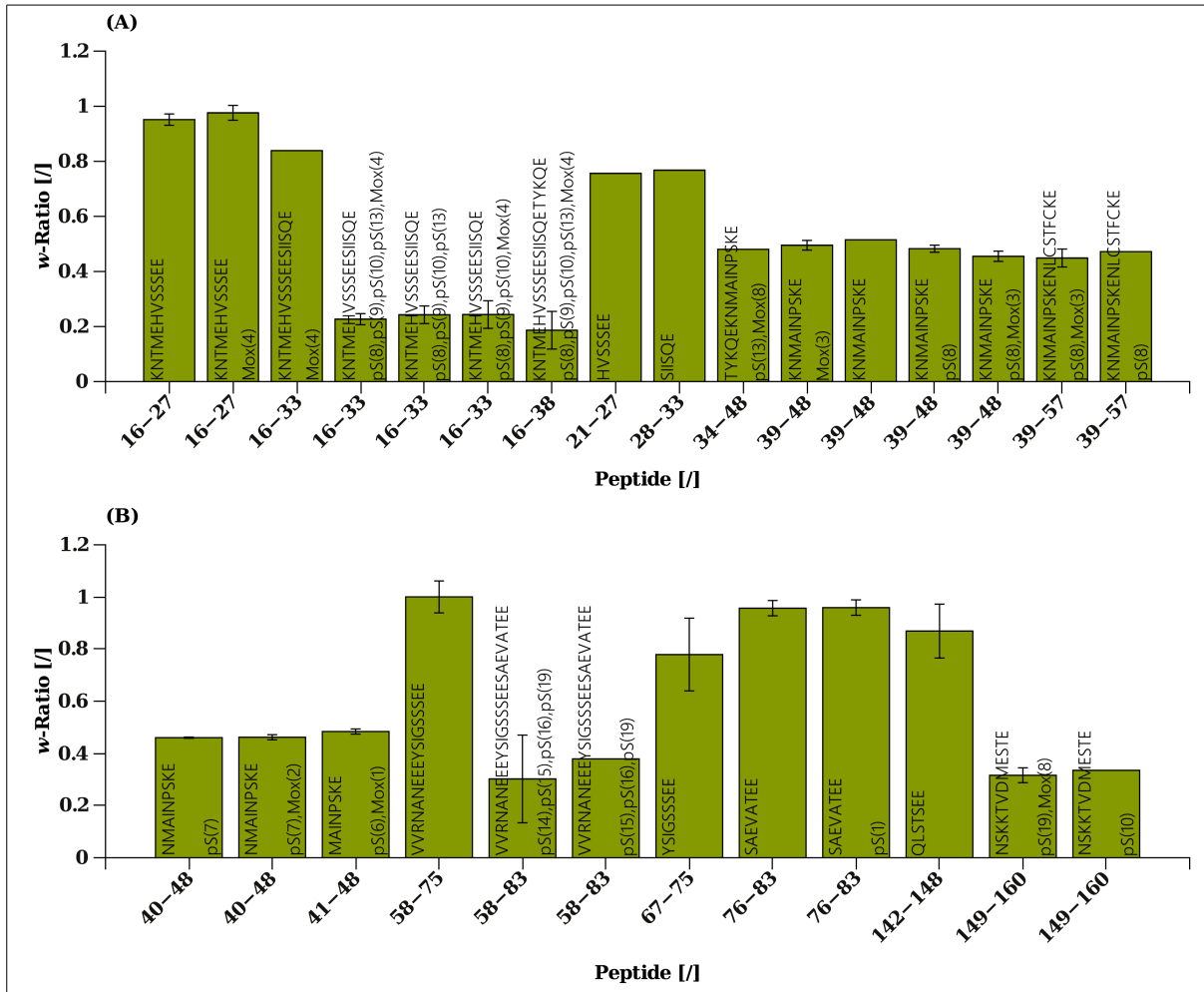


Figure 2-58:  $\alpha$ -S2-casein GluC derived peptides covering a potential phosphorylation site were extracted from the identified PSM list and further split into the related proteases according to their C-terminus and the iTRAQ reporter ions signal. Finally, after re-grouping of unique peptides the median and the standard deviation for each peptide group was calculated. For dephosphorylated peptides, a w-Ratio towards 1 is expected, whereas for the phosphorylated counterpart a w-Ratio of 0.3 should be observed.

## 2.4 MECi – Targeted Phosphoprotein Quantification

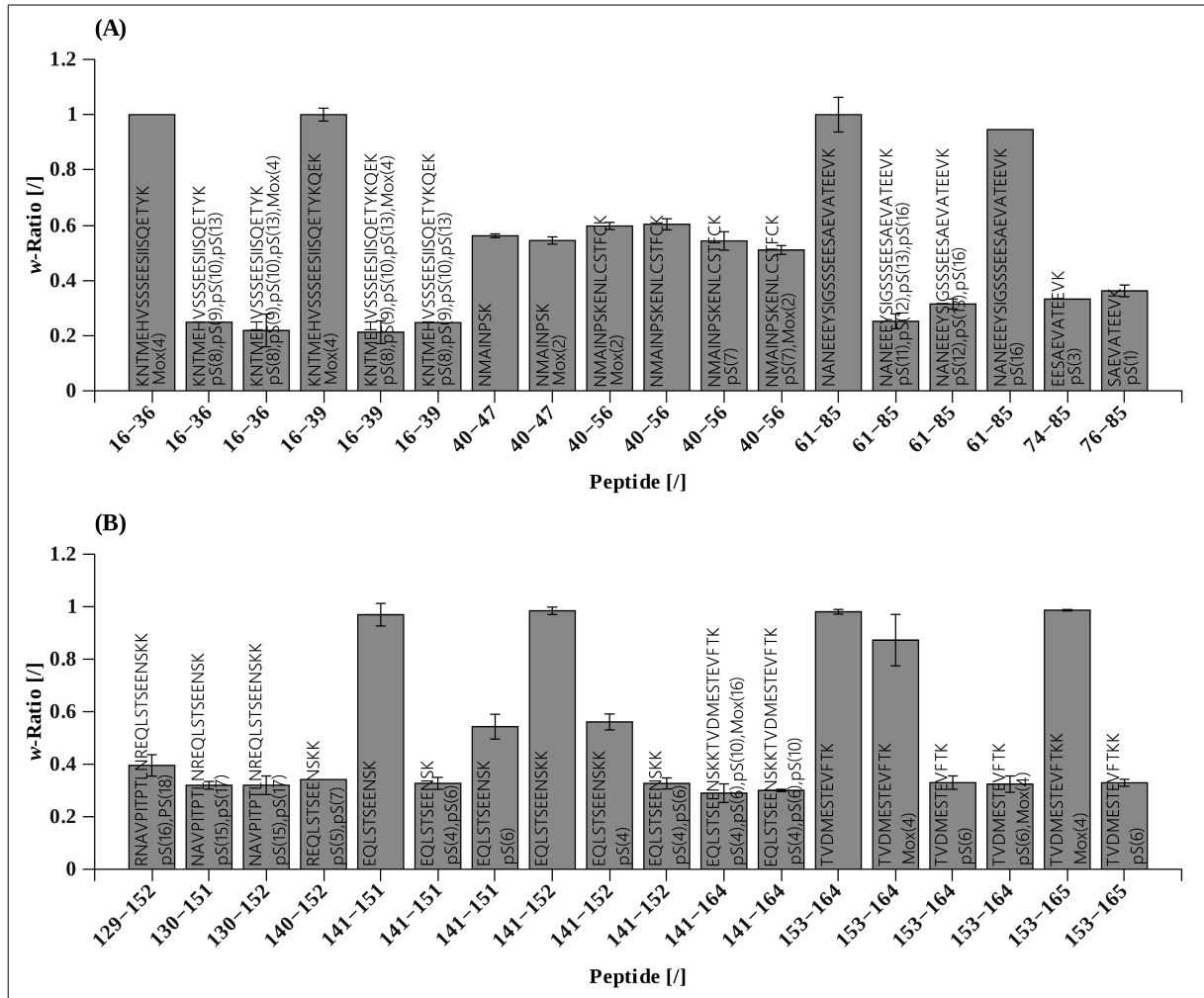


Figure 2-59:  $\alpha$ -S2-casein trypsin derived peptides covering a potential phosphorylation site were extracted from the identified PSM list and further split into the related proteases according to their C-terminus and the iTRAQ reporter ion signals. Finally, after re-grouping of unique peptides the median and the standard deviation for each peptide group was calculated. For dephosphorylated peptides, a w-Ratio towards 1 is expected, whereas for the phosphorylated counterpart a w-Ratio of 0.3 should be observed.

## 2 Developing a Platform for relative Quantification using iTRAQ

### 2.4.4 Conclusion

To accurately identify and quantify phosphorylation sites between two different biological conditions, a multi-protease approach for phosphoprotein analysis published earlier (302) was extended here by the implementation of the possibility for relative quantification.

In this study, iTRAQ-8Plex was not applied in its routinely form to compare eight different states in parallel. In contrast, the experimental setup can be described as an iTRAQ-4x2Plex labeling. This overcomes the requirements to generate eight intense reporter ions for quantification, as commonly only two different peptide species (from one protease) were selected simultaneously for fragmentation. The phosphopeptide enrichment protocol was adjusted to be fully compatible to iTRAQ labeling and phosphopeptide enrichment using TiO<sub>2</sub>. As both biological states and each of the four different digestions were combined, potential sample-handling effects introduced by parallel enrichment were at least reduced. This included the utilization of (an experimentally optimized) Protein:TiO<sub>2</sub> ratio of 1:20 for casein-proteins separated by SDS-PAGE. Using higher amounts of titanium dioxide material increased the robustness of the enrichment procedure. Consequently, not only the number of unique peptide species originated from different protease digestions increased, but also the number of phosphorylated PSMs.

The identification of a numerous complementary non- and phosphorylated peptide species was achieved using a CID-IT/HCD approach, which benefited from optimal measurement settings (see chapter 2.1). For multi-protease studies, identified peptides should be validated using the semi-supervised machine learning code Percolator (chapter 1.2.4), as this improved the number of PSMs for both CID-IT and HCD fragmentation. It is speculated that Percolator benefits from the similar but different peptide sequences, their corresponding retention times in combination to peptide features which are used to dynamically learn to distinguish between false and true positive peptide hits.

By combination of relative quantification with the multi-protease approach, more accurate identification and quantification was achieved. The number of unique phosphopeptides increased significantly, which was particularly true for proteases with low-specificity. These generated a very high number of unique peptides with overlapping sequences. An increased number of unique peptides also assisted in filtering out false positive identifications. In addition, by analyzing all fractions after enrichment, complementary information on the non-phosphorylated peptides was retained. This can be used to normalize quantitative data to compensate any bias introduced during sample processing. Finally, the quantitative multi-protease approach shown here is not limited to the analysis of phosphorylation sites. Principally, any kind of dynamic modification can be analyzed and relatively quantified using this method in a very accurate and precise manner.

## 2.5 Discussion and Outlook

The identification and quantification of iTRAQ labeled (phospho)peptides is crucial to characterize and understand proteome dynamics. To obtain optimal results, process parameters, such as Normalized Collision Energies for peptide fragmentation, have to be evaluated critically. It was the major goal of this thesis, to develop an analytical platform for relative quantification using iTRAQ on an LTQ Orbitrap Velos to analyze comprehensively phosphorylated and non-phosphorylated proteins. In summary, this was successfully realized by combining four different research areas within proteome science; beginning with the evaluation of fragmentation conditions to identify and quantify (phospho)peptides.

### (1<sup>st</sup>) Optimized Fragmentation Conditions for iTRAQ Labeled (Phospho)Peptides

The parallel acquisition of CID-(MSA)-IT and HCD spectra provided an almost optimal trade-off for identification and quantification of iTRAQ labeled (phospho)peptides on an LTQ Orbitrap Velos.

CID-(MSA)-IT fragmentation was both beneficial (fast spectra acquisition due to the ion trap) and detrimental (low mass accuracy and resolution of tandem MS spectra). It has to be highlighted, that phosphopeptide spectra upon CID-IT decomposition were still dominated by signals obtained by neutral losses of phospho-moiety rather than by signals derived by iTRAQ fragmentation. To analyze iTRAQ labeled phosphopeptides, the multistage activation feature of CID should be activated, as the identification and localization of the phosphorylation site was strongly enhanced by this technique, and even outperformed HCD. This was rather unexpected, as high resolution mass spectra upon HCD excitation were claimed to be superior for phosphopeptide identification (181).

CID-IT MS/MS spectra were diminished by the  $1/3^{\text{rd}}$  rule of ion traps by which they were not optimal for iTRAQ quantification. Therefore, CID-(MSA)-IT spectra should exclusively be used for identification of peptides by applying optimized NCE values. In previous studies, HCD spectra were predominantly used for quantification at elevated NCE values to generate intense iTRAQ reporter ions. However, by fine-tuning of fragmentation parameters, an optimal trade-off was evaluated in this study to benefit of HCDs high resolution and mass accuracy spectra for identification and quantification, respectively.

A suggestive dual fragmentation (CID-IT/HCD) measurement cycle on LTQ Orbitrap Velos instruments, starts with the MS prescan at low resolution (7,500), followed by a full scan at high resolution in the Orbitrap (FT-MS). In parallel, with the beginning of the high resolution full scan, CID-(MSA)-IT spectra can already be acquired in the ion trap, taking advantage of its sensitivity and speed of acquisition. After the MS full scan has completed, HCD spectra are recorded in the Orbitrap mass analyzer (see Figure 1-12). Importantly, the application of a dual fragmentation mechanisms does not lead to significantly prolonged cycle times.

However, optimal NCE values were charge-state dependent for HCD fragmentation. The current instrument setup of the LTQ Orbitrap Velos did not allow the application of *on-the-fly* decision-tree approaches, which would greatly increase the performance of HCD. Features for the decision-tree could be: (i) different NCE values in dependence of the precursor mass, (ii) adjustment of NCE values and the number of  $\mu$ -scans in combination

## 2 Developing a Platform for relative Quantification using iTRAQ

with dynamic in- & exclusion lists to improve the signal-to-noise ratios, (iii) variation of the precursor mass isolation windows to minimize *co-isolation* effects, (iv) perform *on-the-fly* identification of peptides by parallelized processing using graphic processing units (GPUs) to exclude repeating fragmentation of the same precursor or (v) selection of the same peptide ion with different charge states for more accurate identification and quantification.

In this study, peptides were *re-identified* across all LC-ESI MS runs for data evaluation by a combination of three features (retention time, precursor m/z and charge state), a process similar to the *re-match between runs* option in MaxQuant. It is noteworthy that this method of data-extraction and data-analysis are unique and have not been performed in any other iTRAQ evaluation study published (177,255,273,274,277,317,318).

By applying this re-mapping approach, data points from elevated Normalized Collision Energies (> 65%), such as the iTRAQ reporter ions intensity, were also included in data-analysis and re-mapped to a specific peptide. As search engine scores were suppressed at higher NCE values due to the limited number of significant b- and y-ions alongside intense (and non-sequence informative) iTRAQ reporter ion signals (211), these data were not reallocated to a specific peptide species in previous reports published. In this study, by combining the precursors m/z value, its retention time and charge state, these data could unambiguously be reassigned to a specific peptide. Although the focus was set to optimize collision energies for any kind of iTRAQ labeled peptides, the data of each (phospho)peptide can also be used to identify sequence specific motifs which might help to predict peptide fragmentation patterns more precisely.

The computational proteomics tools used for process optimization in this study, are planned to be published under the GNU General Public License. This allows for studying, sharing and modifying the software or directly re-using the implemented tools to perform optimization processes for MS instruments before beginning with large scale (phospho)proteomics studies. Beside that, the .raw files of this study should be made public to the proteomics community. Open-source initiatives, such as X!Hunter from The Global Proteome Machine Organization, publish spectra libraries for reliable peptide identification. For this purpose, information about the shape of tandem MS spectra from a particular peptide are normalized via several and independent instruments. These reference spectra could be improved further by the incorporation of spectra-series, which were recorded at various NCE values.

### (2<sup>nd</sup>) Refinement of Precursor m/z Calculation of iTRAQ Labeled Peptides

Przybylski et al. (274) described the discrepancy to accurately calculate the precursor mass (m/z value) of iTRAQ-4Plex labeled peptides; which resulted in a mass shift of up to 3.8 ppm for peptides labeled exclusively with iTRAQ<sub>114</sub>. Przybylski et al. (274) advised to increase the precursor mass tolerance window from 5 to 10 ppm in database search settings.

The effect of inaccurate precursor mass calculation was analyzed systematically in this thesis. It was demonstrated that the mass accuracy was depending on the number of iTRAQ modifications per peptide. This directly affected database search results as different minimal precursor mass tolerance windows were required to obtain comparable protein identification results.

## 2.5 Discussion and Outlook

To circumvent this, a newly post-database search algorithm was compiled which used the iTRAQ reporter ion intensities of tandem MS spectra to calculate the precursor mass ( $m/z$  value) more precisely. This was an important advantage, as the precursor mass of exclusively iTRAQ<sub>114</sub> or iTRAQ<sub>115</sub> labeled peptides could be corrected from 4.3 to 0.0 ppm.

The next step is to integrate this algorithm into database search suites. This can for example be performed by creating four peptide precursor lists during peptide indexing. Three of them are used for iTRAQ<sub>114</sub>, iTRAQ<sub>115</sub> and iTRAQ<sub>116&117</sub> peptide mass calculation; while the last one is a linear combination of the observed iTRAQ intensities in each tandem MS spectrum, multiplied by the different peptide masses from the remaining lists. Although this process has to be performed for every single MS/MS spectrum, hence it will increase the search time, the results and accuracy of precursor mass calculations after database search will improve.

### **(3<sup>rd</sup>) Improving the Identification of iTRAQ Tyrosine Labeled Peptides**

To further improve database search results obtained for iTRAQ labeled peptides, a new immonium ion for iTRAQ labeled tyrosine peptides was discovered for both iTRAQ-4Plex and iTRAQ-8Plex upon beam-type CID fragmentation. This modification influenced the tandem MS fragmentation pattern and, additionally the retention time in reversed phase chromatography.

Both features can be used, amongst others, in semi-supervised machine learning codes to further improve the validation process of peptides by applying more than one peptide feature, as it is done for FDR calculation.

### **(4<sup>th</sup>) Quantitative Phosphoproteomics Using a Multi-Protease-Approach**

Finally, an analytical workflow for relative quantification of phosphoproteins was established. It combined the benefit of increasing sequence coverage and the likelihood to localize the exact site of phosphorylation by applying a multi-protease approach (302) with subsequent iTRAQ labeling for relative quantification, with a selective choice of buffers and chemicals to ensure the compatibility with SDS-PAGE and phosphopeptide enrichment using TiO<sub>2</sub>.

The number of unique and quantifiable (phospho)peptides increased due to the combination of two low and two high specific proteases and the higher sample amount of iTRAQ labeled peptides which can be used for LC MS analysis. Increasing the sample amount does not overload the analytical column, as peptides exhibit a wide range of physiochemical properties (size, charge, hydrophobicity) which lead to large differences (ionization efficiency and retention time) in LC-ESI MS experiments.

To apply the newly implemented strategy to whole cell lysate, peptide fractionation is a prerequisite to simplify the complexity of the sample. This can, for example, be a combination of a 3D-LC approaches with an additional phosphopeptide enrichment (e.g., TiO<sub>2</sub>) step (i) to increase the number of identified and quantifiable peptides, (ii) validate the localized site of phosphorylation by more than one spectra and one phosphopeptide-sequence and (iii) to receive a higher number of (unique) peptides per protein to improve the accuracy of protein quantification.

Further, it is suggested to use the data of the quantitative multi-protease approach to measure the site-

## **2 Developing a Platform for relative Quantification using iTRAQ**

specific degree of phosphorylations. A recently published work by Steen et al. (319) addressed this issue by incorporation of a *flyability* factor to compare the intensities of phosphorylated and the non-phosphorylated counterparts. As their strategy was build on a label-free approach, the XICs of the non- and phosphorylated peptides were *directly* available. For the relative quantification strategy using iTRAQ, these XICs are only *in-directly* available, as the precursor ions are isobaric and appear as a single signal in MS scan (see Figure 1-4). However, the content of each XIC from each biological entity is directly proportional to the iTRAQ reporter ion distribution in the tandem MS spectra, and the partial XIC for every biological entity can be calculated. To finally establish and examine this technique, further LC MS experiments are required.

### **Final Statement**

The development of an analytical platform for relative quantification using iTRAQ on an LTQ Orbitrap Velos to analyze comprehensively phosphorylated and non-phosphorylated proteins was the major goal of this dissertation. This task was successfully accomplished and the newly established and developed methods can be used directly to analyze proteins from biological material. This can be biological samples derived from whole cell lysate to identify dis-regulated signaling pathways by relative quantification using iTRAQ under disease states (e.g., cancer or tuberculosis) to elevate the feasibility to discover new therapeutic treatment options.



## 3 Development of Computational Proteomics Pipelines for Automated Data Processing

### 3.1 Paleoproteomics – Studying the Iceman's Brain Tissue<sup>2</sup>

The ice mummy *Oetzi* was discovered in 1991 by two German mountaineers in an Italian alpine glacier. By freeze drying, the iceman's corpus was well preserved, making it a scientifically and historically interesting research project (320,321). Genome sequencing of this mummy has allowed researches to get an idea about his phenotype and origin (322). The analysis of lipids and proteins on the other hand, which are preserved better than DNA (323), might uncover new scientific breakthroughs, such as the reason of his death.

It is known from gel-based paleoproteomics studies on mummy tissues that degradation processes of fibrous (collagen) and globular proteins into smaller peptides occurs (324,325). These degradation processes can be caused by a variety of sources during and/ or after the mummification process. Rapid pH drops, post mortem formation of reactive oxygen species (ROS) in addition to intracellular and microbial enzymes are some examples of processes that induce protein denaturation and degradation (326–328). Additionally, environmental factors play an important role: e.g., temperature changes between seasons, which can increase the degradation processes; facts limiting the degradation process and support protein preservation are rapid desiccation, as it was the case for the iceman (329).

Studies published so far have focused on collagen in ancient material. This protein complex seems to be well preserved and could even be identified in dinosaur fossils. Additionally, other peptidic remains have been identified in grape seeds, archaeological potsherd, and in a Neanderthal bone (330–332). The most successful of these studies in terms of protein identifications has been published after analyzing a 43,000 years old mammoth bone samples (333).

The focus of this paleoproteomics project is to identify proteins from two soft tissue samples of the iceman *Oetzi*. The cooperative work between the first authors (F. Maixner, T. Overath and D. Linke) is shared equally into biological, experimental (sample preparation and mass spectrometry) and computational proteomics sections. The development and establishment of a workflow to combine results from multi-database searches is the major aim this chapters (3.1.1 & 3.1.2). Further, the performance of each database search engine is evaluated. Additionally, the use of such a strategy to accurately identify proteins for a 5,300 years old wet-mummy, the Tyrolean iceman *Oetzi*, will be highlighted.

---

2 Parts of the chapter 3.1 were published in Cell. Mol. Life Sci. (2013) 70:3709–3722

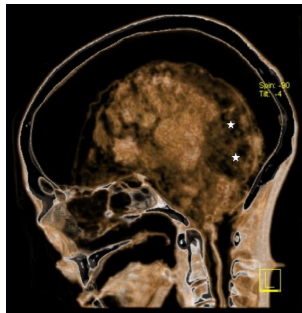
## 3 Development of Computational Proteomics Pipelines for Automated Data Processing

### 3.1.1 Material and Methods

Chemical and instrumental description can be found in the chapter 4.1.

#### 3.1.1.1 Sample Preparation

Two tissue biopsies (denoted as 1024 and 1025 in the following) from the dark area of the occipital lobe of the iceman *Oetzi* were extracted (Figure 3-1). Briefly summarized, a two step extraction process was applied for both tissue samples, which is illustrated in Figure 3-2 (A). After protein extraction, samples of strategy A (1024-A, 1025-A) were separated by SDS-PAGE (Figure 3-2 (B)), whereas the supernatant and cell derbies were subjected to a second extraction process with subsequent in-solution digestion. A more detailed description can be found in (310).



*Figure 3-1: “Transverse CT section through the skull shows an irregular area (actual sampling site) of increased radiographic transparency in the posterior cerebral regions (asterisk). The meninges have become detached from the skull vault and surround the shrunken, inhomogenously disintegrated brain” (310).*

### 3.1 Paleoproteomics – Studying the Iceman's Brain Tissue

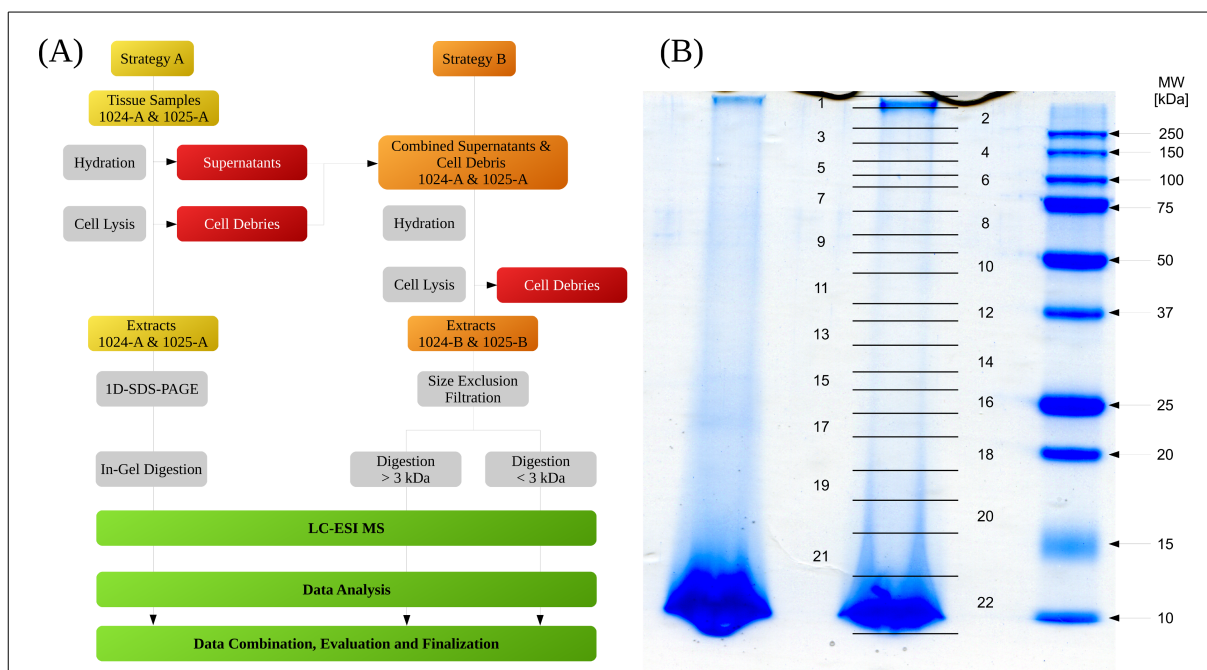


Figure 3-2: (A) Sample preparation workflow applied to process both, sample 1024 and 1025, for protein extraction, digestion and sample separation. All samples were analyzed by online LC-ESI MS, with a subsequent multi-dimensional database search. Finally protein identifications were combined for biological evaluation of proteomics data. (B) 1D-SDS-PAGE of two protein extracts from the Iceman's brain tissue samples. Both samples (approx. 22  $\mu\text{g}$  for sample 1024-A lane 1 and approx. 11  $\mu\text{g}$  for sample 1025-A lane 2) were separated on a 12% SDS-PAGE and colloidal coomassie stained. (B left side) shows the gel without excision labels; (B right side) shows the gel slices excised for tryptic digestion.

#### 3.1.1.2 Mass Spectrometry

Samples obtained from in-gel and in-solution digest (Figure 3-2) were reconstituted in 12  $\mu\text{L}$  loading buffer (0.1% (v/v) TFA, 3% (v/v) ACN in ddH<sub>2</sub>O) and analyzed by LC-ESI MS using reversed phase liquid chromatography coupled online to a LTQ Orbitrap Velos (see chapter 2 and chapter 4.1 for additional information). For each sample injection, 10  $\mu\text{L}$  of redissolved peptides were loaded and desalted for 6 min (flow rate: 30  $\mu\text{L}/\text{min}$ ). Peptides were separated at a flow rate of 300 nL/min using eluent A (0.05% (v/v) formic acid (FA)) and eluent B (0.1 FA, 80% ACN). The gradient, which was used to separate peptides, varied: for the in-gel digested samples (band 1-20) and in-solution digested samples, a 110 min linear gradient from 0 to 50% B was used. For the gel band 21 & 22, the gradient was prolonged to 150 min (linear increase from 0 to 50% B).

XCalibur (Version 2.1) was used to create a TOP 15 method for the LTQ Orbitrap Velos. Full scans were required in the Orbitrap mass analyzer ( $m/z$  300 – 1600) at a resolution of 60,000. The 15 most intense ions (signal intensity  $\geq 500$ ; charge state  $> 1^+$ ) were selected for CID-IT fragmentation acquired in the ion trap (IT). CID-IT settings were set to default (NCE: 35%, activation Q: 25; activation time: 10 ms). Dynamic exclusion was enabled (isolation width  $\pm 10$  ppm; exclusion duration: 60 sec, maximum list size: 500).

### 3 Development of Computational Proteomics Pipelines for Automated Data Processing

#### 3.1.1.3 Database Searches

Recorded .raw files were sliced and re-calibrated prior to the main-database search using the Re-Call Offline feature of XCalibur. Preliminary database searches were performed using SEQUEST (settings see below) to evaluate the highest and lowest retention times observed (FDR < 10%) used for raw file slicing. The polysiloxane signal at m/z 445.120026 was used for re-calibration of MS data.

The main database evaluation was performed using four different search engines: SEQUEST (implemented in Proteome Discoverer v.1.2), Mascot (Matrix Science, v.2.2.04), OMSSA (Open Mass Spectrometry Search Algorithm, v.2.1.9) and X!Tandem (v. CYCLONE 2010.12.01). SEQUEST and Mascot database searches were performed using the software Proteome Discoverer (v.1.2), whereas OMSSA and X!Tandem were part of the cross-platform Java based programs SearchGUI (v.1.7.3) and PeptideShaker (v.0.16.2). The latter two used .mgf as input files. These were generated by the spectrum exporter node (default settings) within Proteome Discoverer 1.2.

All settings (e.g., ppm window, modifications added) were *in-house* standardized for this project and used for all four database searches. Due to spectra re-calibration, precursors ppm tolerance window was set to 5 ppm. For the size exclusion flow-through samples, the precursor ppm window was lowered to 3 ppm. Fragment ions in CID-IT MS/MS spectra were matched with a tolerance of 0.5 Da. Two missed cleavages for trypsin digestion were allowed. Carbamidomethylation of cysteine was set as a fixed modification, while oxidation of methionine was set as a variable modification. Additional database searches were performed including deamidation of N and Q as variable modification (only used in chapter 3.1.2.2, *Deamidation of Proteins due to Protein Aging Processes*).

Peptides were classified as high confident with an FDR < 1% and as medium confident with 1% < FDR < 5%. Proteins were reported as identified if within one search engine approach at least 2 high or 3 medium peptides were reported. Additionally, database searches without protease specification (*no-enzyme*) were performed to allow for the identifications of peptides derived from non-tryptic protein degradation.

#### 3.1.2 Results and Discussion

The identification of proteins obtained from an ice-mummy is a challenging task. Not only does the degradation of proteins over years (5,300 years for the iceman *Oetzi*) hinder the application of classical experimental proteomics approaches (e.g., sample extraction and SDS-PAGE separation), but it also forces computational proteomics analysis workflows to be adopted against this (e.g., selection of the *correct* protease specificity). To incorporate these obstacles, a highly specified data analysis has to be performed.

Therefore, a computational workflow was developed and established which (i) combined four different search engines, (ii) analyzed the degree of non-tryptic peptides, (iii) identified (potential degraded) proteins and (iv) estimated the degree of protein deamidation; an either artificial or biological induced dynamic modification.

In the following chapter 3.1.2.1, the benefit of using a two-stage extraction process will be described. Further, the overlap of protein identification by combining four search engines is presented, and the performance of each search algorithm to identify proteins is evaluated.

In chapter 3.1.2.2, protein aging processes (e.g., truncated proteins) will analyzed and presented. Although the main aim of the project was to develop a computational proteomics pipeline, the implementation of the developed tools is not described, instead biological meaningful results are presented to demonstrate the requirement of developing such a highly specialized data evaluation pipeline.

##### 3.1.2.1 Protein Identification Using a Multi-Database Search Engine Approach

###### Overlap in Protein Identification for Samples 1024 & 1025

Combining the identified proteins from both samples, 1024 and 1025, as well for the in-gel and in-solution digestion methods, a total of 502 proteins were identified. From these, 98% were identified within sample 1024 while only 44% were identified within sample 1025. In calculating the overlap of identified proteins (Figure 3-3) it was appeared that sample 1025 represented a subset of sample 1024 with only 14 additional and unique identified proteins. The lower number of identified proteins for sample 1025 was not unexpected due to lower amount of starting material available (6.5 mg for 1024-A, 3.5 mg for 1025-A). This also resulted in a decreased number of PSMs (approx. 23% less) for sample 1025-A compared to 1024-A.

Analyzing the number of proteins identified by the different experimental setups applied (in-gel versus in-solution digest), the highest number (and percentage in relation to the total number of identified proteins) was observed for the in-gel digestion method. Here, 38% (187/488) and 51% (113/222) of the total identified proteins were identified for sample 1024-A and 1025-A, respectively (Figure 3-3 (B)). The overlap of identified proteins between in-gel and in-solution digestion approach was 42% (207/488) for sample 1024 and 43% (95/222) for sample 1025, respectively. For the in-solution digest of 1025, the number of identified proteins decreased significantly (6% of total proteins identified (14/222)). For sample 1024, 19% (94/488) of the proteins were only identified after in-solution digestion. Although the effect of a combined two-step extraction process (first in-gel, second in-solution) was limited for sample 1025, the information about the doubly identified proteins is of high importance. Being independently searched and validated, these proteins represented a set of repeatedly identified

### 3 Development of Computational Proteomics Pipelines for Automated Data Processing

proteins, even though only every second protein was identified in both experimental setups.

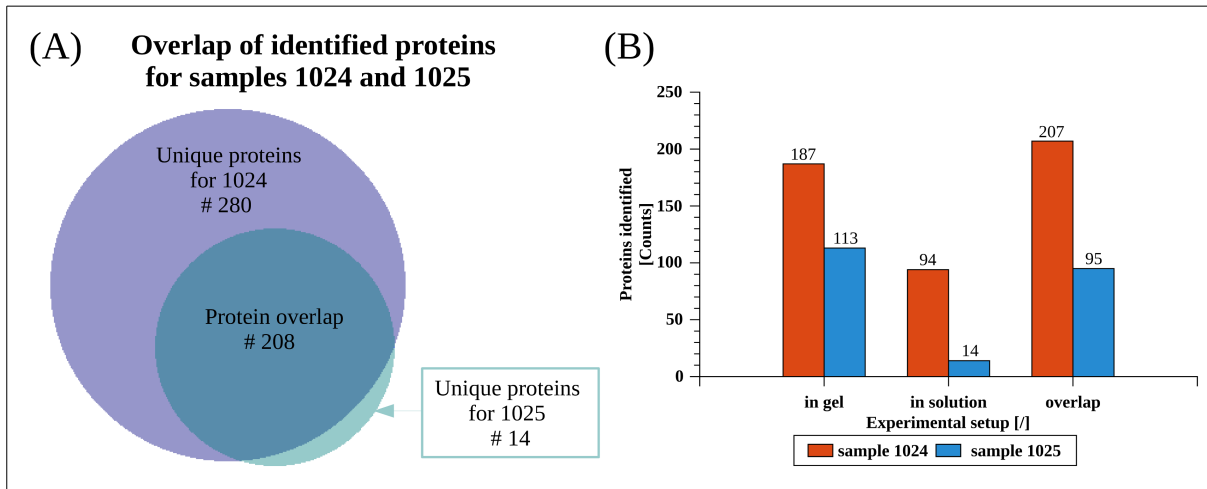


Figure 3-3: Overlap of identified proteins from sample 1024 and 1025 (A), and the number of proteins identified for the in-gel & in-solution digested sample. (B) Overlap between in-gel and in-solution for both samples.

#### Analyzing the Effect of a Multi-Database Search

The protein identifications presented so far were based on the application and combination of four different search algorithms. Proteins were categorized as confidential identified if at least two high or three medium confident peptides could be mapped back to the protein within one search. To further analysis this data, the proteins were grouped according to number of search engines by which the protein was identified (Table 3-1).

It revealed, that by combining and averaging the percentage distributions of all samples from Table 3-1, about 27.4% of the proteins were identified with four search engines, 12.6% with three, 25.6% with two and only 34.4% of the proteins were identified with only one search engine.

These distributions were reflected by the majority of the individual samples in Table 3-1. However, the *no-enzyme* searches of samples 1024-B, 1025-A and 1025-B showed a high number of proteins identified (45 – 58%) by only one search engine, whereas the five remaining database searches followed the overall trend.

Table 3-1: Number of proteins identified for the in-gel and in-solution digested samples 1024 and 1025 with trypsin (Tr) and without enzyme specification (*no-enzyme*, NoE). Additionally, the percentage distribution of the search engine hits for each sample is shown (% of total Proteins).

Sample	1 Search Engine		2 Search Engines		3 Search Engines		4 Search Engines		
	# Proteins	% of total Proteins	# Proteins	% of total Proteins	# Proteins	% of total Proteins	# Proteins	% of total Proteins	
<b>In-Gel</b>	1024 Tr	84	24.3	73	21.2	51	14.8	137	39.7
	1024 NoE	82	23.3	120	34	69	19.5	82	23.2
	1025 Tr	51	26.3	47	24.2	26	13.4	70	36.1
	1025 NoE	75	50	30	20	16	10.7	29	19.3
<b>In-Solution</b>	1024 Tr	60	20.3	79	26.8	37	12.5	119	40.3
	1024 NoE	114	45.8	54	21.7	34	13.7	47	18.9
	1025 Tr	28	26.9	39	37.5	13	12.5	24	23.1
	1025 NoE	44	57.9	15	19.7	3	3.9	14	18.4
<b>Median</b>	-	34.4	-	26.6	-	12.6	-	27.4	

### 3.1 Paleoproteomics – Studying the Iceman's Brain Tissue

By eliminating duplicates entries and merging the identified proteins from the different approaches (in-gel, in-solution; tryptic and *no-enzyme*), 502 proteins were identified as illustrated in Figure 3-3. Of these proteins, more than 82% were identified with at least two search engines (62% with four, 12% with three and 13% with two). Only 13% were identified by one of the four search engines, which clearly showed the advantage of using multiple search algorithms. Not only the number of protein identifications raised, but also the confidence of identifications increases, as each search engine applies its own algorithm to identify Peptide-Spectrum-Matches.

#### Analyzing the Performance of Each Search Engine

The performance of each search engine to identify proteins was critically evaluated using data sets of the four in-gel digested samples (1024-A & 1025-A, with trypsin and without enzyme specificity). It revealed, that X!Tandem identified the highest number of proteins (264 proteins), followed by Mascot (258 proteins), SEQUEST (206 proteins) and OMSSA (203 proteins). The corresponding distribution is illustrated in Figure 3-4 (A).

Further, it was analyzed which combination of three search engines was optimal to obtain a high number of protein identifications. It revealed, that the overlap was highest for Mascot, OMSSA and X!Tandem (referred to as *MAOMXT*) followed by SEQUEST, Mascot and X!Tandem (*SEMAXT*; as indicated in Figure 3-4 (B)). This result was expected, as X!Tandem and Mascot identified most of the proteins. For the combination of two search engines, solely the overlap between SEQUEST and Mascot was significant (*SEMA*). The remaining did not show clear trends. The results of the three remaining samples showed similar distributions, which are summarized in Table 3-2.

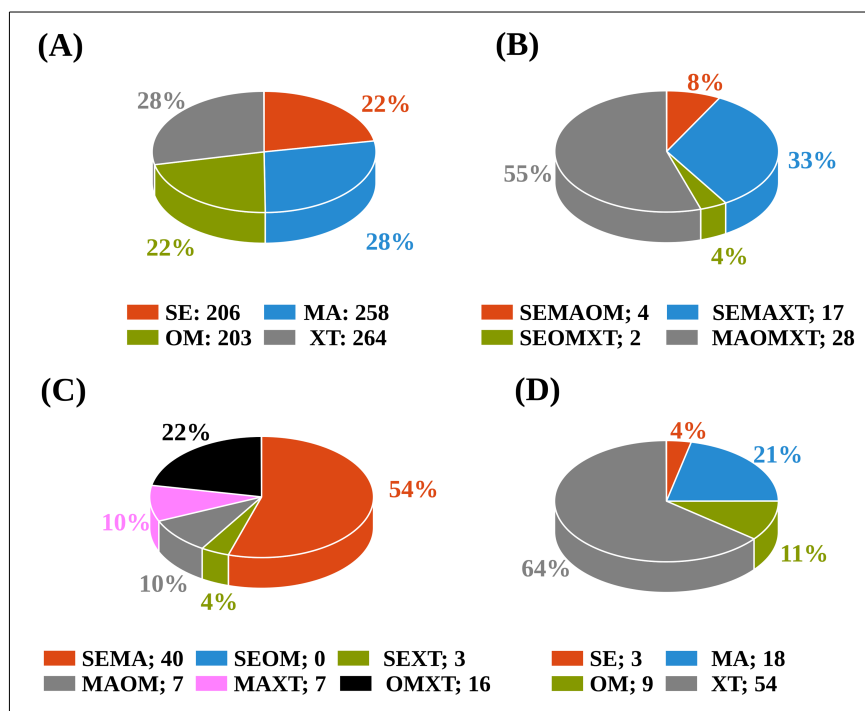


Figure 3-4: The number of proteins identified for sample 1024-A with trypsin specificity: (A) the total number of proteins by each search engine; (B) the overlap of three search engines; (C) the overlap of two search engines; (D) proteins identified with only one search engines. SE: SEQUEST, MA: Mascot, OM: OMSSA, XT: X!Tandem.

### 3 Development of Computational Proteomics Pipelines for Automated Data Processing

Table 3-2: Number of proteins identified for the samples 1024-A and 1025-A with overlapping search engine identifications (4 – 1 Search Engines) and the number of identified proteins by each search engines (last row). SE: SEQUEST, MA: Mascot, OM: OMSSA, XT: X!Tandem.

Overlap	Search Engines	1024		1025	
		Trypsin	No-Enzyme	Trypsin	No-Enzyme
4 Search Engines	SEMAOMXT	137	82	70	29
	SEMAOM	4	14	2	2
3 Search Engines	SEMAXT	17	46	8	7
	SEOMXT	2	5	0	0
	MAOMXT	28	4	16	7
	SEMA	40	69	25	7
2 Search Engines	SEOM	0	1	0	0
	SEXT	3	12	4	1
	MAOM	7	7	5	1
	MAXT	7	18	13	7
	OMXT	16	13	0	14
	SE	3	1	0	0
1 Search Engines	MA	18	7	38	14
	OM	9	4	0	0
	XT	54	70	13	61
	SE	206	230	109	46
Total Identifications by each Search Engine	MA	258	247	177	74
	OM	203	130	93	53
	XT	264	250	124	126
	SE	206	230	109	46

#### 3.1.2.2 Protein Aging Processes

Mummies frozen in ice for several thousands of years were exposed to environmental and microbial factors which might lead to protein degradation or the introduction of modifications, such as deamidation. Degradation and modification of proteins hampers the identification by database searches and to take these into account, database searches with and without the specification of an enzyme or the incorporation of deamidation of asparagine and glutamine as variable modifications were performed.

At this point, it is noteworthy that the .raw files of the in-gel digested samples were not searched in a MudPIT approach but in batch mode, that means every .raw file was searched separately. This helped to eliminate protein identifications caused by random false positive hits. These hits were more likely to be observed for protein identifications by combining PSMs from multiple gel slices. Moreover, the *no-enzyme* search offered the chance to identify truncated proteins.



### Truncated Proteins

Testing for *no-enzyme* specificities was an important aspect, as protein degradation and denaturation processes were previously reported for mummy tissues (327). Therefore, the overlap of identified proteins with and without enzyme specificity will be compared to evaluate how many proteins might partially be truncated.

For sample 1024-A (in-gel), the overlap of identified proteins with tryptic and *no-enzyme* specificity was high (78%); the number of unique proteins for the tryptic (10%) and *no-enzyme* (12%) database search were similar (Figure 3-5 (A)). The overlap of identified proteins (66%) was significantly lower for the in-gel digested sample 1025-A (Figure 3-5 (C)). Here, the number of proteins identified with tryptic specificity was high (28%), whereas only a few additional protein were identified with no enzyme specificity (6% *no-enzyme*).

For the in-solution digested samples (Figure 3-5 (B) & (D)), only 2% and 3% unique proteins (1024-B and 1025-B, respectively) were identified using a *no-enzyme* specificity. The relatively high number of identified proteins with a *no-enzyme* approach for the in-gel samples might be an indicator for protein degradation over 5,300 years. The higher number of *no-enzyme* identified proteins for the in-gel sample compared to the in-solution digested sample can be explained; it is more likely to extract truncated proteins in the first (in-gel) rather than in the second extraction process (in-solution).

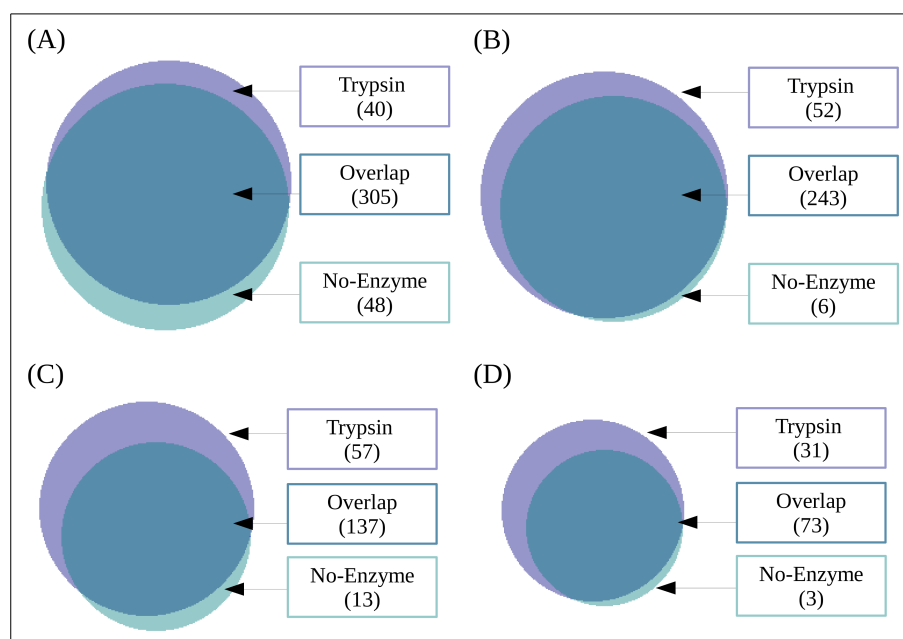


Figure 3-5: Overlap of identified proteins with (trypsin) and without enzyme specificity for the two samples (1024 & 1025) digested in-gel and in-solution. (A) 1024-A in-gel, (B) 1024-B in-solution, (C) 1025-A in-gel, (D) 1025-B in-solution.

### 3 Development of Computational Proteomics Pipelines for Automated Data Processing

#### **Detection of Truncated Proteins**

Protein degradation can lead to the presence of a variety of different protein forms, which might only be identified by an unspecific (*no-enzyme*) database search. The computational proteomics pipeline developed for this project also considered truncated protein forms; by returning a boolean value if the protein was or was not identified within a specific gel band.

Various proteins were identified in more than one band. One of these, *synaptosomal-associated protein 25* (SNAP-25), was identified in 7 gel bands. SNAP-25, which has a mass of 23,315 Da, was expected to be identified in gel bands #14 – #16 (as determined by molecular weight marker; Figure 3-2 (B)). Based on the number of PSMs for each gel band, the protein was most abundant in bands #18 – #21, which might be an indicator for protein degradation. A similar observation was made for the protein dermicin (11,284 Da), for which peptides were randomly identified across 10 of the 22 bands.

Although these proteins might contribute to the observed smearing effect after protein staining (Figure 3-2 (B)), other proteins, e.g., dermokine, were confidently (all four search engines) identified in only a single gel band.

An other interesting candidate represented copine-5 (65,734 Da). This protein was only identified in the bands #20 – #22, which might be an indicator for protein degradation.

#### **Proteolipid Protein – a Truncated Protein**

Another protein, which might have undergone degradation is the myelin proteolipid protein (PLP, 30,077 kDa). Peptides of this protein were identified in a multiple of gel bands. These are summarized in Table 3-3 for the gel bands #1 – #19, and Table 3-4 for the gel bands #20 – #22.

As it was present in all but three gel bands, this protein might have been undergone degradation. To verify this, each peptide sequences identified in relation to the .raw file was used to map back the position of the first peptide amino acid within the protein. Therefore, the amino acid N-terminal to the cleavage site was extracted from the FASTA file, and grouped into *non-tryptic* (no R or K), *semi-tryptic* (one R or K) or *fully tryptic* (two R or K) peptides based on the amino acids N-terminal to the cleavage site (extracted from the FASTA file) and the C-Terminal amino acid of the identified peptide sequence, respectively.

In all gel bands, none of the peptides was identified as non-tryptic. However, for the gel bands #1 – #19, 32 semi and 39 fully tryptic peptides were found. This was rather unexpected as trypsin is known to be a very specific protease (67).

In the gel bands #20 – #22, the number of semi-tryptic peptides was significantly higher with 48 semi- and 22 fully tryptic peptides. In addition, sequence latter such as RMYGVXPWNAFPGK, MYGVXPWN, MYGVXPWNAF, MYGVXPWNAFPG, MYGVXPWNAFPGK and MYGVXPWNAFPGKV were observed in band #22.

Unfortunately, these identified peptides did not automatically prove a degradation effect, e.g., by

### 3.1 Paleoproteomics – Studying the Iceman's Brain Tissue

exopeptidases. But, the intense smearing, a similar effect for other proteins (e.g., cytochrom c subunit 2; bands #15 – #21: 90% tryptic cleavage sites; band #22: 5 of 10 identified peptides with non-tryptic cleavage sites) were a good hint for environmental, bacterial or enzymatic protein degradation, as it was expected for 5,300 years old tissue samples.

#### **Deamidation of Proteins Due to Protein Aging Processes**

The deamidation of asparagine and glutamine is a modification process, which can be induced during either sample preparation or through biological processes, and is known to occur in protein aging (334). To identify this effect, additional database searches were performed in a MudPIT approach for both, the in-gel and in-solution digested samples. Here, peptides identified from all gel bands were merged together for peptide validation. At the peptide level, 20% of the peptides identified for sample 1024-A (in-gel) and 16% in 1025-A (in-gel) showed deamidation reactions. For the in-solution digested samples, about 28% deamidated peptides were identified. However, the overall number of identified proteins did not change after taking this modification into account.

### 3 Development of Computational Proteomics Pipelines for Automated Data Processing

Table 3-3: Peptides identified from myelin proteolipid protein (PLP). The amino acid in P1 and the C-terminal amino acid were extracted to allow for fully-, semi-, no tryptic cleavage sites, respectively, via the different gel bands (#1 – #19).

P1	Peptide Sequence	C-Term	Gel Band	P1	Peptide Sequence	C-Term	Gel Band
M	GXXECCA	R	1	K	GXSATVTGGQ	K	14
K	GXSATVTGGQ	K	1	K	TSASXGXCADA	R	14
K	TSASXGXCADA	R	1	T	SASXGXCADA	R	14
T	SASXGXCADA	R	1	K	GXSATVTGGQ	K	15
A	SXGXCADA	R	1	T	SASXGXCADA	R	15
R	MYGVXPWNAFPG	K	1	R	MYGVXPWNAFPG	K	15
G	SNXXSXC	K	1	K	GXSATVTGGQ	K	16
K	XXETYFS	K	2	K	TSASXGXCADA	R	16
K	GXSATVTGGQ	K	2	T	SASXGXCADA	R	16
K	TSASXGXCADA	R	2	R	MYGVXPWNAFPG	K	16
T	SASXGXCADA	R	2	K	GXSATVTGGQ	K	17
A	SXGXCADA	R	2	T	SASXGXCADA	R	17
K	GXSATVTGGQ	K	3	A	SXGXCADA	R	17
K	TSASXGXCADA	R	3	M	GXXECCA	R	18
T	SASXGXCADA	R	3	K	XXETYFS	K	18
R	MYGVXPWNAFPG	K	3	K	XXETYFSK	N	18
K	TSASXGXCADA	R	4	R	QXFGDY	K	18
T	SASXGXCADA	R	4	R	QXFGDYK	T	18
A	SXGXCADA	R	4	K	GXSATVTGGQ	K	18
K	TSASXGXCADA	R	5	K	TSASXGXCADA	R	18
T	SASXGXCADA	R	5	T	SASXGXCADA	R	18
A	SXGXCADA	R	5	A	SXGXCADA	R	18
K	GXSATVTGGQ	K	8	R	MYGVXPWNAFPG	K	18
K	TSASXGXCADA	R	8	K	VCGSNXXSXC	K	18
T	SASXGXCADA	R	8	M	GXXECCA	R	19
R	MYGVXPWNAFPG	K	8	R	QXFGDY	K	19
K	TSASXGXCADA	R	9	R	QXFGDYK	T	19
A	SXGXCADA	R	9	K	GXSATVTGGQ	K	19
K	TSASXGXCADA	R	11	K	TSASXGXCADA	R	19
K	TSASXGXCADA	R	12	T	SASXGXCADA	R	19
T	SASXGXCADA	R	12	S	ASXGXCADA	R	19
K	GXSATVTGGQ	K	13	A	SXGXCADA	R	19
K	TSASXGXCADA	R	13	R	MYGVXPWNAFPG	K	19
T	SASXGXCADA	R	13	K	VCGSNXXSXC	K	19
A	SXGXCADA	R	13	G	SNXXSXC	K	19
R	MYGVXPWNAFPG	K	13				

### 3.1 Paleoproteomics – Studying the Iceman's Brain Tissue

Table 3-4: Peptides identified from myelin proteolipid protein (PLP). The amino acid in P1 and the C-terminal amino acid were extracted to allow for fully-, semi-, no tryptic cleavage sites, respectively, via the different gel bands (#20 – #22).

P1	Peptide Sequence	C-Term	Gel Band	P1	Peptide Sequence	C-Term	Gel Band
M	GXXECCA	R	20	R	CXVGAPFASXVAT	G	22
K	XXETYFS	K	20	R	CXVGAPFASXVATG	X	22
K	XXETYFSKN	Y	20	G	CGHEAXTGTE	K	22
R	QXFGDY	K	20	C	GHEAXTGTE	K	22
R	QXFGDYK	T	20	G	HEAXTGTE	K	22
R	QXFGDYKTTXCG	K	20	G	TEKXXETYFS	K	22
K	GXSATVTGGQ	K	20	T	EKXXETYFS	K	22
K	GXSATVTGGQKG	R	20	K	XXETYFS	K	22
K	GXSATVTGGQKGR	G	20	K	XXETYFSK	N	22
W	TTCQSXAAPS	K	20	G	AXXXAEGFYTTGAV	R	22
P	SKTSASXGXCADA	R	20	X	AEGFYTTGAV	R	22
K	TSASXGXCADA	R	20	R	QXFGDY	K	22
T	SASXGXCADA	R	20	R	QXFGDYK	T	22
S	ASXGXCADA	R	20	R	QXFGDYKT	T	22
A	SXGXCADA	R	20	K	GXSATVTGGQ	K	22
R	MYGVXPWNAFPG	K	20	K	GXSATVTGGQK	G	22
K	VCGSNXXSXC	K	20	K	GXSATVTGGQKG	R	22
G	HEAXTGTE	K	21	K	GXSATVTGGQKGR	G	22
G	TEKXXETYFS	K	21	W	TTCQSXAAPS	K	22
T	EKXXETYFS	K	21	P	SKTSASXGXCADA	R	22
K	XXETYFS	K	21	K	TSASXGXCADA	D	22
K	XXETYFSK	N	21	K	TSASXGXCADA	R	22
K	XXETYFSKN	Y	21	T	SASXGXCADA	R	22
R	QXFGDY	K	21	S	ASXGXCADA	R	22
K	GXSATVTGGQ	K	21	A	SXGXCADA	R	22
K	GXSATVTGGQK	G	21	S	XGXCADA	R	22
K	GXSATVTGGQKG	R	21	A	RMYGVXPWNAFPG	K	22
K	GXSATVTGGQKGR	G	21	R	MYGVXPW	N	22
W	TTCQSXAAPS	K	21	R	MYGVXPWNA	F	22
P	SKTSASXGXCADA	R	21	R	MYGVXPWNAFP	G	22
K	TSASXGXCADA	R	21	R	MYGVXPWNAFPG	K	22
T	SASXGXCADA	R	21	R	MYGVXPWNAFPGK	V	22
S	ASXGXCADA	R	21	Y	GVXPWNAFPG	K	22
A	SXGXCADA	R	21	K	VCGSNXXSXC	K	22
R	MYGVXPWNAFPG	K	21	G	SNXXSXC	K	22
K	VCGSNXXSXC	K	21				

### 3 Development of Computational Proteomics Pipelines for Automated Data Processing

#### 3.1.3 Conclusion

Two dark spots of the Tyrolean iceman's brain were subjected to protein analysis by mass spectrometry. As protein identification of a 5,300 years old sample was expected to be a complicated task, a two-step extraction process was performed to increase the number of identifiable proteins. The aim of this project was to establish a computational proteomics pipeline to identify a maximal number of proteins by using multi-database search results as input-files to finally combine them (e.g., the number of clearly identified proteins) into spreadsheet like file for data evaluation and visualization.

The combination of several database search approaches not only increased the number of proteins identified, but also their confidence was higher if a protein was identified by more than one search engine. So-called one-hit wonders were excluded and only proteins with at least two high or three medium confident peptides were considered as identified. This represented a conservative data-analysis approach, which was necessary to exclude, or at least reduce the number of, false positive protein identifications. This was important, as datasets of more than one search engine were combined.

Data were solely acquired by CID-IT, which is a sensitivity and fast fragmentation technique, but lacking the high mass accuracy of Orbitrap mass analyzers. Under these conditions, X!Tandem and Mascot performed best. From previous measurements, the enhanced performance of Mascot compared to SEQUEST to match peptides from CID-IT data was known. The overall excellent performance of X!Tandem might be a result of its algorithm (335). Principally, it performs two different database searches. The first uses the complete list of digested proteins (including fixed and dynamic modifications), whereas the second search utilizes only proteins being identified in the previous search. As the search space decreases significantly, variable modifications (e.g., posttranslational modifications), sequence polymorphisms and semi-tryptic peptides can be included without an increase in the search space. Finally, probability scores are calculated (*Hyperscore* and *E-Value*) to validate Peptide-Spectrum-Matches. It can only be speculated, that including unexpected protein variations in the second analysis step resulted in similar results for both, the tryptic and *no-enzyme* database search.

In the end, by using the newly implemented computational proteomics workflow combining four search engines, 502 confident and validated protein were identified. This represents a relatively high number of protein identifications for a 5,300 years old wet-mummy, highlighting the profit to use more than one search engine for protein identification.

## 3.2 PCSP – Computational Proteomics Tools for Protease Cleavage Site Determination<sup>3</sup>

Proteolysis is an irreversible posttranslational protein modification and involved in various biological processes. The determination of protease cleavage site preferences is an important aspect to understand biochemical pathways triggered by proteolysis activity.

A number of methods for the identification of protease cleavage sites have been developed in the past. Two such methods are the PICS (Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites; (336)) and Q-PICS (Quantitative Protease Cleavage Site Profiling using Tandem-Mass-Tag Labeling; (312)) approach. Both use a proteomics library of peptides which is obtained by digestion of a model proteome (e.g., from cell culture) with a first (so-called) *work-protease*. After N-termini protection, the peptide library is subjected to a second digest with a *target-protease*, which is the protease under investigation. In the PICS approach, newly generated N-termini are biotinylated, enriched and identified by LC-ESI MS. This allows the characterization of the specificity of the target-protease by comparing the P1 and P1' sites via the complete peptide library (see Figure 3-8). The P1' site is directly available upon peptide identification, whereas the P1 site information is encoded in the protein sequence.

The Q-PICS approach represents a further development of the PICS approach by incorporation of isobaric labeling for simultaneous identification and quantification of cleavage sites. The Q-PICS approach has been further refined by (i) integration of technical replicates, (ii) measuring the intact peptide library in addition to the digested (target) peptide library, and (iii) the application of both, LC-ESI MS and LC-MALD MS (147).

In order to process the identified peptides after database search in a fast, reproducible and automated way, two software tools are required for cleavage site identification. The focus of the first program is (i) to filter-out low confident peptides and (ii) remove peptides which are present in both the intact and digested peptide library, and (iii) perform data-picking. The second tool is used (i) to select only unique peptides, (ii) to remove sequence homologies and (iii) to calculate the relative distribution of each amino acid in relation to each position (P1-P5 and P1'-P5', respectively (see Figure 3-8)).

In the following sub chapters, the implementation of these two tools is described separately. It should be noted that the two tools are linked, *PeptideMerger* generates the input files for the second tool, *AminoAcidFinder*. The implementation will be described with ADAM10 as target- and GluC as work-protease.

---

3 The tools described were part of the publication J. Proteome Res., 2014, 13 (4), pp 2205–2214

### 3 Development of Computational Proteomics Pipelines for Automated Data Processing

#### 3.2.1 Implementation of PeptideMerger

*PeptideMerger* is a GUI based application, that accepts Proteome Discoverer Peptide-Spectrum-Match (PSM) list files in .csv file format and exports processed lists into a tab-separated .txt file, which is directly compatible to the *AminoAcidFinder*.

In the first steps, the user imports the PSM file(s) for the target-protease digested peptide library (ADAM 10), followed by the PSM file(s) for the work-protease (GluC) via the Tap *Import Blank PSM List* (Figure 3-6). To include only confident peptide identifications, a selection rule can be applied; e.g., the peptide has to be found at least twice within three technical injected LC MS runs. These settings are adjusted in the *Replica Selection* tab.

To increase the flexibility of the tool, the first line of the comma separated PSM files containing the header information is imported and displayed in a user controlled pull down menu. Here, the entry containing the .raw file information has to be selected. This selection menu allows also other database search file results (e.g., derived from PeptideShaker) to be imported. The *Peptide Identification Criteria* tab is used to assign the peptide sequence and protein accession number, respectively. Both values have to be assigned as these are the information which will be exported. To include additional information, these can be activated under *Appendix* tab. Theoretically, all information can be selected which then will be joined together by a semicolon in the final tab-separated .txt file.

By starting the script (*Save & Start* tab), the PSM lists of both the target and work-protease are imported. During the import, the number of different .raw file name(s) are culled and stored temporary. Within each PSM list, the peptides passing the minimum replica selection number, defined in the *Replica Selection* tab, are marked as *true*. Finally, the PSM lists from the target-protease (ADAM 10) are compared versus the work-protease (GluC) PSM list. Peptides found in both blank and target-protease lists are not exported. PSMs identified in only the target-protease PSM list are exported with the sequence, protein accession number and the appendix entries.

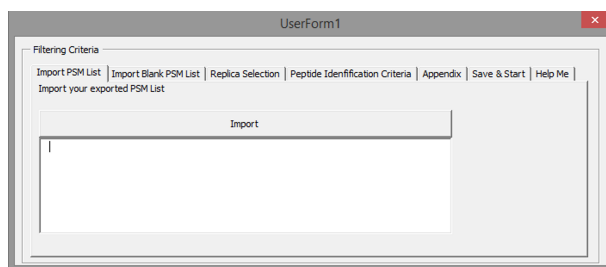


Figure 3-6: Graphical User Interface (GUI) for data import. The first tab, which is shown here, imports the list of peptide identifications. The remaining tabs are grouped into different criteria (e.g., blank) to separate these and improve the lucidity.



### 3.2.2 Implementation of AminoAcidFinder

*AminoAcidFinder* is a fork of the Excel and Perl published tools (336). This was demanded in order to be more flexible in terms of usability (interactive GUI; Figure 3-7) data filtering, source code manipulation and reproducibility. Running the same data set repeatedly with those tools, discriminative result files were obtained by each repetition.

Three different input-files are required for the *AminoAcidFinder*: (i) a FASTA file in UniProt formatting containing the protein sequences of the organism the peptide library originates from. (ii) an additional FASTA file containing the sequence information about the target-protease (and/ or other to be excluded protein sequences) is imported. This is in addition and in contrast to the previously published tools. During processing, these sequences will not be included in the final result file to compensate potential biases introduced. (iii) a tab separated .txt file, which can be created by the *PeptideMerger* is imported. To be flexible, all other .txt files fulfilling the input characteristic are accepted, too.

Figure 3-7: Graphical User Interface (GUI) to load in data calculated previously using *PeptideMerger*. The options for file loading are color-coded (green) and required. Via a pull down menu, the number of amino acids in P and P' positions are selected.

The option to chose either “accession number” or “description” within the FASTA file is not supported at the point of release; as the FASTA file format is limited to those of UniProt (Figure 3-7). In addition the input

### 3 Development of Computational Proteomics Pipelines for Automated Data Processing

option for “handling of protein amount” is still under development. This function will allow the user to chose between a variety of decision trees methods how redundant peptides will be handled by calculating protein FDR values alongside a protein score. These are used to decide from which protein a shared peptide is more likely to belong to, based on the assumption that higher scoring proteins are higher abundant in the sample.

The last user input deals with the number of amino acids positions to be reported (referred to as  $P_{\text{maxPosition}}$ ; default: 5) and to which position sequences should be compared (default: 5). Best results are received if both values are set to the same integer value.

By starting the tool (*Start Compare* button), first, the identified peptide list is read in. Due to a software issue in Proteome Discoverer 1.4 when used in combination with Mascot (Version 2.2.07), peptide hits with randomly missing accession numbers are frequently exported. For peptide sequences with missing protein accession numbers a function is called which scans the FASTA file for this specific peptide sequence(s) and attaches the corresponding protein accession number to the peptide sequence in the list.

For demonstration purposes the peptide *SQLSLTLDVE* from  $\beta$ -casein has been used in this section and is highlighted in orange (Table 3-5). This peptide is compared to the protein sequence in the FASTA file. If the sequence was found within the protein sequence and *not* within the working protease list simultaneously, the amino acids in the prime (P; *EPFTE*, highlighted in gray) and non-prime (P'; *SQLSLTLDVE*) positions, up to  $P_{\text{maxPosition}}$ , are extracted, temporary joined and stored. For the cases in which the number of amino acids to be extracted ( $P_{\text{maxPosition}}$ ) is lager than the number of positions available in the protein sequence, the number of assigned amino acids is adjusted accordingly. Additionally, amino acids N-terminal to a work-protease cleavage site (e.g., D, E for GluC) in P position are removed (*PKYPV*, highlighted in cyan). The next step is to remove duplicate sequences with identical identified P (*EPFTE*) and P' (*SQLSLTLDVE*) sequences, which can (e.g.) be obtained from protein isoforms. As an intermediate result, the frequency of each amino acid in relation to a specific position (e.g, lysine in P1') is calculated and exported into a tab-separated .txt file.

Table 3-5: Protein sequence of  $\beta$ -casein. The peptide sequence identified by LC MS is highlighted in orange, its corresponding prime sequence in gray, and the sequence N-terminal to a work-protease cleavage site in cyan. Only the gray (P) and orange (P') labeled sequences are used, whereas the cyan labeled sequence is removed.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50		
M	K	V	L	I	L	A	C	L	V	A	L	A	L	A	R	E	L	E	E	L	N	V	P	G	E	I	V	E	S	L	S	S	S	E	E	S	I	T	R	I	N	K	K	I	E	K	F	Q	S		
E	E	Q	Q	Q	T	E	D	E	L	Q	D	K	I	H	P	F	A	Q	T	Q	S	L	V	Y	P	F	P	G	P	I	P	N	S	L	P	Q	N	I	P	P	L	T	Q	T	P	V	V	V	P		
P	F	L	Q	P	E	V	M	G	V	S	K	V	K	E	A	M	A	P	K	H	K	E	M	P	F	P	K	Y	P	V	E	P	F	T	E	S	Q	S	L	T	L	T	D	V	E	N	L	H	L		
P	L	P	L	L	Q	S	W	M	H	Q	P	H	Q	P	L	P	P	T	V	M	F	P	P	Q	S	V	L	S	L	S	Q	S	K	V	L	P	V	P	Q	K	A	V	P	Y	P	Q	R	D	M		
P	I	Q	A	F	L	L	Y	Q	E	P	V	L	G	P	V	R	G	P	F	P	I	I	V																												

The next step is to remove identical peptide sequences (P', e.g. *SQLSLTLDVE*) with non-identical prime sequences (P), which occurs due to natural variations such as the exchange of glutamic acid with glutamine at position 132 for  $\beta$ -casein, resulting in *EPFTE* and *PKYPVQPFTE*.

If multiple non-identical prime sequences for the same peptide sequence (non-prime) are found, the prime sequences (*EPFTE* and *PKYPVQPFTE*) are checked for similarity, and only identical amino acids at a specific position are kept, with the first amino acid that differs replaced with an X (e.g., *XPFTE*). The intermediate results are again exported, the peptide list is sorted by the sequence name, and the frequency of each amino acid, in

### 3.2 PCSP – Computational Proteomics Tools for Protease Cleavage Site Determination

relation to a specific position, is exported into a tab-separated .txt file (see Table 3-6).

Table 3-6: Tab-separate .txt file containing the frequency of each of the twenty amino acids (AA) from the prime site 5 (P5) to the non-prime site 5 (P5').

AA	P5	P4	P3	P2	P1	'P1	'P2	'P3	'P4	'P5
A	29	17	38	39	49	21	17	45	22	24
C	5	1	2	0	3	1	1	2	1	2
D	11	6	2	18	2	0	0	7	14	11
E	16	20	3	27	10	0	2	4	7	20
F	8	10	3	7	14	6	15	4	5	7
G	15	5	7	16	10	0	10	32	31	14
H	0	6	1	4	5	0	5	3	5	2
I	9	10	13	0	0	25	27	7	9	15
K	0	0	0	0	0	0	0	4	0	1
L	12	20	16	35	11	58	12	10	25	21
M	0	2	5	7	4	1	3	0	3	2
N	5	10	4	3	6	2	6	12	9	11
P	21	15	24	0	24	0	2	1	18	20
Q	1	6	7	6	9	1	16	10	5	8
R	0	9	0	16	22	0	19	10	7	11
S	12	12	14	14	12	6	13	23	10	4
T	12	19	12	0	5	9	14	11	10	8
V	23	15	34	4	0	71	32	14	18	20
W	0	2	0	1	0	2	1	1	0	0
Y	9	9	9	0	14	0	8	3	4	2

The result file (see Table 3-6) can be imported directly in any spreadsheet related programs to create Heat-Maps using RGB scale (similar to those presented in chapter 4.1) to demonstrate the specificity of a target-protease, as outline in Figure 3-8.

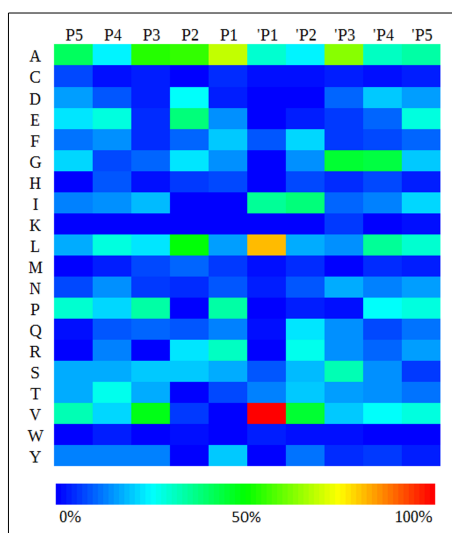


Figure 3-8: Heat-Map created by in-house developed VBA script using minimum (blue) and maximum (red) as delimiter to calculate the appropriate RGB code. Values used as input were created by AminoAcidFinder.

### 3 Development of Computational Proteomics Pipelines for Automated Data Processing

#### 3.2.3 Conclusion

In contrast to previously published tools (336), the newly developed *PeptideMerger* allows (i) a specific filtering of the input data in terms of minimum number of identification within technical injections and (ii) creation of “;”-joined strings which contain user demanded informations. The exported files, which are directly compatible to *AminoAcidFinder*, can also be used for alternative purposes, e.g., filtering of data for exclusively for identification purposes.

The tool *AminoAcidFinder* is developed in an open and easily customized way. At any point within the source code, new algorithms (ideally added as a new function) can be added. Moreover, in contrast to the previously published tools (336), all intermediate results are exported. This allows (i) to analyze the functions used, (ii) find potential software issues or (iii) to use these intermediate results file in alternative strategies to test, develop or modify the currently available tool(s).

### 3.3 Discussion and Outlook

Computational proteomics is a key factor in semi- or large-scale experiments to process MS .raw files. Classically, it covered the steps of protein identification, modification determination, quantification and sample comparison (337). By the rapid improvement of mass spectrometers (e.g., mass accuracy and MS/MS scan cycle speed), the application, possibility and importance to automate tasks grew exponentially. Generally, handling thousands of peptide identifications (e.g., obtained by the biological sample used for NCE optimization, chapter 2.1.3.3) is an error-prone and time-consuming process if it is performed manually. This is especially crucial as non-systematical (random) errors are introduced.

Nowadays, the research subject of computational proteomics needs to be enlarged, as upstream processing (that is after database searches) are more important, as the number of peptides (or proteins) identified increased significantly. Upstream processing of identification results was also the subject of chapter 3, as two independent project tasks were answered by the assistance of computational tools.

These tools were written *in-house*, which allowed modification and enhancement of previously described workflows. For the paleoproteomics study, results from four different search engines were combined, which significantly increased the number of identified proteins. Moreover, the application of a multi-database search algorithm approach demonstrated the importance of selecting the *right* search algorithm for each task. X!Tandem was found to perform the best, which is speculated to be a consequence of its two step database search approach, including the *variations* amino acids polymorphisms, semi-tryptic peptides and unexpected modifications in the second database search procedure.

To calculate the number of non-, semi- or fully tryptic peptides in the low mass region of SDS-gels is a time consuming and error-prone process. For each protein and peptide sequence the amino acids in -1 position to the N-terminus have to be extracted from the FASTA file. Principally, the same problem was solved within the project PCSP (chapter 3.2), as the information about the amino acid sequences N-terminal to the new cleavage-site were demanded. This task was fully fulfilled by compiling two tools to filter, analyze and export the relative frequency of amino acids in the P1-P5 and P1'-P5' position of the new-N-terminus.

Both projects cannot be grouped into *bioinformatics*, as they were lacking statistical data evaluation. However, these upstream tools aided in processing data, as random errors were minimized and new data can be processed using exactly the same workflow as before. One additional feature was to modify existing algorithms at specific positions and to re-run the script to generate a second output file, which subsequently allows for more efficient and reproducible evaluation of proteomics data.

## 4 Appendix

4.1 Chemicals and Instrumentation

4.2 Amino Acids

4.3 cRAP FASTA File

4.4 Supplemental Tables for Chapter 2.4

4.5 Figure Index

4.6 Equation Index

4.7 Table Index

4.8 Copyright

4.9 References

## 4.1 Chemicals and Instrumentation

### Chemicals

iTRAQ Reagents 4/8plex Kit including tetraethylammonium bromide (TEAB), tris(2-carboxyethyl)phosphine (TCEP) and ethanol (Et-OH) were purchased from AB SCIEX (Darmstadt, Germany). Acetonitrile (ACN), formic acid (FA), trifluoroacetic acid (TFA), triethylammonium bicarbonate (TEAB), methanol (Me-OH), phosphoric acid (PA) glycolic Acid (GA), ethylenediaminetetraacetic acid (EDTA), iodoacetamide (IAA) and coomassie blue were purchased from Sigma-Aldrich (Munich, Germany). The proteases chymotrypsin (V106A), elastase (V189A), Staphylococcus aureus V-8 Protease (GluC; V165A) and trypsin (V5111) were purchased from Promega Corporation (Madison, WI, USA). Alkaline phosphatase (AP, EC 3.1.3.1) was purchased from Roch (Roche Diagnostics GmbH, Roche Applied Science, Mannheim, Germany). Ammonium sulfate was purchased from Merck (Darmstadt, Germany), titanium dioxide (TiO<sub>2</sub>) from GL Sciences (via MZ-Analysentechnik, Mainz, Germany). ZipTip C18 was purchased from Millipore (Darmstadt, Germany).

### Sample Material

Phosphopeptide standard (chapter 2.1) was purchased from Intavis (Cologne, Germany). Proteins used for the ten protein-mix are summarized in Table 4-1 and were purchased from Sigma-Aldrich (Munich, Germany).

Table 4-1: Proteins used for the ten protein-mix.

<b>Protein</b>	<b>Species</b>	<b>SwissProt Accession Number</b>
albumin (BSA)	bovine	P02769
carbonic anhydrase	bovine	P00921
catalase	bovine	P00432
cytochrome C	equine	P00004
α-Casein	bovine	P02662
β-Casein	bovine	P02666
α-lactalbumin	bovine	P00711
β-Lactoglobulin	bovine	P02754
lysozym	chicken	B8YK79
myoglobin	equine	P68082
ribonuclease A	bovine	P61823
transferrin	human	P02787

### Buffers and Eluents not supplied by Vendors

- iTRAQ-4plex labeling buffer: 160 μL H<sub>2</sub>O, 160 μL iTRAQ-4plex dissolution buffer and 80 μL ethanol
- AP-buffer (dephosphorylation) was diluted 1:10 with dd H<sub>2</sub>O
- Loading buffer (LC-ESI & LC-MALDI MS): 3% ACN, 0.05% TFA
- Eluent A (LC-ESI MS): 0.05% FA in H<sub>2</sub>O

## 4 Appendix

- Eluent B (LC-ESI MS): 80% ACN, 0.04% FA in H<sub>2</sub>O
- Eluent A (LC-MALDI MS): 0.05% FA in H<sub>2</sub>O
- Eluent B (LC-MALDI MS): 80% ACN, 0.04% FA in H<sub>2</sub>O

### **Instrumentation**

- Sample drying was performed via SpeedVac (Eppendorf, Hamburg, Germany).
- Water used in all experiments was purified by an arium 611VF system (Satorious, Göttingen, Germany).
- SDS-PAGE: Biorad Mini Protean Tetrasystem & Power Pack Universal
- Ultrasonic bath: Elmasonic SH40
- Centrifuge: Thermo Scientific Heraeus Fesco 21 & Biofuge Statos

### **Mass Spectrometry Instrumentation**

#### *LTQ Orbitrap Velos*

NanoLC-ESI MS experiments were performed on a Dionex U3000 (Dionex, Idstein, Germany) nano-HPLC system coupled online to a LTQ Orbitrap Velos MS equipped with ETD support (ThermoFisher, Bremen, Germany). Samples were trapped on an Acclaim Pepmap C18, 300 $\mu$ m x 10mm, 5 $\mu$ m, 100 Å trap column and separated on a an AcclaimPepMap 100 analytical column (3  $\mu$ m, 75  $\mu$ m x 150 mm, Dionex, Idstein, Germany).

#### *Synapt G2s*

NanoLC-ESI MS experiments were performed on a Dionex U3000 (Dionex, Idstein, Germany) nano HPLC system coupled online to a Synapt G2s (Waters, Milford, MA). Samples were trapped on an Acclaim Pepmap C18, 300 $\mu$ m x 10mm, 5 $\mu$ m, 100 Å trap column and separated on a an AcclaimPepMap 100 analytical column (3  $\mu$ m, 75  $\mu$ m x 150 mm, Dionex, Idstein, Germany).

#### *AB SCIEX MALDI TOF/TOF 5800*

Prior to MS analysis with a MALDI TOF/TOF 5800 mass spectrometer (AB Sciex, Darmstadt, Germany), samples were separated on a Dionex U3000 nano-LC system (Dionex, Idstein, Germany, including samples trapping on an Acclaim Pepmap C18, 300 $\mu$ m x 10mm, 5 $\mu$ m, 100 Å trap column and separated on a an AcclaimPepMap 100 analytical column (3  $\mu$ m, 75  $\mu$ m x 150 mm, Dionex, Idstein, Germany) and subsequently spotted on a MALDI target (AB Sciex) using a Probot microfraction collector (Dionex).



### Software

#### *Protein Identification Suites & Nodes*

- Proteome Discoverer (Thermo- Fisher, San Jose, CA): Version 1.2, 1.3 and 1.4
- MS2-Spectrum Processor (<http://ms.imp.ac.at>): Version v. 0.9
- phosphoRS (299): Version 3.1
- Percolator (236): Version 2.04 (supplied by Proteome Discoverer)
- SearchGUI: Version 1.16.2 (chapter 2.3) & 1.7.3 (chapter 3.1) (294)
- PeptideShaker: Version 0.25.2 (chapter 2.3) & 0.16.2 (chapter 3.1) (295)

#### *Search Algorithms*

- SEQUEST & SEQUEST HT: supplied with each Proteome Discoverer Version
- Mascot: Version 2.2.0.7 (chapter 2.3) & 2.2.0.4 (chapter 3.1)
- OMSSA (supplied by SearchGUI): Version 2.1.9 (chapter 2.3) & (chapter 3.1)
- X!Tandem (supplied by SearchGUI): Version 2.2.0.7 (chapter 2.3) & 2.2.0.4 (chapter 3.1)
- TS2Mascot software (version 1.0.0, Matrix Science, London, UK)

#### *Raw-Data*

- XCalibur (Thermo Fischer): Version 2.1 (32 bit Windows), Version 2.2 SP1.48 (64 bit Windows)
- MassLynx (Waters): Version SCN 883
- TOF/TOF Explorer (AB SCIEX): Version 4.0.0 (build 8)
- Proteowizard: Version 3.0.5759 (293)
- mmass: Version 5.5.0 (338)

#### *Programming Languages & Plotting Tools*

- GAMBAS: Version 3.5.3
- Visual Basic for Applications: included in Microsoft Office Excel 2007, 2010 & 2013
- QtiPlot: Version 0.9.8.9 svn 2288

## 4 Appendix

### Paragon Mod Feature Set

```
<MOD_FEATURE_SET xml:id="MOD_FEATURE_SET:71" name="Std iTRAQ 4plex set">
  <MOD_FEATURE mod="iTRAQ4plex">
    <OCCURRENCE target="Lysine" prob="0.993"/>
    <OCCURRENCE target="Tyrosine" prob="0.036"/>
    <OCCURRENCE target="Histidine" prob="0.0053"/>
    <OCCURRENCE target="Serine" prob="0.011"/>
    <OCCURRENCE target="Threonine" prob="0.0027"/>
  </MOD_FEATURE>
  <MOD_FEATURE mod="Terminal iTRAQ4plex">
    <OCCURRENCE target="" term_spec="PepNTerm" prob="0.932"/>
  </MOD_FEATURE>
</MOD_FEATURE_SET>

<MOD_FEATURE_SET xml:id="MOD_FEATURE_SET:11" name="Standard biological mod set">
  <MOD_FEATURE mod="Phospho(Ser,Thr)">
    <OCCURRENCE target="Serine" prob="0.002"/>
    <OCCURRENCE target="Threonine" prob="0.0007"/>
  </MOD_FEATURE>
  <MOD_FEATURE mod="Phospho">
    <OCCURRENCE target="Tyrosine" prob="0.00001"/>
    <OCCURRENCE target="Lysine" prob="0.0001"/>
    <OCCURRENCE target="Histidine" prob="0.0001"/>
    <OCCURRENCE target="Aspartic Acid" prob="0.0001"/>
    <OCCURRENCE target="Arginine" prob="0.0001"/>
    <OCCURRENCE target="Cysteine" prob="0.00001"/>
  </MOD_FEATURE>
</MOD_FEATURE_SET>
```

### Function heatMapColor

*minValue*: minimal number of amino acid identifications; *maxValue* as the maximal number of amino acid identifications; *currentValue*: represents the number of the amino acid being investigated. The corresponding color code (in RGB style) is returned as *heatMapColor*. Code is adopted from Bruce McPherson, <http://ramblings.mcpher.com>.

**Private Function** heatMapColor(*minValue* As Variant, *maxValue* As Variant, *currentValue* As Variant) As Long

**Dim** spread As Double, ratio As Double, red As Double, green As Double, blue As Double

spread = *maxValue* - *minValue*

ratio = (*currentValue* - *minValue*) / spread

If ratio < 0.25 Then

blue = 1

green = 4 \* ratio

Elseif ratio < 0.5 Then

green = 1

blue = 1 + 4 \* (*minValue* - *currentValue* + 0.25 \* spread) / spread

Elseif ratio < 0.75 Then

green = 1

red = 4 \* (*currentValue* - *minValue* - 0.5 \* spread) / spread

Else

red = 1

green = 1 + 4 \* (*minValue* - *currentValue* + 0.75 \* spread) / spread

End If

heatMapColor = RGB(red \* 255, green \* 255, blue \* 255)

**End Function**

## 4.2 Amino Acids

Table 4-2: The chemical composition, the 3 and 1 letter code and the monoisotopic and average mass for each amino acid.

Residue	3-letter code	1-letter code	Mono-isotopic mass	Average mass
Alanine C <sub>3</sub> H <sub>5</sub> NO	Ala	A	71.037114	71.0779
Arginine C <sub>6</sub> H <sub>12</sub> N <sub>4</sub> O	Arg	R	156.101111	156.1857
Asparagine C <sub>4</sub> H <sub>6</sub> N <sub>2</sub> O <sub>2</sub>	Asn	N	114.042927	114.1026
Aspartic acid C <sub>4</sub> H <sub>5</sub> NO <sub>3</sub>	Asp	D	115.026943	115.0874
Cysteine C <sub>3</sub> H <sub>5</sub> NOS	Cys	C	103.009185	103.1429
Glutamic acid C <sub>5</sub> H <sub>7</sub> NO <sub>3</sub>	Glu	E	129.042593	129.114
Glutamine C <sub>5</sub> H <sub>8</sub> N <sub>2</sub> O <sub>2</sub>	Gln	Q	128.058578	128.1292
Glycine C <sub>2</sub> H <sub>3</sub> NO	Gly	G	57.021464	57.0513
Histidine C <sub>6</sub> H <sub>7</sub> N <sub>3</sub> O	His	H	137.058912	137.1393
Isoleucine C <sub>6</sub> H <sub>11</sub> NO	Ile	I	113.084064	113.1576
Leucine C <sub>6</sub> H <sub>11</sub> NO	Leu	L	113.084064	113.1576
Lysine C <sub>6</sub> H <sub>12</sub> N <sub>2</sub> O	Lys	K	128.094963	128.1723
Methionine C <sub>5</sub> H <sub>9</sub> NOS	Met	M	131.040485	131.1961
Phenylalanine C <sub>9</sub> H <sub>9</sub> NO	Phe	F	147.068414	147.1739
Proline C <sub>5</sub> H <sub>7</sub> NO	Pro	P	97.052764	97.1152
Serine C <sub>3</sub> H <sub>5</sub> NO <sub>2</sub>	Ser	S	87.032028	87.0773
Threonine C <sub>4</sub> H <sub>7</sub> NO <sub>2</sub>	Thr	T	101.047679	101.1039
Tryptophan C <sub>11</sub> H <sub>10</sub> N <sub>2</sub> O	Trp	W	186.079313	186.2099
Tyrosine C <sub>9</sub> H <sub>9</sub> NO <sub>2</sub>	Tyr	Y	163.06332	163.1733
Valine C <sub>5</sub> H <sub>9</sub> NO	Val	V	99.068414	99.1311

## 4 Appendix

### 4.3 cRAP FASTA File

Table 4-3: Proteins contained in the cRAP like FASTA file. All proteins used originated from UniProtKB/Swiss-Prot.

Unique Identifier	EntryName	Unique Identifier	EntryName	Unique Identifier	EntryName
Q13515	BFSP2_HUMAN	Q14532	K1H2_HUMAN	P60411	KR109_HUMAN
P31944	CASPE_HUMAN	P35908	K22E_HUMAN	P60014	KR10A_HUMAN
Q9UMD9	COHA1_HUMAN	Q3TTY5	K22E_MOUSE	P60412	KR10B_HUMAN
P15924	DESP_HUMAN	Q01546	K22O_HUMAN	P60413	KR10C_HUMAN
P11532	DMD_HUMAN	Q7Z794	K2C1B_HUMAN	Q8IUC1	KR111_HUMAN
O75190	DNJB6_HUMAN	P04264	K2C1_HUMAN	P59990	KR121_HUMAN
Q9UW98	DPP5_TRIRU	P04104	K2C1_MOUSE	Q9Z287	KR121_MOUSE
Q03001	DYST_HUMAN	A5A6M6	K2C1_PANTR	P59991	KR122_HUMAN
Q92817	EVPL_HUMAN	P12035	K2C3_HUMAN	P60328	KR123_HUMAN
Q8TES7	FBF1_HUMAN	P19013	K2C4_HUMAN	P60329	KR124_HUMAN
P20930	FILA_HUMAN	P13647	K2C5_HUMAN	Q8IUC0	KR131_HUMAN
Q8N117	IL31R_HUMAN	A5A6M8	K2C5_PANTR	Q52LG2	KR132_HUMAN
P07476	INVO_HUMAN	P02538	K2C6A_HUMAN	Q3SY46	KR133_HUMAN
P16144	ITB4_HUMAN	P50446	K2C6A_MOUSE	Q3LI77	KR134_HUMAN
A6NCN2	K121P_HUMAN	Q4FZU2	K2C6A_RAT	Q3LI76	KR151_HUMAN
P13645	K1C10_HUMAN	P04259	K2C6B_HUMAN	A8MUX0	KR161_HUMAN
Q99456	K1C12_HUMAN	Q9Z331	K2C6B_MOUSE	Q9BYP8	KR171_HUMAN
P13646	K1C13_HUMAN	P48668	K2C6C_HUMAN	Q8IUB9	KR191_HUMAN
P02533	K1C14_HUMAN	Q3SY84	K2C71_HUMAN	Q3LHN2	KR192_HUMAN
P19012	K1C15_HUMAN	Q14CN4	K2C72_HUMAN	Q7Z4W3	KR193_HUMAN
P08779	K1C16_HUMAN	Q86Y46	K2C73_HUMAN	Q3LI73	KR194_HUMAN
Q04695	K1C17_HUMAN	Q7RTS7	K2C74_HUMAN	Q3LI72	KR195_HUMAN
A5A6M0	K1C17_PANTR	O95678	K2C75_HUMAN	Q3LI70	KR196_HUMAN
P05783	K1C18_HUMAN	Q8N1N4	K2C78_HUMAN	Q3SYF9	KR197_HUMAN
P05784	K1C18_MOUSE	Q5XKE5	K2C79_HUMAN	Q3LI54	KR198_HUMAN
P08727	K1C19_HUMAN	P08729	K2C7_HUMAN	Q3LI63	KR201_HUMAN
P35900	K1C20_HUMAN	Q9DCV7	K2C7_MOUSE	Q3LI61	KR202_HUMAN
Q9C075	K1C23_HUMAN	A5A6N0	K2C7_PANTR	Q3LI60	KR203_HUMAN
Q2M2I5	K1C24_HUMAN	Q8HZR5	K2C7_POTTR	Q3LI62	KR204_HUMAN
Q7Z3Z0	K1C25_HUMAN	Q6KB66	K2C80_HUMAN	Q3LI58	KR211_HUMAN
A5A6N2	K1C25_PANTR	P05787	K2C8_HUMAN	Q3LI59	KR212_HUMAN
Q9BGM5	K1C25_SHEEP	Q02156	KPCE_HUMAN	Q3LHN1	KR213_HUMAN
Q7Z3Y9	K1C26_HUMAN	P60331	KR101_HUMAN	Q3MIV0	KR221_HUMAN
Q7Z3Y8	K1C27_HUMAN	P60368	KR102_HUMAN	Q3LI68	KR222_HUMAN
Q7Z3Y7	K1C28_HUMAN	P60369	KR103_HUMAN	A1A580	KR231_HUMAN
Q6A163	K1C39_HUMAN	P60372	KR104_HUMAN	Q3LI83	KR241_HUMAN
Q6A162	K1C40_HUMAN	P60370	KR105_HUMAN	Q3LHN0	KR251_HUMAN
P35527	K1C9_HUMAN	P60371	KR106_HUMAN	Q6PEX3	KR261_HUMAN
Q15323	K1H1_HUMAN	P60409	KR107_HUMAN	Q3LI81	KR271_HUMAN
A5A6M5	K1H1_PANTR	P60410	KR108_HUMAN	A8MX34	KR291_HUMAN

Table 4-3 (continued): Proteins contained in the cRAP like FASTA file. All proteins used originated from UniProtKB/Swiss-Prot.

Unique Identifier	EntryName	Unique Identifier	EntryName
Q9BYQ6	KRA411_HUMAN	Q9BYQ0	KRA98_HUMAN
Q9BQ66	KRA412_HUMAN	Q9BYP9	KRA99_HUMAN
Q6L8G5	KR510_HUMAN	O76011	KRT34_HUMAN
Q6L8G4	KR511_HUMAN	Q92764	KRT35_HUMAN
Q07627	KRA11_HUMAN	O76013	KRT36_HUMAN
Q8IUG1	KRA13_HUMAN	O76014	KRT37_HUMAN
P0C5Y4	KRA14_HUMAN	O76015	KRT38_HUMAN
Q9BYS1	KRA15_HUMAN	Q14533	KRT81_HUMAN
Q9BYU5	KRA21_HUMAN	Q9NSB4	KRT82_HUMAN
Q9BYT5	KRA22_HUMAN	P78385	KRT83_HUMAN
P0C7H8	KRA23_HUMAN	Q9NSB2	KRT84_HUMAN
Q9BYR9	KRA24_HUMAN	P78386	KRT85_HUMAN
Q9BYR8	KRA31_HUMAN	O43790	KRT86_HUMAN
Q9BYR7	KRA32_HUMAN	P97861	KRT86_MOUSE
Q9BYR6	KRA33_HUMAN	Q8N1A0	KT222_HUMAN
Q9BYQ7	KRA41_HUMAN	O76009	KT33A_HUMAN
Q9BYR5	KRA42_HUMAN	A5A6P3	KT33A_PANTR
Q9BYR4	KRA43_HUMAN	Q14525	KT33B_HUMAN
Q9BYR3	KRA44_HUMAN	Q8N6L1	KTAP2_HUMAN
Q9BYR2	KRA45_HUMAN	Q5QHG5	LAP1_TRIRU
Q9BYQ5	KRA46_HUMAN	Q5QHG6	LAP2_TRIRU
Q9BYR0	KRA47_HUMAN	A6XGK3	MCPA_TRIRU
Q9BYQ9	KRA48_HUMAN	Q8NIB6	MEP1_TRIRU
Q9BYQ8	KRA49_HUMAN	Q6WIH8	MEP3_TRIRU
Q6L8H4	KRA51_HUMAN	Q8NIJ4	MEP4_TRIRU
Q701N4	KRA52_HUMAN	Q29983	MICA_HUMAN
Q6L8H2	KRA53_HUMAN	P07197	NFM_HUMAN
Q6L8H1	KRA54_HUMAN	Q99650	OSMR_HUMAN
Q701N2	KRA55_HUMAN	Q8TEW0	PARD3_HUMAN
Q6L8G9	KRA56_HUMAN	P41219	PERI_HUMAN
Q6L8G8	KRA57_HUMAN	Q13835	PKP1_HUMAN
O75690	KRA58_HUMAN	Q2G2B2	SASG_STAA8
P26371	KRA59_HUMAN	Q69F58	SUB1_TRIRU
Q3LI64	KRA61_HUMAN	Q69F56	SUB3_TRIRU
Q3LI66	KRA62_HUMAN	Q69F35	SUB4_TRIRU
Q3LI67	KRA63_HUMAN	Q69F34	SUB5_TRIRU
Q8IUC3	KRA71_HUMAN	Q9NZ50	SYUG_BOVIN
Q8IUC2	KRA81_HUMAN	O76070	SYUG_HUMAN
A8MXZ3	KRA91_HUMAN	Q9BT92	TCHP_HUMAN
Q9BYQ4	KRA92_HUMAN	Q15628	TRADD_HUMAN
A5A6P5	KRA92_PANTR	Q07283	TRHY_HUMAN
Q9BYQ3	KRA93_HUMAN	P06922	VE4_HPVI16
Q9BYQ2	KRA94_HUMAN	P06459	VE4_HPVI6B
A8MVA2	KRA96_HUMAN	P08670	VIME_HUMAN
A8MTY7	KRA97_HUMAN		

## 4 Appendix

### 4.4 Supplemental Tables for Chapter 2.4

Table 4-4: Overview of identified and quantified peptides obtained by proteolytic digestion of  $\alpha$ -S1-casein

Protein Accession Number	Protease	Sequence	Modification*	Position	# of q-PSMs	W-Ratio	S.D.	Phospho-Index
P02662	Ch	RPKHPIKHQGLPQEVLE	-	16-34	2	0.63	0.07	-
P02662	Ch	RPKHPIKHQGLPQEVLENNLL	-	16-37	3	0.53	0.27	-
P02662	Ch	QGLPQEVLE	-	24-32	1	0.33	0.00	-
P02662	Ch	QGLPQEVLENNLL	-	24-37	11	0.45	0.02	-
P02662	Ch	GLPQEVLENNLL	-	25-37	3	0.49	0.01	-
P02662	Ch	EVLNENLL	-	29-37	4	0.68	0.17	-
P02662	Ch	NENLLRFF	-	32-40	4	0.51	0.01	-
P02662	Ch	FFVAPFPEVF	-	38-48	6	0.78	0.15	-
P02662	Ch	FVAPFPEVF	-	39-48	20	0.43	0.02	-
P02662	Ch	VAPFPEVF	-	40-48	74	0.44	0.03	-
P02662	Ch	GKEKVNEL	-	48-56	4	0.42	0.00	S1_56
P02662	Ch	AESISSEE	-	77-86	2	0.41	0.00	S1_79
P02662	Ch	HIQKEDVPSERY	-	95-107	1	0.81	0.00	-
P02662	Ch	IQKEDVPSERY	-	96-107	8	0.36	0.02	-
P02662	Ch	KEDVPSERY	-	98-107	9	0.36	0.01	-
P02662	Ch	LGYLEQL	-	107-114	12	0.43	0.02	-
P02662	Ch	LGYLEQLL	-	107-115	18	0.46	0.03	-
P02662	Ch	LRLKKY	-	114-120	3	0.38	0.02	-
P02662	Ch	KVPQLEIVPNSA	S11(Phospho)	120-132	1	0.37	0.00	S1_130
P02662	Ch	KVPQLEIVPNSAEERL	S11(Phospho)	120-136	2	0.37	0.04	S1_130
P02662	Ch	IVPNSAEERLHSMKE	M13(Oxidation)	126-141	1	0.54	0.00	S1_130
P02662	Ch	HSMKEGIHA	-	136-145	2	0.58	0.06	-
P02662	Ch	SMKEGIHAQQKEPM	-	137-151	1	0.41	0.00	-
P02662	Ch	SMKEGIHAQQKEPMIGVQNQEL	M2(Oxidation); M14(Oxidation)	137-158	1	0.53	0.00	-
P02662	Ch	SMKEGIHAQQKEPMIGVQNQELAY	M2(Oxidation)	137-160	2	0.43	0.07	-
P02662	Ch	SMKEGIHAQQKEPMIGVQNQELAYF	M2(Oxidation); M14(Oxidation)	137-161	1	0.52	0.00	-
P02662	Ch	AQQKEPM	-	144-151	1	0.41	0.00	-
P02662	Ch	AQQKEPMIGVQNQELAY	M7(Oxidation)	144-160	8	0.42	0.03	-
P02662	Ch	AQQKEPMIGVQNQELAY	-	144-160	3	0.42	0.02	-
P02662	Ch	AQQKEPMIGVQNQELAYF	M7(Oxidation)	144-161	5	0.42	0.02	-
P02662	Ch	AQQKEPMIGVQNQELAYF	-	144-161	1	0.42	0.00	-
P02662	Ch	IGVNQEL	-	151-158	3	0.49	0.03	-
P02662	Ch	IGVNQELAY	-	151-160	8	0.41	0.02	-
P02662	Ch	IGVNQELAYF	-	151-161	6	0.42	0.01	-
P02662	Ch	AYFYPELF	-	158-166	10	0.45	0.01	-
P02662	Ch	FYPELF	-	160-166	20	0.43	0.02	-
P02662	Ch	YQLDAYPSGAW	-	169-180	6	0.44	0.02	-
P02662	Ch	YQLDAYPSGAWY	-	169-181	4	0.45	0.02	-
P02662	Ch	QLDAYPSGAW	-	170-180	13	0.45	0.02	-
P02662	Ch	QLDAYPSGAWY	-	170-181	12	0.42	0.04	-
P02662	Ch	WYYVPL	-	179-185	1	0.48	0.00	-
P02662	Ch	YYVPLGTQY	-	180-189	10	0.42	0.01	-

## 4.4 Supplemental Tables for Chapter 2.4

Protein Accession Number	Protease	Sequence	Modification*	Position	# of q-PSMs	W-Ratio	S.D.	Phospho-Index
P02662	Ch	YVPLGTQY	-	181-189	6	0.42	0.00	-
P02662	Ch	VPLGTQY	-	182-189	3	0.45	0.01	-
P02662	Ch	QYTDAPSFSDIPNPI	-	187-202	2	0.48	0.01	-
P02662	Ch	TDAPSFSDIPNPIGSENSEKTT MPLW	M23(Oxidation)	189-215	3	0.69	0.02	-
P02662	Ch	TDAPSFSDIPNPIGSENSEKTT MPLW	-	189-215	1	0.75	0.00	-
P02662	Ch	DAPSFSDIPNPI	-	190-202	1	0.44	0.00	-
P02662	Ch	SFSDIPNPIGSENSEKTTMPLW	S1(Phospho); M19(Oxidation)	193-215	7	0.00	0.03	-
P02662	Ch	GSENSEKTTMPLW	M10(Oxidation)	202-215	1	0.48	0.00	-
P02662	EI	RPKHPI	-	16-22	7	0.35	0.01	-
P02662	EI	RPKHPIKHQGLPQEVLENLL	-	16-37	3	0.52	0.06	-
P02662	EI	KHQGLPQEV	-	22-31	13	0.76	0.07	-
P02662	EI	GLPQEV	-	25-31	2	0.40	0.01	-
P02662	EI	RFFVAPFPEVFGKEKV	-	37-53	4	0.39	0.03	-
P02662	EI	VAPFPEVFGKEKV	-	40-53	1	0.39	0.00	-
P02662	EI	NELSKDIGSESTEDQAMEDI	S4(Phospho); S9(Phospho); S11(Phospho)	53-73	1	0.33	0.00	S1_56
P02662	EI	NELSKDIGSESTEDQAMEDI	S11(Phospho); T12(Phospho)	53-73	1	0.30	0.00	S1_56
P02662	EI	NELSKDIGSESTEDQAMEDI	S9(Phospho); S11(Phospho); M17(Oxidation)	53-73	1	0.32	0.00	S1_56
P02662	EI	NELSKDIGSESTEDQAMEDIK QM	S9(Phospho); S11(Phospho); M23(Oxidation)	53-76	5	0.31	0.01	S1_56
P02662	EI	NELSKDIGSESTEDQAMEDIK QM	S9(Phospho); S11(Phospho)	53-76	2	0.30	0.01	S1_56
P02662	EI	NELSKDIGSESTEDQAMEDIK QM	S9(Phospho); S11(Phospho); M17(Oxidation); M23(Oxidation)	53-76	1	0.36	0.00	S1_56
P02662	EI	EAESISSEEIVPNSVEQKHIQ KEDVPSERYL	S4(Phospho); S6(Phospho); S7(Phospho); S8(Phospho); S15(Phospho)	76-108	1	0.38	0.00	S1_79
P02662	EI	KEDVPSERYL	-	98-108	1	0.42	0.00	-
P02662	EI	DVPSERYLGYLE	-	100-112	1	0.64	0.00	-
P02662	EI	GYLEQLL	-	108-115	13	0.35	0.01	-
P02662	EI	YLEQLL	-	109-115	5	0.38	0.08	-
P02662	EI	KKYKVPQLEIVPNSAEERL	S14(Phospho)	117-136	2	0.28	0.02	S1_130
P02662	EI	LEIVPNSAEERL	S7(Phospho)	124-136	8	0.32	0.03	S1_130
P02662	EI	LEIVPNSAEERLHSMKEGI	S7(Phospho); M15(Oxidation)	124-143	8	0.32	0.03	S1_130
P02662	EI	LEIVPNSAEERLHSMKEGIHA	S7(Phospho); M15(Oxidation)	124-145	5	0.33	0.02	S1_130
P02662	EI	SAEERL	-	130-136	1	0.80	0.00	S1_130
P02662	EI	SAEERLHSMKEGIHA	M9(Oxidation)	130-145	1	0.83	0.00	S1_130
P02662	EI	HSMKEGI	M3(Oxidation)	136-143	7	0.48	0.01	-
P02662	EI	HSMKEGIHA	M3(Oxidation)	136-145	13	0.63	0.01	-
P02662	EI	HSMKEGIHA	-	136-145	1	0.59	0.00	-
P02662	EI	SMKEGI	M2(Oxidation)	137-143	1	0.42	0.00	-
P02662	EI	SMKEGIHA	-	137-145	3	0.43	0.01	-
P02662	EI	SMKEGIHA	M2(Oxidation)	137-145	2	0.43	0.00	-

#### 4 Appendix

Protein Accession Number	Protease	Sequence	Modification*	Position	# of q-PSMs	W-Ratio	S.D.	Phosho-Index
P02662	El	HAQQKEPMIG	M8(Oxidation)	143-153	1	0.44	0.00	-
P02662	El	HAQQKEPMIGV	-	143-154	2	0.42	0.03	-
P02662	El	HAQQKEPMIGVQNQELA	M8(Oxidation)	143-159	4	0.44	0.04	-
P02662	El	QQKEPMIG	M6(Oxidation)	145-153	5	0.39	0.01	-
P02662	El	QQKEPMIG	-	145-153	2	0.44	0.01	-
P02662	El	QQKEPMIGV	-	145-154	5	0.37	0.01	-
P02662	El	QQKEPMIGVQNQELA	-	145-159	9	0.37	0.01	-
P02662	El	QQKEPMIGVQNQELA	M6(Oxidation)	145-159	21	0.38	0.02	-
P02662	El	YFYPELF	-	159-166	14	0.42	0.02	-
P02662	El	RQFYQLDAYPS	-	166-177	2	0.39	0.04	-
P02662	El	QFYQLDAYPS	-	167-177	4	0.41	0.01	-
P02662	El	FYQLDAYPS	-	168-177	13	0.40	0.01	-
P02662	El	YQLDAYPS	-	169-177	2	0.40	0.00	-
P02662	El	QLDAYPS	-	170-177	12	0.42	0.01	-
P02662	El	GAWYYVPL	-	177-185	15	0.40	0.03	-
P02662	El	WYYVPL	-	179-185	10	0.39	0.01	-
P02662	El	YVPLGTQY	-	181-189	2	0.55	0.13	-
P02662	El	GTQYTDAPSFSDIPNPI	-	185-202	3	0.35	0.00	-
P02662	El	QYTDAPSF	-	187-196	8	0.41	0.01	-
P02662	El	QYTDAPSFSDIPNPI	-	187-202	24	0.37	0.02	-
P02662	El	QYTDAPSFSDIPNPI	S7(Phospho)	187-202	4	0.29	0.05	-
P02662	El	DAPSFSDIPNPI	-	190-202	7	0.39	0.02	-
P02662	El	SFSDIPNPI	-	193-202	1	0.45	0.00	-
P02662	El	FSDIPNPI	-	194-202	13	0.43	0.02	-
P02662	El	DIPNPI	-	196-202	13	0.43	0.04	-
P02662	El	GSENSEKTTMPL	M10(Oxidation)	202-214	2	0.43	0.01	-
P02662	El	GSENSEKTTMPLW	M10(Oxidation)	202-215	3	0.36	0.00	-
P02662	El	TMPLW	-	210-215	2	0.37	0.02	-
P02662	Gl	RPKHPIKHQGLPQEVLE	-	16-34	13	0.41	0.08	-
P02662	Gl	GLPQEVLE	-	25-34	6	0.44	0.02	-
P02662	Gl	NLLRFFVAPFPE	-	34-46	39	0.52	0.02	-
P02662	Gl	FFVAPFPE	-	38-46	6	0.49	0.01	-
P02662	Gl	FVAPFPE	-	39-46	3	0.42	0.01	-
P02662	Gl	VAPFPE	-	40-46	3	0.45	0.00	-
P02662	Gl	VFGKEKVNE	-	46-55	2	0.45	0.01	-
P02662	Gl	LSKDIGSESTE	-	55-66	5	0.81	0.07	S1_56
P02662	Gl	LSKDIGSESTE	S7(Phospho); S9(Phospho)	55-66	1	0.30	0.00	S1_56
P02662	Gl	LSKDIGSESTEDQAME	S7(Phospho); S9(Phospho); M15(Oxidation)	55-71	1	0.33	0.00	S1_56
P02662	Gl	LSKDIGSESTEDQAMEDIKQME	M15(Oxidation); M21(Oxidation)	55-77	3	1.00	0.00	S1_56
P02662	Gl	LSKDIGSESTEDQAMEDIKQME	M15(Oxidation)	55-77	2	1.00	0.00	S1_56
P02662	Gl	LSKDIGSESTEDQAMEDIKQME	S7(Phospho); S9(Phospho); M15(Oxidation); M21(Oxidation)	55-77	2	0.31	0.05	S1_56
P02662	Gl	DQAMEDIKQME	M4(Oxidation); M10(Oxidation)	66-77	2	0.75	0.10	-
P02662	Gl	AESISSEIIVPNSVE	-	77-93	6	0.87	0.05	S1_79



## 4.4 Supplemental Tables for Chapter 2.4

Protein Accession Number	Protease	Sequence	Modification*	Position	# of q-PSMs	W-Ratio	S.D.	Phosho-Index
P02662	Gl	IVPNSVE	-	86-93	8	0.91	0.04	S1_90
P02662	Gl	IVPNSVE	S5(Phospho)	86-93	4	0.34	0.02	S1_90
P02662	Gl	QKHIQKE	-	93-100	1	0.51	0.00	-
P02662	Gl	KEDVPSERYLGYLE	-	98-112	1	0.42	0.00	-
P02662	Gl	DVPSERYLGYLE	-	100-112	26	0.43	0.01	-
P02662	Gl	RYLGYLE	-	105-112	5	0.45	0.02	-
P02662	Gl	LEIVPNSAEE	S7(Phospho)	124-134	2	0.32	0.02	S1_130
P02662	Gl	LEIVPNSAEERLHSMKE	S7(Phospho); M15(Oxidation)	124-141	6	0.32	0.03	S1_130
P02662	Gl	IVPNSAE	S5(Phospho)	126-133	2	0.37	0.01	S1_130
P02662	Gl	IVPNSAEE	-	126-134	7	0.83	0.06	S1_130
P02662	Gl	IVPNSAEE	S5(Phospho)	126-134	7	0.32	0.01	S1_130
P02662	Gl	IVPNSAEERLHSMKE	M13(Oxidation)	126-141	2	0.86	0.04	S1_130
P02662	Gl	IVPNSAEERLHSMKE	S5(Phospho)	126-141	31	0.32	0.02	S1_130
P02662	Gl	IVPNSAEERLHSMKE	S5(Phospho); M13(Oxidation)	126-141	36	0.32	0.03	S1_130
P02662	Gl	RLHSMKE	M5(Oxidation)	134-141	1	0.72	0.00	-
P02662	Gl	GIHAQQKEPMIGVNQE	M10(Oxidation)	141-157	27	0.43	0.02	-
P02662	Gl	GIHAQQKEPMIGVNQE	-	141-157	16	0.41	0.02	-
P02662	Gl	KMPMIGVNQE	M4(Oxidation)	147-157	2	0.46	0.02	-
P02662	Gl	LAYFYPE	-	157-164	29	0.44	0.03	-
P02662	Gl	LFRQFYQLD	-	164-173	33	0.44	0.02	-
P02662	Gl	AYPSGAWYYVPLGTQYTDAP SFSDIPNPIGSE	-	173-205	4	0.61	0.27	-
P02662	Gl	AYPSGAWYYVPLGTQYTDAP SFSDIPNPIGSENSE	-	173-208	5	0.48	0.36	-
P02662	Tr	HPIKHQGLPQEVLNENLLR	-	19-38	1	0.56	0.00	-
P02662	Tr	HQGLPQEVLNENLLR	-	23-38	41	0.47	0.04	-
P02662	Tr	FFVAPFPEVFGK	-	38-50	78	0.50	0.04	-
P02662	Tr	EKVNELSK	-	50-58	1	0.51	0.00	S1_56
P02662	Tr	EKVNELSKDIGSESTEDQAME DIK	S7(Phospho); S12(Phospho); S14(Phospho); M20(Oxidation)	50-74	4	0.27	0.05	S1_56
P02662	Tr	VNELSKDIGSESTEDQAMED K	S5(Phospho); S10(Phospho); S12(Phospho)	52-74	3	0.28	0.02	S1_56
P02662	Tr	VNELSKDIGSESTEDQAMED K	S5(Phospho); S10(Phospho); S12(Phospho); M18(Oxidation)	52-74	5	0.27	0.02	S1_56
P02662	Tr	VNELSKDIGSESTEDQAMED K	S5(Phospho); S10(Phospho); M18(Oxidation)	52-74	4	0.31	0.03	S1_56
P02662	Tr	VNELSKDIGSESTEDQAMED K	S12(Phospho); T13(Phospho)	52-74	3	0.31	0.02	S1_56
P02662	Tr	DIGSESTEDQAMEDIK	M12(Oxidation)	58-74	6	1.00	0.00	S1_61
P02662	Tr	DIGSESTEDQAMEDIK	-	58-74	1	1.00	0.00	S1_61
P02662	Tr	DIGSESTEDQAMEDIK	S4(Phospho); S6(Phospho); M12(Oxidation)	58-74	10	0.33	0.02	S1_61
P02662	Tr	DIGSESTEDQAMEDIK	S4(Phospho); S6(Phospho)	58-74	12	0.31	0.02	S1_61
P02662	Tr	DIGSESTEDQAMEDIK	S4(Phospho)	58-74	2	0.33	0.01	S1_61
P02662	Tr	DIGSESTEDQAMEDIK	S6(Phospho); M12(Oxidation)	58-74	2	0.37	0.01	S1_61
P02662	Tr	QMEAESISSSEIIVPNSVEQK	-	74-95	1	0.84	0.00	S1_79

## 4 Appendix

Protein Accession Number	Protease	Sequence	Modification*	Position	# of q-PSMs	W-Ratio	S.D.	Phosho-Index
P02662	Tr	HIQKEDVPSEER	-	95-106	27	0.48	0.02	-
P02662	Tr	EDVPSER	-	99-106	9	0.45	0.00	-
P02662	Tr	YLGYLEQLLR	-	106-116	21	0.63	0.05	-
P02662	Tr	YKVPQLEIVPNSAEER	-	119-135	15	0.92	0.05	S1_130
P02662	Tr	YKVPQLEIVPNSAEER	S12(Phospho)	119-135	38	0.33	0.06	S1_130
P02662	Tr	VPQLEIVPNSAEER	-	121-135	13	0.87	0.12	S1_130
P02662	Tr	VPQLEIVPNSAEER	S10(Phospho)	121-135	33	0.34	0.05	S1_130
P02662	Tr	EGIHAQQK	-	140-148	23	0.47	0.02	-
P02662	Tr	EGIHAQQKEPMIGVNVQELAYFYPELFR	M11(Oxidation)	140-167	4	0.64	0.05	-
P02662	Tr	EPMIGVNVQELAYFYPELFR	M3(Oxidation)	148-167	11	0.64	0.05	-
P02662	Tr	EPMIGVNVQELAYFYPELFR	-	148-167	4	0.64	0.02	-
P02662	Tr	FYPELFR	-	160-167	4	0.46	0.01	-

Table 4-5: Overview of identified and quantified peptides obtained by proteolytic digestion of  $\alpha$ -S2-casein

Protein Accession #	Protease	Sequence	Modification*	Position	# of q-PSMs	W-Ratio	S.D.	Phosho-Index
P02663	Ch	KQEKNMAINPSKENL	S11(Phospho); M6(Oxidation)	36-51	1	0.37	0.00	S2_46
P02663	Ch	MAINPSKENL	M1(Oxidation)	41-51	8	0.43	0.03	S2_46
P02663	Ch	MAINPSKENL	-	41-51	1	0.48	0.00	S2_46
P02663	Ch	NEINQFY	-	98-105	13	0.46	0.02	-
P02663	Ch	YQKFPQYL	-	104-112	10	0.46	0.01	-
P02663	Ch	YQKFPQYLQY	-	104-114	23	0.45	0.06	-
P02663	Ch	YQKFPQYLQYLY	-	104-116	3	0.56	0.01	-
P02663	Ch	QKFPQYL	-	105-112	8	0.48	0.02	-
P02663	Ch	QKFPQYLQY	-	105-114	22	0.47	0.02	-
P02663	Ch	QKFPQYLQYLY	-	105-116	3	0.53	0.01	-
P02663	Ch	YLQYLY	-	110-116	1	0.51	0.00	-
P02663	Ch	LYQGPIVL	-	114-122	7	0.46	0.02	-
P02663	Ch	LYQGPIVLNPW	-	114-125	33	0.51	0.05	-
P02663	Ch	YQGPIVLNPW	-	115-125	2	0.55	0.00	-
P02663	Ch	QGPIVLNPW	-	116-125	14	0.48	0.01	-
P02663	Ch	DQVKRNAVPIPTLNREQL	-	125-144	2	0.92	0.00	S2_144
P02663	Ch	NAVPIPTLNREQL	-	130-144	6	0.96	0.06	S2_144
P02663	Ch	NAVPIPTLNREQLSTSEENSK KTVDMESTEVEF	S15(Phospho); S17(Phospho); S21(Phospho); M27(Oxidation)	130-163	7	0.32	0.07	S2_144
P02663	Ch	NAVPIPTLNREQLSTSEENSK KTVDMESTEVEF	T8(Phospho); T16(Phospho); T24(Phospho)	130-163	3	0.30	0.07	S2_144
P02663	Ch	AVPIPTLNREQL	-	131-144	1	0.61	0.00	S2_144
P02663	Ch	AVPIPTLNREQLSTSEENSKK TVDMESTEVEF	S16(Phospho); S20(Phospho); S28(Phospho)	131-163	5	0.30	0.03	S2_144
P02663	Ch	AVPIPTLNREQLSTSEENSKK TVDMESTEVEF	S16(Phospho); S20(Phospho); S28(Phospho); M26(Oxidation)	131-163	9	0.28	0.04	S2_144
P02663	Ch	AVPIPTLNREQLSTSEENSKK TVDMESTEVEF	T15(Phospho); T23(Phospho); M26(Oxidation)	131-163	1	0.36	0.00	S2_144

## 4.4 Supplemental Tables for Chapter 2.4

Protein Accession #	Protease	Sequence	Modification*	Position	# of q-PSMs	W-Ratio	S.D.	Phospho-Index
P02663	Ch	LNREQLSTSEENSKKTVDMES TEVF	S7(Phospho); S9(Phospho); S13(Phospho)	138-163	1	0.35	0.00	S2_144
P02663	Ch	NREQLSTSEENSKKTVDMEST EVF	S6(Phospho); S8(Phospho); S12(Phospho); M18(Oxidation)	139-163	1	0.36	0.00	S2_144
P02663	Ch	STSEENSKKTVDMESTEVF	M13(Oxidation)	144-163	2	0.93	0.08	S2_144
P02663	Ch	TVDMESTEVF	S6(Phospho); M4(Oxidation)	153-163	2	0.37	0.01	S2_158
P02663	Ch	LTEEEKNRLNF	-	168-179	1	0.71	0.00	-
P02663	Ch	TEEEKNRLNF	-	169-179	6	0.45	0.01	-
P02663	Ch	TEEEKNRLNFL	-	169-180	4	0.52	0.04	-
P02663	Ch	FLKKISQRY	-	178-187	2	0.54	0.00	-
P02663	Ch	LKKISQRY	-	179-187	3	0.50	0.04	-
P02663	Ch	YLKTVY	-	194-200	2	0.56	0.01	-
P02663	Ch	KAMKPW	-	203-209	1	0.54	0.00	-
P02663	Ch	IQPKTKVIPY	-	209-219	8	0.48	0.01	-
P02663	El	KNTMEHV	-	16-23	4	0.94	0.04	S2_23
P02663	El	KNTMEHVS	-	16-24	6	0.95	0.05	S2_23
P02663	El	KNTMEHVS	M4(Oxidation)	16-24	14	0.99	0.02	S2_23
P02663	El	KNTMEHVSS	M4(Oxidation)	16-25	3	0.95	0.03	S2_23
P02663	El	KNTMEHVSSEESII	M4(Oxidation)	16-31	2	0.97	0.05	S2_23
P02663	El	SSSEESII	-	23-31	3	0.82	0.04	S2_23
P02663	El	SSEESII	-	24-31	1	0.89	0.00	S2_24
P02663	El	SQETYKQEKMAINPS	M11(Oxidation)	31-47	1	0.91	0.00	S2_31
P02663	El	SQETYKQEKMAINPSKENL CS	S16(Phospho); M11(Oxidation)	31-53	1	1.00	0.00	S2_31
P02663	El	YKQEKMAINPSKENLCS	M7(Oxidation)	35-53	1	0.89	0.00	S2_46
P02663	El	NMAINPS	M2(Oxidation)	40-47	2	0.51	0.00	S2_46
P02663	El	NMAINPSKENLCS	S7(Phospho)	40-53	1	0.63	0.00	S2_46
P02663	El	NMAINPSKENLCS	S7(Phospho); M2(Oxidation)	40-53	2	0.48	0.02	S2_46
P02663	El	MAINPSKENL	-	41-51	1	0.91	0.00	S2_46
P02663	El	MAINPSKENL	S6(Phospho); M1(Oxidation)	41-51	7	0.33	0.01	S2_46
P02663	El	MAINPSKENL	S6(Phospho)	41-51	5	0.34	0.01	S2_46
P02663	El	MAINPSKENLCS	M1(Oxidation)	41-53	5	0.41	0.03	S2_46
P02663	El	MAINPSKENLCS	S6(Phospho); M1(Oxidation)	41-53	8	0.35	0.02	S2_46
P02663	El	MAINPSKENLCS	S6(Phospho)	41-53	4	0.36	0.02	S2_46
P02663	El	AINPSKENL	S5(Phospho)	42-51	3	0.37	0.02	S2_46
P02663	El	AINPSKENLCS	-	42-53	2	0.46	0.01	S2_46
P02663	El	AINPSKENLCS	S5(Phospho)	42-53	10	0.42	0.01	S2_46
P02663	El	INPSKENLCS	S4(Phospho)	43-53	4	0.40	0.01	S2_46
P02663	El	KENLCS	-	47-53	2	0.42	0.02	-
P02663	El	STFCKEVV	-	52-60	6	0.45	0.03	-
P02663	El	TFCKEVV	-	53-60	7	0.43	0.01	-
P02663	El	FCKEVV	-	54-60	2	0.46	0.00	-
P02663	El	RNANEEYSIG	-	60-71	8	0.93	0.03	S2_71
P02663	El	RNANEEYSIG	S9(Phospho)	60-71	1	0.43	0.00	S2_71

#### 4 Appendix

Protein Accession #	Protease	Sequence	Modification*	Position	# of q-PSMs	W-Ratio	S.D.	Phospho-Index
P02663	El	RNANEEEYSIGS	-	60-72	1	0.72	0.00	S2_71
P02663	El	RNANEEEYSIGSSSEESA EVAT EEV	S12(Phospho); S13(Phospho); S14(Phospho); S17(Phospho)	60-85	1	0.00	0.00	S2_71
P02663	El	SIGSSSEESA EVATEEV	S5(Phospho); S6(Phospho); S9(Phospho)	68-85	1	0.27	0.00	S2_71
P02663	El	SSSEESA EVATEEV	-	71-85	8	0.93	0.04	S2_71
P02663	El	SSSEESA EVATEEVKI	T11(Phospho)	71-87	1	0.91	0.00	S2_71
P02663	El	SSEESA EVATEEV	-	72-85	1	0.66	0.00	S2_72
P02663	El	SEESA EVATEEV	-	73-85	4	0.86	0.01	S2_73
P02663	El	KITVDDKHYQKA	-	85-97	12	0.42	0.02	-
P02663	El	TVDDKHYQKA	-	87-97	7	0.43	0.05	-
P02663	El	NQFYQKFPQYL	-	101-112	27	0.46	0.01	-
P02663	El	QYLYQGPI	-	112-120	8	0.45	0.00	-
P02663	El	QYLYQGPIVL	-	112-122	13	0.46	0.03	-
P02663	El	QYLYQGPIVLNPWDQV	-	112-128	21	0.49	0.05	-
P02663	El	YQGPIVL	-	115-122	2	0.48	0.00	-
P02663	El	YQGPIVLNPWDQV	-	115-128	4	0.46	0.01	-
P02663	El	VLNPWDQV	-	120-128	7	0.47	0.01	-
P02663	El	NPWDQV	-	122-128	3	0.47	0.01	-
P02663	El	KRNAVPI	-	128-135	8	0.41	0.01	-
P02663	El	TPTLNREQLSTSEENS	T11(Phospho)	135-151	1	0.50	0.00	S2_144
P02663	El	LNREQLS	-	138-145	7	0.91	0.01	S2_144
P02663	El	LNREQLSTS	-	138-147	4	0.71	0.10	S2_144
P02663	El	LNREQLSTSEENS	S7(Phospho); S9(Phospho)	138-151	3	0.41	0.04	S2_144
P02663	El	LNREQLSTSEENS	S9(Phospho)	138-151	1	0.47	0.00	S2_144
P02663	El	KLTEEEKNRL	-	167-177	3	0.45	0.02	-
P02663	El	KLTEEEKNRLNFL	-	167-180	4	0.46	0.04	-
P02663	El	TEEEKNRL	-	169-177	3	0.44	0.02	-
P02663	El	SQRYQKFA	-	183-191	3	0.50	0.01	-
P02663	El	SQRYQKFALPQYL	-	183-196	5	0.39	0.01	-
P02663	El	VYQHQA	-	198-205	20	0.47	0.02	-
P02663	El	YQHQA	-	199-205	10	0.75	0.01	-
P02663	El	KVIPYV	-	214-220	1	0.52	0.00	-
P02663	Gl	KNTMEHVSSSEE	-	16-28	4	0.95	0.02	S2_23
P02663	Gl	KNTMEHVSSSEE	M4(Oxidation)	16-28	10	0.98	0.03	S2_23
P02663	Gl	KNTMEHVSSSEESIISQE	M4(Oxidation)	16-34	1	0.84	0.00	S2_23
P02663	Gl	KNTMEHVSSSEESIISQE	S8(Phospho); S9(Phospho); S10(Phospho); S13(Phospho); M4(Oxidation)	16-34	13	0.23	0.02	S2_23
P02663	Gl	KNTMEHVSSSEESIISQE	S8(Phospho); S9(Phospho); S10(Phospho); S13(Phospho)	16-34	4	0.24	0.03	S2_23
P02663	Gl	KNTMEHVSSSEESIISQE	S8(Phospho); S9(Phospho); S10(Phospho); M4(Oxidation)	16-34	2	0.24	0.05	S2_23

## 4.4 Supplemental Tables for Chapter 2.4

Protein Accession #	Protease	Sequence	Modification*	Position	# of q-PSMs	W-Ratio	S.D.	Phospho-Index
P02663	Gl	KNTMEHVSSEESIISQETYKQ E	S8(Phospho); S9(Phospho); S10(Phospho); S13(Phospho); M4(Oxidation)	16-39	2	0.19	0.07	S2_23
P02663	Gl	HVSSSEE	-	21-28	1	0.76	0.00	S2_23
P02663	Gl	SIISQE	-	28-34	1	0.77	0.00	S2_31
P02663	Gl	TYKQEKMAINPSKE	S13(Phospho); M8(Oxidation)	34-49	1	0.48	0.00	S2_46
P02663	Gl	KNMAINPSKE	M3(Oxidation)	39-49	22	0.50	0.02	S2_46
P02663	Gl	KNMAINPSKE	-	39-49	1	0.52	0.00	S2_46
P02663	Gl	KNMAINPSKE	S8(Phospho)	39-49	4	0.48	0.01	S2_46
P02663	Gl	KNMAINPSKE	S8(Phospho); M3(Oxidation)	39-49	13	0.46	0.02	S2_46
P02663	Gl	KNMAINPSKENLCSTFCKE	S8(Phospho); M3(Oxidation)	39-58	14	0.45	0.03	S2_46
P02663	Gl	KNMAINPSKENLCSTFCKE	S8(Phospho)	39-58	1	0.47	0.00	S2_46
P02663	Gl	NMAINPSKE	S7(Phospho)	40-49	2	0.46	0.00	S2_46
P02663	Gl	NMAINPSKE	S7(Phospho); M2(Oxidation)	40-49	6	0.46	0.01	S2_46
P02663	Gl	MAINPSKE	S6(Phospho); M1(Oxidation)	41-49	2	0.48	0.01	S2_46
P02663	Gl	NLCSTFCKE	-	49-58	22	0.52	0.01	-
P02663	Gl	VVRNANEE	-	58-66	4	0.70	0.01	-
P02663	Gl	VVRNANEE	-	58-67	20	0.94	0.08	-
P02663	Gl	VVRNANEEYSIGSSSEE	-	58-76	9	1.00	0.06	S2_71
P02663	Gl	VVRNANEEYSIGSSSEESAEV ATEE	S14(Phospho); S15(Phospho); S16(Phospho); S19(Phospho)	58-84	13	0.30	0.17	S2_71
P02663	Gl	VVRNANEEYSIGSSSEESAEV ATEE	S15(Phospho); S16(Phospho); S19(Phospho)	58-84	1	0.38	0.00	S2_71
P02663	Gl	YSIGSSSEE	-	67-76	8	0.78	0.14	S2_71
P02663	Gl	SAEVATEE	-	76-84	5	0.96	0.03	S2_76
P02663	Gl	SAEVATEE	T6(Phospho)	76-84	8	0.96	0.03	S2_76
P02663	Gl	VKITVDDKHYQKALNE	-	84-100	18	0.47	0.03	-
P02663	Gl	QLSTSEE	-	142-149	6	0.87	0.10	S2_144
P02663	Gl	NSKKTVDMESTE	S10(Phospho); M8(Oxidation)	149-161	7	0.32	0.03	S2_158
P02663	Gl	NSKKTVDMESTE	S10(Phospho)	149-161	1	0.34	0.00	S2_158
P02663	Tr	KNTMEHVSSEESIISQETYK	M4(Oxidation)	16-37	5	1.00	0.00	S2_23
P02663	Tr	KNTMEHVSSEESIISQETYK	S8(Phospho); S9(Phospho); S10(Phospho); S13(Phospho)	16-37	1	0.25	0.00	S2_23
P02663	Tr	KNTMEHVSSEESIISQETYK	T3(Phospho); S10(Phospho); S13(Phospho); S16(Phospho); M4(Oxidation)	16-37	4	0.22	0.06	S2_23
P02663	Tr	KNTMEHVSSEESIISQETYKQ EK	M4(Oxidation)	16-40	10	1.00	0.02	S2_23
P02663	Tr	KNTMEHVSSEESIISQETYKQ EK	S8(Phospho); S9(Phospho); S10(Phospho); S13(Phospho); M4(Oxidation)	16-40	8	0.21	0.04	S2_23
P02663	Tr	KNTMEHVSSEESIISQETYKQ EK	S9(Phospho); S10(Phospho); S13(Phospho); S16(Phospho)	16-40	1	0.25	0.00	S2_23

## 4 Appendix

Protein Accession #	Protease	Sequence	Modification*	Position	# of q-PSMs	W-Ratio	S.D.	Phosho-Index
P02663	Tr	NMAINPSK	-	40-48	12	0.56	0.01	S2_46
P02663	Tr	NMAINPSK	M2(Oxidation)	40-48	13	0.54	0.01	S2_46
P02663	Tr	NMAINPSKENLCSTFCK	M2(Oxidation)	40-57	13	0.60	0.01	S2_46
P02663	Tr	NMAINPSKENLCSTFCK	-	40-57	6	0.60	0.02	S2_46
P02663	Tr	NMAINPSKENLCSTFCK	S7(Phospho)	40-57	7	0.54	0.03	S2_46
P02663	Tr	NMAINPSKENLCSTFCK	S7(Phospho); M2(Oxidation)	40-57	17	0.51	0.02	S2_46
P02663	Tr	ENLCSTFCK	-	48-57	19	0.56	0.01	-
P02663	Tr	NANEEEEYSIGSSSEESAEVATE EVK	-	61-86	16	1.00	0.06	S2_71
P02663	Tr	NANEEEEYSIGSSSEESAEVATE EVK	S11(Phospho); S12(Phospho); S13(Phospho); S16(Phospho)	61-86	28	0.25	0.03	S2_71
P02663	Tr	NANEEEEYSIGSSSEESAEVATE EVK	S12(Phospho); S13(Phospho); S16(Phospho)	61-86	2	0.31	0.02	S2_71
P02663	Tr	NANEEEEYSIGSSSEESAEVATE EVK	T21(Phospho)	61-86	1	0.95	0.00	S2_71
P02663	Tr	EESAEVATEEVK	S3(Phospho)	74-86	1	0.33	0.00	S2_76
P02663	Tr	SAEVATEEVK	S1(Phospho)	76-86	2	0.36	0.02	S2_76
P02663	Tr	ITVDDK	-	86-92	4	0.54	0.01	-
P02663	Tr	ITVDDKHYQK	-	86-96	12	0.55	0.02	-
P02663	Tr	ALNEINQFYQK	-	96-107	54	0.54	0.03	-
P02663	Tr	FPQYLQYLYQGPIVLNPWDQ VK	-	107-129	7	0.79	0.03	-
P02663	Tr	LYQGPIVLNPWDQVK	-	114-129	11	0.53	0.03	-
P02663	Tr	QGPIVLNPWDQVK	-	116-129	1	0.49	0.00	-
P02663	Tr	RNAVPIPTLNR	-	129-141	5	0.57	0.02	-
P02663	Tr	RNAVPIPTLNREQLSTSEENS KK	T9(Phospho); T17(Phospho)	129-153	2	0.40	0.04	S2_144
P02663	Tr	NAVPIPTLNR	-	130-141	27	0.57	0.02	-
P02663	Tr	NAVPIPTLNREQLSTSEENSK	S15(Phospho); S17(Phospho)	130-152	2	0.32	0.01	S2_144
P02663	Tr	NAVPIPTLNREQLSTSEENSK K	S15(Phospho); S17(Phospho)	130-153	11	0.32	0.04	S2_144
P02663	Tr	AVPIPTLNR	-	131-141	2	0.56	0.02	-
P02663	Tr	REQLSTSEENSKK	S5(Phospho); S7(Phospho)	140-153	1	0.34	0.00	S2_144
P02663	Tr	EQLSTSEENSK	-	141-152	18	0.97	0.04	S2_144
P02663	Tr	EQLSTSEENSK	S4(Phospho); S6(Phospho)	141-152	7	0.33	0.02	S2_144
P02663	Tr	EQLSTSEENSK	S6(Phospho)	141-152	12	0.54	0.05	S2_144
P02663	Tr	EQLSTSEENSKK	-	141-153	31	0.98	0.01	S2_144
P02663	Tr	EQLSTSEENSKK	S4(Phospho)	141-153	23	0.56	0.03	S2_144
P02663	Tr	EQLSTSEENSKK	S4(Phospho); S6(Phospho)	141-153	26	0.33	0.02	S2_144
P02663	Tr	EQLSTSEENSKKTVDMESTE VFTK	T5(Phospho); T13(Phospho); S18(Phospho); M16(Oxidation)	141-165	8	0.29	0.04	S2_144
P02663	Tr	EQLSTSEENSKKTVDMESTE VFTK	S4(Phospho); S6(Phospho); S10(Phospho)	141-165	2	0.30	0.00	S2_144
P02663	Tr	TVDMESTEVEFTK	-	153-165	13	0.98	0.01	S2_158
P02663	Tr	TVDMESTEVEFTK	M4(Oxidation)	153-165	17	0.87	0.10	S2_158
P02663	Tr	TVDMESTEVEFTK	S6(Phospho)	153-165	24	0.33	0.03	S2_158
P02663	Tr	TVDMESTEVEFTK	T7(Phospho); M4(Oxidation)	153-165	38	0.32	0.03	S2_158
P02663	Tr	TVDMESTEVEFTKK	M4(Oxidation)	153-166	2	0.99	0.00	S2_158
P02663	Tr	TVDMESTEVEFTKK	S6(Phospho)	153-166	4	0.33	0.01	S2_158

## 4.4 Supplemental Tables for Chapter 2.4

Protein Accession #	Protease	Sequence	Modification*	Position	# of q-PSMs	W-Ratio	S.D.	Phospho-Index
P02663	Tr	LTEEEKNR	-	168-176	22	0.57	0.01	-
P02663	Tr	LNFLKK	-	176-182	1	0.60	0.00	-
P02663	Tr	FALPQYLK	-	189-197	28	0.57	0.01	-
P02663	Tr	TVYQHQQ	-	197-204	27	0.56	0.01	-
P02663	Tr	AMKPWIQPK	M2(Oxidation)	204-213	8	0.56	0.02	-
P02663	Tr	AMKPWIQPK	-	204-213	9	0.59	0.03	-

Table 4-6: Overview of identified and quantified peptides obtained by proteolytic digestion of  $\beta$ -casein

Protein Accession #	Protease	Sequence	Modification*	Position	# of q-PSMs	W-Ratio	S.D.	Phospho-Index
P02666	Ch	KIEKFQSEEQQTTEDELQDKI HPF	S7(Phospho)	44-68	5	1.00	0.37	beta_50
P02666	Ch	QSEEQQTTEDELQDKIHPF	-	49-68	2	0.79	0.04	beta_50
P02666	Ch	SLPQNIPPLTQTPVVVPPFLQP E	-	84-107	1	0.49	0.00	-
P02666	Ch	SLPQNIPPLTQTPVVVPPFLQP EVM	M25(Oxidation)	84-109	9	0.45	0.26	-
P02666	Ch	SLPQNIPPLTQTPVVVPPFLQP EVM	-	84-109	5	0.42	0.21	-
P02666	Ch	NIPPLTQTPVVVPPFLQPEVM	M21(Oxidation)	88-109	9	0.45	0.09	-
P02666	Ch	NIPPLTQTPVVVPPFLQPEVM	-	88-109	7	0.46	0.22	-
P02666	Ch	NIPPLTQTPVVVPPFLQPEVM GVS	M21(Oxidation)	88-112	1	0.62	0.00	-
P02666	Ch	HKEMPFKYPVEPFTESSQL	-	121-141	1	0.82	0.00	-
P02666	Ch	SQSLTLTDVE	-	137-147	1	0.50	0.00	-
P02666	Ch	TLTDVENL	-	141-149	4	0.44	0.02	-
P02666	Ch	TLTDVENLHPLPL	-	141-155	21	0.52	0.03	-
P02666	Ch	KAVPYPQRDMPIQAF	M10(Oxidation)	191-206	23	0.42	0.02	-
P02666	Ch	KAVPYPQRDMPIQAF	-	191-206	13	0.41	0.04	-
P02666	Ch	KAVPYPQRDMPIQAFL	M10(Oxidation)	191-207	8	0.41	0.02	-
P02666	Ch	KAVPYPQRDMPIQAFL	-	191-207	4	0.40	0.02	-
P02666	Ch	PQRDMPIQAF	M5(Oxidation)	196-206	2	0.46	0.01	-
P02666	Ch	LLYQEPVLGPVRGPFPI	-	206-223	1	0.47	0.00	-
P02666	Ch	LLYQEPVLGPVRGPFPIIV	-	206-225	13	0.51	0.05	-
P02666	Ch	LYQEPVLGPVRGPFPIIV	-	207-225	7	0.45	0.03	-
P02666	Ch	YQEPVLGPVRGPFPIIV	-	208-225	3	0.46	0.02	-
P02666	Ch	QEPVLGPVRGPFPIIV	-	209-225	7	0.44	0.02	-
P02666	Ch	RGPFPII	-	217-224	2	0.23	0.02	-
P02666	El	RELEELNVPGEIVESL	-	16-32	2	0.84	0.22	beta_30
P02666	El	NVPGEIVESL	-	22-32	4	0.89	0.03	beta_30
P02666	El	NVPGEIVESL	S9(Phospho)	22-32	1	0.38	0.00	beta_30
P02666	El	NKKIEKFQSEEQQTTEDELQD KIHPF	S9(Phospho)	42-68	1	0.26	0.00	beta_50
P02666	El	EKFQSEEQQQ	S5(Phospho)	46-56	2	0.34	0.00	beta_50
P02666	El	EKFQSEEQQTTEDELQDKIHP F	S5(Phospho)	46-68	14	0.31	0.03	beta_50
P02666	El	QSLVYPPFGPI	-	71-82	10	0.38	0.01	-
P02666	El	QSLVYPPFGPIPN	-	71-84	14	0.39	0.04	-
P02666	El	SLVYPPFGPI	-	72-82	8	0.40	0.01	-
P02666	El	LVYPPFGPI	-	73-82	7	0.41	0.03	-
P02666	El	LVYPPFGPIPN	-	73-84	8	0.39	0.01	-
P02666	El	VYPPFGPI	-	74-82	11	0.40	0.02	-

## 4 Appendix

Protein Accession #	Protease	Sequence	Modification*	Position	# of q-PSMs	W-Ratio	S.D.	Phospho-Index
P02666	EI	VYFPGPIP	-	74-84	9	0.38	0.01	-
P02666	EI	YFPGPI	-	75-82	7	0.41	0.02	-
P02666	EI	YFPGPIP	-	75-84	8	0.39	0.01	-
P02666	EI	SLPQNIPPLTQTPV	-	84-98	20	0.39	0.05	-
P02666	EI	SLPQNIPPLTQTPVVVPPFLQP EVMGV	M25(Oxidation)	84-111	2	0.79	0.30	-
P02666	EI	NIPPLTQTPV	-	88-98	7	0.42	0.02	-
P02666	EI	NIPPLTQTPVVVPPFLQPEVM GV	M21(Oxidation)	88-111	7	0.52	0.05	-
P02666	EI	VVPPFLQPEVM	-	98-109	11	0.38	0.02	-
P02666	EI	VVPPFLQPEVMGV	M11(Oxidation)	98-111	24	0.40	0.03	-
P02666	EI	VVPPFLQPEVMGV	-	98-111	4	0.37	0.03	-
P02666	EI	FLQPEVM	-	102-109	2	0.44	0.01	-
P02666	EI	FLQPEVMGV	M7(Oxidation)	102-111	3	0.42	0.00	-
P02666	EI	MAPKHKEMPFKYPVEPF	M1(Oxidation); M8(Oxidation)	117-135	1	0.38	0.00	-
P02666	EI	TESQSLTLTDVENL	-	135-149	2	0.42	0.03	-
P02666	EI	SLTLTDVENL	-	139-149	4	0.39	0.01	-
P02666	EI	TLTDVENL	-	141-149	5	0.46	0.03	-
P02666	EI	LTDVENL	-	142-149	8	0.41	0.01	-
P02666	EI	DVENLHLLPL	-	144-155	5	0.42	0.04	-
P02666	EI	LQSWMHQPHQLPPTV	M5(Oxidation)	155-171	11	0.43	0.05	-
P02666	EI	LQSWMHQPHQLPPTV	-	155-171	11	0.38	0.02	-
P02666	EI	LQSWMHQPHQLPPTVMFPP QSV	M17(Oxidation)	155-178	1	0.42	0.00	-
P02666	EI	SWMHQPHQLPPTV	M3(Oxidation)	157-171	1	0.53	0.00	-
P02666	EI	MFPPQSV	M1(Oxidation)	171-178	1	0.49	0.00	-
P02666	EI	MFPPQSV	-	171-178	20	0.43	0.02	-
P02666	EI	MFPPQSVL	M1(Oxidation)	171-179	11	0.57	0.11	-
P02666	EI	MFPPQSVL	-	171-179	15	0.42	0.02	-
P02666	EI	SQSKVLPVPQ	-	181-191	9	0.38	0.05	-
P02666	EI	QSKVLPVPQ	-	182-191	15	0.36	0.02	-
P02666	EI	KVLPVPQ	-	184-191	3	0.38	0.01	-
P02666	EI	KAVPYPQ	-	191-198	6	0.38	0.01	-
P02666	EI	KAVPYPQRDMPI	M10(Oxidation)	191-203	3	0.41	0.04	-
P02666	EI	KAVPYPQRDMPIQA	M10(Oxidation)	191-205	5	0.36	0.02	-
P02666	EI	KAVPYPQRDMPIQAF	M10(Oxidation)	191-206	1	0.45	0.00	-
P02666	EI	VPYPQRDMPI	M8(Oxidation)	193-203	5	0.40	0.03	-
P02666	EI	VPYPQRDMPI	-	193-203	11	0.37	0.04	-
P02666	EI	VPYPQRDMPIQA	M8(Oxidation)	193-205	3	0.39	0.02	-
P02666	EI	VPYPQRDMPIQA	-	193-205	3	0.36	0.01	-
P02666	EI	RDMPIQA	M3(Oxidation)	198-205	2	0.38	0.01	-
P02666	EI	DMPIQAFLL	M2(Oxidation)	199-208	1	0.46	0.00	-
P02666	EI	FLLYQEPV	-	205-213	14	0.37	0.05	-
P02666	EI	FLLYQEPVLPV	-	205-217	27	0.38	0.03	-
P02666	EI	LLYQEPVLPVVR	-	206-218	2	0.56	0.03	-
P02666	EI	LYQEPVLPV	-	207-217	4	0.40	0.01	-
P02666	EI	RGPFPI	-	217-223	1	0.38	0.00	-
P02666	EI	RGPFPII	-	217-224	1	0.46	0.00	-
P02666	EI	RGPFPIIV	-	217-225	10	0.38	0.09	-
P02666	GI	RELEELNVGPEIVESLSSSEESI TR	-	16-41	1	0.42	0.00	beta_30



## 4.4 Supplemental Tables for Chapter 2.4

Protein Accession #	Protease	Sequence	Modification*	Position	# of q-PSMs	W-Ratio	S.D.	Phospho-Index
P02666	Gl	LNVPGEIVE	-	21-30	8	0.92	0.00	beta_30
P02666	Gl	KFQSEE	S4(Phospho)	47-53	3	0.32	0.01	beta_50
P02666	Gl	KFQSEEQQTE	S4(Phospho)	47-58	2	0.37	0.01	beta_50
P02666	Gl	KFQSEEQQTEDE	-	47-60	4	0.91	0.01	beta_50
P02666	Gl	KFQSEEQQTEDE	S4(Phospho)	47-60	4	0.32	0.00	beta_50
P02666	Gl	LQDKIHFAQTQ	-	60-72	16	0.44	0.01	-
P02666	Gl	SLVYFPFGPIPNLSLPQNIPPLTQ TPVVVPPFLQPE	-	72-107	2	1.00	0.00	-
P02666	Gl	SLPQNIPPLTQTPVVVPPFLQP E	-	84-107	6	0.43	0.02	-
P02666	Gl	NIPPLTQTPVVVPPFLQPE	-	88-107	1	0.42	0.00	-
P02666	Gl	NIPPLTQTPVVVPPFLQPEVM	-	88-109	1	1.00	0.00	-
P02666	Gl	VMGVSKVKE	M2(Oxidation)	107-116	1	0.50	0.00	-
P02666	Gl	VMGVSKVKE	-	107-116	4	0.41	0.01	-
P02666	Gl	AMAPKHKE	-	116-124	4	0.41	0.01	-
P02666	Gl	MPFPKYVPEPFTE	M1(Oxidation)	124-137	2	0.43	0.03	-
P02666	Gl	MPFPKYVPEPFTE	-	124-137	8	0.42	0.01	-
P02666	Gl	SQSLTLTDVE	-	137-147	12	0.45	0.02	-
P02666	Gl	NLHLPLLLQ	-	147-157	13	0.41	0.03	-
P02666	Gl	LQSWMHQPHQLPPTV	M5(Oxidation)	155-171	3	0.44	0.01	-
P02666	Tr	RELEELNVPGEIVESLSSEESI TR	-	16-41	2	1.00	0.00	beta_30
P02666	Tr	RELEELNVPGEIVESLSSEESI TR	S17(Phospho); S18(Phospho); S19(Phospho); S22(Phospho)	16-41	3	0.00	0.14	beta_30
P02666	Tr	IEKFQSEEQQTEDELQDK	S6(Phospho)	45-64	4	0.31	0.03	beta_50
P02666	Tr	FQSEEQQTEDELQDK	-	48-64	4	0.92	0.10	beta_50
P02666	Tr	FQSEEQQTEDELQDK	S3(Phospho)	48-64	17	0.33	0.02	beta_50
P02666	Tr	SLPQNIPPLTQTPVVVPPFLQP EVMGVSK	M25(Oxidation)	84-113	12	0.49	0.06	-
P02666	Tr	SLPQNIPPLTQTPVVVPPFLQP EVMGVSK	-	84-113	1	0.37	0.00	-
P02666	Tr	VKEAMAPK	M5(Oxidation)	113-121	1	0.47	0.00	-
P02666	Tr	EAMAPK	M3(Oxidation)	115-121	2	0.46	0.01	-
P02666	Tr	EAMAPK	-	115-121	2	0.43	0.01	-
P02666	Tr	HKEMPFK	M4(Oxidation)	121-129	2	0.45	0.00	-
P02666	Tr	EMPFK	-	123-129	9	0.43	0.01	-
P02666	Tr	VLPVPQK	-	185-192	13	0.46	0.01	-
P02666	Tr	AVPYPQR	-	192-199	11	0.46	0.00	-
P02666	Tr	DMPIQAFLLYQEPVLPVVR	M2(Oxidation)	199-218	23	0.52	0.08	-
P02666	Tr	DMPIQAFLLYQEPVLPVVR	-	199-218	6	0.52	0.11	-
P02666	Tr	LLYQEPVLPVVR	-	206-218	14	0.45	0.02	-

## 4.5 Figure Index

- Figure 1-1: Mass spectrometry workflow describing the principle steps to obtain an experimental (upper) and theoretical (lower) spectrum. Both spectra are used for peptide identification by a search engine. Protein images used were published under the CC-BY-SA-3.0 and GNU Free Documentation License, respectively (30,31).....2
- Figure 1-2: Summary of absolute (yellow) and relative (orange) quantification methods commonly applied in mass spectrometry based proteome research. Relative quantification can be further subdivided into MS (blue) and MS/MS (green) methods depending at which level relative quantification is performed. Abbreviations: AQUA (Absolute Quantification); SISCAPA (Stable Isotope Standard Capture with Anti-Peptide Antibodies); QconCAT (Quantification Concatamer); IDi (Isotope Dilution); SILAC (Stable Isotope Labeling by/with Amino Acids in Cell Culture); ICPL (Isotope-Coded Protein Label); ICAT (Isotope-Coded Affinity Tag); iTRAQ (Isobaric Tags for Relative and Absolute Quantitation); TMT (Tandem Mass Tags).....4
- Figure 1-3: Chemical structure of iTRAQ-4Plex (upper), TMT-6Plex (middle) and iTRAQ-8Plex (lower) and their corresponding isotopes being used and their resulting masses, respectively (right side). Cleavage sites after MS/MS fragmentation are coded red and green for iTRAQ liberation and iTRAQ + balancer group dissociations which are also observable, respectively. The blue line indicates the leaving group while iTRAQ labeling. The exact structure for iTRAQ-8Plex is not published. Illustration is adapted by (63).....6
- Figure 1-4: Representative MS (left) and MS/MS (right) spectrum showing the combinatorial effect of iTRAQ-4Plex multiplexing. The four samples are non-distinguishable in MS spectra (same precursor m/z due to iTRAQ reporter ion and balancer group). Upon fragmentation, the four iTRAQ labelings liberate unique iTRAQ reporter ions in the low m/z region (demonstrated by orange, blue, green and gray signals).....7
- Figure 1-5: From iTRAQ labeled samples to biological pathways. In iTRAQ-4Plex experiments, the four samples are first reduced, alkylated and digested (usually trypsin), followed by iTRAQ labeling, each with a unique iTRAQ labeling reagent. After labeling, samples are combined and analyzed by 1D- or 2D-LC ESI or MALDI MS. Upon precursor ion fragmentation, reporter ion are formed together with sequence specific fragment ions. Intensities (or areas) of the iTRAQ reporter ions are extracted from MS/MS spectra. Protein identification and quantification results from database search are combined to calculate the significance of differentially abundant proteins. For visualization, protein networks are a common approach to highlight differentially regulated pathways (string-db; (71)). Workflow was adapted by (69) and further modified, images used were published under the CC-BY-SA-3.0 License (31).....9
- Figure 1-6: Schematic diagram of the fundamental parts of a mass spectrometer with the ion source, mass analyzer and detector being part of the instrument and the inlet system and data storage system which are attached by interfaces.....14
- Figure 1-7: Ion formation during the ESI process. According to the polarity of the capillary, either positively or negatively charged analyte ions are obtained. Typically, molecules > 1000 Da are observed as multiple charged ions. Image is adapted by (158,159) and further modified.....16
- Figure 1-8: Schematic diagram of the LTQ Orbitrap Velos equipped with ETD. ©2009 Thermo Fisher Scientific .....17
- Figure 1-9: Nomenclature to describe peptide backbone fragmentation leading to a-, b-, c- and x-, y-, z-ions depending on the charge remaining N- or C-terminal. Illustration is adapted by (84,169).....17
- Figure 1-10: Structure of the commonly observed a-, b-, c- and x-, y- and z-ions after peptide fragmentation.....18
- Figure 1-11: Cross section of the Orbitrap mass analyzer. ©2009 Thermo Fisher Scientific.....22
- Figure 1-12: Schematic overview of a dual fragmentation method utilizing CID-FT and HCD. After a prescan, the five most intense precursor ions are being fragmented by CID and analyzed in the ion trap mass analyzer. In parallel, the full scan is completed in the Orbitrap. After both processes have finished, the same precursor ions are fragmented using HCD and subsequently analyzed in the Orbitrap mass analyzer.....23
- Figure 2-1: Workflow utilized for database search in Proteome Discoverer. The default setting of each node (e.g., Spectrum Files) was used if not otherwise specified: (0) Spectrum Files to load in the .raw files; (1) Spectrum Selector (pre-slicing and filtering of MS and MS/MS data); (2) MS2-Spectrum Processor (deisotope and deconvolute), (3) & (7) Scan Event Filter (to split into CID and HCD fragmentation data); (4) & (8) SEQUEST

HT (search algorithm settings, see below); (5) phosphorS 3.0 (an algorithm to assign phosphorylation sites); (9) Reporter Ions Quantifier (iTRAQ-4Plex (Thermo Scientific Instruments; Integration Tolerance 20 ppm; Most Confident Centroid as Integration Method; Additional Settings: Show the Raw Quan Values; Apply Value Corrections, use All Peptides enabled and deactivated Experimental Bias); (10) Event Detector (filter for precursor ion settings); (11) Precursor Ions Area Detector (reports XICs of identified peptides); and (12) Percolator (for peptide validation and filtering).....	43
Figure 2-2: Offline MS/MS measurements in Orbitrap of the iTRAQ labeled phosphopeptide i115-SpTFHAGQLR. (A) CID, 2+ charged precursor at m/z 620.8; (B) CID, 3+-charged precursor at m/z 414.2. Fragments labeled with an asterisk (*) have lost a charge. The fragment ion b92+-P shares the same mass as the signal derived from the precursor including a neutral loss of the phospho-moiety [M-P] and a subsequent loss of water (18/z).....	47
Figure 2-3: Offline MS/MS measurements in Orbitrap of the iTRAQ labeled phosphopeptide i115-SpTFHAGQLR. (A) CID-MSA, 2+ charged precursor at m/z 620.8; (B) CID-MSA, 3+ charged precursor at m/z 414.2. Fragments labeled with an asterisk (*) have lost a charge.....	48
Figure 2-4: Offline MS/MS measurements in Orbitrap of the iTRAQ labeled phosphopeptide i115-SpTFHAGQLR. (A) HCD, 2+ charged precursor at m/z 620.8; (B) HCD, 3+ charged precursor at m/z 414.2. Fragments labeled with an asterisk (*) have lost a charge. The green labeled signal at m/z 110 (left to iTRAQ-115 (B)) belongs to the histidine immonium ion.....	50
Figure 2-5: Dependence of ion intensities and XCorrs of labeled phosphopeptide i115-SpTFHAGQLR acquired in offline nano-ESI-Orbitrap MS on the NCE. (A) CID, 2+ charged precursor at m/z 620.8; (B) CID 3+ charged precursor at m/z 414.2; (C) CID-MSA, 2+; (D) CID-MSA, 3+; (E) HCD, 2+; (F) HCD, 3+.....	51
Figure 2-6: Offline MS/MS measurement of tow iTRAQ labeled, 2+ charged pY-peptides i115-VpYELMR (A, C, E), m/z 517.7; and i114-VIEDNEpYTAR (B, D, F) at m/z 717.3. (A, B): CID; (C, D): CID-MSA; (E, F): HCD.....	52
Figure 2-7: Comparison of the number of peptide identifications (ID) of mix-2 between correct (blue; lower number) and incorrect mapped phosphorylations sites (orange; upper number) at different NCE values. (A) CID-MSA, 2+ charged precursors; (B) CID-MSA, 3+; (C) HCD, 2+; (D) HCD, 3+. The first number represents correct and the second number the incorrect identifications.....	54
Figure 2-8: Quantitative analysis of the peptide i114-117-SpTFHAGQLR in dependence of NCE applied in HCD measurements; (A) 2+ and (B) 3+ precursor ion charge state.....	55
Figure 2-9: Online LC-ESI MS measurements of the iTRAQ labeled phosphopeptide i114-117-SpTFHAGQLR fragmented by HCD in dependence on the NCE. (A, B) relative intensities; (C, D) normalized intensities; for the 2+ charged precursor m/z 620.8 (A, C) and 3+ charged precursor (B, D) at m/z 414.2.....	56
Figure 2-10: Comparison of normalized iTRAQ reporter ion intensities, the protonated non-dissociated isobaric tag (iTRAQ + balancer group) at m/z 145.10 and the corresponding iTRAQ ratios (reporter ion 114 as denominator) depending on NCEs for HCD fragmentation for the 2+ charged precursors of peptides (A) i114-117-VpYELMR and (B) i114-117 -VIEDNEpYTAR.....	56
Figure 2-11: Online LC-ESI-MS/MS measurements of the iTRAQ labeled phosphopeptide i114-117-GHLpSEGLVTK fragmented by HCD. (A, B) relative intensities; (C, D) normalized intensities; (E, F) the corresponding iTRAQ ratios (reporter ion 114 as denominator), for the 2+ charged precursor at m/z 704.9 (A, C, E) and the 3+ charged precursor (B, D, F) at m/z 470.3.....	58
Figure 2-12: Comparison and overlap of the number of peptide identifications of mix-2 at different NCE. Green bars: peptide identified by both, CID-MSA-IT and HCD; blue bars: peptide only identified by CID; orange bars: peptide only identified by HCD. (A) 2+ charged precursors; (B) 3+ charged precursors.....	59
Figure 2-13: Comparison of online LC-ESI MS measurements of the iTRAQ labeled i114/i116-phosphopeptide mix-2 fragmented by HCD at various NCE. Normalized intensities are shown, averaged over all identified peptides. (A): 51 pS-peptides, 2+ charged; (B): 52 pS-peptides, 3+ charged; (C): 14 pT-peptides, 2+ charged; (D) 15 pT-peptides, 3+ charged. (E): 30 pY-peptides, 2+ charged; (F) 25 pY-peptides, 3+ charged.....	60
Figure 2-14: Comparison of SEQUESTs XCorr (Whiskers-Box-Plots) and the number of unique (x10) and total number of peptides identified via different NCEs applied for CID (A, C, E) and HCD (B, D, F) for 2+ (A, B), 3+ (C, D) and 4+ (E, F) charged precursor ions.....	62

## 4 Appendix

Figure 2-15: Averaged lengths of identified peptides for CID-IT and HCD in dependence of NCE applied for 2+, 3+ and 4+ charged precursor ions (A) and the median raw intensity of iTRAQ reporter ion (B).....	64
Figure 2-16: Whiskers-Box-Plots of iTRAQ reporter ion intensities after HCD fragmentation for the the protein-mix in dependence of NCE applied for 2+ (A, D, G), 3+ (B, E, H) and 4+ (C, F, I) charged precursor ions. (A - C) Normalized Intensities; (D - F) Relative Intensities; (G - I) Intensities (raw values).....	65
Figure 2-17: Whiskers-Box-Plot for both technical replicates TR1 (gray, green) and TR2 (blue, orange) showing the number of MS (gray and blue) and MS/MS scans (green and orange). The diversity of different MS/MS scans was related to the dynamic exclusion settings, as the number of repeating scans was lowered from 3 (TR1) to 1 (TR2) to increase the possibility for selecting low abundant precursor ions and precursors of different charge states, respectively.....	66
Figure 2-18: Total Ion Chromatogram of two technical replicate (TR1 (orange) and TR2 (blue)) demonstrating the effect of shifted retention times. The samples Normalized Collision Energy settings were: CID-IT 20%, HCD 50%.....	67
Figure 2-19: Extracted Ion Chromatograms (XICs) of five different precursor masses (m/z values) being roughly equally distributed via the LC separation from minute 40 to 100 to highlight the extend of retention time shifting. ....	68
Figure 2-20: XY-scatter plot for sample CID 40 – HCD 70 for both the 1st and 2nd technical LC run measured. For sample normalization, a linear fit model was applied, as samples (=peptides) elution profile was shifted in total rather than changing the LC profile. The linear fit function was exported for each of the 36 runs and used for recalculation of retention times.....	69
Figure 2-21: Comparison of SEQUESTs XCorr (Whiskers-Box-Plots) and the number of unique and total number of peptides identified via different Normalized Collision Energies applied for CID (A, C, E) and HCD (B, D, F) for 2+ (A, B), 3+ (C, D) and 4+ (E, F) charged precursor ions.....	71
Figure 2-22: Whiskers-Box-Plots of iTRAQ reporter ion intensities after HCD fragmentation for the the protein-mix in dependence of NCE applied for 2+(A, D, G), 3+ (B, E, H) and 4+ (C, F, I) charged precursor ions. (A - C) Normalized Intensities; (D - F) Relative Intensities; (G - I) Intensities (raw values).....	72
Figure 2-23: The frequency of peptides identified within the bins of -10 to +10 ppm for the samples with increasing content of iTRAQ114 labeling (A, C & E) (left side (A)) and with increasing content of iTRAQ115 (B, D, & F) (right sight (B)).....	80
Figure 2-24: The frequency of peptides identified within the bins of -10 to +10 ppm for the samples iTRAQ-114 to iTRAQ-117 (A) and in-depths analyzing the number of iTRAQ labelings attached for sample iTRAQ-114 (B). ....	80
Figure 2-25: Whiskers-Box-Plots for the ten samples analyzed regarding the in-silico introduced mass shift. Rectangular: median value; triangle: 1%, 99%; x: maximal values. Box-Plots represent values of 25, 50 and 75%.....	81
Figure 2-26: Number of identified peptides for the samples iTRAQ-114 to iTRAQ-117 with a low (0.01 ppm) to a high (20 ppm) predefined precursor mass tolerance.....	82
Figure 2-27: Workflow applied in this study to re-calculate the peptides precursor mass for every spectrum separately. (A) represents the acquired and annotated HCD MS/MS spectra; (B) a zoom-in of the lower m/z region with the four iTRAQ reporter ions and their raw intensities; (C) pie-chart demonstrating the impact of each iTRAQ reporter ion to the sum of all four; (D) the specific iTRAQpart values and their final, newly calculated iTRAQ mass.....	84
Figure 2-28: The frequency of peptides identified within the bins of -10 to +10 ppm for the samples A(1:1:1:1) to F(4:1:1:1) were plotted with (iTRAQcorrected, red color) and without (iTRAQ116&117, blue color) utilization of the re-calculation algorithm.....	85
Figure 2-29: The frequency of peptides identified within the bins of -10 to +10 ppm for the samples iTRAQ-114 to iTRAQ-117 (A - D) were plotted with (iTRAQcorrected, red) and without (iTRAQ116&117, blue) utilization of the re-calculation algorithm.....	86
Figure 2-30: Whiskers-Box-Plots for the ten samples analyzed regarding the in-silico introduced mass shift. Rectangular: median value; triangle: 1%, 99%; x: maximal values. Box-Plots represent values of 25, 50 and 75%.....	87

- Figure 2-31: Structure and m/z value of the immonium ions generated after MS/MS fragmentation for the native (-H), phosphorylated (H<sub>3</sub>PO<sub>4</sub>) and iTRAQ labeled tyrosine.....95
- Figure 2-32: Annotated MS/MS spectra after HCD fragmentation (LTQ Orbitrap Velos) shows the b- and y-ions matched for the peptide VGINYWLAHK in its typical iTRAQ labeled form iVGINYWLAHiK (A) and with an additional iTRAQ labeling at tyrosine iVGINiYWLAHiK (B). The additional iTRAQ modification is observable by a mass increase of +144.10 Da for the b<sub>5+</sub> and y<sub>6+</sub> (and following) ions. At m/z 280.17, the Y-iTRAQ(-4Plex) immonium ion is labeled and only present in the modified form (B). The peaks labeled with an asterisk (\*) belong to the intact precursor ion, which has lost one (charged) iTRAQ labeling (+ balancer group).....96
- Figure 2-33: Annotated MS/MS spectra after CID fragmentation (Synapt G2s, ESI mode) shows the b- and y-ions matched for the peptide VGINYWLAHK in its common iTRAQ labeled form iVGINYWLAHiK (A) and with an additional iTRAQ labeling on the tyrosine iVGINiYWLAHiK (B). The additional iTRAQ modification is observable by a mass increase of +144.10 Da for the b<sub>5+</sub> and y<sub>6+</sub> (and following) ions. For the latter, the Y-iTRAQ immonium is present at m/z 280.17, whereas the tyrosine immonium ion (m/z 136.07) shows a similar intensity for the doubly iTRAQ derivatized form (A).....97
- Figure 2-34: Annotated MS/MS spectra after CID fragmentation (AB SCIEX 5800 MALDI TOF/TOF) with higher (A, B) and medium gas pressure (C, D) for the peptide VGINYWLAHK in its native iTRAQ labeled form iVGINYWLAHiK (A,C) and with the additional iTRAQ labeling on tyrosine iVGINiYWLAHiK (B, D). The Y-iTRAQ immonium at m/z 280.17 is only present in the Y-iTRAQ modified form and more prominent after fragmentation with higher gas pressure (B). Matrix used for MALDI measurements was CHCA.....98
- Figure 2-35: Annotated MS/MS spectra after CID fragmentation (Orbitrap Velos, recorded in ion trap) shows the b- and y-ions matched for the Peptide VGINYWLAHK in its typical iTRAQ labeled form iVGINYWLAHiK (A) and with an additional iTRAQ labeling on the tyrosine iVGINiYWLAHiK (B). The additional iTRAQ modification is observable by a mass increase of +144.10 Da for the b<sub>5+</sub> and y<sub>6+</sub> (and following) ions.....100
- Figure 2-36: Annotated MS/MS spectra after PQD fragmentation (LTQ Orbitrap Velos) shows the b and y-ions matched for the Peptide VGINYWLAHK in its common iTRAQ labeled form iVGINYWLAHiK (A) and with an additional iTRAQ labeling on the tyrosine iVGINiYWLAHiK (B).....101
- Figure 2-37: The annotated MS/MS spectra after ETD fragmentation (Orbitrap Velos, spectra recorded in Orbitrap mass analyzer) for the peptide VGINYWLAHK in its native iTRAQ labeled form iVGINYWLAHiK (A) and with the additional iTRAQ labeling on tyrosine iVGINiYWLAHiK (B).....102
- Figure 2-38: The relative intensity (see Material and Methods for description) was calculated for the precursor [M], the non-dissociated iTRAQ reporter ions (iTRAQ + balancer group), iTRAQ reporter ions iTRAQ114-iTRAQ117, tyrosine iTRAQ-4Plex immonium ion (iY-iTRAQ-4Plex) and the tyrosine immonium ion (iY) at different collision energies (NCEs & CEs) for the peptide VGINYWLAHK. Ion properties were calculated for the common iTRAQ modified (iVGINYWLAHiK, (A, C)) and the Y-iTRAQ modified form (iVGINiYWLAHiK, (B, D)) for 2+ charged precursors. HCD data (LTQ Orbitrap Velos; A,B) and CID data (Synapt G2S; C, D) were compared to demonstrate similar fragmentation pattern.....104
- Figure 2-39: Annotated MS/MS spectra after HCD fragmentation (LTQ Orbitrap Velos) shows the b- and y-ions matched for the Staphylococcus aureus V-8 Protease (GluC) derived peptide LFRQFYQLD (A, B) and the tryptic peptide EPMIGVNQELAYFYPELFR in its common iTRAQ labeled form iLFRQFYQLD (A), iEPMIGVNQELAiYFYPELFR (C) and with an additional iTRAQ labeling on the tyrosine, iLFRQFiYQLD (B), iEPMIGVNQELAiYFYPELFR (D). The additional iTRAQ modification mass shift is obvious for the b<sub>6+</sub> (and following) ions. For the latter, the iY-iTRAQ immonium is present at m/z 440.28, which equals the mass of tyrosine immonium ion (m/z 136.07) plus an iTRAQ-8Plex labeling (+304.20 Da for iTRAQ115, 118, 119 & 121 and +304.21 Da for iTRAQ113, 114, 116 & 117; according to UniMod Accession numbers 730 and 731).  
.....106
- Figure 2-40: Overlay of Extracted Ion Chromatograms (XICs) for three tryptic peptides (iTRAQ-4Plex labeled) being modified on the ε-amino and N-terminal amino group (upper row, unmodified) or carrying an additional modification on tyrosine (lower row, Y-modified).....108
- Figure 2-41: The iTRAQ-4Plex labeled peptide YSLTVAK was identified to be present in three forms: iYSLTVAViK (A), with an additional phosphorylation at the N-terminus i<sub>p</sub>YSLTVAViK (B) and in a N-terminal doubly iTRAQ labeled form i<sub>i</sub>YSLTVAViK (C). For iYSLTVAViK (A), the Y-iTRAQ specific immonium ion at m/z 280.17 is equaling the a<sub>1+</sub> ion. After phosphorylation (B), the pY-iTRAQ immonium ion (= a<sub>1+</sub> ion) was detected at m/z 360.14. Replacing the phospho-moiety with an iTRAQ group, the Y-2x(iTRAQ) immonium ion

## 4 Appendix

was detected at  $m/z$  424.27. All MS/MS spectra were acquired after HCD fragmentation in the Orbitrap mass analyzer. For the latter peptide (i,iYSLTVAViK), the additional iTRAQ modification had an influence on the b- and a- ion series, which were predominantly present in the doubly charged form.....114

Figure 2-42: Peptide fragmentation of the iTRAQ-4Plex labeled peptide YSLTVAVK was shown before (Figure 2-41) for the three observed modification forms: iYSLTVAViK (A), with an additional phosphorylation at the N-terminus i,pYSLTVAViK (B) and in a N-terminal doubly iTRAQ labeled form i,iYSLTVAViK (C). To demonstrate the impact/presence of the observed specific immonium ions, the MS/MS spectra after HCD fragmentation at NCE = 75% are shown. For iYSLTVAVK, the Y-iTRAQ immonium ion was not detectable. For the phosphorylated form, the tyrosine- (iY,  $m/z$  136.07), the phosphotyrosine- (pY,  $m/z$  216.04) and the phosphotyrosine-iTRAQ immonium ion (pY-iTRAQ,  $m/z$  360.14) were detectable. For the doubly iTRAQ labeled N-terminal tyrosine (i,iYSLTVAViK), the tyrosine- (iY,  $m/z$  136.07) and Y-iTRAQ immonium (Y-iTRAQ,  $m/z$  280.17) were present. The doubly iTRAQ labeled immonium ion (Y-2x(iTRAQ),  $m/z$  424.27) was not detectable.....115

Figure 2-43: Schematic diagram of the multi-protease workflow for relative quantification using iTRAQ-8Plex. The control sample ("P"hosphorylated) was directly used for iTRAQ labeling, whereas the treated sample ("D"ephosphorylated) was a 1:1 mixture of the phosphorylated and dephosphorylated sample. This simulated two biologically differently phosphorylated samples. Afterwards, both samples are digested separately using four enzymes (chymotrypsin, elastase, Staphylococcus aureus V-8 Protease (GluC) and trypsin). Each digest is labeled with a unique iTRAQ-8Plex channel. The pooled sample is then enriched for phosphopeptides in a batch process using TiO<sub>2</sub> (ratio 1:20). Samples labeled with gray dashes ("Single Protease Analysis") are only used for method development and not part of the workflow for analysis. All samples were subsequently analyzed by LC-ESI MS.....118

Figure 2-44: For the two workflows using a small FASTA file (sF) and a 10x bigger FASTA file (bF), respectively, the protein sequence coverage with a classical FDR approach and the improvement after percolating the results are demonstrated for the protein coverage (A), the number of unique peptides (B), the PSMs (C) and the protein area (D). For data generation the casein sample after iTRAQ-8Plex labeling and enrichment for phosphorylated peptides was used (sample CF, see workflow). These cover the non-phosphorylated- (NP), wash- (WF) and phosphorylated fraction (PF) and were searched in MudPIT approach. ....126

Figure 2-45: For the two workflows using a small FASTA file (sF) and a 10x bigger FASTA file (bF), respectively, (A) PSMs were split into CID-IT and HCD fragmentation and (B) for non- and phosphorylated peptides. For data generation the casein sample after iTRAQ-8Plex labeling and enrichment for phosphorylated peptides was used (sample CF, see workflow). These cover the non-phosphorylated- (NP), wash- (WF) and phosphorylated Fraction (PF) and were searched in MudPIT approach.....128

Figure 2-46: To visualize the differences in protein coverage (A), the number of unique peptides (B) and the number of PSMs, these values were plotted for the four singly used proteases (chymotrypsin (Ch), elastase (El), GluC (Gl) and trypsin (Tr); each sample D + P in MudPIT approach), for the in-silico combination of the four proteases and the phosphopeptide enrichment fractions (CF).....129

Figure 2-47: Heat-Map in RGB style to visualize the frequency of each amino acid found within all unique peptide groups identified by Proteome Discoverer for  $\alpha$ -S1-casein using the reference workflow. The information about the four singly used proteases chymotrypsin, elastase, GluC and trypsin, the in-silico combined sample (CETG) are compared to the enriched sample (CF). The enriched fraction covers the samples non-phosphorylated- (NP), wash- (WF) and phosphopeptide-fraction (PF). The same data sets (unique peptides) were used. However, for (A) the highest number of amino acid occurrence was identified within each experiment, whereas for (B), this was performed over all experiments. Especially for the phosphorylated amino acids pS56, pS61, pS63 and pS130 the positive effect of phosphopeptide enrichment using TiO<sub>2</sub> is clearly demonstrate able. ....131

Figure 2-48: Heat-Map in RGB style to visualize the frequency of each amino acid found within all unique peptide groups identified by Proteome Discoverer for  $\beta$ -casein using the reference workflow. Here, the number of PSMs within a peptide groups was not taken into account. The information about the four singly used proteases chymotrypsin (Ch), elastase (El), GluC (Gl) and trypsin (Tr), the combination of all four (ME) as well as the in-silico combined sample (CETG) are compared to the enriched sample (CF). The enriched fraction covers the samples non-phosphorylated- (NP), wash- (WF) and phosphopeptide-fraction (PF). The same data sets (unique peptides) were used.....132

Figure 2-49: Heat-Map in RGB style to visualize the frequency of each amino acid found within all unique peptide groups identified by Proteome Discoverer for  $\alpha$ -S2-casein using the reference workflow. Here, the number of PSMs within a peptide groups was not taken into account. The information about the four singly used proteases chymotrypsin (Ch), elastase (El), GluC (Gl) and trypsin (Tr), the combination of all four (ME) as well as the in-silico combined sample (CETG) are compared to the enriched sample (CF). The enriched fraction covers the samples non-phosphorylated- (NP), wash- (WF) and phosphopeptide-fraction (PF). The same data sets (unique peptides) were used.....133

Figure 2-50: Number of unique identified phosphopeptide sequences (A, C, E) and their total Peptide-Spectra-Matches (PSMs; (B, D, F)) for  $\alpha$ -S1-casein (A, B),  $\beta$ -casein (C, D) and  $\alpha$ -S2-casein (E, F) identified by Proteome Discoverer 1.4 using the reference workflow (see Material and Methods for further details). The column diagrams contain the information about the four singly used proteases chymotrypsin (Ch), elastase (El), GluC (Gl) and trypsin (Tr). Additionally, the in-silico combined sample (CETG) is compared to the enriched sample (CF). The enriched fraction covers the samples non-phosphorylated- (NP), wash- (WF) and phosphopeptide-fraction (PF).....135

Figure 2-51: The identified phosphorylated peptides within the protein  $\beta$ -casein (A) and  $\alpha$ -S2-casein (B) were split into protease groups according to the C-terminal amino acid. Chymotrypsin and GluC (Ch-Gl) share two P1' and chymotrypsin and elastase (Ch-El) one P1' cleavage sites, respectively, making them indistinguishable without further bioinformatics tools (147). For the remaining cleavage sites (chymotrypsin, Ch; elastase, El; GluC, Gl; trypsin, Tr) only unique C-terminal cleavage sites are known. Other C-terminal amino acids were also summed up and shown (unspecific, un). The amino acids threonine (un-T) and histidine (un-H) were highly present and shown additionally. For the peptide group 130, all unspecific amino acids were also a part of the Histidine group (un-H).....137

Figure 2-52: The annotated HCD MS/MS spectrum (filter: max fragment ion deviation:  $\pm$  20ppm, signal-intensity > 5% of most intense signal in spectra) for the iTRAQ-8Plex labeled peptide iLEIVPNpSAEERL showing an almost complete sequence coverage (b- and y-ion series combined). Additionally, fragment ions formed by neutral loss events similar to [M-P], were observed (e.g., y6+-P). Although this peptide can be generated by elastase or chymotrypsin cleavage (I & L can be set to equal in database searches, e.g., MaxQuant (300)) the iTRAQ reporter ion intensities (without impurity correction) clearly shows its origin by elastase digestion (intense iTRAQ113 and iTRAQ119 signal).....138

Figure 2-53: All tryptic peptides identified and quantified for  $\alpha$ -S1-casein. After re-grouping of unique peptides the median and the standard deviation for each peptide group was calculated. For dephosphorylated peptides, a w-Ratio towards 1 is expected, whereas for the phosphorylated counterpart a w-Ratio of 0.3 should be observed. ....139

Figure 2-54:  $\alpha$ -S1-casein peptides covering a potential phosphorylation site were extracted from the identified PSM list and further split into the related proteases according to their C-terminus and the iTRAQ reporter ion signals for easier visualization (chymotrypsin in yellow, elastase in cyan, GluC in olive and trypsin in gray). Finally, after re-grouping of unique peptides the median and the standard deviation for each peptide group was calculated. For dephosphorylated peptides, a w-Ratio towards one is expected, whereas for the phosphorylated counterpart a w-Ratio of 0.3 should be observed.....141

Figure 2-55:  $\beta$ -casein peptides covering a potential phosphorylation site were extracted from the identified PSM list and further split into the related proteases according to their C-terminus and the iTRAQ reporter ion signals for easier visualization (chymotrypsin in yellow, elastase in cyan, GluC in olive and trypsin in gray). Finally, after re-grouping of unique peptides the median and the standard deviation for each peptide group was calculated. For dephosphorylated peptides, a w-Ratio towards 1 is expected, whereas for the phosphorylated counterpart a w-Ratio of 0.3 should be observed.....143

Figure 2-56:  $\alpha$ -S2-casein chymotryptic peptides covering a potential phosphorylation site were extracted from the identified PSM list and further split into the related proteases according to their C-terminus and the iTRAQ reporter ion signals. Finally, after re-grouping of unique peptides the median and the standard deviation for each peptide group was calculated. For dephosphorylated peptides, a w-Ratio towards 1 is expected, whereas for the phosphorylated counterpart a w-Ratio of 0.3 should be observed.....144

Figure 2-57:  $\alpha$ -S2-casein elastase derived peptides covering a potential phosphorylation site were extracted from the identified PSM list and further split into the related proteases according to their C-terminus and the iTRAQ reporter ion signals. Finally, after re-grouping of unique peptides the median and the standard deviation for each peptide group was calculated. For dephosphorylated peptides, a w-Ratio towards 1 is expected, whereas for the

## 4 Appendix

- phosphorylated counterpart a w-Ratio of 0.3 should be observed.....145
- Figure 2-58:  $\alpha$ -S2-casein GluC derived peptides covering a potential phosphorylation site were extracted from the identified PSM list and further split into the related proteases according to their C-terminus and the iTRAQ reporter ions signal. Finally, after re-grouping of unique peptides the median and the standard deviation for each peptide group was calculated. For dephosphorylated peptides, a w-Ratio towards 1 is expected, whereas for the phosphorylated counterpart a w-Ratio of 0.3 should be observed.....146
- Figure 2-59:  $\alpha$ -S2-casein trypsin derived peptides covering a potential phosphorylation site were extracted from the identified PSM list and further split into the related proteases according to their C-terminus and the iTRAQ reporter ion signals. Finally, after re-grouping of unique peptides the median and the standard deviation for each peptide group was calculated. For dephosphorylated peptides, a w-Ratio towards 1 is expected, whereas for the phosphorylated counterpart a w-Ratio of 0.3 should be observed.....147
- Figure 3-1: “Transverse CT section through the skull shows an irregular area (actual sampling site) of increased radiographic transparency in the posterior cerebral regions (asterisk). The meninges have become detached from the skull vault and surround the shrunken, inhomogenously disintegrated brain” (310).....154
- Figure 3-2: (A) Sample preparation workflow applied to process both, sample 1024 and 1025, for protein extraction, digestion and sample separation. All samples were analyzed by online LC-ESI MS, with a subsequent multi-dimensional database search. Finally protein identifications were combined for biological evaluation of proteomics data. (B) 1D-SDS-PAGE of two protein extracts from the Iceman’s brain tissue samples. Both samples (approx. 22  $\mu$ g for sample 1024-A lane 1 and approx. 11  $\mu$ g for sample 1025-A lane 2) were separated on a 12% SDS-PAGE and colloidal coomassie stained. (B left side) shows the gel without excision labels; (B right side) shows the gel slices excised for tryptic digestion.....155
- Figure 3-3: Overlap of identified proteins from sample 1024 and 1025 (A), and the number of proteins identified for the in-gel & in-solution digested sample. (B) Overlap between in-gel and in-solution for both samples.....158
- Figure 3-4: The number of proteins identified for sample 1024-A with trypsin specificity: (A) the total number of proteins by each search engine; (B) the overlap of three search engines; (C) the overlap of two search engines; (D) proteins identified with only one search engines. SE: SEQUEST, MA: Mascot, OM: OMSSA, XT: X! Tandem.....159
- Figure 3-5: Overlap of identified proteins with (trypsin) and without enzyme specificity for the two samples (1024 & 1025) digested in-gel and in-solution. (A) 1024-A in-gel, (B) 1024-B in-solution, (C) 1025-A in-gel, (D) 1025-B in-solution.....161
- Figure 3-6: Graphical User Interface (GUI) for data import. The first tab, which is shown here, imports the list of peptide identifications. The remaining tabs are grouped into different criteria (e.g., blank) to separate these and improve the lucidity.....168
- Figure 3-7: Graphical User Interface (GUI) to load in data calculated previously using PeptideMerger. The options for file loading are color-coded (green) and required. Via a pull down menu, the number of amino acids in P and P' positions are selected.....169
- Figure 3-8: Heat-Map created by in-house developed VBA script using minimum (blue) and maximum (red) as delimiter to calculate the appropriate RGB code. Values used as input were created by AminoAcidFinder.....171



## 4.6 Equation Index

Equation 1-1.....	20
Equation 1-2.....	20
Equation 1-3.....	21
Equation 1-4.....	21
Equation 1-5.....	21
Equation 1-6.....	25
Equation 1-7.....	25
Equation 1-8.....	25
Equation 2-1.....	42
Equation 2-2.....	42
Equation 2-3.....	44
Equation 2-4.....	44
Equation 2-5.....	64
Equation 2-6.....	78
Equation 2-7.....	78
Equation 2-8.....	78
Equation 2-9.....	122
Equation 2-10.....	124
Equation 2-11.....	124

## 4.7 Table Index

Table 1-1: Overview of the main groups used to relatively quantify proteins. Metabolic labeling includes techniques such as SILAC. Chemical labeling on protein or peptide level cover techniques like iTRAQ, TMT or ICAT. For SPC, the number of identified spectra is summarized (MS/MS*).....	11
Table 1-2: Overview of phosphopeptide enrichment techniques commonly applied in proteomics (130). Abbreviation: [P] phospho.....	13
Table 2-1: Peptides used for the semi-complex mix-1 (online measurement) with a tandem CID-MSA (ion trap) and HCD (Orbitrap) method. After labeling, the ratio of the four different iTRAQ channels was 1:1:1:1.3 for 114:115:116:117. Peptides in bold were used for offline measurements with CID, CID-MSA and HCD in Orbitrap.....	35
Table 2-2: Peptides in complex mixture (mix-2) used for online measurements with a tandem CID-MSA (ion trap) and HCD (Orbitrap) method. After labeling, the ratio of the two different iTRAQ channels was 1:0.75 for 114:116.....	36
Table 2-2 (continued): Peptides in complex mixture (mix-2) used for online measurements with a tandem CID-MSA (ion trap) and HCD (Orbitrap) method. After labeling, the ratio of the two different iTRAQ channels was 1:0.75 for 114:116. (continued).....	37
Table 2-3: Summary of the iTRAQ labeling applied for the biological sample.....	39
Table 2-4: The linear curve fitting ( $y = A*x + B$ ) parameters A, B and $R^2$ related to each LC run used for retention time recalculation in .mgf files.....	70
Table 2-5: Mixing schema of the ten protein-mix to obtain samples with (simulated) different peptide ratios. iTRAQ-114 to iTRAQ-117 represent the starting material after labeling, which were also measured.....	76
Table 2-6: Median retention times [min] and the number of features (counts) were calculated for the iTRAQ-4Plex labeled ten protein-mix. Based on the total number of Peptide-Spectrum-Matches of 2x 18 LC runs, the retention time differences for tyrosine iTRAQ-4Plex (Y-iTRAQ) labeled and in its native form (unmodified, without iTRAQ modification on tyrosine). Except one peptide (VAGTWYSLAMAASDISLLDAQSAPLR), the retention time for the Y-iTRAQ labeled peptide species was usually lower compared to the native form.....	108
Table 2-6 (continued): Median retention times [min] and the number of features (counts) were calculated for the iTRAQ-4Plex labeled ten protein-mix. Based on the total number of Peptide-Spectrum-Matches of 2x 18 LC runs, the retention time differences for tyrosine iTRAQ-4Plex (Y-iTRAQ) labeled and in its native form (unmodified, without iTRAQ modification on tyrosine). Except one peptide (VAGTWYSLAMAASDISLLDAQSAPLR), the retention time for the Y-iTRAQ labeled peptide species was usually lower compared to the native form.....	109
Table 2-7: Median retention times [min] and the number of features (counts) were calculated for the iTRAQ-8Plex labeled 10 protein-mix. Based on the total number of Peptide-Spectrum-Matches of 9 LC runs, the retention time differences for tyrosine iTRAQ-8Plex (Y-iTRAQ) labeled and in its native form (without iTRAQ modification on tyrosine). Except four peptides (bold), the retention time for the Y-iTRAQ labeled peptide species was usually smaller compared to the native form.....	110
Table 2-7 (continue): Median retention times [min] and the number of features (counts) were calculated for the iTRAQ-8Plex labeled 10 protein-mix. Based on the total number of Peptide-Spectrum-Matches of 9 LC runs, the retention time differences for tyrosine iTRAQ-8Plex (Y-iTRAQ) labeled and in its native form (without iTRAQ modification on tyrosine). Except four peptides (bold), the retention time for the Y-iTRAQ labeled peptide species was usually smaller compared to the native form.....	111
Table 2-8: Each of the two samples (control (P) and treated sample (D)) were digested with one of the four proteases applied and labeled with one of the eight iTRAQ-8Plex reagents, respectively. By shuffling of the iTRAQ reporter ion channels, samples can also be analyzed by HCD (as performed in this study) and ETD, which gives complementary fragment ions (c- and z-ions).....	120
Table 3-1: Number of proteins identified for the in-gel and in-solution digested samples 1024 and 1025 with trypsin (Tr) and without enzyme specification (no-enzyme, NoE). Additionally, the percentage distribution of the search engine hits for each sample is shown (% of total Proteins).....	158

## 4.7 Table Index

Table 3-2: Number of proteins identified for the samples 1024-A and 1025-A with overlapping search engine identifications (4 – 1 Search Engines) and the number of identified proteins by each search engines (last row). SE: SEQUEST, MA: Mascot, OM: OMSSA, XT: X!Tandem.....	160
Table 3-3: Peptides identified from myelin proteolipid protein (PLP). The amino acid in P1 and the C-terminal amino acid were extracted to allow for fully-, semi-, no tryptic cleavage sites, respectively, via the different gel bands (#1 – #19).....	164
Table 3-4: Peptides identified from myelin proteolipid protein (PLP). The amino acid in P1 and the C-terminal amino acid were extracted to allow for fully-, semi-, no tryptic cleavage sites, respectively, via the different gel bands (#20 – #22).....	165
Table 3-5: Protein sequence of $\beta$ -casein. The peptide sequence identified by LC MS is highlighted in orange, its corresponding prime sequence in gray, and the sequence N-terminal to a work-protease cleavage site in cyan. Only the gray (P) and orange (P') labeled sequences are used, whereas the cyan labeled sequence is removed. 170	170
Table 3-6: Tab-separate .txt file containing the frequency of each of the twenty amino acids (AA) from the prime site 5 (P5) to the non-prime site 5 (P5').....	171
Table 4-1: Proteins used for the ten protein-mix.....	II
Table 4-2: The chemical composition, the 3 and 1 letter code and the monoisotopic and average mass for each amino acid.....	VI
Table 4-3: Proteins contained in the cRAP like FASTA file. All proteins used originated from UniProtKB/Swiss-Prot.....	VII
Table 4-3 (continued): Proteins contained in the cRAP like FASTA file. All proteins used originated from UniProtKB/Swiss-Prot.....	VIII
Table 4-4: Overview of identified and quantified peptides obtained by proteolytic digestion of $\alpha$ -S1-casein.....	IX
Table 4-5: Overview of identified and quantified peptides obtained by proteolytic digestion of $\alpha$ -S2-casein....	XIII
Table 4-6: Overview of identified and quantified peptides obtained by proteolytic digestion of $\beta$ -casein.....	XVIII

## 4.8 Copyright

### 4.8.1 Journal of Proteome Research



Type here to search on Elsevier

Advanced search

Follow us: [f](#) [in](#) [t](#) [v](#)

[Help & Contact](#)

[Journals & books](#)

[Solutions](#)

[Authors, editors & reviewers](#)

[About Elsevier](#)

[Community](#)

[Store](#)

#### For Authors

[Journal authors' home](#)

#### Author Rights

[Ethics](#)

[Agreements](#)

[Open access](#)

[Author services](#)

[Early career researchers](#)

[Authors' Update](#)

[Book authors' home](#)

[Sharing your article](#)

[Journal and article metrics](#)

#### Author Rights

Elsevier supports the need for authors to share, disseminate and maximize the impact of their research. We take our responsibility as stewards of the online record seriously, and work to ensure our policies and procedures help to protect the integrity of scholarly works.

Author's rights to reuse and post their own articles published by Elsevier are defined by Elsevier's copyright policy. For our proprietary titles, the type of copyright agreement used depends on the author's choice of publication:

**For subscription articles:** These rights are determined by a copyright transfer, where authors retain scholarly rights to post and use their articles.

**For open access articles:** These rights are determined by an exclusive license agreement, which applies to all our open access content.

In both cases, the fundamental rights needed to publish and distribute an article remain the same and Elsevier authors will be able to use their articles for a wide range of scholarly purposes.

Details on how authors can reuse and post their own articles are provided below.

#### Help and support

For reuse and posting not detailed below, please see our [posting policy](#), or for authors who would like to:

- Include material from other sources in your work being published by Elsevier, please visit: [Permission seeking guidelines for Elsevier authors](#).
- Obtain permission to re-use material from Elsevier books, journals, databases, or other products, please visit: [Obtaining permission to reuse Elsevier material](#)
- Or if you are an Elsevier author and are contacted by a requestor who wishes to re-use all or part of your article or chapter, please also refer them to our [Obtaining Permission to Re-Use Elsevier Material page](#).
- See our [FAQ on posting and copyright queries](#).
- Contact us directly, please email our [Permissions Help Desk](#).

Author Use	Author Posting	Definitions								
<p><b>How authors can use their own journal articles</b></p> <p>Authors can use their articles for a wide range of scholarly, non-commercial purposes as outlined below. These rights apply for all Elsevier authors who publish their article as either a subscription article or an open access article.</p> <p>We require that all Elsevier authors always include a full acknowledgement and, if appropriate, a link to the final published version hosted on Science Direct.</p> <p>For open access articles these rights are separate from how readers can reuse your article as defined by the author's choice of <a href="#">Creative Commons user license options</a>.</p>										
<p><b>Authors can use either their <a href="#">accepted author manuscript</a> or <a href="#">final published article</a> for:</b></p> <table border="1"> <tbody> <tr> <td></td> <td>Use at a conference, meeting or for teaching purposes</td> </tr> <tr> <td></td> <td>Internal training by their company</td> </tr> <tr> <td></td> <td>Sharing individual articles with colleagues for their research use* (also known as 'scholarly sharing')</td> </tr> <tr> <td></td> <td>Use in a subsequent compilation of the author's works</td> </tr> </tbody> </table>				Use at a conference, meeting or for teaching purposes		Internal training by their company		Sharing individual articles with colleagues for their research use* (also known as 'scholarly sharing')		Use in a subsequent compilation of the author's works
	Use at a conference, meeting or for teaching purposes									
	Internal training by their company									
	Sharing individual articles with colleagues for their research use* (also known as 'scholarly sharing')									
	Use in a subsequent compilation of the author's works									

Authors can use either their [accepted author manuscript](#) or [final published article](#) for:

	Inclusion in a thesis or dissertation
	Reuse of portions or extracts from the article in other works
	Preparation of derivative works (other than for <a href="#">commercial purposes</a> )

\*Please note this excludes any [systematic or organized distribution](#) of published articles.



[Choose language](#) [Industries](#) [Advertising](#) [Careers](#) [Feedback](#) [Site Map](#) [Elsevier Websites](#) [A Reed Elsevier Company](#)

Copyright © 2014 Elsevier B.V. [Privacy Policy](#) [Terms & Conditions](#)

Cookies are set by this site. To decline them or learn more, visit our [Cookies page](#).

## 4 Appendix

### 4.8.2 Cellular and Molecular Life Sciences

Rightslink Printable License

https://s100.copyright.com/App/PrintableLicenseFrame.jsp...

#### SPRINGER LICENSE TERMS AND CONDITIONS

Apr 30, 2014

---

---

This is a License Agreement between Dennis Linke ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

License Number	3378810325947
License date	Apr 30, 2014
Licensed content publisher	Springer
Licensed content publication	Cellular and Molecular Life Sciences
Licensed content title	Paleoproteomic study of the Iceman's brain tissue
Licensed content author	Frank Maixner
Licensed content date	Jan 1, 2013
Volume number	70
Issue number	19
Type of Use	Thesis/Dissertation
Portion	Excerpts
Author of this Springer article	Yes and you are a contributor of the new work
Order reference number	None
Title of your thesis / dissertation	Method Development for Qualitative, Quantitative and Computational Proteomics
Expected completion date	Jul 2014
Estimated size(pages)	250
Total	0.00 EUR
Terms and Conditions	

#### Introduction

The publisher for this copyrighted material is Springer Science + Business Media. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing

and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

### Limited License

With reference to your request to reprint in your thesis material on which Springer Science and Business Media control the copyright, permission is granted, free of charge, for the use indicated in your enquiry.

Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process.

This License includes use in an electronic form, provided its password protected or on the university's intranet or repository, including UMI (according to the definition at the Sherpa website: <http://www.sherpa.ac.uk/romeo/>). For any other electronic use, please contact Springer at ([permissions.dordrecht@springer.com](mailto:permissions.dordrecht@springer.com) or [permissions.heidelberg@springer.com](mailto:permissions.heidelberg@springer.com)).

The material can only be used for the purpose of defending your thesis limited to university-use only. If the thesis is going to be published, permission needs to be re-obtained (selecting "book/textbook" as the type of use).

Although Springer holds copyright to the material and is entitled to negotiate on rights, this license is only valid, subject to a courtesy information to the author (address is given with the article/chapter) and provided it concerns original material which does not carry references to other sources (if material in question appears with credit to another source, authorization from that source is required as well).

Permission free of charge on this occasion does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

### Altering/Modifying Material: Not Permitted

You may not alter or modify the material in any manner. Abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of the author(s) and/or Springer Science + Business Media. (Please contact Springer at ([permissions.dordrecht@springer.com](mailto:permissions.dordrecht@springer.com) or [permissions.heidelberg@springer.com](mailto:permissions.heidelberg@springer.com)))

### Reservation of Rights

Springer Science + Business Media reserves all rights not specifically granted in the

## 4 Appendix

Rightslink Printable License

<https://s100.copyright.com/App/PrintableLicenseFrame.jsp...>

combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

### Copyright Notice:Disclaimer

You must include the following copyright and permission notice in connection with any reproduction of the licensed material: "Springer and the original publisher /journal title, volume, year of publication, page, chapter/article title, name(s) of author(s), figure number(s), original copyright notice) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media"

### Warranties: None

Example 1: Springer Science + Business Media makes no representations or warranties with respect to the licensed material.

Example 2: Springer Science + Business Media makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

### Indemnity

You hereby indemnify and agree to hold harmless Springer Science + Business Media and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

### No Transfer of License

This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without Springer Science + Business Media's written permission.

### No Amendment Except in Writing

This license may not be amended except in a writing signed by both parties (or, in the case of Springer Science + Business Media, by CCC on Springer Science + Business Media's behalf).

### Objection to Contrary Terms

Springer Science + Business Media hereby objects to any terms contained in any



purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer Science + Business Media (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

#### Jurisdiction

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in The Netherlands, in accordance with Dutch law, and to be conducted under the Rules of the 'Netherlands Arbitrage Instituut' (Netherlands Institute of Arbitration). **OR:**

**All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of Germany, in accordance with German law.**

#### Other terms and conditions:

##### v1.3

**If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number 501291363. Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.**

**Make Payment To:  
Copyright Clearance Center  
Dept 001  
P.O. Box 843006  
Boston, MA 02284-3006**

**For suggestions or comments regarding this order, contact RightsLink Customer Support: [customer@copyright.com](mailto:customer@copyright.com) or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.**

**Gratis licenses (referencing \$0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.**

### 4.9 References

1. Consortium IHGS. Finishing the euchromatic sequence of the human genome. *Nature*. 2004 Oct 21;431(7011):931–45.
2. Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF, et al. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev*. 1996;13:19–50.
3. Pray LA. *Eukaryotic Genome Complexity*. 2008;1(1).
4. Godovac-Zimmermann J, Brown LR. Perspectives for mass spectrometry and functional proteomics. *Mass Spectrom Rev*. 2001 Feb;20(1):1–57.
5. Arora PS, Yamagiwa H, Srivastava A, Bolander ME, Sarkar G. Comparative evaluation of two two-dimensional gel electrophoresis image analysis software applications using synovial fluids from patients with joint disease. *J Orthop Sci*. 2005 Mar 31;10(2):160–6.
6. Yates JR, Ruse CI, Nakorchevsky A. Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. *Annu Rev Biomed Eng*. 2009;11(1):49–79.
7. Karas M, Bachmann D, Bahr U, Hillenkamp F. Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *Int J Mass Spectrom Ion Process*. 1987 Sep 24;78:53–68.
8. Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem*. 1988 Oct 1;60(20):2299–301.
9. Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, Yoshida T, et al. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom*. 1988 Aug 1;2(8):151–3.
10. Yamashita M, Fenn JB. Electrospray ion source. Another variation on the free-jet theme. *J Phys Chem*. 1984 Sep 1;88(20):4451–9.
11. Whitelegge J. Intact protein mass spectrometry and top-down proteomics. *Expert Rev Proteomics*. 2013 Apr;10(2):127–9.
12. Borchers CH, Thapar R, Petrotchenko EV, Torres MP, Speir JP, Easterling M, et al. Combined top-down and bottom-up proteomics identifies a phosphorylation site in stem-loop-binding proteins that contributes to high-affinity RNA binding. *Proc Natl Acad Sci U S A*. 2006 Feb 28;103(9):3094–9.
13. Han X, Jin M, Breuker K, McLafferty FW. Extending Top-Down Mass Spectrometry to Proteins with Masses Greater Than 200 Kilodaltons. *Science*. 2006 Jun 10;314(5796):109–12.
14. Whitelegge J, Halgand F, Souda P, Zabrouskov V. Top-down mass spectrometry of integral membrane proteins. *Expert Rev Proteomics*. 2006 Dec;3(6):585–96.
15. Zhang Y, Fonslow BR, Shan B, Baek M-C, Yates JR. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem Rev*. 2013 Apr 10;113(4):2343–94.
16. Mostovenko E, Hassan C, Rattke J, Deelder AM, van Veelen PA, Palmblad M. Comparison of peptide and protein fractionation methods in proteomics. *EuPA Open Proteomics*. 2013;1:30–7.
17. Lee MS, Kerns EH. LC/MS applications in drug development. *Mass Spectrom Rev*. 1999 Jan 1;18(3-4):187–279.
18. Nagele E, Vollmer M, Horth P. Improved 2D Nano-LC/MS for Proteomics Applications: A Comparative Analysis Using Yeast Proteome. *J Biomol Tech JBT*. 2004 Jun;15(2):134–43.
19. Davis MT, Beierle J, Bures ET, McGinley MD, Mort J, Robinson JH, et al. Automated LC-LC-MS-MS platform using binary ion-exchange and gradient reversed-phase chromatography for improved proteomic analyses. *J Chromatogr B Biomed Sci App*. 2001 Mar 10;752(2):281–91.
20. Delmotte N, Lasaosa M, Tholey A, Heinzle E, Huber CG. Two-dimensional reversed-phase x ion-pair reversed-phase HPLC: an alternative approach to high-resolution peptide separation for shotgun proteome analysis. *J Proteome Res*. 2007 Nov;6(11):4363–73.

21. Gilar M, Olivova P, Daly AE, Gebler JC. Orthogonality of Separation in Two-Dimensional Liquid Chromatography. *Anal Chem*. 2005 Oct 1;77(19):6426–34.
22. Schirmer EC, Yates JR, Gerace L. MudPIT: A powerful proteomics tool for discovery. *Discov Med*. 2003 Oct;3(18):38–9.
23. Zhang H, Ge Y. Comprehensive Analysis of Protein Modifications by Top-Down Mass Spectrometry. *Circ Cardiovasc Genet*. 2011 Jan 12;4(6):711–711.
24. Piersma SR, Warmoes MO, de Wit M, de Reus I, Knol JC, Jiménez CR. Whole gel processing procedure for GeLC-MS/MS based proteomics. *Proteome Sci*. 2013;11(1):17.
25. López-Ferrer D, Cañas B, Vázquez J, Lodeiro C, Rial-Otero R, Moura I, et al. Sample treatment for protein identification by mass spectrometry-based techniques. *TrAC Trends Anal Chem*. 2006 Nov;25(10):996–1005.
26. 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(3):231–43.
27. O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem*. 1975 May 25;250(10):4007–21.
28. Pappin DJ, Hojrup P, Bleasby AJ. Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol CB*. 1993 Jun 1;3(6):327–32.
29. Parker KC. Scoring methods in MALDI peptide mass fingerprinting: ChemScore, and the ChemApplex program. *J Am Soc Mass Spectrom*. 2002 Jan;13(1):22–39.
30. File:Various AMPs.png [Internet]. Wikipedia, the free encyclopedia. [cited 2014 Apr 8]. Available from: [https://en.wikipedia.org/wiki/File:Various\\_AMPs.png](https://en.wikipedia.org/wiki/File:Various_AMPs.png)
31. Yoshikawa S, Shinzawa-Itoh K, Nakashima R, Yaono R, Yamashita E, Inoue N, et al. Redox-coupled crystal structural changes in bovine heart cytochrome c oxidase. *Science*. 1998 Jul 9;280:1723–9.
32. Meissner F, Mann M. Quantitative shotgun proteomics: considerations for a high-quality workflow in immunology. *Nat Immunol*. 2014 Feb;15(2):112–7.
33. Anderson NL, Taylor J, Scandora AE, Coulter BP, Anderson NG. The TYCHO system for computer analysis of two-dimensional gel electrophoresis patterns. *Clin Chem*. 1981 Nov;27(11):1807–20.
34. Posch A, Franz T, Hartwig S, Knebel B, Al-Hasani H, Passlack W, et al. 2D-ToGo workflow: increasing feasibility and reproducibility of 2-dimensional gel electrophoresis. *Arch Physiol Biochem*. 2013 Jul;119(3):108–13.
35. Unlü M, Morgan ME, Minden JS. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis*. 1997 Oct;18(11):2071–7.
36. Wu WW, Wang G, Baek SJ, Shen R-F. Comparative Study of Three Proteomic Quantitative Methods, DIGE, cICAT, and iTRAQ, Using 2D Gel- or LC-MALDI TOF/TOF. *J Proteome Res*. 2006 Mar 1;5(3):651–8.
37. Jin M, Diaz PT, Bourgeois T, Eng C, Marsh CB, Wu HM. Two-dimensional gel proteome reference map of blood monocytes. *Proteome Sci*. 2006 Sep 1;4(1):16.
38. Beynon RJ, Doherty MK, Pratt JM, Gaskell SJ. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat Methods*. 2005 Aug;2(8):587–9.
39. Zinn N, Hahn B, Pipkorn R, Schwarzer D, Lehmann WD. Phosphorus-based absolutely quantified standard peptides for quantitative proteomics. *J Proteome Res*. 2009 Oct;8(10):4870–5.
40. Schmidt C, Grønberg M, Deckert J, Bessonov S, Conrad T, Lührmann R, et al. Mass spectrometry-based relative quantification of proteins in precatalytic and catalytically active spliceosomes by metabolic labeling (SILAC), chemical labeling (iTRAQ), and label-free spectral count. *RNA* [Internet]. 2014 Jan 21 [cited 2014 Jun 20]; Available from: <http://rnajournal.cshlp.org/content/early/2014/01/20/rna.041244.113>
41. Krijgsveld J, Ketting RF, Mahmoudi T, Johansen J, Artal-Sanz M, Verrijzer CP, et al. Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics. *Nat Biotechnol*. 2003 Aug;21(8):927–31.
42. Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by

#### 4 Appendix

- amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics MCP*. 2002 May;1(5):376–86.
43. Schmidt A, Kellermann J, Lottspeich F. A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics*. 2005 Jan;5(1):4–15.
  44. Marcus K, editor. *Quantitative Methods in Proteomics* [Internet]. Totowa, NJ: Humana Press; 2012 [cited 2014 Jun 20]. Available from: [http://www.springerprotocols.com/Abstract/doi/10.1007/978-1-61779-885-6\\_11](http://www.springerprotocols.com/Abstract/doi/10.1007/978-1-61779-885-6_11)
  45. Shiio Y, Aebersold R. Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry. *Nat Protoc*. 2006 Jun;1(1):139–45.
  46. Wong JWH, Cagney G. An overview of label-free quantitation methods in proteomics by mass spectrometry. *Methods Mol Biol Clifton NJ*. 2010;604:273–83.
  47. Zhou W, Liotta LA, Petricoin EF. The Spectra Count Label-free Quantitation in Cancer Proteomics. *Cancer Genomics Proteomics*. 2012;9(3):135–42.
  48. Lundgren DH, Hwang S-I, Wu L, Han DK. Role of spectral counting in quantitative proteomics. *Expert Rev Proteomics*. 2010 Feb;7(1):39–53.
  49. Zhang Y, Wen Z, Washburn MP, Florens L. Effect of dynamic exclusion duration on spectral count based quantitative proteomics. *Anal Chem*. 2009 Aug 1;81(15):6317–26.
  50. Asara JM, Christofk HR, Freimark LM, Cantley LC. A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *PROTEOMICS*. 2008 Mar 1;8(5):994–9.
  51. Bond NJ, Shliaha PV, Lilley KS, Gatto L. Improving Qualitative and Quantitative Performance for MSE-based Label-free Proteomics. *J Proteome Res*. 2013 Jun 7;12(6):2340–53.
  52. Nagaraj N, Kulak NA, Cox J, Neuhauser N, Mayr K, Hoerning O, et al. System-wide Perturbation Analysis with Nearly Complete Coverage of the Yeast Proteome by Single-shot Ultra HPLC Runs on a Bench Top Orbitrap. *Mol Cell Proteomics*. 2012 Jan 3;11(3):M111.013722.
  53. Nahnsen S, Bielow C, Reinert K, Kohlbacher O. Tools for Label-free Peptide Quantification. *Mol Cell Proteomics MCP*. 2013 Mar;12(3):549–56.
  54. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008 Dec;26(12):1367–72.
  55. Zhu W, Smith JW, Huang C-M. Mass Spectrometry-Based Label-Free Quantitative Proteomics. *BioMed Res Int*. 2009 Nov 10;2010:e840518.
  56. Greis KD, Zhou S, Burt TM, Carr AN, Dolan E, Easwaran V, et al. MALDI-TOF MS as a Label-Free Approach to Rapid Inhibitor Screening. *J Am Soc Mass Spectrom*. 2006 Jun;17(6):815–22.
  57. Benk AS, Roesli C. Label-free quantification using MALDI mass spectrometry: considerations and perspectives. *Anal Bioanal Chem*. 2012 Sep 1;404(4):1039–56.
  58. Molloy MP, Brzezinski EE, Hang J, McDowell MT, VanBogelen RA. Overcoming technical variation and biological variation in quantitative proteomics. *PROTEOMICS*. 2003 Oct 1;3(10):1912–9.
  59. Karp NA, Lilley KS. Design and Analysis Issues in Quantitative Proteomics Studies. *PROTEOMICS*. 2007 Sep 1;7(S1):42–50.
  60. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, et al. Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents. *Mol Cell Proteomics*. 2004 Jan 12;3(12):1154–69.
  61. Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, et al. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal Chem*. 2003 Apr 1;75(8):1895–904.
  62. Evans C, Noirel J, Ow SY, Salim M, Pereira-Medrano AG, Couto N, et al. An insight into iTRAQ: where do we stand now? *Anal Bioanal Chem*. 2012 Sep 1;404(4):1011–27.
  63. Pichler P, Kocher T, Holzmann J, Mazanek M, Taus T, Ammerer G, et al. Peptide Labeling with Isobaric

## 4.9 References

- Tags Yields Higher Identification Rates Using iTRAQ 4-Plex Compared to TMT 6-Plex and iTRAQ 8-Plex on LTQ Orbitrap. *Anal Chem.* 2010 Aug 1;82(15):6549–58.
64. Matthiesen R. Methods, algorithms and tools in computational proteomics: A practical point of view. *PROTEOMICS.* 2007 Aug 1;7(16):2815–32.
  65. Regnier FE, Julka S. Primary amine coding as a path to comparative proteomics. *PROTEOMICS.* 2006 Jul 1;6(14):3968–79.
  66. Matthiesen R, Carvalho AS. Methods and Algorithms for Relative Quantitative Proteomics by Mass Spectrometry. In: Matthiesen R, editor. *Bioinformatics Methods in Clinical Research* [Internet]. Humana Press; 2010 [cited 2014 Apr 1]. p. 187–204. Available from: [http://link.springer.com/protocol/10.1007/978-1-60327-194-3\\_10](http://link.springer.com/protocol/10.1007/978-1-60327-194-3_10)
  67. Olsen JV, Ong S-E, Mann M. Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues. *Mol Cell Proteomics.* 2004 Jan 6;3(6):608–14.
  68. Nabetani T, Makino A, Hullin-Matsuda F, Hirakawa T, Takeoka S, Okino N, et al. Multiplex analysis of sphingolipids using amine-reactive tags (iTRAQ). *J Lipid Res.* 2011 Jun 1;52(6):1294–302.
  69. Evers CE, Gaskell S, Domon B. *Quantitative Proteomics.* S.l.: Royal Soc of Chemistry; 2014. 390 p.
  70. Hill EG, Schwacke JH, Comte-Walters S, Slate EH, Oberg AL, Eckel-Passow JE, et al. A STATISTICAL MODEL FOR iTRAQ DATA ANALYSIS. *J Proteome Res.* 2008 Aug;7(8):3091–101.
  71. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, et al. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* 2009 Jan;37(Database issue):D412–416.
  72. Fraterman S, Zeiger U, Khurana TS, Rubinstein NA, Wilm M. Combination of peptide OFFGEL fractionation and label-free quantitation facilitated proteomics profiling of extraocular muscle. *Proteomics.* 2007 Sep;7(18):3404–16.
  73. Ow SY, Salim M, Noirel J, Evans C, Rehman I, Wright PC. iTRAQ underestimation in simple and complex mixtures: 'the good, the bad and the ugly'. *J Proteome Res.* 2009 Nov;8(11):5347–55.
  74. Gehrig PM, Hunziker PE, Zahariev S, Pongor S. Fragmentation pathways of NG-methylated and unmodified arginine residues in peptides studied by ESI-MS/MS and MALDI-MS. *J Am Soc Mass Spectrom.* 2004 Feb;15(2):142–9.
  75. Zhang R, Regnier FE. Minimizing resolution of isotopically coded peptides in comparative proteomics. *J Proteome Res.* 2002 Apr;1(2):139–47.
  76. Zhang R, Sioma CS, Wang S, Regnier FE. Fractionation of isotopically labeled peptides in quantitative proteomics. *Anal Chem.* 2001 Nov 1;73(21):5142–9.
  77. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem.* 2007 Oct 1;389(4):1017–31.
  78. Hubner NC, Bird AW, Cox J, Splettstoesser B, Bandilla P, Poser I, et al. Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *J Cell Biol.* 2010 May 17;189(4):739–54.
  79. Abdallah C, Sergeant K, Guillier C, Dumas-Gaudot E, Leclercq CC, Renaut J. Optimization of iTRAQ labelling coupled to OFFGEL fractionation as a proteomic workflow to the analysis of microsomal proteins of *Medicago truncatula* roots. *Proteome Sci.* 2012 Jun 6;10(1):37.
  80. Lau E, Lam MPY, Siu SO, Kong RPW, Chan WL, Zhou Z, et al. Combinatorial use of offline SCX and online RP-RP liquid chromatography for iTRAQ-based quantitative proteomics applications. *Mol Biosyst.* 2011 May;7(5):1399–408.
  81. Liu H, Sadygov RG, Yates JR. A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Anal Chem.* 2004 Jul 1;76(14):4193–201.
  82. Luo R, Colangelo CM, Sessa WC, Zhao H. Bayesian Analysis of iTRAQ Data with Nonrandom Missingness: Identification of Differentially Expressed Proteins. *Stat Biosci.* 2009 Nov 1;1(2):228–45.
  83. Luo R, Zhao H. Protein quantitation using iTRAQ: Review on the sources of variations and analysis of nonrandom missingness. *Stat Interface.* 2012 Jan 1;5(1):99–107.

#### 4 Appendix

84. Roepstorff P, Fohlman J. Letter to the editors. *Biol Mass Spectrom.* 1984 Nov 1;11(11):601–601.
85. Olsen JV, Godoy LMF de, Li G, Macek B, Mortensen P, Pesch R, et al. Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. *Mol Cell Proteomics.* 2005 Jan 12;4(12):2010–21.
86. Wenger CD, Lee MV, Hebert AS, McAlister GC, Phanstiel DH, Westphall MS, et al. Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nat Methods.* 2011 Nov;8(11):933–5.
87. Ting L, Rad R, Gygi SP, Haas W. MS3 eliminates ratio distortion in isobaric labeling-based multiplexed quantitative proteomics. *Nat Methods.* 2011 Oct 2;8(11):937–40.
88. Gonzalez-Galarza FF, Qi D, Fan J, Bessant C, Jones AR. A tutorial for software development in quantitative proteomics using PSI standard formats. *Biochim Biophys Acta.* 2014 Jan;1844(1 Pt A):88–97.
89. Dengjel J, Akimov V, Olsen JV, Bunkenborg J, Mann M, Blagoev B, et al. Quantitative proteomic assessment of very early cellular signaling events. *Nat Biotechnol.* 2007 Jun;25(5):566–8.
90. Feramisco JR, Kemp BE, Krebs EG. Phosphorylation of hydroxyproline in a synthetic peptide catalyzed by cyclic AMP-dependent protein kinase. *J Biol Chem.* 1979 Oct 8;254(15):6987–90.
91. Sickmann A, Meyer HE. Phosphoamino acid analysis. *PROTEOMICS.* 2001 Feb 1;1(2):200–6.
92. Lorocho S, Dickhut C, Zahedi RP, Sickmann A. Phosphoproteomics--more than meets the eye. *Electrophoresis.* 2013 Jun;34(11):1483–92.
93. Eyrich B, Sickmann A, Zahedi RP. Catch me if you can: mass spectrometry-based phosphoproteomics and quantification strategies. *Proteomics.* 2011 Feb;11(4):554–70.
94. Hunter T. Signaling--2000 and beyond. *Cell.* 2000 Jan 7;100(1):113–27.
95. Morandell S, Stasyk T, Grosstessner-Hain K, Roitinger E, Mechtler K, Bonn GK, et al. Phosphoproteomics strategies for the functional analysis of signal transduction. *Proteomics.* 2006 Jul;6(14):4047–56.
96. Reinders J, Sickmann A. Modificomics: Posttranslational modifications beyond protein phosphorylation and glycosylation. *Biomol Eng.* 2007 Jun;24(2):169–77.
97. Tholey A, Pipkorn R, Bossemeyer D, Kinzel V, Reed J. Influence of myristoylation, phosphorylation, and deamidation on the structural behavior of the N-terminus of the catalytic subunit of cAMP-dependent protein kinase. *Biochemistry (Mosc).* 2001 Jan 9;40(1):225–31.
98. Shi R, Kumar C, Zougman A, Zhang Y, Podtelejnikov A, Cox J, et al. Analysis of the mouse liver proteome using advanced mass spectrometry. *J Proteome Res.* 2007 Aug;6(8):2963–72.
99. Foster LJ, de Hoog CL, Zhang Y, Zhang Y, Xie X, Mootha VK, et al. A mammalian organelle map by protein correlation profiling. *Cell.* 2006 Apr 7;125(1):187–99.
100. Chen M, Ying W, Song Y, Liu X, Yang B, Wu S, et al. Analysis of human liver proteome using replicate shotgun strategy. *Proteomics.* 2007 Jul;7(14):2479–88.
101. Gilchrist A, Au CE, Hiding J, Bell AW, Fernandez-Rodriguez J, Lesimple S, et al. Quantitative proteomics analysis of the secretory pathway. *Cell.* 2006 Dec 15;127(6):1265–81.
102. Kislinger T, Cox B, Kannan A, Chung C, Hu P, Ignatchenko A, et al. Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell.* 2006 Apr 7;125(1):173–86.
103. Barford D. Molecular mechanisms of the protein serine/threonine phosphatases. *Trends Biochem Sci.* 1996 Nov;21(11):407–12.
104. M Fardilha SLCE. The physiological relevance of protein phosphatase 1 and its interacting proteins to health and disease. *Curr Med Chem.* 2010;17(33):3996–4017.
105. Manning G, Plowman GD, Hunter T, Sudarsanam S. Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci.* 2002 Oct;27(10):514–20.
106. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science.* 2002 Dec 6;298(5600):1912–34.

107. Hunter T. Protein kinase classification. *Methods Enzymol.* 1991;200:3–37.
108. Holmes CFB. A new method for the selective isolation of phosphoserine-containing peptides. *FEBS Lett.* 1987 May 4;215(1):21–4.
109. Erickson AK, Payne DM, Martino PA, Rossomando AJ, Shabanowitz J, Weber MJ, et al. Identification by mass spectrometry of threonine 97 in bovine myelin basic protein as a specific phosphorylation site for mitogen-activated protein kinase. *J Biol Chem.* 1990 Nov 15;265(32):19728–35.
110. Hunter T, Sefton BM. Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. *Proc Natl Acad Sci U S A.* 1980 Mar;77(3):1311–5.
111. Puttick J, Baker EN, Delbaere LTJ. Histidine phosphorylation in biological systems. *Biochim Biophys Acta.* 2008 Jan;1784(1):100–5.
112. Cieśla J, Frączyk T, Rode W. Phosphorylation of basic amino acid residues in proteins: important but easily missed. *Acta Biochim Pol.* 2011;58(2):137–48.
113. Olsen JV, Blagoev B, Gnäd F, Macek B, Kumar C, Mortensen P, et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell.* 2006 Nov 3;127(3):635–48.
114. Zhang G, Neubert TA. Comparison of Three Quantitative Phosphoproteomic Strategies to Study Receptor Tyrosine Kinase Signaling. *J Proteome Res.* 2011 Dec 2;10(12):5454–62.
115. Engholm-Keller K, Larsen MR. Technologies and challenges in large-scale phosphoproteomics. *PROTEOMICS.* 2013 Mar 1;13(6):910–31.
116. Schmelzle K, White FM. Phosphoproteomic approaches to elucidate cellular signaling networks. *Curr Opin Biotechnol.* 2006 Aug;17(4):406–14.
117. Steen H, Kuster B, Fernandez M, Pandey A, Mann M. Tyrosine phosphorylation mapping of the epidermal growth factor receptor signaling pathway. *J Biol Chem.* 2002 Jan 11;277(2):1031–9.
118. Andersson L, Porath J. Isolation of phosphoproteins by immobilized metal (Fe<sup>3+</sup>) affinity chromatography. *Anal Biochem.* 1986 Apr;154(1):250–4.
119. Sano A, Nakamura H. Titania as a chemo-affinity support for the column-switching HPLC analysis of phosphopeptides: application to the characterization of phosphorylation sites in proteins by combination with protease digestion and electrospray ionization mass spectrometry. *Anal Sci Int J Jpn Soc Anal Chem.* 2004 May;20(5):861–4.
120. Pinkse MWH, Uitto PM, Hilhorst MJ, Ooms B, Heck AJR. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem.* 2004 Jul 15;76(14):3935–43.
121. Kokubu M, Ishihama Y, Sato T, Nagasu T, Oda Y. Specificity of immobilized metal affinity-based IMAC/C18 tip enrichment of phosphopeptides for protein phosphorylation analysis. *Anal Chem.* 2005 Aug 15;77(16):5144–54.
122. Thingholm TE, Jensen ON, Robinson PJ, Larsen MR. SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. *Mol Cell Proteomics MCP.* 2008 Apr;7(4):661–71.
123. Thingholm TE, Jørgensen TJD, Jensen ON, Larsen MR. Highly selective enrichment of phosphorylated peptides using titanium dioxide. *Nat Protoc.* 2006;1(4):1929–35.
124. Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, Jørgensen TJD. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics MCP.* 2005 Jul;4(7):873–86.
125. Jensen SS, Larsen MR. Evaluation of the impact of some experimental procedures on different phosphopeptide enrichment techniques. *Rapid Commun Mass Spectrom RCM.* 2007;21(22):3635–45.
126. Endler A, Reiland S, Gerrits B, Schmidt UG, Baginsky S, Martinoia E. In vivo phosphorylation sites of barley tonoplast proteins identified by a phosphoproteomic approach. *Proteomics.* 2009 Jan;9(2):310–21.
127. Villén J, Beausoleil SA, Gerber SA, Gygi SP. Large-scale phosphorylation analysis of mouse liver. *Proc Natl Acad Sci U S A.* 2007 Jan 30;104(5):1488–93.

#### 4 Appendix

128. Wagner PD, Vu ND. Histidine to aspartate phosphotransferase activity of nm23 proteins: phosphorylation of aldolase C on Asp-319. *Biochem J.* 2000 Mar 15;346 Pt 3:623–30.
129. Ficarro SB, Zhang Y, Carrasco-Alfonso MJ, Garg B, Adelmant G, Webber JT, et al. Online nanoflow multidimensional fractionation for high efficiency phosphopeptide analysis. *Mol Cell Proteomics MCP.* 2011 Nov;10(11):O111.011064.
130. Beltran L, Cutillas PR. Advances in phosphopeptide enrichment techniques for phosphoproteomics. *Amino Acids.* 2012 Sep 1;43(3):1009–24.
131. Amorim Madeira PJ, A. P, M. C. High Resolution Mass Spectrometry Using FTICR and Orbitrap Instruments. In: Salih S, editor. *Fourier Transform - Materials Analysis* [Internet]. InTech; 2012 [cited 2014 Apr 2]. Available from: <http://www.intechopen.com/books/fourier-transform-materials-analysis/high-resolution-mass-spectrometry-using-fticr-and-orbitrap-instruments>
132. Wien K. 100 years of ion beams: Willy Wien's canal rays. *Braz J Phys.* 1999 Sep;29(3):401–14.
133. Thomson JJ. XL. Cathode Rays. *Philos Mag Ser 5.* 1897;44(269):293–316.
134. Thomson JJ. XLVII. On rays of positive electricity. *Philos Mag Ser 6.* 1907;13(77):561–75.
135. *Proceedings of the American Physical Society.* *Phys Rev.* 1946 Jun 1;69(11-12):674–674.
136. Sommer H, Thomas HA, Hipple JA. The Measurement of eM by Cyclotron Resonance. *Phys Rev.* 1951 Jun 1;82(5):697–702.
137. Paul W. Electromagnetic traps for charged and neutral particles. *Rev Mod Phys.* 1990 Jul 1;62(3):531–40.
138. Ryhage R. Use of a Mass Spectrometer as a Detector and Analyzer for Effluent Emerging from High Temperature Gas Liquid Chromatography Columns. *Anal Chem.* 1964 Apr 1;36(4):759–64.
139. Watson JT, Biemann K. Direct recording of high resolution mass spectra of gas chromatographic effluents. 1965. *J Mass Spectrom JMS.* 1998 Feb;33(2):109–17.
140. Morris HR, Panico M, Barber M, Bordoli RS, Sedgwick RD, Tyler A. Fast atom bombardment: A new mass spectrometric method for peptide sequence analysis. *Biochem Biophys Res Commun.* 1981 Jul 30;101(2):623–31.
141. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science.* 1989 Oct 6;246(4926):64–71.
142. Jennings KR. Collision-induced decompositions of aromatic molecular ions. *Int J Mass Spectrom Ion Phys.* 1968 Aug;1(3):227–35.
143. Makarov A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal Chem.* 2000 Mar 1;72(6):1156–62.
144. Hager JW. A new linear ion trap mass spectrometer. *Rapid Commun Mass Spectrom.* 2002 Mar 30;16(6):512–26.
145. Bodnar WM, Blackburn RK, Krise JM, Moseley MA. Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage. *J Am Soc Mass Spectrom.* 2003 Sep;14(9):971–9.
146. Stapels MD, Barofsky DF. Complementary Use of MALDI and ESI for the HPLC-MS/MS Analysis of DNA-Binding Proteins. *Anal Chem.* 2004 Sep 1;76(18):5423–30.
147. Tucher J, Linke D, Koudelka T, Cassidy L, Tredup C, Wichert R, et al. LC-MS Based Cleavage Site Profiling of the Proteases ADAM10 and ADAM17 Using Proteome-Derived Peptide Libraries. *J Proteome Res.* 2014 Mar 17;
148. Ham BM. Posttranslational Modification (PTM) of Proteins. *Proteomics of Biological Systems* [Internet]. John Wiley & Sons, Inc.; 2011 [cited 2014 Jun 23]. p. 1–58. Available from: <http://onlinelibrary.wiley.com/doi/10.1002/9781118137048.ch1/summary>
149. *Encyclopedia of Analytical Chemistry* [Internet]. [cited 2014 Apr 2]. Available from: <http://onlinelibrary.wiley.com/book/10.1002/9780470027318>
150. Cotter RJ. Time-of-flight mass spectrometry for the structural analysis of biological molecules. *Anal*



- Chem. 1992 Nov 1;64(21):1027A–1039A.
151. Chait BT, Kent SB. Weighing naked proteins: practical, high-accuracy mass measurement of peptides and proteins. *Science*. 1992 Sep 25;257(5078):1885–94.
  152. Hillenkamp F, Karas M, Beavis RC, Chait BT. Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers. *Anal Chem*. 1991 Dec 1;63(24):1193A–1203A.
  153. Karas M, Glückmann M, Schäfer J. Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *J Mass Spectrom JMS*. 2000 Jan;35(1):1–12.
  154. Kononikhin AS, Nikolaev EN, Frankevich V, Zenobi R. Multiply charged ions in matrix-assisted laser desorption/ionization generated from electrosprayed sample layers. *Eur J Mass Spectrom Chichester Engl*. 2005;11(3):257–9.
  155. Trimpin S, Inutan ED, Herath TN, McEwen CN. Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Method for Selectively Producing Either Singly or Multiply Charged Molecular Ions. *Anal Chem*. 2010 Jan 1;82(1):11–5.
  156. Lee J, Hong J, Kim T, Kim J. Optimization in Detecting Multiply-charged Protein Ions using MALDI TOF-MS. *Mass Spectrom Lett*. 2013 Mar 29;4(1):21–3.
  157. Szájli E, Fehér T, Medzihradzky KF. Investigating the Quantitative Nature of MALDI-TOF MS. *Mol Cell Proteomics*. 2008 Jan 12;7(12):2410–8.
  158. Lehmann WD. *Massenspektrometrie in der Biochemie*. Spektrum, Akad. Verlag; 1996. 419 p.
  159. Banerjee S, Mazumdar S. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *Int J Anal Chem [Internet]*. 2012 Mar 12 [cited 2014 Apr 2];2012. Available from: <http://www.hindawi.com/journals/ijac/2012/282574/abs/>
  160. Labowsky M, Whitehouse C, Fenn JB. Three-dimensional deconvolution of multiply charged spectra. *Rapid Commun Mass Spectrom*. 1993 Jan 1;7(1):71–84.
  161. Wang G, Cole RB. Effect of Solution Ionic Strength on Analyte Charge State Distributions in Positive and Negative Ion Electrospray Mass Spectrometry. *Anal Chem*. 1994 Nov 1;66(21):3702–8.
  162. Guevremont R, Siu KW, Le Blanc JC, Berman SS. Are the electrospray mass spectra of proteins related to their aqueous solution chemistry? *J Am Soc Mass Spectrom*. 1992 Mar;3(3):216–24.
  163. Kebarle P, Tang L. From ions in solution to ions in the gas phase - the mechanism of electrospray mass spectrometry. *Anal Chem*. 1993 Nov 1;65(22):972A–986A.
  164. Cole RB. Some tenets pertaining to electrospray ionization mass spectrometry. *J Mass Spectrom*. 2000 Jul 1;35(7):763–72.
  165. Cech NB, Enke CG. Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom Rev*. 2001 Jan 1;20(6):362–87.
  166. Frese CK, Altelaar AFM, Hennrich ML, Nolting D, Zeller M, Griep-Raming J, et al. Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J Proteome Res*. 2011 May 6;10(5):2377–88.
  167. Tonya Pekar Second JDB. Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures. *Anal Chem*. 2009;81(18):7757–65.
  168. Thermo Scientific. Orbitrap Velos Training Course Material. Thermo Fisher Scientific;
  169. Johnson RS, Martin SA, Biemann K, Stults JT, Watson JT. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal Chem*. 1987 Nov 1;59(21):2621–5.
  170. Bender ML, Ladenheim H, Chen MC. Acylium Ion Formation in the Reactions of Carboxylic Acid Derivatives. II. The Hydrolysis and Oxygen Exchange of Methyl Mesitoate in Sulfuric Acid1a. *J Am Chem Soc*. 1961 Jan 1;83(1):123–7.
  171. Wysocki VH, Tsaprailis G, Smith LL, Brei LA. Mobile and localized protons: a framework for understanding peptide dissociation. *J Mass Spectrom*. 2000 Dec 1;35(12):1399–406.

#### 4 Appendix

172. Paizs B, Suhai S. Fragmentation pathways of protonated peptides. *Mass Spectrom Rev.* 2005 Jul 1;24(4):508–48.
173. Hunt DF, Yates JR 3rd, Shabanowitz J, Winston S, Hauer CR. Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci U S A.* 1986 Sep;83(17):6233–7.
174. Wells JM, McLuckey SA. Collision-induced dissociation (CID) of peptides and proteins. *Methods Enzymol.* 2005;402:148–85.
175. Soares R, Pires E, M. A, Santos R, Gomes R, Koi K, et al. Tandem Mass Spectrometry of Peptides. In: Prasain J, editor. *Tandem Mass Spectrometry - Applications and Principles* [Internet]. InTech; 2012 [cited 2014 Apr 3]. Available from: <http://www.intechopen.com/books/tandem-mass-spectrometry-applications-and-principles/tandem-mass-spectrometry-of-peptides>
176. Gogichaeva NV, Williams T, Alterman MA. MALDI TOF/TOF tandem mass spectrometry as a new tool for amino acid analysis. *J Am Soc Mass Spectrom.* 2007 Feb;18(2):279–84.
177. Loïc Dayon CP. Combining low- and high-energy tandem mass spectra for optimized peptide quantification with isobaric tags. *J Proteomics.* 2009;73(4):769–77.
178. Wu WW, Wang G, Insel PA, Hsiao C-T, Zou S, Maudsley S, et al. Identification of Proteins and Phosphoproteins Using Pulsed Q Collision Induced Dissociation (PQD). *J Am Soc Mass Spectrom.* 2011 Oct;22(10):1753–62.
179. 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 MCP.* 2008 Sep;7(9):1702–13.
180. Michalski A, Neuhauser N, Cox J, Mann M. A Systematic Investigation into the Nature of Tryptic HCD Spectra. *J Proteome Res.* 2012 Nov 2;11(11):5479–91.
181. Boersema PJ, Mohammed S, Heck AJR. Phosphopeptide fragmentation and analysis by mass spectrometry. *J Mass Spectrom JMS.* 2009 Jun;44(6):861–78.
182. Summerfield SG, Whiting A, Gaskell SJ. Intra-ionic interactions in electrosprayed peptide ions. *Int J Mass Spectrom Ion Process.* 1997 Mar;162(1–3):149–61.
183. Dongré AR, Jones JL, Somogyi Á, Wysocki VH. Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *J Am Chem Soc.* 1996 Jan 1;118(35):8365–74.
184. Griffin TJ, Xie H, Bandhakavi S, Popko J, Mohan A, Carlis JV, et al. iTRAQ reagent-based quantitative proteomic analysis on a linear ion trap mass spectrometer. *J Proteome Res.* 2007 Nov;6(11):4200–9.
185. McLuckey SA, Goeringer DE. SPECIAL FEATURE:TUTORIAL Slow Heating Methods in Tandem Mass Spectrometry. *J Mass Spectrom.* 1997 May 1;32(5):461–74.
186. Want EJ, Cravatt BF, Siuzdak G. The expanding role of mass spectrometry in metabolite profiling and characterization. *ChemBiochem Eur J Chem Biol.* 2005 Nov;6(11):1941–51.
187. Schwartz JC, Senko MW, Syka JEP. A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom.* 2002 Jun;13(6):659–69.
188. Jones AW, Cooper HJ. Dissociation techniques in mass spectrometry-based proteomics. *Analyst.* 2011 Aug 8;136(17):3419–29.
189. Xia Y, Liang X, McLuckey SA. Ion trap versus low-energy beam-type collision-induced dissociation of protonated ubiquitin ions. *Anal Chem.* 2006 Feb 15;78(4):1218–27.
190. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods.* 2007 Sep;4(9):709–12.
191. McAlister GC, Phanstiel DH, Brumbaugh J, Westphall MS, Coon JJ. Higher-energy collision-activated dissociation without a dedicated collision cell. *Mol Cell Proteomics MCP.* 2011 May;10(5):O111.009456.
192. Horner J, Remes P, Biringer R, Huhmer A, Specht A. Achieving Increased Coverage in Global Proteomics Survey Experiments Using Higher-Energy Collisional Dissociation (HCD) on a Linear Ion Trap Mass Spectrometer. Thermo Fisher Scientific; 2011.

## 4.9 References

193. Pulsed Q Collision Induced Dissociation (PQD) on Linear Ion Trap Mass Spectrometers. Thermo Fisher Scientific;
194. Bakhtiar R, Guan Z. Electron capture dissociation mass spectrometry in characterization of post-translational modifications. *Biochem Biophys Res Commun*. 2005 Aug 19;334(1):1–8.
195. Zubarev RA. Electron-capture dissociation tandem mass spectrometry. *Curr Opin Biotechnol*. 2004 Feb;15(1):12–6.
196. Axelsson J, Palmblad M, Håkansson K, Håkansson P. Electron capture dissociation of substance P using a commercially available Fourier transform ion cyclotron resonance mass spectrometer. *Rapid Commun Mass Spectrom*. 1999 Mar 30;13(6):474–7.
197. Zubarev RA, Horn DM, Fridriksson EK, Kelleher NL, Kruger NA, Lewis MA, et al. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem*. 2000 Feb 1;72(3):563–73.
198. Breuker K, Oh H, Cerda B, Horn D, McLafferty F. Hydrogen atom loss in electron-capture dissociation: a Fourier transform-ion cyclotron resonance study with single isotopomeric ubiquitin ions. *Eur J Mass Spectrom*. 2002;8(1):177.
199. Mikesch LM, Ueberheide B, Chi A, Coon JJ, Syka JEP, Shabanowitz J, et al. The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta*. 2006 Dec;1764(12):1811–22.
200. Swaney DL, McAlister GC, Wirtala M, Schwartz JC, Syka JEP, Coon JJ. Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal Chem*. 2007 Jan 15;79(2):477–85.
201. Molina H, Matthiesen R, Kandasamy K, Pandey A. Comprehensive Comparison of Collision Induced Dissociation and Electron Transfer Dissociation. *Anal Chem*. 2008 Jul 1;80(13):4825–35.
202. Swaney DL, McAlister GC, Coon JJ. Decision Tree-Driven Tandem Mass Spectrometry for Shotgun Proteomics. *Nat Methods*. 2008 Nov;5(11):959–64.
203. Yufeng Shen NT. Effectiveness of CID, HCD, and ETD with FT MS/MS for degradomic-peptidomic analysis: comparison of peptide identification methods. *J Proteome Res*. 2011;10(9):3929–43.
204. Lottspeich F, Engels JW. *Bioanalytik*. Elsevier, Spektrum Akad. Verlag; 2006. 1119 p.
205. Yates JR 3rd. Mass spectral analysis in proteomics. *Annu Rev Biophys Biomol Struct*. 2004;33:297–316.
206. Roboz J. *Mass Spectrometry in Cancer Research*. CRC Press; 2002. 580 p.
207. March RE, Todd JFJ. *Quadrupole Storage Mass Spectrometry*. Wiley; 1989. 504 p.
208. Douglas DJ, Frank AJ, Mao D. Linear ion traps in mass spectrometry. *Mass Spectrom Rev*. 2005 Jan 1;24(1):1–29.
209. Kingdon KH. A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Phys Rev*. 1923 Apr 1;21(4):408–18.
210. Perry RH, Cooks RG, Noll RJ. Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrom Rev*. 2008 Dec;27(6):661–99.
211. Köcher T, Pichler P, Schutzbier M, Stingl C, Kaul A, Teucher N, et al. High precision quantitative proteomics using iTRAQ on an LTQ Orbitrap: a new mass spectrometric method combining the benefits of all. *J Proteome Res*. 2009 Oct;8(10):4743–52.
212. McAlister GC, Phanstiel D, Wenger CD, Lee MV, Coon JJ. Analysis of tandem mass spectra by FTMS for improved large-scale proteomics with superior protein quantification. *Anal Chem*. 2010 Jan 1;82(1):316–22.
213. Makarov A, Denisov E, Kholomeev A, Balschun W, Lange O, Strupat K, et al. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem*. 2006 Apr 1;78(7):2113–20.
214. Scherl A, Shaffer SA, Taylor GK, Hernandez P, Appel RD, Binz P-A, et al. On the Benefits of Acquiring Peptide Fragment Ions at High Measured Mass Accuracy. *J Am Soc Mass Spectrom*. 2008 Jun;19(6):891–901.

#### 4 Appendix

215. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom.* 1994 Nov 1;5(11):976–89.
216. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *ELECTROPHORESIS.* 1999 Dec 1;20(18):3551–67.
217. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, et al. Open Mass Spectrometry Search Algorithm. *J Proteome Res.* 2004 Oct 1;3(5):958–64.
218. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. *Bioinforma Oxf Engl.* 2004 Jun 12;20(9):1466–7.
219. Balgley BM, Laudeman T, Yang L, Song T, Lee CS. Comparative evaluation of tandem MS search algorithms using a target-decoy search strategy. *Mol Cell Proteomics MCP.* 2007 Sep;6(9):1599–608.
220. Henzel WJ, Billeci TM, Stults JT, Wong SC, Grimley C, Watanabe C. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci U S A.* 1993 Jun 1;90(11):5011–5.
221. Yates JR 3rd, Speicher S, Griffin PR, Hunkapiller T. Peptide mass maps: a highly informative approach to protein identification. *Anal Biochem.* 1993 Nov 1;214(2):397–408.
222. Mann M, Højrup P, Roepstorff P. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol Mass Spectrom.* 1993 Jun;22(6):338–45.
223. Sadygov RG, Cociorva D, Yates JR 3rd. Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. *Nat Methods.* 2004 Dec;1(3):195–202.
224. Moore RE, Young MK, Lee TD. Qscore: an algorithm for evaluating SEQUEST database search results. *J Am Soc Mass Spectrom.* 2002 Apr;13(4):378–86.
225. Jiang X, Jiang X, Han G, Ye M, Zou H. Optimization of filtering criterion for SEQUEST database searching to improve proteome coverage in shotgun proteomics. *BMC Bioinformatics.* 2007 Aug 31;8(1):323.
226. Diamant BJ, Noble WS. Faster SEQUEST searching for peptide identification from tandem mass spectra. *J Proteome Res.* 2011 Sep 2;10(9):3871–9.
227. MacCoss MJ, Wu CC, Yates JR 3rd. Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal Chem.* 2002 Nov 1;74(21):5593–9.
228. Nesvizhskii AI. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J Proteomics.* 2010 Oct 10;73(11):2092–123.
229. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods.* 2007 Mar;4(3):207–14.
230. Jeong K, Kim S, Bandeira N. False discovery rates in spectral identification. *BMC Bioinformatics.* 2012 Nov 5;13(Suppl 16):S2.
231. Käll L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods.* 2007 Nov;4(11):923–5.
232. Käll L, Storey JD, MacCoss MJ, Noble WS. Assigning Significance to Peptides Identified by Tandem Mass Spectrometry Using Decoy Databases. *J Proteome Res.* 2008 Jan 1;7(1):29–34.
233. Käll L, Storey JD, Noble WS. Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. *Bioinformatics.* 2008 Aug 15;24(16):i42–i48.
234. Brosch M, Yu L, Hubbard T, Choudhary J. Accurate and Sensitive Peptide Identification with Mascot Percolator. *J Proteome Res.* 2009 Jun 5;8(6):3176–81.
235. Spivak M, Weston J, Bottou L, Käll L, Noble WS. Improvements to the Percolator Algorithm for Peptide Identification from Shotgun Proteomics Data Sets. *J Proteome Res.* 2009 Jul 6;8(7):3737–45.
236. Käll L, Storey JD, MacCoss MJ, Noble WS. Posterior Error Probabilities and False Discovery Rates: Two Sides of the Same Coin. *J Proteome Res.* 2008 Jan 1;7(1):40–4.
237. Frank AM. Predicting Intensity Ranks of Peptide Fragment Ions. *J Proteome Res.* 2009 May 1;8(5):2226–

- 40.
238. Craig R, Cortens JC, Fenyo D, Beavis RC. Using annotated peptide mass spectrum libraries for protein identification. *J Proteome Res.* 2006 Aug;5(8):1843–9.
239. Deutsch EW. Tandem Mass Spectrometry Spectral Libraries and Library Searching. In: Hamacher M, Eisenacher M, Stephan C, editors. *Data Mining in Proteomics* [Internet]. Humana Press; 2011 [cited 2014 Mar 31]. p. 225–32. Available from: [http://link.springer.com/protocol/10.1007/978-1-60761-987-1\\_13](http://link.springer.com/protocol/10.1007/978-1-60761-987-1_13)
240. Muth T, Weilnböck L, Rapp E, Huber CG, Martens L, Vaudel M, et al. DeNovoGUI: An Open Source Graphical User Interface for de Novo Sequencing of Tandem Mass Spectra. *J Proteome Res.* 2014 Feb 7;13(2):1143–6.
241. He L, Han X, Ma B. *DE NOVO SEQUENCING WITH LIMITED NUMBER OF POST-TRANSLATIONAL MODIFICATIONS PER PEPTIDE.* *J Bioinform Comput Biol.* 2013 Aug;11(04):1350007.
242. Frank AM, Savitski MM, Nielsen ML, Zubarev RA, Pevzner PA. De Novo Peptide Sequencing and Identification with Precision Mass Spectrometry. *J Proteome Res.* 2007 Jan 1;6(1):114–23.
243. Dancík V, Addona TA, Clauser KR, Vath JE, Pevzner PA. De novo peptide sequencing via tandem mass spectrometry. *J Comput Biol J Comput Mol Cell Biol.* 1999 Fall-Winter;6(3-4):327–42.
244. Taylor JA, Johnson RS. Sequence database searches via de novo peptide sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom RCM.* 1997;11(9):1067–75.
245. Frank A, Pevzner P. PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal Chem.* 2005 Feb 15;77(4):964–73.
246. Hughes C, Ma B, Lajoie GA. De Novo Sequencing Methods in Proteomics. In: Hubbard SJ, Jones AR, editors. *Proteome Bioinformatics* [Internet]. Humana Press; 2010 [cited 2014 Aug 11]. p. 105–21. Available from: [http://link.springer.com/protocol/10.1007/978-1-60761-444-9\\_8](http://link.springer.com/protocol/10.1007/978-1-60761-444-9_8)
247. Palumbo AM, Smith SA, Kalcic CL, Dantus M, Stemmer PM, Reid GE. Tandem mass spectrometry strategies for phosphoproteome analysis. *Mass Spectrom Rev.* 2011 Jul 1;30(4):600–25.
248. DeGnore J, Qin J. Fragmentation of phosphopeptides in an ion trap mass spectrometer. *J Am Soc Mass Spectrom.* 1998 Nov;9(11):1175–88.
249. Tholey A, Reed J, Lehmann WD. Electrospray tandem mass spectrometric studies of phosphopeptides and phosphopeptide analogues. *J Mass Spectrom JMS.* 1999 Feb;34(2):117–23.
250. Palumbo AM, Tepe JJ, Reid GE. Mechanistic insights into the multistage gas-phase fragmentation behavior of phosphoserine- and phosphothreonine-containing peptides. *J Proteome Res.* 2008 Feb;7(2):771–9.
251. Reid GE, Simpson RJ, O’Hair RA. Leaving group and gas phase neighboring group effects in the side chain losses from protonated serine and its derivatives. *J Am Soc Mass Spectrom.* 2000 Dec;11(12):1047–60.
252. Moyer SC, Cotter RJ, Woods AS. Fragmentation of phosphopeptides by atmospheric pressure MALDI and ESI/Ion trap mass spectrometry. *J Am Soc Mass Spectrom.* 2002 Mar;13(3):274–83.
253. Moyer SC, VonSeggern CE, Cotter RJ. Fragmentation of cationized phosphotyrosine containing peptides by atmospheric pressure MALDI/Ion trap mass spectrometry. *J Am Soc Mass Spectrom.* 2003 Jun;14(6):581–92.
254. Metzger S, Hoffmann R. Studies on the dephosphorylation of phosphotyrosine-containing peptides during post-source decay in matrix-assisted laser desorption/ionization. *J Mass Spectrom JMS.* 2000 Oct;35(10):1165–77.
255. Zhang Y, Ficarro SB, Li S, Marto JA. Optimized Orbitrap HCD for quantitative analysis of phosphopeptides. *J Am Soc Mass Spectrom.* 2009 Aug;20(8):1425–34.
256. Jedrychowski MP, Huttlin EL, Haas W, Sowa ME, Rad R, Gygi SP. Evaluation of HCD- and CID-type Fragmentation Within Their Respective Detection Platforms For Murine Phosphoproteomics. *Mol Cell Proteomics MCP* [Internet]. 2011 Dec [cited 2014 Apr 24];10(12). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3237076/>

#### 4 Appendix

257. Kelstrup CD, Hekmat O, Francavilla C, Olsen JV. Pinpointing phosphorylation sites: Quantitative filtering and a novel site-specific x-ion fragment. *J Proteome Res.* 2011 Jul 1;10(7):2937–48.
258. Palumbo AM, Reid GE. Evaluation of gas-phase rearrangement and competing fragmentation reactions on protein phosphorylation site assignment using collision induced dissociation-MS/MS and MS3. *Anal Chem.* 2008 Dec 15;80(24):9735–47.
259. Benschop JJ, Mohammed S, O’Flaherty M, Heck AJR, Slijper M, Menke FLH. Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis. *Mol Cell Proteomics MCP.* 2007 Jul;6(7):1198–214.
260. Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villén J, Li J, et al. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A.* 2004 Aug 17;101(33):12130–5.
261. Gruhler A, Olsen JV, Mohammed S, Mortensen P, Faergeman NJ, Mann M, et al. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics MCP.* 2005 Mar;4(3):310–27.
262. Olsen JV, Mann M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc Natl Acad Sci U S A.* 2004 Sep 14;101(37):13417–22.
263. Schroeder MJ, Shabanowitz J, Schwartz JC, Hunt DF, Coon JJ. A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. *Anal Chem.* 2004 Jul 1;76(13):3590–8.
264. Sweet SMM, Bailey CM, Cunningham DL, Heath JK, Cooper HJ. Large scale localization of protein phosphorylation by use of electron capture dissociation mass spectrometry. *Mol Cell Proteomics MCP.* 2009 May;8(5):904–12.
265. Mohammed S, Lorenzen K, Kerkhoven R, van Breukelen B, Vannini A, Cramer P, et al. Multiplexed proteomics mapping of yeast RNA polymerase II and III allows near-complete sequence coverage and reveals several novel phosphorylation sites. *Anal Chem.* 2008 May 15;80(10):3584–92.
266. Chi A, Huttenhower C, Geer LY, Coon JJ, Syka JEP, Bai DL, et al. Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry. *Proc Natl Acad Sci U S A.* 2007 Feb 13;104(7):2193–8.
267. Molina H, Horn DM, Tang N, Mathivanan S, Pandey A. Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A.* 2007 Feb 13;104(7):2199–204.
268. Wan Y, Cripps D, Thomas S, Campbell P, Ambulos N, Chen T, et al. PhosphoScan: a probability-based method for phosphorylation site prediction using MS2/MS3 pair information. *J Proteome Res.* 2008 Jul;7(7):2803–11.
269. Kim S, Mischerikow N, Bandeira N, Navarro JD, Wich L, Mohammed S, et al. The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search. *Mol Cell Proteomics MCP.* 2010 Dec;9(12):2840–52.
270. Macek B, Mann M, Olsen JV. Global and site-specific quantitative phosphoproteomics: principles and applications. *Annu Rev Pharmacol Toxicol.* 2009;49:199–221.
271. Olsen JV, Schwartz JC, Griep-Raming J, Nielsen ML, Damoc E, Denisov E, et al. A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol Cell Proteomics MCP.* 2009 Dec;8(12):2759–69.
272. Thingholm TE, Palmisano G, Kjeldsen F, Larsen MR. Undesirable Charge-Enhancement of Isobaric Tagged Phosphopeptides Leads to Reduced Identification Efficiency. *J Proteome Res.* 2010 Aug 6;9(8):4045–52.
273. Wu J, Warren P, Shakey Q, Sousa E, Hill A, Ryan TE, et al. Integrating titania enrichment, iTRAQ labeling, and Orbitrap CID-HCD for global identification and quantitative analysis of phosphopeptides. *Proteomics.* 2010 Jun;10(11):2224–34.
274. Przybylski C, Jünger MA, Aubertin J, Radvanyi F, Aebersold R, Pflieger D. Quantitative analysis of protein complex constituents and their phosphorylation states on a LTQ-Orbitrap instrument. *J Proteome Res.* 2010 Oct 1;9(10):5118–32.

## 4.9 References

275. Shilov IV, Seymour SL, Patel AA, Loboda A, Tang WH, Keating SP, et al. 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 Jan 9;6(9):1638–55.
276. Savitski MM, Lemeer S, Boesche M, Lang M, Mathieson T, Bantscheff M, et al. Confident phosphorylation site localization using the Mascot Delta Score. *Mol Cell Proteomics MCP*. 2011 Feb;10(2):M110.003830.
277. Boja ES, Phillips D, French SA, Harris RA, Balaban RS. Quantitative mitochondrial phosphoproteomics using iTRAQ on an LTQ-Orbitrap with high energy collision dissociation. *J Proteome Res*. 2009 Oct;8(10):4665–75.
278. Kutuk O, Letai A. Regulation of Bcl-2 family proteins by posttranslational modifications. *Curr Mol Med*. 2008 Mar;8(2):102–18.
279. Zhao R, Yang FT, Alexander DR. An oncogenic tyrosine kinase inhibits DNA repair and DNA-damage-induced Bcl-xL deamidation in T cell transformation. *Cancer Cell*. 2004 Jan;5(1):37–49.
280. He Z, Yu W. Improving peptide identification with single-stage mass spectrum peaks. *Bioinforma Oxf Engl*. 2009 Nov 15;25(22):2969–74.
281. Lemeer S, Kunold E, Klaeger S, Raabe M, Towers MW, Claudes E, et al. Phosphorylation site localization in peptides by MALDI MS/MS and the Mascot Delta Score. *Anal Bioanal Chem*. 2012 Jan;402(1):249–60.
282. Beausoleil SA, Villén J, Gerber SA, Rush J, Gygi SP. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol*. 2006 Oct;24(10):1285–92.
283. De Graaf EL, Altelaar AFM, van Breukelen B, Mohammed S, Heck AJR. Improving SRM assay development: a global comparison between triple quadrupole, ion trap, and higher energy CID peptide fragmentation spectra. *J Proteome Res*. 2011 Sep 2;10(9):4334–41.
284. Breitwieser FP, Müller A, Dayon L, Köcher T, Hainard A, Pichler P, et al. General statistical modeling of data from protein relative expression isobaric tags. *J Proteome Res*. 2011 Jun 3;10(6):2758–66.
285. Geiger T, Cox J, Mann M. Proteomics on an Orbitrap Benchtop Mass Spectrometer Using All-ion Fragmentation. *Mol Cell Proteomics MCP*. 2010 Oct;9(10):2252–61.
286. Nepomuceno AI, Gibson RJ, Randall SM, Muddiman DC. Accurate Identification of Deamidated Peptides in Global Proteomics Using a Quadrupole Orbitrap Mass Spectrometer. *J Proteome Res*. 2014 Feb 7;13(2):777–85.
287. Kapp EA, Schütz F, Connolly LM, Chakel JA, Meza JE, Miller CA, et al. An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: Sensitivity and specificity analysis. *PROTEOMICS*. 2005 Aug 1;5(13):3475–90.
288. Von Haller PD, Yi E, Donohoe S, Vaughn K, Keller A, Nesvizhskii AI, et al. The application of new software tools to quantitative protein profiling via isotope-coded affinity tag (ICAT) and tandem mass spectrometry: I. Statistically annotated datasets for peptide sequences and proteins identified via the application of ICAT and tandem mass spectrometry to proteins copurifying with T cell lipid rafts. *Mol Cell Proteomics MCP*. 2003 Jul;2(7):426–7.
289. Tang WH, Halpern BR, Shilov IV, Seymour SL, Keating SP, Loboda A, et al. Discovering known and unanticipated protein modifications using MS/MS database searching. *Anal Chem*. 2005 Jul 1;77(13):3931–46.
290. Chalkley RJ, Baker PR, Huang L, Hansen KC, Allen NP, Rexach M, et al. Comprehensive analysis of a multidimensional liquid chromatography mass spectrometry dataset acquired on a quadrupole selecting, quadrupole collision cell, time-of-flight mass spectrometer: II. New developments in Protein Prospector allow for reliable and comprehensive automatic analysis of large datasets. *Mol Cell Proteomics MCP*. 2005 Aug;4(8):1194–204.
291. Wiktorowicz JE, English RD, Wu Z, Kurosky A. Model studies on iTRAQ modification of peptides: sequence-dependent reaction specificity. *J Proteome Res*. 2012 Mar 2;11(3):1512–20.
292. Kishani Ambihapathy TY. Pathways to Immonium Ions in the Fragmentation of Protonated Peptides. *J Mass Spectrom*. 1997;32(2):209 – 215.

#### 4 Appendix

293. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol.* 2012 Oct;30(10):918–20.
294. Vaudel M, Barsnes H, Berven FS, Sickmann A, Martens L. SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. *Proteomics.* 2011 Mar;11(5):996–9.
295. Barsnes H, Vaudel M, Colaert N, Helsens K, Sickmann A, Berven FS, et al. compomics-utilities: an open-source Java library for computational proteomics. *BMC Bioinformatics.* 2011;12:70.
296. Liu N, Chan W, Lee K-C, Cai Z. A Method to Enhance a1 Ions and Application for Peptide Sequencing and Protein Identification. *J Am Soc Mass Spectrom.* 2009 Jun;20(6):1214–23.
297. Ernoult E, Gamelin E, Guette C. Improved proteome coverage by using iTRAQ labelling and peptide OFFGEL fractionation. *Proteome Sci.* 2008 Oct 13;6:27.
298. Hohmann LJ, Eng JK, Gemmill A, Klimek J, Vitek O, Reid GE, et al. Quantification of the Compositional Information Provided by Immonium Ions on a Quadrupole-Time-of-Flight Mass Spectrometer. *Anal Chem.* 2008 Jul 15;80(14):5596–606.
299. Taus T, Köcher T, Pichler P, Paschke C, Schmidt A, Henrich C, et al. Universal and confident phosphorylation site localization using phosphoRS. *J Proteome Res.* 2011 Dec 2;10(12):5354–62.
300. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res.* 2011 Apr 1;10(4):1794–805.
301. Marx H, Lemeer S, Schliep JE, Matheron L, Mohammed S, Cox J, et al. A large synthetic peptide and phosphopeptide reference library for mass spectrometry-based proteomics. *Nat Biotechnol.* 2013 Jun;31(6):557–64.
302. Schlosser A, Vanselow JT, Kramer A. Mapping of Phosphorylation Sites by a Multi-Protease Approach with Specific Phosphopeptide Enrichment and NanoLC-MS/MS Analysis. *Anal Chem.* 2005;77(16):5243–50.
303. Gauci S, Helbig AO, Slijper M, Krijgsveld J, Heck AJR, Mohammed S. Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. *Anal Chem.* 2009 Jun 1;81(11):4493–501.
304. Swaney DL, Wenger CD, Coon JJ. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J Proteome Res.* 2010 Mar 5;9(3):1323–9.
305. Tran BQ, Hernandez C, Waridel P, Potts A, Barblan J, Lisacek F, et al. Addressing trypsin bias in large scale (phospho)proteome analysis by size exclusion chromatography and secondary digestion of large post-trypsin peptides. *J Proteome Res.* 2011 Feb 4;10(2):800–11.
306. Wiśniewski JR, Mann M. Consecutive proteolytic digestion in an enzyme reactor increases depth of proteomic and phosphoproteomic analysis. *Anal Chem.* 2012 Mar 20;84(6):2631–7.
307. Gilmore JM, Kettenbach AN, Gerber SA. Increasing phosphoproteomic coverage through sequential digestion by complementary proteases. *Anal Bioanal Chem.* 2012 Jan;402(2):711–20.
308. Mertins P, Udeshi ND, Clauser KR, Mani DR, Patel J, Ong S, et al. iTRAQ labeling is superior to mTRAQ for quantitative global proteomics and phosphoproteomics. *Mol Cell Proteomics MCP.* 2012 Jun;11(6):M111.014423.
309. Nilsson CL. Advances in quantitative phosphoproteomics. *Anal Chem.* 2011;84(2):735–46.
310. Maixner F, Overath T, Linke D, Janko M, Guerriero G, van den Berg BHJ, et al. Paleoproteomic study of the Iceman's brain tissue. *Cell Mol Life Sci CMLS.* 2013 Oct;70(19):3709–22.
311. Zarei M, Sprenger A, Gretzmeier C, Dengjel J. Rapid combinatorial ERLIC-SCX solid-phase extraction for in-depth phosphoproteome analysis. *J Proteome Res.* 2013 Dec 6;12(12):5989–95.
312. Jakoby T, van den Berg BH, Tholey A. Quantitative protease cleavage site profiling using tandem-mass-tag labeling and LC-MALDI-TOF/TOF MS/MS analysis. *J Proteome Res.* 2012 Mar 2;11(3):1812–20.
313. Michalski A, Damoc E, Hauschild J-P, Lange O, Wiegand A, Makarov A, et al. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics MCP.* 2011 Sep;10(9):M111.011015.



314. Wang B, Malik R, Nigg EA, Körner R. Evaluation of the Low-Specificity Protease Elastase for Large-Scale Phosphoproteome Analysis. *Anal Chem*. 2008 Dec 15;80(24):9526–33.
315. Sachon E, Mohammed S, Bache N, Jensen ON. Phosphopeptide quantitation using amine-reactive isobaric tagging reagents and tandem mass spectrometry: application to proteins isolated by gel electrophoresis. *Rapid Commun Mass Spectrom RCM*. 2006;20(7):1127–34.
316. Fox PF, McSweeney PLH. *Advanced Dairy Chemistry: Volume 1: Proteins, Parts A&B: Protein*. Springer; 2003. 656 p.
317. Martínez-Esteso MJ, Casado-Vela J, Sellés-Marchart S, Elortza F, Pedreño MA, Bru-Martínez R. iTRAQ-based profiling of grape berry exocarp proteins during ripening using a parallel mass spectrometric method. *Mol Biosyst*. 2011 Mar;7(3):749–65.
318. Pichler P, Köcher T, Holzmann J, Möhring 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 Feb 15;83(4):1469–74.
319. Steen H, Jebanathirajah JA, Springer M, Kirschner MW. Stable isotope-free relative and absolute quantitation of protein phosphorylation stoichiometry by MS. *Proc Natl Acad Sci U S A*. 2005 Mar 15;102(11):3948–53.
320. Lynnerup N. Mummies. *Am J Phys Anthropol*. 2007;Suppl 45:162–90.
321. Spindler K. *The Man In The Ice*. Hachette UK; 2013. 259 p.
322. Keller A, Graefen A, Ball M, Matzas M, Boisguerin V, Maixner F, et al. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat Commun*. 2012 Feb 28;3:698.
323. The Scientific Study of Mummies [Internet]. [cited 2014 Apr 24]. Available from: [https://books.google.com/books/about/The\\_Scientific\\_Study\\_of\\_Mummies.html?id=P\\_xj3QTHHvoC](https://books.google.com/books/about/The_Scientific_Study_of_Mummies.html?id=P_xj3QTHHvoC)
324. Barraco RA. Preservation of proteins in mummified tissues. *Am J Phys Anthropol*. 1978 May;48(4):487–91.
325. Schmidt-Schultz TH, Schultz M. Bone protects proteins over thousands of years: extraction, analysis, and interpretation of extracellular matrix proteins in archeological skeletal remains. *Am J Phys Anthropol*. 2004 Jan;123(1):30–9.
326. Ferrer I, Santpere G, Arzberger T, Bell J, Blanco R, Boluda S, et al. Brain protein preservation largely depends on the postmortem storage temperature: implications for study of proteins in human neurologic diseases and management of brain banks: a BrainNet Europe Study. *J Neuropathol Exp Neurol*. 2007 Jan;66(1):35–46.
327. Paczkowski S, Schutz S. Post-mortem volatiles of vertebrate tissue. *Appl Microbiol Biotechnol*. 2011 Aug;91(4):917–35.
328. Vass AA, Barshick S-A, Sega G, Caton J, Skeen JT, Love JC, et al. Decomposition chemistry of human remains: a new methodology for determining the postmortem interval. *J Forensic Sci*. 2002 May;47(3):542–53.
329. Ortner DJ, Aufderheide AC, International Congress of Anthropological and Ethnological Sciences, editors. *Human paleopathology: current syntheses and future options*. Washington: Smithsonian Institution Press; 1991.
330. Nielsen-Marsh CM, Richards MP, Hauschka PV, Thomas-Oates JE, Trinkaus E, Pettitt PB, et al. Osteocalcin protein sequences of Neanderthals and modern primates. *Proc Natl Acad Sci U S A*. 2005 Mar 22;102(12):4409–13.
331. Solazzo C, Fitzhugh WW, Rolando C, Tokarski C. Identification of protein remains in archaeological potsherds by proteomics. *Anal Chem*. 2008 Jun 15;80(12):4590–7.
332. Cappellini E, Gilbert MTP, Geuna F, Fiorentino G, Hall A, Thomas-Oates J, et al. A multidisciplinary study of archaeological grape seeds. *Naturwissenschaften*. 2010 Feb;97(2):205–17.
333. Cappellini E, Jensen LJ, Szklarczyk D, Ginolhac A, da Fonseca RAR, Stafford TW, et al. Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. *J*

#### 4 Appendix

Proteome Res. 2012 Feb 3;11(2):917–26.

334. Lindner H, Helliger W. Age-dependent deamidation of asparagine residues in proteins. *Exp Gerontol*. 2001 Sep;36(9):1551–63.
335. Bjornson RD, Carriero NJ, Colangelo C, Shifman M, Cheung K-H, Miller PL, et al. X!Tandem, an Improved Method for Running X!Tandem in Parallel on Collections of Commodity Computers. *J Proteome Res* [Internet]. 2008 Jan [cited 2014 Apr 29];7(1). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3863625/>
336. Schilling O, Overall CM. Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat Biotechnol*. 2008 Jun;26(6):685–94.
337. Colinge J, Bennett KL. Introduction to Computational Proteomics. *PLoS Comput Biol* [Internet]. 2007 Jul [cited 2014 Apr 30];3(7). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1933459/>
338. Niedermeyer THJ, Strohal M. mMass as a Software Tool for the Annotation of Cyclic Peptide Tandem Mass Spectra. *PLoS ONE*. 2012 Sep 13;7(9):e44913.