



**Wissenschaftszentrum Weihenstephan
für Ernährung, Landnutzung und Umwelt**

**Entwicklung, Implementierung und Etablierung einer modularen
Auswertesoftware zur Untersuchung von Enzymfunktion und
-regulation mittels Massenspektrometrie**

Michael Krappmann

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. D. Langosch

Prüfer der Dissertation:

1. Priv.-Doz. Dr. Th. Letzel
2. Univ.-Prof. Dr. B. Küster
3. Prof. Dr. F. Leßke
(Hochschule Weihenstephan-Triesdorf)

Die Dissertation wurde am 03.06.2015 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 17.11.2015 angenommen.

Inhaltsverzeichnis

Abkürzungen	iii
Abbildungsverzeichnis	v
Tabellenverzeichnis	vii
Zusammenfassung	viii
Summary	ix
1 Einleitung	1
1.1 Die Proteomik	2
1.2 Enzymologie - eine Teilmenge der "funktionellen" Proteomik	4
1.3 Analysemethoden der (funktionellen) Proteomik	9
1.4 Software-Landschaft in der (Prote)omik-Forschung	16
2 Material und Methoden	28
2.1 Chemikalien	28
2.1.1 Enzyme	28
2.1.2 Substrate	28
2.1.3 Inhibitoren und Extrakte	28
2.1.4 Interne Standards	29
2.1.5 Weitere Chemikalien	29
2.2 Kontinuierliche Enzym-Assays mittels (LC-)MS	30
2.2.1 PKA-Assay: On-line Monitoring der Enzymaktivität und der Stabilität des Messsystems bei sich ändernden physikalischen Eigenschaften	30
2.2.2 HEWL-Assay: Nachweis der Bildung von nichtkovalenten Komplexen und simultane Reaktionsinhibition	31
2.2.3 iAP und AChE/Hausstaub-Assay: Monitoring der Enzymaktivität und Screening nach unbekanntem Inhibitoren in komplexen Proben	33
2.3 Datenanalyse	36
2.4 Softwareimplementierung	38

3	Ergebnisse.....	44
3.1	Achroma: a software strategy for analyzing (a-)typical mass spectrometric data.....	44
3.2	Achroma Software - High-Quality Policy in (a-)Typical Mass Spectrometric Data Handling and Applied Functional Proteomics.....	58
3.3	Mass Spectrometric Real-Time Monitoring of an Enzymatic Phosphorylation Assay using Internal Standards and Data Handling Freeware.....	68
3.4	Functional Analysis of Proteins, Including LC-MS and Special Freeware.....	95
4	Diskussion	111
5	Ausblick	119
6	Referenzen.....	120
	Danksagung	126
	Liste der Publikationen, Vorträge und Posterbeiträge.....	127
	Curriculum Vitae.....	129
	Appendix	132

Abkürzungen

2DE	2-D Gelelektrophorese
4-HBA	4-hydroxybenzoyl-Conenzym A
(GlcNAc) ₃	N, N', N''-Triacetylchitotriose
(GlcNAc) ₄	Tetra-N-acetylchitotetraose
(GlcNAc) ₆	Hexa-N-acetylchitohexaose
ACh	Acetylcholin
AChCl	Acetylcholinchlorid
AChE	Acetylcholinesterase
ADO	Adenosin
ADP	Adenosin 5'-diphosphat
AMP	Adenosin 5'-monophosphat
ASCII	American Standard Code for Information Interchange
ATP	Adenosin 5'-triphosphat
cAMP	Adenosin 3':5'-zyklisches Monophosphat
CPU	Central Processing Unit
CSV	Comma-Separated Values (oder: Character-Separated Values)
EIC	Extrahiertes Ionenchromatogramm
ESI	Elektrosprayionisation
GC	Gaschromatographie
GC-MS	Gaschromatographie-Massenspektrometrie
HDF5	Hierarchical Data Format
HEWL	Lysozym aus dem Hühnereiweiß
HPLC	High Performance Liquid Chromatography
HT	High Temperature
iAP	intestinale alkaline Phosphatase
ICP-MS	Induktiv gekoppeltes Plasma-Massenspektrometrie (Inductively Coupled Plasma-Mass Spectrometry)
IDE	Integraded Development Environment
IS1	Interner Standard 1
IS2	Interner Standard 2

iTRAQ	isobaric Tags for Relative and Absolute Quantitation
Kem	Kemptide
LC	Flüssigchromatographie (Liquid Chromatography)
LC-MS	Flüssigchromatographie-Massenspektrometrie
Mal	Malantide
MALDI-MS	Matrix-unterstützte Laser Desorption Ionisation-Massenspektrometrie
MeOH	Methanol
MFC	Microsoft Foundation Class
MgATP	Adenosin 5'-Triphosphat (Magnesiumsalz)
MgSO ₄	Magnesiumsulfat
MRM	Multi Reaction Monitoring
MS	Massenspektrometrie
Mw	Molekulargewicht (Molecular weight)
m/z	Massen-zu-Ladungsverhältnis
NH ₄ Ac	Ammoniumacetat
PKA	Proteinkinase A
PEEK	Polyetheretherketon
PubMed	PubMed.gov (http://www.ncbi.nlm.nih.gov/pubmed/ ; US National Library of Medicine National Institutes of Health)
Q	Quadrupol
RAM	Random Access Memory
SaaS	Software as a Service
SILAC	Stable Isotope Labeling by Amino acids in Cell culture
SRM	Single Reaction Monitoring
TOF	Time Of Flight
XML	Extended Markup Language

Abbildungsverzeichnis

Abbildung 1:

Abfrage bei PubMed zu den Begriffen "Proteomic" (blau), "Mass Spectrometry" (rot), "Software" (grün) mit den jeweiligen UND-Verknüpfungen (AND) zwischen "Proteomic" und den beiden anderen Begriffen (violett) in "All Fields" 2

Abbildung 2:

Generelle Klassifizierung der Proteomik; die grün hinterlegten Forschungsfelder sind Teil dieser Arbeit. 5

Abbildung 3:

Pubmed-Abfrage zum Stichwort "Enzyme" (alle Felder)..... 6

Abbildung 4:

Zeitlicher Verlauf der Entwicklung der Analysetechnologien. Je dunkler die Farbe wird, umso stärker ist der Einfluss auf die Proteomik. 9

Abbildung 5:

Darstellung der "State of the Art"-Methoden der Proteomik am Versuchsablauf [36]. 10

Abbildung 6:

Abfrage bei PubMed nach Veröffentlichungen in den Feldern "Titel/Zusammenfassung" nach den Begriffen "Enzyme" ODER "Enzyme Assay" UND "Liquid Chromatography Mass Spectrometry" (grün) bzw. UND "Mass Spectrometry" (blau). 11

Abbildung 7:

Schema eines kontinuierlichen Enzym-Assay mittels MS mit einer vorgeschalteten HPLC zur Auftrennung von komplexen Mischungen. 13

Abbildung 8:

Kontinuierliche Produktmessung eines MS-basierten Enzym-Assays. A) Produktpur ohne Inhibition und B) Produktpur, C) Substratpur und D) Inhibitorspur zum Zeitpunkt der Inhibition. 15

Abbildung 9:

Darstellung der Relationen verschiedener Analysetechniken zu den einzelnen Softwarewerkzeugen und Datenbanken [61]. 17

Abbildung 10:

Abfrage bei PubMed zu den Begriffen "Enzyme" AND "Software" AND "Mass Spectrometry" als Und-Verknüpfung (AND) in "Title/Abstract". 19

Abbildung 11:

Ablauf einer automatisierten Abarbeitungsliste der verschiedenen Softwares (Spiegelstriche) am Beispiel von TPP. 20

Abbildung 12:

Plug-In Technologie am Beispiel der Eclipse-Plattform; A) schematische Darstellung der Verknüpfung von Plug-ins; B) Darstellung eines Fließbilds neuer Module einer Analytiksoftware, über die Eclipsemodule hin zur darunterliegenden plattformunabhängigen "Java Virtual Machine" und dem Betriebssystem. 23

Abbildung 13:

On-line Monitoring System von Enzymaktivitäten mittels ESI-MS 31

Abbildung 14:

Schematische Darstellung des kontinuierlichen Flusssystems mittels (LC-)MS zum Nachweis der Enzymaktivität von iAP und zum Inhibitor-Screening in Hausstaubproben..... 34

Abbildung 15:

Schematische Darstellung der Document-View Architektur von Achroma. 38

Abbildung 16:

Aufbau der graphischen Oberfläche der Achroma Software. 40

Tabellenverzeichnis

Tabelle 1:

Klassifizierung der Enzyme, Nennung wichtiger Subklassen und Art der Reaktion..... 7

Tabelle 2:

Auflistung aktueller Software-Werkzeuge und -Plattformen für die Auswertung von Proteomik-Daten, die mittels massenspektrometrischer Detektion gewonnen wurden. 25

Tabelle 3:

Mengenangaben der Inhibitorinjektionen und Verhältnis der Menge an Inhibitor zu Substrat und Enzym. 33

Tabelle 4:

Detektierte m/z Verhältnisse für alle Substrate, Produkte und Inhibitoren der untersuchten Enzym Assays..... 36

Tabelle 5:

Inhibitormenge am Übergang zum nächsthöheren Enzym-Inhibitor Komplex. 96

Zusammenfassung

Die Enzymologie ist ein wichtiges Forschungsgebiet der funktionellen Proteomik. Enzymreaktionen und deren Untersuchung finden in vielen Bereichen, wie der Lebensmitteltechnologie, der Biotechnologie, der Pharmazie, der Umwelttechnologie oder der Waschmittelindustrie Anwendung. Die massenspektrometrische Detektion kontinuierlich durchgeführter Enzym-Assays hat sich in den letzten Jahren zu einer etablierten Technik entwickelt.

Neben der Detektionstechnologie ist auch die Software zur Datenauswertung ein wichtiger Baustein für qualitativ hochwertige Ergebnisse. Diese hat sich in den letzten Jahren stark entwickelt und es besteht aufgrund der fortschreitenden Technologieentwicklung weiterhin großer Bedarf nach spezialisierten Softwarelösungen.

Die vorliegende Arbeit thematisiert die Datenauswertung kontinuierlicher Flüssigchromatographie-Massenspektrometrie (LC-MS) Enzym-Assays und die Entwicklung einer Software-Strategie zur Auswertung für (un-)typische Daten, die aus den durchgeführten Versuchen gewonnen werden, sowie die Implementierung einer modularen, jederzeit erweiterbaren Software.

Ausgehend von verschiedenen Fragestellungen der Enzymologie, wie Enzymaktivität, Inhibition oder konkurrierender Substrate in Verbindung mit neuen technologischen Setups kontinuierlicher Flusssysteme wurden die gewonnenen Daten auf die Möglichkeit der Auswertung mit bereits gängiger Software untersucht. Dies führte dazu, dass die Auswertung teilweise überhaupt nicht möglich oder wenn doch, dann sehr aufwendig war. Gerade bei kontinuierlichen LC-MS Enzym-Assays entstehen untypische massenspektrometrische Daten, die gängigen Softwarelösungen Schwierigkeiten bereiten. Es wurde während dieser Arbeit mit *Achroma* eine Software entwickelt, die diese untypischen Daten teilautomatisiert verarbeiten und auswerten kann. *Achroma* wurde in der Sprache C++ entwickelt, ist modular aufgebaut und folgt der Model-View-Controller Architektur. Jedes der fünf umgesetzten Module von *Achroma* selbst folgt ebenfalls dieser Struktur und ist in sich selbst eine gekapselte Anwendung.

Es wurden enzymatische Fragestellungen, wie Enzymaktivität, Hemmung, Inhibition und der Nachweis des Übergangs von einfachen zum nächsthöheren nichtkovalenten Enzym-Inhibitor-Komplex mit *Achroma* untersucht.

Es konnte gezeigt werden, dass *Achroma* als modulare Software einen großen Beitrag zur verbesserten Datenauswertung leisten kann und somit für die eingangs erwähnten Forschungsbereiche von Nutzen ist.

Summary

Enzymology is an important field of research in functional proteomics. Studying enzymatic reactions is of interest in many industrial sectors like food technology, biotechnology, pharmaceuticals, environmental technology or detergents industry. In recent years mass spectrometric detection of continuous enzymatic assays has been developed to an established technology.

Besides the detection technology the software plays a key role for getting high quality results. Software development has been strongly developed during the last years and due to the gradual technology development there will be additional a need for specialized software solutions.

This work broaches the issue of data evaluation of continuous LC-MS enzymatic assays and the development of a software strategy for evaluation of (a-)typical data, which were measured during the assays. In addition, a modular software was implemented which can be extended any time.

Different questions of enzymology, like enzyme activity, inhibition or alternative substrates were evaluated with new experimental setups of continuous flow systems and the received raw data were analyzed and evaluated with existing software. The evaluation was thus partially impossible or at least cumbersome. Continuous LC-MS enzymatic assays deliver a-typical massspectrometric data, which are often difficult to manage with established software solutions. During this work, *Achroma* was developed, which can manage and evaluate this a-typical data partly automatically. *Achroma* was implemented in C++, is modular and follows the Model-View-Controller architecture. Each of the five implemented *Achroma* modules itself has the same architecture and is an encapsulated application.

Enzymatic questions, like enzyme activity, inhibition and the switchover to the next higher not covalent enzyme-inhibitor complex were studied with *Achroma*.

It could be shown that *Achroma* - as a modular software - contributes to an improved data evaluation for the research fields of the industrial sectors mentioned at the beginning.

1 Einleitung

Allgemein steht die Wortendung "omik" im weitesten Sinne für Forschungsfelder, die sich mit der Untersuchung von biologischen oder verwandten Systemen in ihrer natürlichen Umgebung beschäftigen. Zur Zeit sind bei frei zugänglichen wissenschaftlichen Glossaren 140 ^[1] bzw. 250 ^[2] verschiedene Begriffe gelistet, die auf „ome“ bzw. „omik“ enden. Beeindruckend ist dabei der rasante Anstieg der „omik“-Forschung, respektive der Proteomik in den letzten 14 Jahren (Abbildung 1). Mit zunehmender technischer Entwicklung, was die Analytik, aber auch die Informations- und Kommunikationstechnologie betrifft, konnten immer präzisere Versuche durchgeführt werden. Gerade die Revolution in der Computertechnologie mit immer größeren Speicherkapazitäten, enormer Steigerung der Rechengeschwindigkeiten und der parallelen Verarbeitung von Daten hat es möglich gemacht, die immense Datenflut aus der Proteomforschung überhaupt in adäquater Zeit auswerten zu können. Gleichzeitig konnten mit dem stetigen Wachstum der Protein- und Peptidbanken die Vorhersagen und somit wiederum die Versuchsdurchführungen verbessert werden. Zusätzlich hat sich die (mehrdimensionale) Massenspektrometrie als Standardtechnologie in der Proteomforschung etabliert und sorgte mit der gleichzeitigen Detektion von mehreren hundert bis tausend Protein-/Peptidfragmenten für eine weitere Beschleunigung der Proteomforschung.

Eine Abfrage bei PubMed ^[3] zu den Schlagworten "Proteomic", "Mass Spectrometry" und "Software", sowie einer UND-Verknüpfung des Wortes "Proteomic" mit jeweils den Begriffen "Mass Spectrometry" und "Software" und einer UND-Verknüpfung aller drei Schlagworte zeigt einen stetigen Anstieg der Publikationen in den Bereichen (Abbildung 1). Waren 1997 nur vier Veröffentlichungen zum Thema Proteomik gelistet und jeweils Null zur Kombination mit Massenspektrometrie und Software, so hat die Anzahl im Jahr 2014 mit 7.361 Publikationen zum Thema Proteomik vorerst einen Höhepunkt erreicht. Die Kombination mit Massenspektrometrie hatte mit 2.942 Publikationen in 2012 ihren bisherigen Höhepunkt erreicht und hat einen Anteil von 37,3% an den Gesamtpublikationen im Bereich Proteomik für 2014. Der Anteil von Softwareentwicklung lag im Jahr 2012 bei 419 Publikationen und nahm damit ebenfalls den derzeitigen Höchststand ein. Im Jahr 2013 und 2014 sanken diese auf 330 Publikationen/Jahr. Mit einem Anteil von knapp 5% an den Gesamtveröffentlichungen für die Proteomik ist aber nach wie vor die Relevanz gegeben. Die UND-Verknüpfung der drei Schlagworte erzielte 2014 eine Trefferquote von 187 Veröffentlichungen (bisheriger Höchststand 2011: 268) und einem prozentualen Anteil an den Proteomik-Publikationen für 2014 von 2,5%. Daraus wird ersichtlich, dass neben den biologischen Fragestellungen der Proteomforschung ebenfalls die verbesserten technischen

Voraussetzungen zu einem wichtigen Forschungsbereich in Bezug zur Entschlüsselung des Proteoms geworden sind.

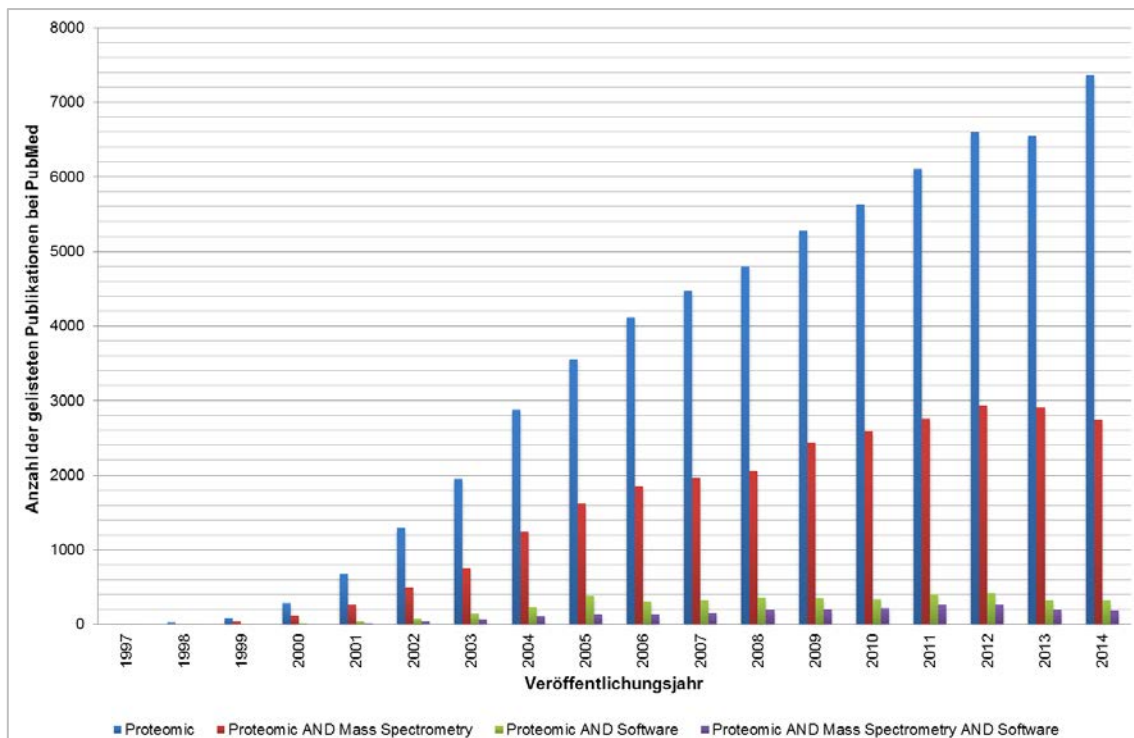


Abbildung 1: Abfrage bei PubMed zu den Begriffen "Proteomic" (blau), "Mass Spectrometry" (rot), "Software" (grün) mit den jeweiligen UND-Verknüpfungen (AND) zwischen "Proteomic" und den beiden anderen Begriffen (violett) in "All Fields".

1.1 Die Proteomik

Die Geschichte der Proteomik-Forschung ist noch sehr jung. Vor gut 20 Jahren, im Jahr 1994, bei der ersten 2D Elektrophorese Konferenz in Sienna in Italien wurde der Begriff "Proteom" geprägt ^[4]. Seit dem ist das Proteom als Protein-Komplement zum Genom definiert. Das Proteom beschreibt die Gesamtheit aller Proteine eines biologischen Systems, z.B. eines Lebewesens, einer Zelle, eines Gewebes oder eines Zellkompartiments, zu einem definierten Zeitpunkt und zu definierten Bedingungen. Die Erforschung des Proteoms wird Proteomik genannt.

Scott D. Patterson und Ruedi Aebersold geben in ihrem Review: "Proteomics the first decade and beyond" (2003) ^[5] eine eindrucksvolle Zusammenfassung über die ersten 10 Jahre der Proteomik und die Technologien zum damaligen Stand. Nach Patterson und Aebersold ist die Proteomik, genauso wie die Genom-Sequenzierung, die Microarray-Analyse und das Metabolit-Profilung, eine sogenannte *Discovery Science*: die Erforschung eines biologischen Systems oder eines Prozesses durch das zählen und/oder benennen seiner einzelnen Elemente, unabhängig von Hypothesen wie das System oder der Prozess funktionieren könnte. Das große Feld der Proteomik ist als logische Konsequenz aus den Ergebnissen der ambitionierten Genomprojekte zur Erforschung der kompletten Genome verschiedener

Spezies und den daraus resultierenden Änderungen in den biologischen Modellen entstanden. Die Theorie hinter diesen systembiologischen Versuchen ist, dass für alle Arten die Menge an möglichen Biomolekülen und deren Funktion und Einbindung in Stoffwechselkreisläufe und Prozesse groß aber endlich ist. Theoretisch kann deshalb ein biologisches System ausreichend beschrieben werden, wenn eine signifikante Anzahl an Beobachtungen aller Elemente dieses Systems möglich ist. Die Proteomik ist daher eine immense Informationsquelle über biologische Systeme, da Proteine an allen biologischen Aktivitäten beteiligt sind und die Erforschung diverser Eigenschaften von Proteinen hervorragend geeignet ist, das Verständnis von biologischen Systemen zu erhöhen ^[5].

Der erste Ansatz der Proteomik war, möglichst schnell alle exprimierten Proteine eines biologischen Systems, z.B. einer Zelle, zu identifizieren. 2014 wurde ein erster Meilenstein dieses Ansatzes erreicht. Unabhängig voneinander haben im Mai 2014 zeitgleich Wissenschaftler der John Hopkins University School of Medicine in Baltimore (Maryland, USA) ^[6] und das Team um Prof. Dr. Küster der Technischen Universität München (Deutschland) ^[7] einen Katalog mit fast allen Proteinen des Menschen vorgelegt. Die amerikanische Forschergruppe hat 17.294 Proteine und damit ca. 84% und die Forschergruppe der TUM hat 18.097 Proteine und damit ca. 92% des menschlichen Proteoms identifiziert. Allerdings zeigen auch diese aktuellen Studien, dass die Proteomik noch lange nicht ihr Ende erreicht hat. So wurden fast 200 Proteine gefunden, die nicht von bekannten proteinkodierenden Genen produziert werden und umgekehrt konnten fast 2.000 Proteine nicht gefunden werden, die laut Genkarte aber existieren sollten. Welche Funktionen die neuentdeckten Proteine haben ist noch unbekannt. Daraus wird ersichtlich, dass neben der Lösung biologischer/biochemischer Fragestellungen ebenso noch viel Potential in der Optimierung der Analysegeräte, wie Massenspektrometer, und der dazugehörigen Computertechnologie, sowie der Softwareentwicklung im Bereich der Proteomik steckt.

Aufgrund der Komplexität der Proteine und ihrer Interaktionen haben sich neben der Protein-Identifizierung auch andere Forschungsfelder der Proteomik gebildet. So spielen die Proteinquantifizierung, die Protein-Modifizierung, die Proteininteraktion mit anderen Proteinen und/oder Molekülen, die Proteinaktivität, die intrazelluläre Verteilung und die Proteinstruktur eine ebenso wichtige Rolle. Aus dieser multidimensionalen Betrachtung der Proteomik heraus, hat sich, neben weiteren Sub-Forschungsfeldern, die "funktionelle" Proteomik entwickelt ^[8-14]. Allerdings ergeben sich bei der Eingrenzung der "funktionellen" Proteomik nach biologischen Gesichtspunkten sowohl Überschneidungen mit anderen Klassifizierungen der Proteomik, z.B. der "strukturellen" Proteomik ^[15, 16], als auch unabhängige Forschungsfelder. Godovac-Zimmermann beschreibt in seinem Review

(2001)^[9] die "funktionelle" Proteomik als den Einsatz von Proteomik-Methoden für das Monitoring und die Analyse der räumlichen und zeitlichen Eigenschaften von molekularen Netzwerken und Flüssen, die eine lebende Zelle ausmachen. Die "funktionelle" Proteomik stellt somit auch Methoden zur Verfügung, wie Simulation von Funktionen, Perturbation oder Isolation einzelner Komponenten vom Netzwerk, um die molekularen Spezies zu identifizieren, welche Teile dieser Netzwerke oder Stoffkreisläufe sind. Naylor und Kumar (2003)^[8] hingegen beschreiben die "funktionelle" Proteomik als die Identifikation aller Protein-Protein Interaktionen in einem definierten System, z.B. einer Zelle. Ebenso zählen hierzu Protein-DNA und Protein-RNA Interaktionen, sowie die Modulation aufgrund von Interaktionen mit Liganden, Medikamenten, Toxinen und Metall-Ionen, die einen Effekt auf die Proteinfunktion haben können. Weitere Autoren klassifizieren die "funktionelle" Proteomik nach "Interaktions-Proteomik, quantitative Proteomik und posttranslationalen Modifikationen"^[10] oder durch "Hefe-Genomik, Affinität des gereinigten Proteinkomplexes oder Knock-out Versuchen mit Mäusen"^[12]. Zusätzliche Definitionen "weichen" die Definition der "funktionellen" Proteomik weiter auf, wie beispielsweise "activity-based proteomics"^[17-20], "kinomics"^[21-23] oder "interactomics"^[24, 25].

Der Fokus dieser Arbeit liegt primär in der Softwareentwicklung für neu entwickelte enzymologische Analysetechniken mit Hilfe der kontinuierlichen Flüssigphasentechnologie-Massenspektrometrie und ist somit stark analytisch/technisch ausgerichtet. Enzymatische Assays können der "aktivitätsbasierten" Proteomik, der Kinomik, der Interaktomik und noch weiteren biologischen Klassifizierungen der "funktionellen" Proteomik zugeordnet werden.

1.2 Enzymologie - eine Teilmenge der "funktionellen" Proteomik

Enzyme spielen eine wichtige Rolle in allen lebenden Organismen. Sie sind sehr spezialisierte Proteine. Die Mehrheit der biochemischen Reaktionen in Organismen wird kontrolliert und katalysiert durch Enzyme. Enzymkatalysierte Reaktionen (Biokatalysen) sind im Gegensatz zu chemischen Katalysen selbst in komplexen Reaktionsgemischen sehr spezifisch. Die Erforschung von enzymatischen Reaktionen, der Aktivität von Enzymen, der Reaktionswege und der Inhibition von Enzymen ist sehr wichtig für viele industrielle Anwendungen, z.B. in der Pharmaindustrie, der Lebensmittelindustrie, der Biotechnologie, der Kosmetikherstellung und der Waschmittelindustrie. Aufgrund der "aktiven" Beteiligung von Enzymen an zahlreichen Stoffwechselreaktionen innerhalb eines Proteoms sind enzymatische Versuche der "funktionellen" Proteomik zuzuordnen (Abbildung 2).

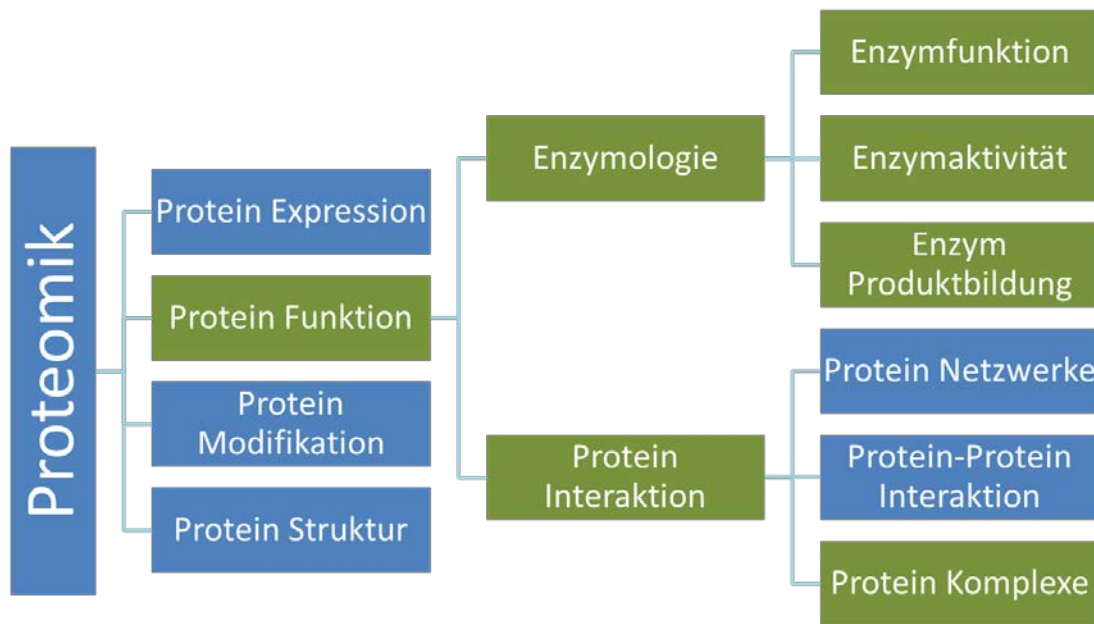


Abbildung 2: Generelle Klassifizierung der Proteomik; die grün hinterlegten Forschungsfelder sind Teil dieser Arbeit.

Die Wirkung von Biokatalysatoren ist schon sehr lange bekannt und wurde bereits geschichtlich sehr früh angewendet, ohne sie jedoch beschreiben zu können und ohne dem Wissen, um welche Enzyme es sich handelt. Ein Beispiel hierfür ist die alkoholische Gärung. Im Jahr 1926 wurde erstmalig das Enzym Urease aus Schwertbohnen von James Summer kristallisiert. 1963 wurde die erste Aminosäuresequenz eines Enzyms aufgeklärt, Ribonuklease A aus dem Rinderpankreas, und 1965 die Sequenz von Lysozym aus dem Hühnereiweiß. Die Enzymologie ist ein wichtiges Forschungsfeld der "funktionellen" Proteomik. Dies wird aus einer Abfrage bei Pubmed ^[26] nach dem Schlagwort "Enzyme" deutlich ersichtlich (Abbildung 3). Die erste Veröffentlichung zum Thema Enzyme ist im Juli 1884 mit dem Titel: "On the Ferments or Enzymes of the Digestive Tract in Fishes." ^[27] von W. Stirling bei Pubmed verzeichnet. Im Jahr 1926 sind bereits 14 Veröffentlichungen mit dem Begriff "Enzyme" bei Pubmed gelistet und 1965 sind es 9.486. 29 Jahre später, im Jahr als das Wort "Proteom" geprägt wurde, war die Anzahl der jährlichen Veröffentlichungen schon auf 59.490 gestiegen. Der Anstieg setzte sich fast linear bis zum Jahr 2013 mit 121.252 (Abbildung 3) gelisteten Veröffentlichungen, die das Schlagwort "Enzyme" enthalten hatten, fort und hat dort ihren Höhepunkt erreicht. Somit hat sich die Anzahl der Veröffentlichungen in einem Zeitraum von knapp 20 Jahren noch einmal verdoppelt. 2014 sind es "nur noch" 95.198. Aber auch diese Zahl zeigt, dass das Forschungsinteresse im Bereich der Enzymologie enorm ist und weiterhin ein Bedarf besteht, neben dem biologischen Interesse, auch die Technologie und die Softwareentwicklung für die Auswertung enzymatischer Versuche zu verbessern.

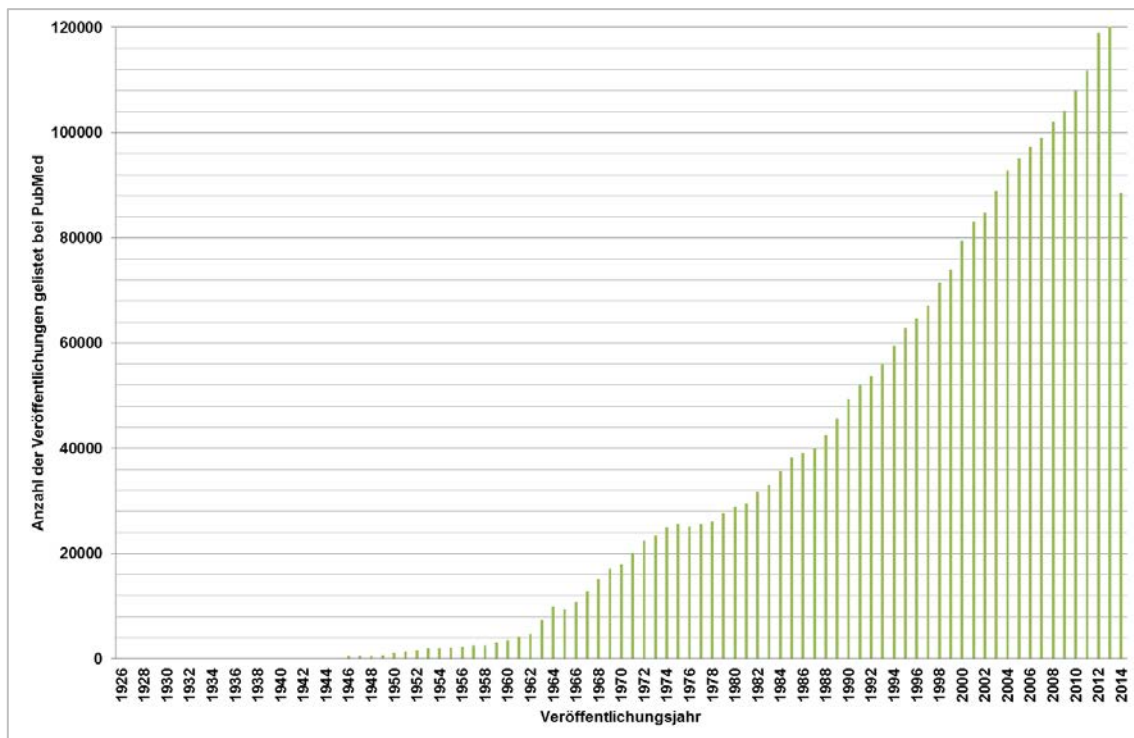


Abbildung 3: Pubmed-Abfrage zum Stichwort "Enzyme" (alle Felder).

Aufgrund der hohen Zahl an in Organismen vorkommenden Enzymen und der vielfältigen Namen neu isolierter Enzyme wurden die Enzyme in sechs Klassen nach ihrer jeweiligen spezifischen Funktion unterteilt ^[28]. Im ersten Report der "Enzyme Commission" ^[29] wurden 1961 712 Enzyme klassifiziert. Im sechsten Report 1992 waren es bereits 3.196 Enzyme. Die BRENDA-Datenbank umfasst aktuell 6.547 Einträge ^[30], was zeigt, dass sich gegenüber 1992 die Anzahl der bekannten und klassifizierten Enzyme mehr als verdoppelt hat und im Vergleich zu 1961 um mehr als das Neunfache gestiegen ist.

Die sechs Hauptklassen der Enzym-Klassifizierung sind: Oxidoreduktasen, Transferasen, Hydrolasen, Lyasen, Isomerasen und Ligasen (Tabelle 1). Diese Hauptgruppen sind unterteilt in Subklassen und diese wiederum in weitere Unterkategorien, um der Komplexität der Enzyme gerecht zu werden. So hat die Gruppe der Oxidoreduktasen beispielsweise 24 Subklassen und jede dieser Subklassen hat wiederum bis zu 15 Unterkategorien.

Tabelle 1: Klassifizierung der Enzyme, Nennung wichtiger Subklassen und Art der Reaktion

Enzymklasse	wichtige Subklassen	Reaktionstyp
Oxidoreduktasen	Dehydrogenasen Oxidasen Peroxidasen Reduktasen	Elektronentransfer: $X_{ox} + Y_{red} \rightleftharpoons X_{red} + Y_{ox}$
Transferasen	Kinasen Polymerasen Glycosyl-, Acethyl-, Methyl-, Amino-Transferasen	Transfer chemischer Gruppen: $X + Y-R \rightleftharpoons X-R + Y$
Hydrolasen	Peptidasen Proteasen Phosphatasen Phospholipasen Esterasen Nukleasen Glycosidasen ATPasen	Hydrolyse; Spaltung unter formaler Anlagerung von Wasser: $X-Y + H_2O \rightleftharpoons X-H + Y-OH$
Lyasen	Decarboxylasen Aldehydlyasen Hydrolyasen Synthasen	Abspalten oder Hinzufügen von Gruppen von Doppelbindungen: $X-Y \rightleftharpoons X + Y$
Isomerasen	Racemasen Epimerasen Isomerasen	Transfer von Gruppen innerhalb von Molekülen: $X \rightleftharpoons X'$
Ligasen	Carboxylasen Synthetasen	Bildung von C-C, C-S, C-O, C-N Bindungen unter NTP*-Abspaltung: $X + Y + NTP \rightleftharpoons X-Y + NDP$ * Nukleosidtriphosphat: A, G, U, C

Alle Enzyme zeichnen sich durch eine hohe Spezifität aus. Im sogenannten *aktiven Zentrum* des Enzyms können nur bestimmte Ausgangsstoffe - Substrate (Edukte) - einer biochemischen Reaktion interagieren, um mit Hilfe des Enzyms als Biokatalysator in ein oder mehrere Produkte umgewandelt zu werden. Es bildet sich ein Enzym-Substrat-Komplex als Zwischenstadium, bevor das (die) Produkt(e) aus dem aktiven Zentrum frei gesetzt werden.

Ebenso können im aktiven Zentrum oder an einer anderen Stelle des Enzyms andere Stoffe binden, die das Substrat daran hindern im aktiven Zentrum zu binden. Es kann somit nicht zur Produktbildung kommen. Diese Stoffe werden Inhibitoren genannt. Konkurriert der Inhibitor mit dem Substrat um das aktive Zentrum, dann wird diese Art der Inhibition *kompetitive Hemmung* genannt. Bindet der Inhibitor im allosterischen Zentrum des Enzyms und es wird die Konformation des aktiven Zentrum so verändert, dass kein Substrat mehr binden kann, dann wird in diesem Fall von einer *nicht-kompetitiven Hemmung* oder *allosterischen Hemmung* gesprochen. Es gibt noch eine dritte Variante der Hemmung: wenn der Inhibitor an den Enzym-Substrat-Komplex bindet und somit die Produktbildung verhindert. Man spricht dann von *unkompetitiver Hemmung*.

Gerade diese spezifischen Eigenschaften der Enzyme: Ausbildung von Enzym-Substrat-Komplexen und anschließende Produktbildung oder die Inhibition durch die Bildung von Enzym-Inhibitor-(Substrat)-Komplexen ohne dass ein Produkt entsteht wird in der Enzymanalytik genutzt, um:

- a) mit Hilfe von bekannten Enzym-Assays neue experimentelle Methoden zu entwickeln und zu etablieren, bzw. die Versuchsparameter (Puffer, pH-Wert, Konzentrationen der Edukte, ...) zu optimieren.
- b) mit etablierten Methoden neue Enzyme, Substrate oder Inhibitoren zu identifizieren, z.B. in komplexen biologischen Proben, wie Pflanzenextrakte.

Gerade mit dem Einzug der Massenspektrometrie (siehe Kapitel 1.3) in die "funktionelle" Proteomik und somit auch in die Analytik von Enzym-Assays haben sich neue Möglichkeiten der Enzymtechnologie aufgetan. Dadurch ergeben sich aber auch neue Ansprüche an die Datenauswertung. Oft können große Hersteller von Analysengeräten nicht mit der Geschwindigkeit der Methodenentwicklung im Labor mithalten und es entsteht dadurch eine Lücke in der Datenauswertung, die durch neue modulare, oftmals Open Source - Softwarelösungen geschlossen werden kann (siehe auch Kapitel 1.4).

In der Implementierung einer Softwarestrategie zur Behebung/Verbesserung dieses Dilemmas ist auch die Motivation zu dieser Arbeit begründet. Es sollte ein modulares Software-Werkzeug geschaffen werden, das:

- a) die Stabilität und Funktionalität neuer Analysetechniken zeigen kann, damit
- b) im Nachgang komplexe biologische Proben z.B. auf unbekannte Inhibitoren untersucht werden können.

1.3 Analysemethoden der (funktionellen) Proteomik

Wie in den obigen Abschnitten bereits erwähnt, hängen die Ergebnisse in der (funktionellen) Proteomikforschung auch stark vom Fortschritt und von der Entwicklung der Analytik- und Informationstechnologie ab.

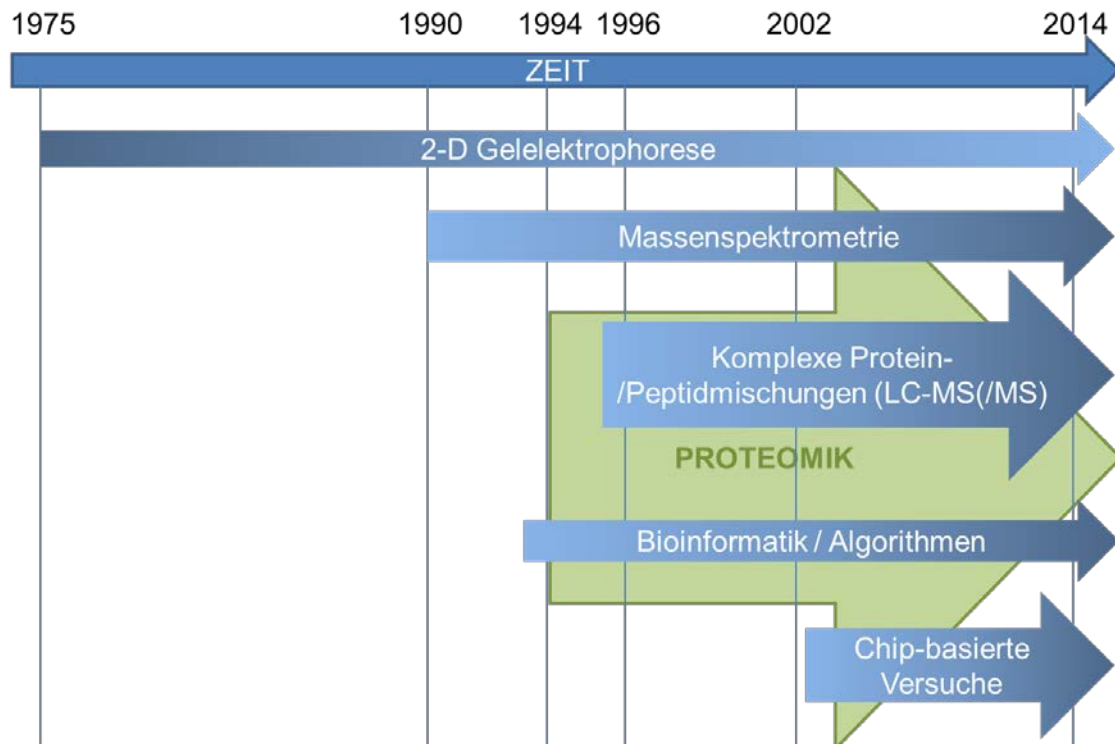


Abbildung 4: Zeitlicher Verlauf der Entwicklung der Analysentechnologien. Je dunkler die Farbe wird, umso stärker ist der Einfluss auf die Proteomik.

Die 2-D Gelelektrophorese ^[31, 32] wurde in den 70er Jahren entwickelt und war bis in die 90er Jahre die Standard-Analysemethode zur Proteom-/Proteinidentifizierung und -quantifizierung. Diese Methode wurde in dieser Zeit immer weiter verbessert. Die Proteine werden nach zwei physikochemischen Eigenschaften voneinander getrennt. Die Trennung in der ersten Dimension erfolgt nach dem isoelektrischen Punkt und in der zweiten Dimension nach der Größe des Proteins. Anfang der 90er Jahre wurde 2DE mit der Massenspektrometrie gekoppelt. Die 2DE-MS stellte eine attraktive Analysetechnologie dar, um tausende von Proteinen voneinander zu trennen, wobei nur geringe Mengen an Probenmaterial notwendig waren ^[33]. Mit der Entwicklung der "Shotgun Proteomics" ^[34, 35], erstmals veröffentlicht durch Appella et al. 1995 ^[34], wurde die erste gelfreie Proteomik-Analysetechnologie vorgestellt (Abbildung 4). Das System bestand aus einer flüssigchromatographischen Trenntechnik und einer MS/MS Detektionseinheit. Mit diesem System konnten erstmals gelfrei komplexe Peptidmischungen analysiert und die Peptide identifiziert werden. Die LC-MS etablierte sich damit endgültig als Standardmethode für die Proteomik-Forschung. Allerdings bedeutet dies nicht, dass die 2DE damit überflüssig geworden wäre. Beide Ansätze haben ihre Vor- und

Nachteile. Eine Kombination aus beiden Ansätzen, in Abhängigkeit vom Versuch, ist hier wohl die beste Strategie.

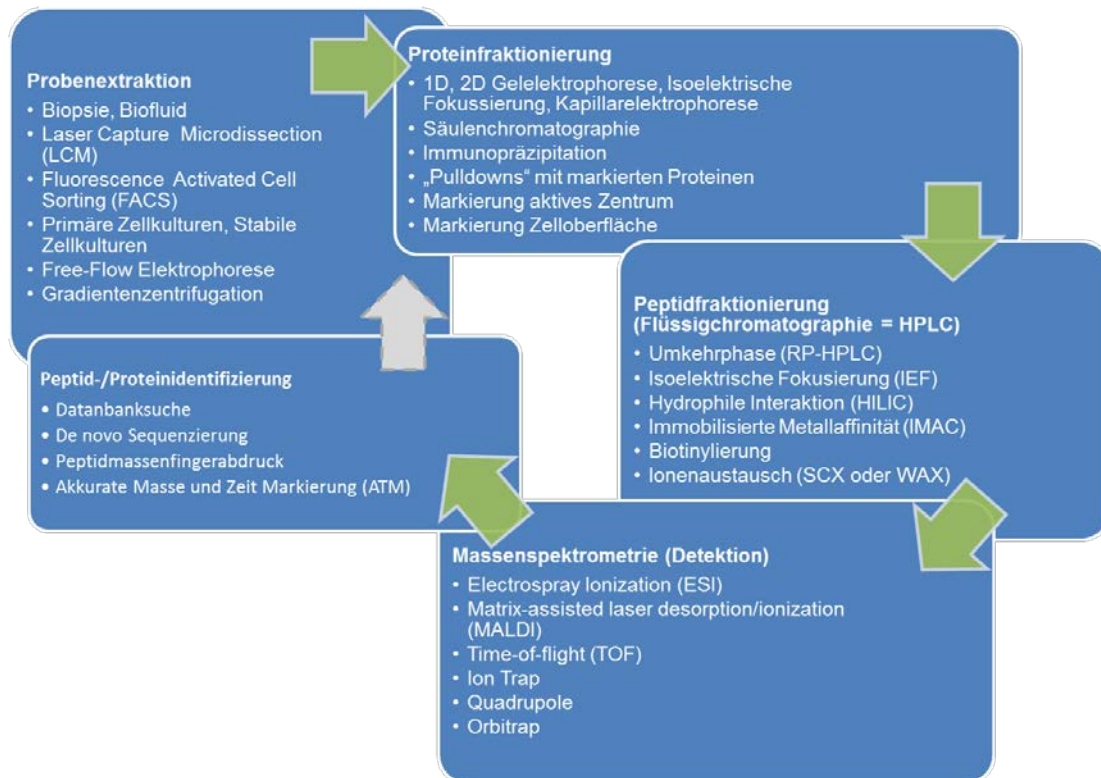


Abbildung 5: Darstellung der "State of the Art"-Methoden der Proteomik am Versuchsablauf ^[36].

Seit Ende der 90er Jahre bis heute hat sich eine Vielzahl von Methoden in der (Prote)omik-Forschung etabliert. Parag Mallick und Bernhard Küster haben mit Ihrem Überblick "Proteomics: a pragmatic perspective" (2010) ^[36] anschaulich die zahlreichen Technologien dargestellt. Die Methoden wurden aufgeteilt nach dem Ablauf einer Proteomanalyse, angefangen bei der Probenextraktion bis hin zur Identifizierung und Quantifizierung einzelner Proteine (Abbildung 5). Die Massenspektrometrie hat sich mittlerweile als Standard-Detektionsmethode in der (funktionellen) Proteomik etabliert.

Nachfolgend wird näher auf die Methoden der "funktionellen" Proteomik eingegangen, speziell auf die der Enzymologie. Die anderen Methoden sind in Abbildung 5 dargestellt und werden nicht näher erläutert. Weiterführende Informationen können dem Overview von Mallick und Küster "Proteomics: a pragmatic perspective" (2010) ^[36] und dessen Anlagen entnommen werden.

Historisch gesehen ist die Enzymologie/Enzymtechnologie von großem Forschungsinteresse seit dem Zeitpunkt an dem herausgefunden wurde, dass Reaktionen durch Proteine katalysiert werden können und es sich bei den meisten Proteinen um Enzyme handelt. Mit dem Aufkommen der "funktionellen" Proteomik, als ein sehr produktives Forschungsfeld ^{[8-11,}

^{13]}, hat die Enzymologie neue Popularität gewonnen. Dies wurde durch die Vorteile der Massenspektrometrie noch verstärkt.

Die ersten Veröffentlichungen in denen im Titel bzw. in der Zusammenfassung die Begriffe "Enzyme", "Enzyme Assay" und "Mass Spectrometry" enthalten waren, sind bei PubMed im Jahr 1973 (3 Artikel) gelistet (Abbildung 6). Mit dem Schlüsselbegriff "Liquid Chromatography Mass Spectrometry" sind es zu diesem Zeitpunkt bereits 14 Artikel. Im Jahr 1995, indem das erste Mal LC-MS/MS in der Proteomik erwähnt ist, sind für die Kombination mit MS 152 und für die Kombination mit LC-MS 170 Artikel zu finden. Die Anzahl der Veröffentlichungen in den nächsten rund 20 Jahren steigert sich für beide Begriffe rapide: MS = 729 Artikel (2014), LC-MS = 512 Artikel (2013).

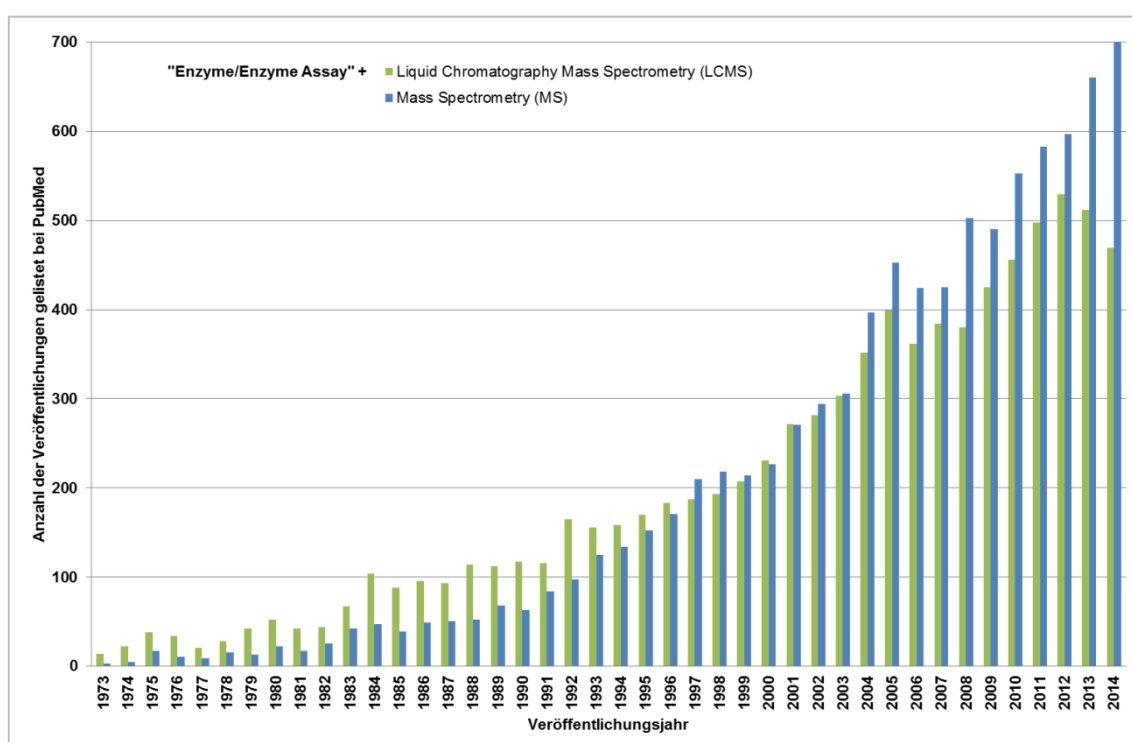


Abbildung 6: Abfrage bei PubMed nach Veröffentlichungen in den Feldern "Titel/Zusammenfassung" nach den Begriffen "Enzyme" ODER "Enzyme Assay" UND "Liquid Chromatography Mass Spectrometry" (grün) bzw. UND "Mass Spectrometry" (blau).

Seit 1973 wurden zusammengefasst 17.384 Artikel bei PubMed mit den genannten Begriffskombinationen gelistet. Welche Bedeutung die Enzymologie in Verbindung mit der LC-MS oder nur der MS dazu gewonnen hat, wird ersichtlich, wenn man die Gesamtzahl der Veröffentlichungen seit 1995 - dem "Startschuss" der LC-MS/MS Technologie in der Proteomik - vergleicht: Zeitraum 1995 bis zum Abfragedatum 24.03.2015 = 14.671 Artikel, d.h. ca. 84% der Artikel wurden in den letzten 20 Jahren veröffentlicht. Seit 2004 wurden 10.580 Artikel gelistet, dass entspricht ca. 60% des Gesamtanteils und ca. 72% des Anteils der letzten 20 Jahre, d.h. fast dreiviertel der Artikel der letzten 20 Jahre sind innerhalb der letzten 10 Jahre veröffentlicht worden. Diese Zahlen zeigen deutlich das gesteigerte und

weiter anhaltende Forschungsinteresse seit der Entwicklung der LC-MS zur Standardtechnologie im Bereich der "funktionellen" Proteomik, respektive der Enzymologie.

In dieser Arbeit liegt der Schwerpunkt auf der Detektion mittels Massenspektrometrie von *in-vitro* Enzym-Assays, sowie der dazugehörigen Softwareentwicklung zur Auswertung der gewonnenen Daten. Die enzymatischen Funktionen von Proteinen werden typischerweise mittels *in-vitro* Enzym-Assays untersucht. Hier gibt es zwei verschiedene Vorgehensweisen:

- a. die Analyse von Aliquoten der enzymatischen Reaktion nach definierten Zeitpunkten.
- b. das kontinuierliche Monitoring der enzymatischen Konversion.

Verschiedene klassische Methoden wurden zur Analyse enzymatischer Reaktionen etabliert: spektroskopische, elektrochemische oder radiometrische ^[37]. Viele Enzym-Assays basieren auf photometrischen oder fluorometrischen Detektionsmechanismen. Zahlreiche spektroskopische Detektionsmethoden sind im Hinblick auf ihrer einfachen Handhabung, die moderaten Kosten und die Genauigkeit der Methoden detailliert beschrieben ^[38]. Hierbei müssen während der Beobachtung der enzymatischen Reaktion Chromophoren oder Fluorophoren dem Substrat hinzugefügt oder davon abgetrennt werden. In vielen Fällen wird die Produktbildung gemessen. Ein großer Nachteil dieser Methoden ist die Notwendigkeit von künstlich modifizierten Substraten, um die Detektionseigenschaften überhaupt möglich zu machen. Diese "künstlichen" Substrate können die Reaktionsrate und/oder die Enzymkinetik beeinflussen ^[39]. Radioisotopische Markierung des Substrats wird aufgrund des Gesundheitsrisikos während der Versuchsdurchführung immer weniger eingesetzt.

In ihrem Review-Artikel mit dem Titel: "Functional proteomics: application of mass spectrometry to the study of enzymology in complex mixtures" aus dem Jahr 2012 geben Graßmann et al. ^[40] sehr detailliert den aktuellen Stand der Technik für Enzym-Assays mittels massenspektrometrischer Detektion wieder. Neben den MS-Methoden für die "klassische" Proteomik ^[5, 36, 41] können speziell für die "funktionelle" Proteomik, wie für die Bildung non-kovalenter Komplexe ^[42, 43] oder der Enzymologie ^[44], massenspektrometrische Ansätze genutzt werden. Verglichen mit den spektroskopischen und den radiometrischen Methoden hat die Massenspektrometrie nicht nur den Vorteil, dass sie in der Durchführung unbedenklich für die Gesundheit ist und natürliche Substrate eingesetzt werden können, sondern sie ermöglicht auch die simultane Detektion von Substrat, Zwischenprodukten, Produkten und in manchen Fällen sogar des Enzyms und seiner Komplexe. Nicht zuletzt ist dieser Grund ursächlich für den Zuwachs der MS-basierten Enzym-Assays in den letzten Jahren ^[45-48].

Die Etablierung neuer Methoden führte auch zu neuen Aspekten der Enzymologie:

- Screening großer Komponenten-Bibliotheken oder komplexer Mixturen (z.B. Pflanzenextrakte) auf potentielle Enzymregulatoren für neue Medikamente.
- Kenntnisse über natürliche Substrate neuer Enzyme zu erlangen, um diese Enzyme einer bekannten Funktionalität zuordnen zu können.
- Identifizierung von Enzymaktivitäten in komplexen Mischungen (z.B. menschliches Blutplasma), um die Ursachen von Enzymdysfunktion und ihrer Relation zu Erkrankungen herauszufinden.

Graßmann et. al. unterteilen in ihrem Review die MS-basierten Enzym-Assays in Assays mit einfachen Enzymlösungen und Enzym-Assays komplexer Mischungen. Die meisten "einfachen" Ansätze nutzen den klassischen Ansatz des partitionierten *in-vitro* Enzym-Assays. Auf diesen wird nicht näher eingegangen und nur der gängige Ablauf skizziert:

- Starten der Reaktion durch Mischen der Komponenten,
- Entnehmen von Aliquoten nach bestimmten Zeitintervallen,
- Stoppen der Reaktion in jedem Aliquot,
- Trennen der Reaktionsprodukte,
- Detektion der Komponenten (Quantifizierung) ^[44].

Weiterführende Informationen können dem Review ^[40] und den dort genannten Referenzen entnommen werden ^[45-48].

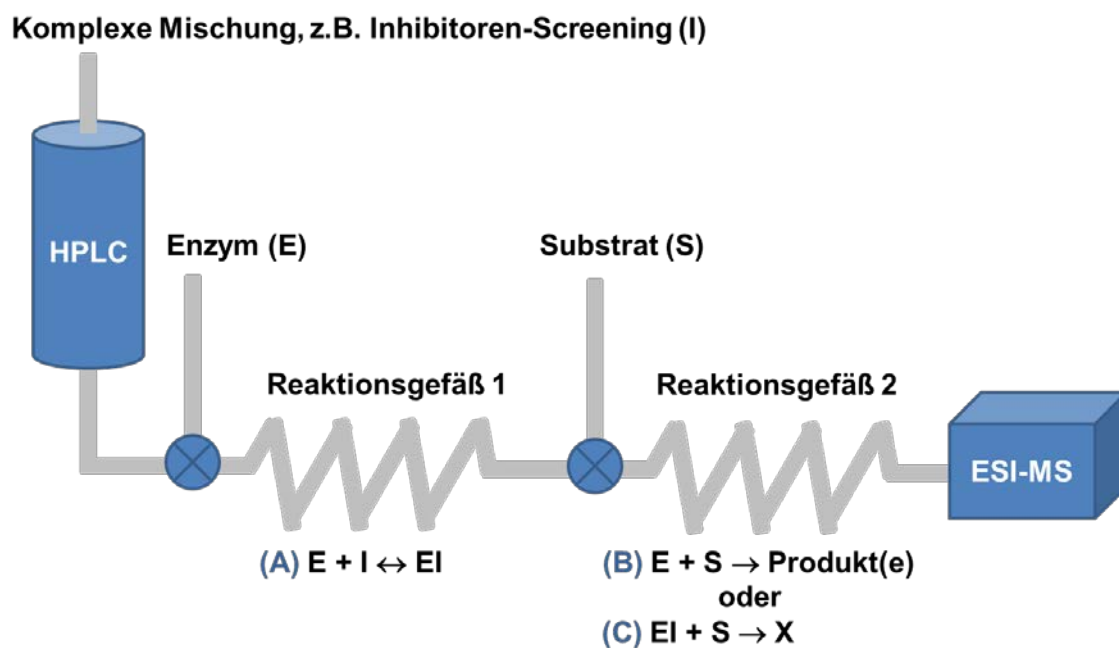


Abbildung 7: Schema eines kontinuierlichen Enzym-Assay mittels MS mit einer vorgeschalteten HPLC zur Auftrennung von komplexen Mischungen.

Neben obiger Methode hat sich der kontinuierliche Enzym-Assay in den letzten Jahren rasant entwickelt. Diese noch relativ neue Methode ermöglicht das Monitoring von enzymatischen Reaktionen in Echtzeit ^[44, 49, 50]. Lee et al. hat 1989 das erste System vorgestellt, indem über eine Injektionspumpe kontinuierlich die Reaktionsmischung in das MS eingebracht wird. Die Autoren nutzten das System, um Laktase, α -Chymotrypsin und Leucin Aminopeptidase in Echtzeit zu monitoren ^[51]. Die Grundidee ist, dass dem System ein konstanter Fluss von Enzym- und Substratlösung zugeführt und zur Detektion ins MS eingebracht wird (Abbildung 7 - ohne HPLC). Dieser Ansatz ist weitaus zeitsparender als der konventionelle Ansatz und ermöglicht die Determination von Enzymkinetiken, Substrat(en), Produkt(en) und in einem erweiterten Ansatz (Abbildung 7, inkl. HPLC) die Auswirkungen von zusätzlich zugeführten Inhibitoren in einem einzigen Durchlauf.

Mit der in Abbildung 7 vereinfacht dargestellten Methode konnten verschiedene hydrolysierende Enzyme untersucht werden ^[52-58]. So konnte am Beispiel der Hydrolyse von hexa-N-Acetylchitohexaose durch Lysozym in Echtzeit der Einfluss eines Inhibitors gemessen werden ^[54]. Weitere Beispiele für den Erfolg des kontinuierlichen Ansatzes in verschiedenen Variationen sind:

- Untersuchung der Aktivität von Glycosidasen, Chitosanase, Chitinase und verschiedenen Mutanten ^[56].
- Monitoring der Substrat- und Produktkomplexe des Enzyms 4-hydroxybenzoyl-Coenzym A (4-HBA) Thioesterase und dessen Substrat führten zur Annahme, dass während der Katalyse das Produkt 4-HBA früher aus dem aktiven Zentrum entlassen wird als das Coenzym A ^[59].

Diese und andere Beispiele zeigen die Bandbreite auf, die durch die Echtzeit-MS im Bereich der Enzym-Forschung ermöglicht wird. Neben "einfachen" Ansätzen konnte in Studien gezeigt werden, dass sich der Versuchsansatz in Abbildung 7 auch für komplexe Mischungen, wie z.B. das Inhibitor-Screening in Pflanzenextrakten ^[60], eignet. Die Einfachheit und die Vorteile des MS-basierten Ansatzes werden klar ersichtlich, wenn man sich am Beispiel des Inhibitor-Screenings den Ablauf ansieht. So wird die komplexe Mischung, z.B. das Pflanzenextrakt, über eine HPLC separiert. In Reaktionskapillare 1 (Abbildung 7) wird kontinuierlich Enzymlösung gepumpt. Diese vermischt sich mit den Eluenten aus der HPLC und es kann zu einer Reaktion mit einem eluierten (unbekannten) Inhibitor kommen. In Reaktionskapillare 2 (Abbildung 7) wird kontinuierlich Substrat und das Gemisch aus Reaktionskapillare 1 zugeführt, bevor es zur Detektion in das MS weitergeleitet wird. Im Massenspektrometer können gleichzeitig alle Reaktionen anhand des Massen-zu-Ladungs-Verhältnisses erfasst werden: Enzym, Inhibitor(en), Substrat, Enzym-Inhibitor-

Komplex(e) und Produkt(e). Man beobachtet kontinuierlich die Produktpur im MS über die Zeit, also das Chromatogramm.

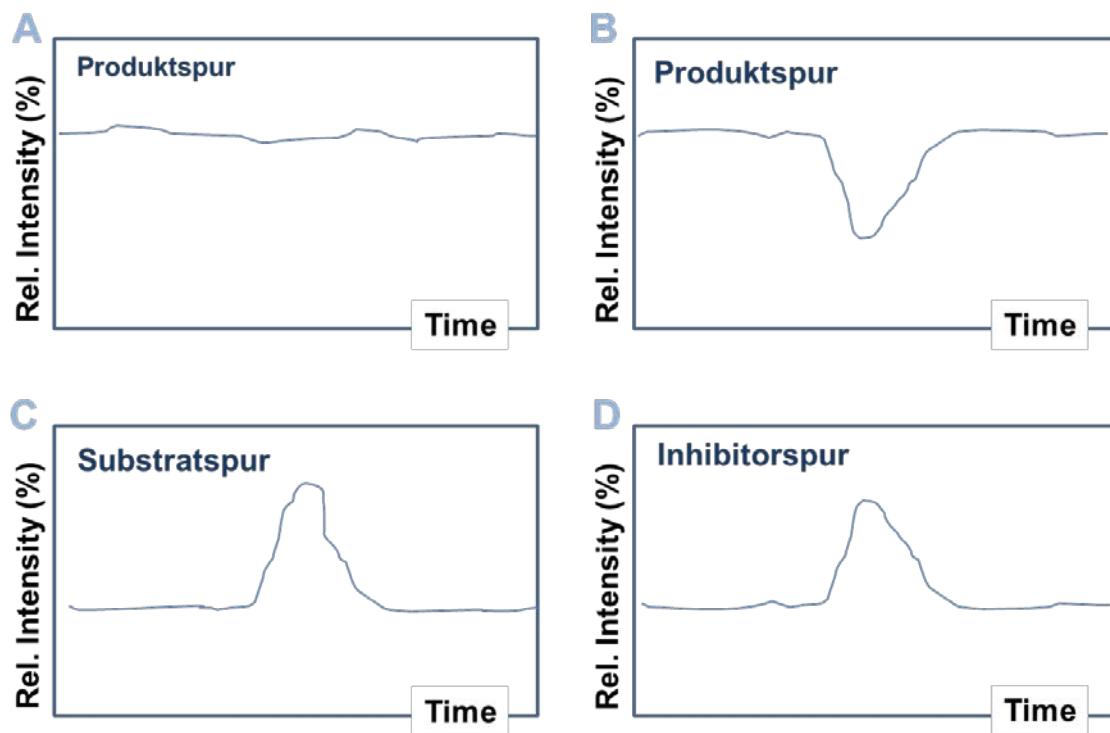


Abbildung 8: Kontinuierliche Produktmessung eines MS-basierten Enzym-Assays. A) Produktpur ohne Inhibition und B) Produktpur, C) Substratspur und D) Inhibitorspur zum Zeitpunkt der Inhibition.

Reagiert das Enzym nicht mit einem Eluenten aus der HPLC, bleibt die Produktpur eine Parallele zur Zeitachse (Abbildung 8A). Wird aber ein Inhibitor eluiert, dann gibt es eine Reaktion in Gefäß 1 und es kommt in Gefäß 2 voraussichtlich nicht zu einer Reaktion mit dem Substrat. Folglich erfährt die detektierte Produktpur an diesem Zeitpunkt im MS einen Einbruch - ein negatives Signal (Abbildung 8B). Vergleicht man jetzt zum gleichen Zeitpunkt die Substratspur und die Spur des Inhibitors (sofern bekannt), dann müssten beide ein positives Signal (Abbildung 8C und D) aufweisen, da die Anzahl beider, im Vergleich zu den Produkt-Ionen, im Detektor zugenommen hat. Ist der Inhibitor nicht bekannt, so muss man über das ganze gemessene Spektrum screenen und feststellen, wo zu dem gemessenen Einbruch des Produkts ein Anstieg einer anderen gemessenen Substanz zu sehen ist, die nicht Substrat ist. Ist diese Stelle gefunden, so können über das Massen-zu-Ladungs-Verhältnis Rückschlüsse auf den gefundenen unbekanntem Inhibitor gezogen werden.

In alternativen Ansätzen, die auf anderen Abläufen basieren, muss die Probe erst in vielen zeitaufwendigen Schritten getrennt und die einzelnen Komponenten, wie potentielle Liganden oder Komplexe, getrennt voneinander detektiert werden. Das macht solche alternative Verfahren zeitaufwendig und kompliziert. Die Abläufe von MS-basierten Enzym-Assays mit zwei oder mehreren Schritten können dem Review ^[40] von Graßmann et al. und den dort genannten Referenzen entnommen werden.

Ebenso notwendig wie eine gute Analysetechnik ist die Auswertesoftware, die am Ende der Versuche benötigt wird, um die Daten sinnvoll verarbeiten zu können. Wie sich die Software-"Landschaft" mit dem Aufkommen der Proteomik in den letzten Jahren entwickelt hat, wird im nächsten Kapitel beschrieben.

1.4 Software-Landschaft in der (Prote)omik-Forschung ¹

Wie komplex die Verknüpfungen von Biologie, Analytik und Informatik sind, wird in Abbildung 9 ersichtlich. Ausgehend vom Mikroorganismus werden die Relationen zwischen Analytik und Informatik für die Forschungsfelder Genomik, Proteomik, Proteogenomik, Interaktomik, Transkriptomik und Metabolomik dargestellt, wobei die Relationen nicht abschließend sind ^[61]. So kann ein Proteom eines Mikroorganismus analysiert werden, indem beispielsweise die Proteine mit einer LC oder 2DE aufgetrennt werden. Die Ergebnisse können dann mit Softwares, wie Melanie, PDQuest oder 2DHunt interpretiert und daraus weitere Schlüsse auf das vorliegende Proteom gezogen werden. Ein anderer Ansatz wäre die Fragmentierung der Proteine (z.B. als tryptische Peptide) in der Massenspektrometrie (LC-MS) und der Einsatz von Softwares wie MASCOT ^[62], PEPSEA ^[63, 64] oder ProQuant ^[65], um Informationen zum Proteom zu erhalten. Üblicherweise werden in einem Versuchsablauf nicht alle Schritte durchlaufen, sondern abhängig vom angestrebten Ziel nur ein Set davon. Ein typischer Ablauf eines solchen Versuchs ist bereits in Abbildung 7 in Form des kontinuierlichen LC-MS Enzym-Assays aus dem Bereich der funktionellen Proteomik dargestellt:

1. HPLC zur Auftrennung der biologischen Probe (z.B. Pflanzenextrakt),
2. ESI-MS zur Detektion der Proteine bzw. Proteinfragmente (z.B. Inhibitoren),
3. Achroma-Software (=Teil dieser Arbeit) zum Monitoren nach unbekanntem Inhibitoren,
4. MASCOT ^[62] Software, um die unbekanntem Inhibitoren zu identifizieren.

Die Bioinformatik hat hier ein breites Aufgabenspektrum. Angefangen beim ursprünglichen Bedarf nach Datenbanken zur Archivierung der gewonnenen Ergebnisse aus der Proteomikforschung reicht das Spektrum über Werkzeuge zur Identifizierung von Proteinen und deren Quantifizierung bis hin zu Software, die Modelle für die Vorhersage von Reaktionen errechnet. Es muss unterschieden werden, dass es ganze Software Systeme und (kleine) Software-Werkzeuge für bestimmte Aufgabenstellungen bis hin zu einzelnen spezifischen Algorithmen, z.B. Suchalgorithmen für Proteindatenbanken, gibt. Viele Firmen

¹ Die Einleitung wurde teilweise publiziert:

Michael Krappmann, Marco Luthardt, Frank Leßke und Thomas Letzel: "The software-landscape in (prote)omic research", Journal of Proteomics & Bioinformatics.

siehe Appendix

bieten ganze Softwarepakete an, die es über eine graphische Benutzeroberfläche ermöglichen, mehrere Schritte mit einer Software zu erledigen.

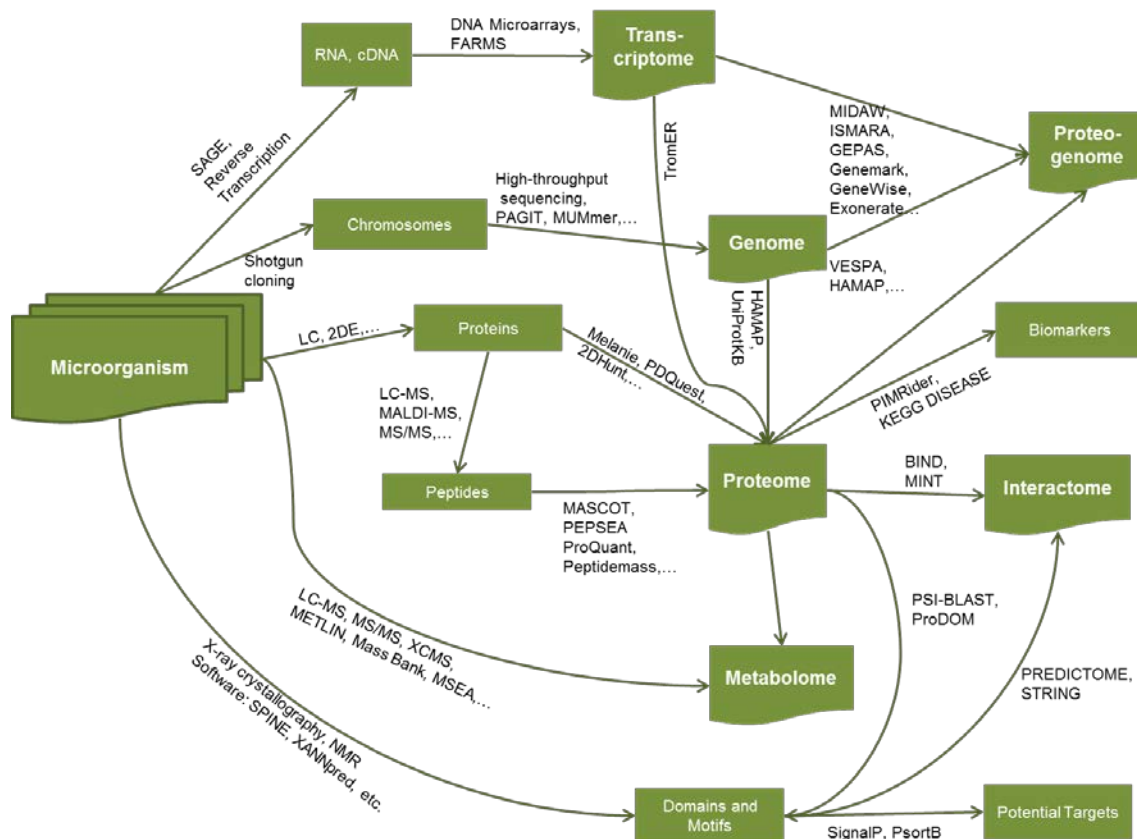


Abbildung 9: Darstellung der Relationen verschiedener Analysetechniken zu den einzelnen Softwarewerkzeugen und Datenbanken ^[61].

1.4.1 Vielfalt der Software-Landschaft

Einen Überblick über die Vielfältigkeit der Software-Landschaft in der Proteomik gibt die Proteomikwebseite bei ExPASy ^[66]. Dort sind unter dem Begriff Proteomik 31 Datenbanken und 240 Software-Werkzeuge in acht Kategorien gelistet:

- protein sequence and identification,
- mass spectrometry and 2-DE data,
- protein characterisation and function,
- families, patterns and profiles,
- post-translational modifications,
- protein structure,
- protein-protein interaction,
- similarity search/alignment.

Aber auch ExPASy spiegelt noch nicht die gesamte Bandbreite wieder. Neben den Softwaresystemen der Gerätehersteller, z.B. von Massenspektrometern, gibt es mittlerweile eine große "Open Source"-Gemeinschaft und viele frei benutzbare Software-Werkzeuge

(Freeware), die oft in Forschungsinstituten oder Universitäten ihren Ursprung haben, aber nicht bei ExPASy gelistet sind, weil sie sehr speziell sind oder noch nicht den Bekanntheitsgrad erreicht haben, der notwendig ist, um über die Webseite verlinkt zu werden. Ebenso gibt es mittlerweile eigene Rubriken in der freien Web-Enzyklopädie Wikipedia, die ausschließlich Proteomik-Software listen. Ein Beispiel hierfür ist die "List of mass spectrometry software". Dort werden in drei Hauptkategorien ("Proteomics software", "MS/MS peptide quantification" und "Other software") 76 kommerzielle und freie Softwares gelistet. Viele Software-Werkzeuge wurden zum Thema Protein-/Peptididentifizierung bzw. der Quantifizierung entwickelt, dem Hauptthema der Proteomik. Diese im Einzelnen aufzuführen würde den Rahmen dieser Arbeit überschreiten, deshalb ist am Ende dieses Kapitels eine ausführliche aber nicht abschließende Auflistung in Tabelle 2 von kommerzieller und freier Softwareplattformen und -Werkzeugen dargestellt. Die in Tabelle 2 gelisteten Softwares sind nur zum Teil bei ExPASy geführt und erweitern somit die Vielfalt der Softwarelösungen zur Auswertung von Proteomikversuchen.

1.4.2 Datenformate und -Konverter

Alle großen Gerätehersteller von Massenspektrometern bieten heutzutage auch Softwarelösungen zur Auswertung von Proteomikversuchen zusammen mit ihren Geräten an (Tabelle 2 - "2. Kommerzielle Software"). Die kommerziellen Softwarelösungen arbeiten in der Regel mit geschlossenen Datenformaten. Es haben sich aber auch einige spezielle für die Massenspektrometrie entwickelte offene Datenformate etabliert. Auf diesen Trend haben bereits viele Hersteller von Geräten und Entwickler von kommerziellen Softwarelösungen reagiert und bieten meist ein Austauschformat auf XML-Basis an oder liefern mit ihrer Software einen sogenannten Datenkonverter aus, mit dem Daten in ein anderes nicht herstellerspezifisches Format exportiert werden können.

Die Proteomic Standard Initiative (PSI) der Human Proteome Organization (HUPO) hat in den letzten Jahren einige Standards auf Basis der "Extensible Markup Language" (XML), einer textbasierten Sprache zur Darstellung von Hierarchien, entwickelt. Durch den strukturierten Aufbau dieser Textdateien durch definierte Elemente mit Anfangs- und Endmarkierung sind XML-Formate sehr gut zum Austausch von Daten geeignet. In Tabelle 2 sind unter erstens einige XML-basierte Datenformate aufgeführt.

Diese Notwendigkeit ist aus der Tatsache entstanden, dass in vielen Labors und Instituten immer neue Methoden entwickelt wurden, um MS- oder MS/MS-Daten auf neue Weise auszuwerten, die die Standardsoftware der Gerätehersteller noch nicht abdecken konnten. Es wurden neue Algorithmen und spezielle Software-Werkzeuge entwickelt, aber die Daten konnten nicht von der Gerätesoftware auf andere Softwarelösungen übertragen werden. Die geschlossenen Datenformate der Gerätehersteller wurden z.T. mittels *Reverse Engineering*

direkt von Softwareentwicklern entschlüsselt. Durch die stetige Weiterentwicklung sind freie Daten-Konverter entstanden, wie *ReAdW* (.Raw-Konverter), *mzWiff* (.wiff-Konverter), *MassWolf* (.raw-Konverter) oder *Trapper* (Agilent .d- Daten-Konverter). Diese sind zwischenzeitlich alle in die Software *mconvert* überführt worden, welche Teil von ProteoWizard^[67] (Tabelle 2) ist. Auch andere "Open Source"-Lösungen wie OpenChrom^[68] bieten viele herstellerspezifische Formate an.

1.4.3 Software-Bedarf in der funktionellen Proteomik/Enzymologie

Der Fokus dieser Arbeit liegt in der "funktionellen" Proteomik, speziell der Auswertung von Daten, die mit kontinuierlichen Enzym-Assays gewonnen wurden. Hier besteht noch großer Bedarf der Auswertungsoptimierung. Eine Abfrage bei PubMed mit den Begriffen: "Enzyme" AND "Software" AND "Mass Spectrometry" ergab insgesamt 56 Treffer seit 2002 (Abbildung 10). Wobei der bisherige Höhepunkt 2014 mit 13 Veröffentlichungen erreicht wurde und 39 von den 56 Veröffentlichungen seit 2009 erfasst worden sind. Die Abfrage zeigt aber auch, dass die Softwareentwicklung speziell für MS-basierte Enzym-Assays ein noch junges Forschungsfeld ist, das nicht im Hauptfokus der Proteomforschung steht.

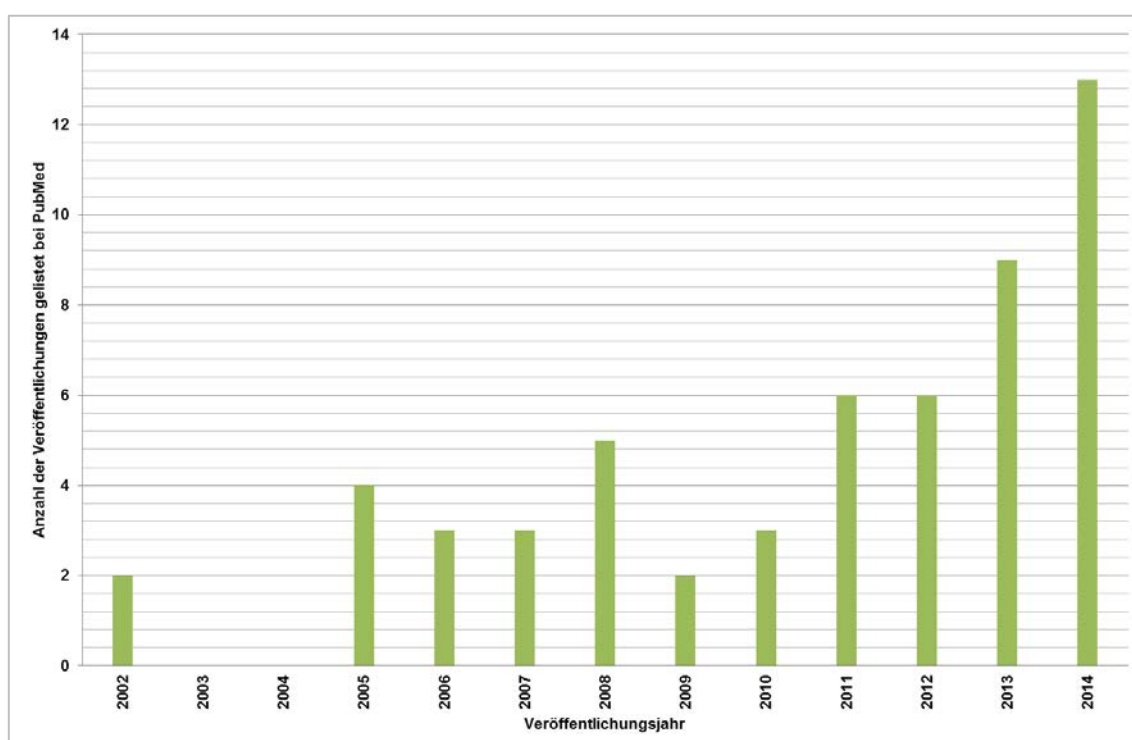


Abbildung 10: Abfrage bei PubMed zu den Begriffen "Enzyme" AND "Software" AND "Mass Spectrometry" als Und-Verknüpfung (AND) in "Title/Abstract".

Der Bedarf neuer Software ergibt sich oft daraus, dass sich einzelne Forschergruppen mit sehr speziellen Themen beschäftigen und nach automatisierten Lösungen zur Auswertung der Daten suchen. Hierfür werden dann sehr spezielle Software-Werkzeuge/-Algorithmen

entwickelt, die nicht allgemein angewendet werden können. Die Kombination aus bereits bekannten und diesen speziellen Softwares/Algorithmen könnte durchaus wertvoll sein. Darin besteht auch eine der Aufgaben für die Zukunft: Interoperabilität zwischen den vielen Softwareplattformen und -Werkzeugen herzustellen.

1.4.4 Interoperabilität und Workflow

Zu einem Teil ist die Interoperabilität auf Datenebene durch die Datenaustauschformate (Tabelle 2 - "1. Datenformate") gewährleistet. Wie kann diese aber unter den einzelnen und vielfältigen Software Systemen hergestellt werden? Aufgrund der Vielzahl der verschiedenen Lösungen (Expasy: 271, Wikipedia: 76 und Tabelle 2: 56 - z.T. mit Überschneidungen der drei genannten Quellen) ist es nicht möglich alle Lösungen in einer Plattform zu vereinen, d.h. dass für komplexe Auswertungen unter Umständen zwischen verschiedenen Software-Systemen gewechselt werden muss. Somit ist die Interoperabilität auf Software-Plattform-Ebene nicht stark ausgeprägt.

Andererseits gibt es natürlich auch viele Lösungen, die gleiche bzw. ähnliche Ansätze zur Auswertung der Daten anbieten, sich von den Algorithmen her ähneln oder sich die Lösungsansätze überschneiden. Unter der Vielzahl von Softwares finden sich auch Lösungen die als reine "Pipelines" dienen, wie z. B. die Trans-Proteomic Pipeline (TPP)-Software ^[69] (Abbildung 11) oder TOPP ^[70]. Diese "Pipelines" sind darauf spezialisiert bereits bestehende Werkzeuge zu einer Abarbeitungskette zu verbinden, die automatisiert durchlaufen wird. Dies bringt den Vorteil, dass nicht manuell die Daten zwischen den einzelnen Softwares transportiert werden müssen, sondern die Pipeline-Software erledigt dies und liefert die Endergebnisse zurück.

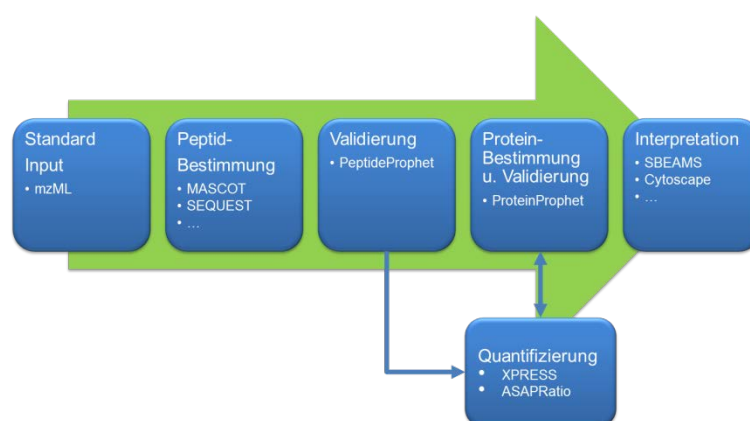


Abbildung 11: Ablauf einer automatisierten Abarbeitungsliste der verschiedenen Softwares (Spiegelstriche) am Beispiel von TPP.

1.4.5 Programmiersprachen und -techniken der Proteomik

Aus den bisherigen Beschreibungen wird ersichtlich, dass es für alle Bereiche der Informatik in Bezug zur Proteomik bereits diverse Softwarelösungen gibt, angefangen beim Austausch-Datenformat, über entsprechende Konverter und Lösungen zum Management von

Proteomikdaten, bis hin zu speziell für die Proteomik entwickelten Lösungen und Software, die mehrere Softwares zu einer Abarbeitungsliste zusammen fügt.

So heterogen die Software-Landschaft auch sein mag, so haben sich überwiegend zwei Programmiersprachen zur Entwicklung der Software-Lösungen etabliert, zum einen die Sprachen der C-Familie: C, C++ und C#, wobei C++ hier dominiert und zum anderen die neuere Programmiersprache Java.

Historisch bedingt ist ein überwiegender Teil der kommerziellen Lösungen in C und C++ programmiert, da diese meist ihren Ursprung bei den Geräteherstellern und in der klassischen Massenspektrometrie haben. C und C++ waren lange Zeit Standard für gerätenahe Anwendungsentwicklung und kommen dort auch heute noch zur Anwendung. Somit wurden viele Anwendungen der ersten Generation in C und C++ entwickelt. Diese Anwendungen haben meist den Nachteil, dass sie nur unter Microsoft Windows oder Microsoft Server Betriebssystemen laufen.

Mit Entwicklung der "Proteomics Community" hat sich ebenfalls eine große Gemeinschaft etabliert, die freie oder sogar "Open Source" Software zur Verfügung stellt. Bei Letzterem wird der gesamte Quellcode mit zur Verfügung gestellt, damit möglichst viele daran partizipieren und mitentwickeln können. Die frei zugänglichen Software-Lösungen sind häufig plattformunabhängig, bzw. laufen mindestens auf Linux- und Windowsbetriebssystemen. Das resultiert aus der Tatsache, dass zum Entwickeln oft freie Werkzeuge verwendet werden. Aus Tabelle 2 "3. Freie und Open Source Plattformen und Software-Werkzeuge" wird klar ersichtlich, dass sich in diesem Bereich immer mehr die Programmiersprache Java und deren Hilfsmittel (13 Softwares gelistet) zum Standard entwickelt, obwohl C++ (10 Nennungen) nach wie vor stark vertreten ist. Aber gerade die neueren Software-Lösungen seit 2010 wurden oft in Java entwickelt. Solche Entscheidungen hängen zudem auch vom informationstechnischen Hintergrund der Forschergruppe ab. So ist z.B. Python (5 Nennungen) eine weitverbreitete Programmiersprache im naturwissenschaftlichen Bereich, da Python sehr gut im Batch-Modus einsetzbar ist und sich für die Programmierung von Pipelines deshalb sehr gut eignet. Mit wxPython bietet die Programmiersprache zwischenzeitlich auch Werkzeuge an, um graphische Benutzeroberflächen umzusetzen. Ebenfalls mit 5 Nennungen in Tabelle 2 unter 3. sind webbasierte Lösungen zu finden, die wohl in Zukunft und gerade mit der Weiterentwicklung der webbasierten Programmierung (Web 2.0) und von *Cloud*-Lösungen mehr Interesse von der Bioinformatik/Proteomik-Gemeinschaft erhalten werden.

Cloud Computing rückt erst langsam in den Fokus der Bioanalytik/Proteomik. Das Cloud Computing wurde vom National Institute of Standards and Technology (NIST) wie folgt

definiert: "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.", was bedeutet, dass Cloud Computing ein Modell darstellt, welches es ermöglicht jederzeit, komfortabel und nach Bedarf Zugriff auf einen gemeinsamen Pool von konfigurierbaren Computerressourcen zu haben, die mit minimalem Managementaufwand vorgehalten und zur Verfügung gestellt werden können. Dies können Netzwerke, Server, Speicher, Applikationen oder Services sein. Es sind drei Service-Modelle und vier Deployment-Modelle beschrieben, wobei das Service-Modell den Service beschreibt, der zum Nutzer gebracht wird und das Deployment-Modell somit wer den Service nutzen darf und wie. Der Vorteil der Cloud liegt darin, dass die in der Proteomik anfallenden großen und komplexen Datenmengen in einem externen Rechenzentrum schnell und effizient verarbeitet werden und die Ergebnisse anschließend einfach über Internet mit vielen Forschern und/oder Forschungsinstituten weltweit geteilt werden können. Das hierfür am besten geeignete Service-Modell heißt: "Software as a Service (SaaS)". Komplexe Auswertungen laufen innerhalb der Cloud und die Ergebnisse werden im Client, z.B. im Webbrowser, dargestellt. Das Deployment-Modell beschreibt den Status der Cloud. Die drei Haupt-Modelle sind die private Cloud, die Gemeinschafts-Cloud und die öffentliche Cloud. Die private Cloud ist begrenzt auf eine einzelne Organisation, z. B. eine Forschungsabteilung oder ein Unternehmen. Die Gemeinschafts-Cloud bietet einem größeren Kreis Zugriff, welcher z.B. gemeinsame Interessen teilt. Die öffentliche Cloud ist nicht zugriffsbegrenzt und jeder kann partizipieren. Unabhängig vom Deployment-Modell einer Cloud bestehen, gerade in der Analytik, noch große Bedenken zur Datensicherheit, wenn diese extern berechnet werden. Wer hat Zugriff? Wem gehören sie? Der größte Vorteil des Cloud Computing für die Proteomik ist der Zugriff auf größere Rechnerleistungen für komplexe Analysen und das Teilen von Informationen innerhalb einer großen Forschergemeinschaft, was zu schnellerem Ergebniserfolg und wahrscheinlich neuen Erkenntnissen und Ansätzen führen kann. Jedoch muss der rechtliche Umgang mit den Daten gut geklärt sein, damit sich die Cloud-Technologie in der Bioanalytik/Proteomik etablieren wird. Als Beispiel für den Einsatz neuer Technologien der Informatik soll ein kürzlich veröffentlichter Übersichtsartikel aus dem Bereich der Metabolomik dienen. Der Übersichtsartikel beschreibt die Abhängigkeit der Metabolomik von der Bioinformatik und nennt unter anderem auch neue Ansätze, wie die Cloud-basierte Metabolomik^[71].

Im Java-Bereich haben sich in den letzten Jahren zwei konkurrierende Entwicklungsplattformen stark weiterentwickelt und gelten als Standard-Entwicklungsumgebungen für Java-Anwendungen: Eclipse^[72] und NetBeans^[73]. Beide Plattformen sind sogenannte "Integrated Development Environments" (IDE), also integrierte

Entwicklungsumgebungen. Das bedeutet, dass es sich bei diesen Plattformen um ein quelloffenes Programmierwerkzeug für Software in den verschiedensten Programmiersprachen und Art handelt. Man kann mit beiden Werkzeugen auch Anwendungen in anderen Programmiersprachen entwickeln. Da aber beide Plattformen überwiegend in Java entwickelt wurden und damit die Plattformen selbst sehr gut erweiterbar sind, werden die meisten Applikationen in Java entwickelt und beide Plattformen gelten somit als Standard-Entwicklungsumgebungen für Java-Applikationen. Größter Vorteil beider Plattformen ist, dass sie aufgrund der Java-Technologie betriebssystemunabhängig betrieben werden können und dass die Entwicklungsplattform selbst die Basis für die zu entwickelnde Applikation darstellt und mittels einer Plug-In-Technologie (Schlüssel-Schloss-Prinzip) die einzelnen Anwendungsteile als Plug-Ins oder Module darin integriert werden.

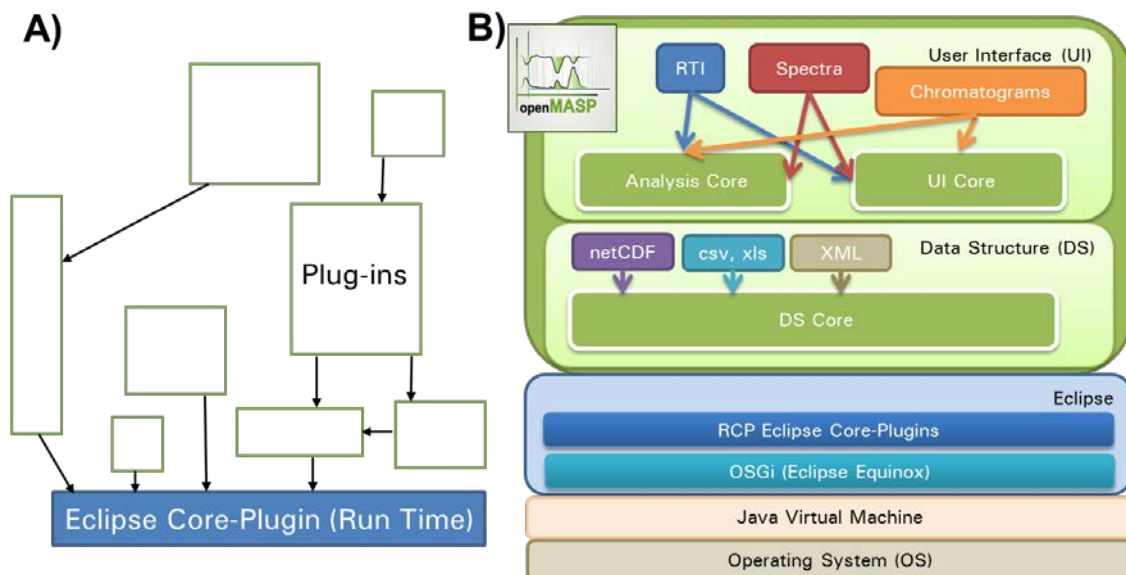


Abbildung 12: Plug-In Technologie am Beispiel der Eclipse-Plattform; A) schematische Darstellung der Verknüpfung von Plug-ins; B) Darstellung eines Fließbilds neuer Module einer Analytiksoftware, über die Eclipsemodule hin zur darunterliegenden plattformunabhängigen "Java Virtual Machine" und dem Betriebssystem.

Die Programmierumgebung stellt bereits alles zur Verfügung, um ein modulares Projekt zu generieren. Abbildung 12 veranschaulicht dies am Beispiel der Eclipse-Plattform. Auf der linken Seite (Abbildung 12 A) wird schematisch dargestellt, wie eine Plug-In Applikation aufgebaut ist. Selbst der Startpunkt der Applikation ist ein Plug-in, nämlich das Eclipse Kern-Plug-In. Abbildung 12 B) zeigt das Schema einer Projektidee zu einer Applikationsplattform, die während dieser Arbeit entstanden ist: der grüne Bereich zeigt die frei programmierbaren Plug-Ins, wobei in der obersten Ebene die Benutzerschnittstelle umgesetzt wird und in der darunterliegenden Ebene die Datenstruktur. Der mittlere blaue Bereich zeigt die Verbindung zu den Kern-Plug-Ins von Eclipse und darüber die Verbindung zur "Java Virtual Machine" die den interpretierten Java-Code ausführt. Die "Java Virtual Machine" läuft wiederum auf einem Betriebssystem. Diese kann auf verschiedenen Betriebssystemen installiert werden und

somit sind diese sogenannten "Cross-Plattform"-Anwendungen betriebssystemunabhängig. Diese beiden IDEs können neben der Programmierung für Java-Programme auch für die Anwendungsentwicklung von webbasierten Systemen, aber auch für Anwendungen der C-Familie verwendet werden. Jedoch gibt es für die Programmierung in C, C++ oder C# auch bereits etablierte z.T. kommerzielle Entwicklungsumgebungen, wie z.B. das Microsoft Visual Studio. Auch hier gibt es Hilfsmittel wie das *.NET-Framework* von Microsoft oder die *Microsoft Foundation Classes*, die nicht nur bei der grafischen Benutzerschnittstellenprogrammierung Vereinfachungen anbieten. Diese Applikationen sind in der Regel nicht Plug-in-basiert sondern folgen anderen Softwarearchitekturen. Die Achroma-Software, die Bestandteil dieser Arbeit ist, wurde mit Hilfe des Microsoft Visual Studio 2003 implementiert. Die Implementierung der Software wird im nächsten Kapitel "Material und Methoden" näher erläutert.

Tabelle 2: Auflistung aktueller Software-Werkzeuge und -Plattformen für die Auswertung von Proteomik-Daten, die mittels massenspektrometrischer Detektion gewonnen wurden.

Software	Hersteller/ Entwickler	Publi- kations- datum	Bezug zur Proteomik	Entwicklungs- basis	Plattform
1. Datenformate					
mzData ^[74]	HUPO PSI	2007	standardisiertes Datenaustauschformat	XML	unabhängig
mzXML ^[75]	Seattle Proteome Center; Institute for Systems Biology	2004	standardisiertes Datenaustauschformat	XML	unabhängig
mzML (Nachfolger von mzXML u. mzData) ^[74]	HUPO PSI	2008	standardisiertes Datenaustauschformat	XML	unabhängig
TraML ^[74]	HUPO PSI	2010	standardisiertes Datenaustauschformat	XML	unabhängig
mzIdentML ^[74]	HUPO PSI	2009	standardisiertes Datenaustauschformat	XML	unabhängig
mzQuantML ^[74]	HUPO PSI	2011	standardisiertes Datenaustauschformat	XML	unabhängig
mz5 ^[76]	Boston Children's Hospital; Steen & Steen Laboratory	2012	Speicheroptimiertes Datenaustauschformat	XML, HDF5	unabhängig
netCDF ^[77]	UCAR	1997	standardisiertes Datenaustauschformat	binär	unabhängig
2. Kommerzielle Software					
Analyst	ABSciex ^[78]	-	Basisplattform, automatisierte MS u. MS/MS Detektion, MRM Monitoring, Datenbanksuche möglich	C, C++	Windows
SWATH Acquisition	ABSciex	-	Datenunabhängige Datenerfassung, MRM, höhere Auflösung, MS/MS, labelfreie Quantifizierung	C, C++	Windows
MetabolitePilot	ABSciex	-	Detektion und Identifikation von Metaboliten (Arzneimittelherstellung)	C, C++	Windows
ProteinPilot	ABSciex	-	Proteinidentifizierung und - quantifizierung	C, C++	Windows
MultiQuant	ABSciex	-	MRM, Quantifizierung	C, C++	Windows
PeakView	ABSciex	-	TripleTOF, Masseninterpretation, akkurate Massen, Proteinstruktur	C, C++	Windows
MassHunter	Agilent ^[79]	-	Basissystem für MS-Detektion von Agilent, Peakdetektion, - Integration	C, C++	Windows
LCMS ChemStation	Agilent	-	Hochdurchsatz Quantifizierung, Peakdetektion, NIST-Suche	C, C++	Windows
OpenLAB	Agilent	-	Datencenter, Vernetzung der Analysegeräte und Software von Agilent	C, C++	Windows
Proteome Discoverer	Thermo Fisher Scientific Inc. ^[80]	-	Erweiterbare Plattform zur qualitativen u. quantitativen Proteomanalyse	C, C++	Windows
Xcalibur	Thermo Fisher Scientific Inc.	-	Basis Datensystem zur Gerätesteuerung und Datenanalyse	C, C++	Windows
ProMass	Thermo Fisher Scientific Inc.	-	Automatisierte Biomolekül Dekonvolution	C, C++	Windows
MassLynx	Waters ^[81]	-	Basisplattform für MS-Detektion von Waters	C, C++	Windows
OpenLynx	Waters	-	Qualitatives Screening und Identifikation	C, C++	Windows
QuanLynx	Waters	-	Automatisierte Quantifizierung	C, C++	Windows
ProteinLynx Global SERVER	Waters	-	Plattform für quantitative u. qualitative Proteomik	C, C++	Windows
Phenyx ^[82]	GeneBio	2006	MS/MS Protein Identifikationssoftware, - quantifizierung	-	webbasiert

Seaquest ^[63, 64]	Yates Lab	1993/ 1994	Algorithmus zur Proteinidentifikation mittels MS-Daten	-	Windows, Linux
PEAKS ^[83, 84]	Bioinformatics Solutions Inc.	2003	De Novo Sequenzierung, Datenbanksuche, PTMs, SPIDER-Suche, Quantifizierung, Protein-/Peptididentifizierung	C, C++	Windows
3. Freie und Open Source Plattformen und Software-Werkzeuge					
Achroma ^[85, 86]	Hochschule Weihenstephan-Triesdorf	2011	LC-MS, Funktionelle Proteomik	C++, C#	Windows
OpenChrom (freie Community Edition) ^[68]	Universität Hamburg, Lablicate	2010	GC/MS, LC/MS, HPLC-MS, ICP-MS, MALDI-MS	Java, Eclipse	unabhängig
Maltcms/Mauai ^[87]	Universität Bielefeld	2012	LC-MS-, GC-MS- u. GC-MS/MS-Hochdurchsatz-Metabolomik	Java, NetBeans	unabhängig
polyXmass / massXpert ^[88]	Filippo Rusconi	2001 / 2009	Simulation biologischer oder chemischer Reaktionen, Fragmentierung im MS (Proteinchemie)	C++, Qt-Bibliotheken	unabhängig
DeconLS ^[89]	Pacific Northwest National Laboratory	2009	Detektion von Moleküleigenschaften in MS-Daten unter der Nutzung von Isotopensignaturen; Dekonvolution	C++, C#, Microsoft NET Framework u. Visual Studio 2003	Windows
BioClipse ^[90, 91]	Dept. of Pharmaceutical Biosciences, Uppsala University; European Bioinformatics Institute	2007	Werkzeug-Bibliothek für Chemo- u. Bioinformatik	Java, Eclipse	unabhängig
BioSunMS ^[92]	Center of Computational Biology, Beijing Institute of Basic Medical Sciences	2009	Untersuchung von Peptid-Profilen via MS	Java, R, SQL	Windows, Linux
openMS ^[93]	Universität Tübingen, Freie Universität Berlin, ETH Zürich	2008	LC-MS Datenmanagement u. -Anlysen, Label-freie Quantifizierung, SILAC, iTRAQ, SRM, de-novo identification, Datenbanksuche	C++	Windows, Linux
MetaQuant ^[94]	Bioinformatics Organization	2006	Automatisierte Quantifizierung von GC-MS basierten Metabolom-Daten	Java	unabhängig
intelliMS ^[95]	Yonsei Proteome Research Center	2008	Management und Visualisierung von Tandem-MS Daten	PHP, MySQL	webbasiert
mMass ^[96]	Martin Strohal	2008	Protein- u Lipid-Identifikation, Protein Modifikation, Dekonvolution, Datenbanksuche, Standard-MS Auswertungen, wie Peakerkennung, Spektrum-Viewer,...	Python, wxPython für die grafische Benutzerschnittstelle	unabhängig
MZMine2 ^[97]	Okinawa Institute of Science and Technology	2010	Peak-Identifikation von MS-Daten, Peak Dekonvolution, 3D-Visualisierung, Statistik, Peak-Identifikation (Datenbanksuche)	Java, Java3D	unabhängig
pymzML ^[98]	Universität Münster	2012	Hochdurchsatz Bioinformatik mit MS-Daten	Python	unabhängig
MS-Spectre ^[99]	Netherlands Proteomics Centre	2007	Visualisierung von Spektren, Filter, Peakdetektion, Peak Matching	Java, Eclipse	unabhängig
ProteoWizard ^[67]	ProteoWizard Software Foundation	2008	Proteomik und MS Software-Werkzeugbox und Bibliothek; bietet mit msconvert ein Konvertierungswerkzeug von Herstellerformaten in mzML an	C++	unabhängig
X!Tandem / X!!Tandem (Parallele Berechnung) ^[100]	The Global Proteome Machine Organization	2004	Datenbank(suchalgorithmus) zum Matching von Tandem-MS mit Peptidsequenzen zur Proteinidentifikation	C	Linux

Mascot (Server) (freie Suche bis 1200 Spektrn und eingeschränkten Funktionen) ^[62, 101]	Matrix Science	1993	Datenbank(suchalgorithmus) zur Proteinidentifikation, - charakterisierung und - quantifizierung mittels MS-Daten	-	Windows, Linux
PRIDE ^[102]	EMBL-EBI	2005	Datenbank(suchalgorithmus) zur Proteinidentifikation	Java, HTML; SQL, XML	Windows, Linux und webbasiert
TOPP ^[70]	Universität Tübingen	2007	OpenMS Proteomik Pipeline - Analysen Workflow Software	C++	Windows, Linux
MapQuant ^[103]	Havard Molecular Technologies	2006	Large-scale Proteinquantifizierung	C++	Windows, Linux
SpecArray ^[104]	Seattle Proteome Center	2005	Vergleich von Peptid-Arrays mit LC-MS Datensätzen	C	Linux
ATAQS ^[105]	Seattle Proteome Center	2011	Automatisierte und zielgerichtete Analyse mit quantitativer SRM/MRM und MS	Java, R, mysql	Linux
Skyline ^[106]	MacCoss Lab Software	2010	Dokument Editor zur Erstellung und Analyse zielgerichteter Proteomik Experimente	C#, Windwos Forms	Windows
msInspect ^[107]	Fred Hutchinson Cancer Research Center	2006	Analyse von komplexen Proteinmischungen mittels LC- MS	Java	unabhängig
SuperHirn ^[108]	Seattle Proteome Center	2007	LC-MS Peptid-/Protein-Profilng	C++	Linux, MacOS X
MassSorter ^[109]	FUGE Bioinformatics Plattform	2006	Analysieren von MS- Proteinexperimenten mit bekannten Aminosäuresequenzen	Java, NetBeans, Java3D	unabhängig
PASSEL ^[110]	Institute of System Biology, Seattle	2012	PeptideAtlas SRM Experiment Bibliothek	-	webbasiert
jTraML ^[111]	CompOmics	2011	Java API für den TraML- Standard der Proteome Standards Initiative, SRM Experimente	Java	unabhängig
HDX-analyzer ^[112]	Texas A&M University	2011	Paket zur statistischen Analyse von Proteinstrukturndynamiken	Python, R	unabhängig
PeptidePicker ^[113]	Genome British Columbia Proteomics Center	2014	Werkzeug zur Selektion von möglichen Peptiden für zielgerichteten Proteomikexperimente	-	webbasiert
TPP (Trans-Proteomic Pipeline) ^[69]	Seattle Proteome Center	2005	MS/MS basierte "shotgun"- Proteomik, Identifikation und Quantifizierung	C++, Python	unabhängig
SpectraST ^[114]	Institute of System Biology, Seattle	2007	Suchalgorithmus für Spektrnbibliotheken zur Identifizierung von Peptiden mittels MS/MS	als Teil der TPP in Python	unabhängig
OMSSA (Open Mass Spectrometry Search Algorithm) ^[115]	National Center of Biotechnology Information	2004	Datenbanksuchalgorithmus zur Peptididentifikation mittels MS/MS	-	unabhängig
Andromeda (Bestandteil von MaxQuant) ^[116]	Max Planck Institute of Biochemistry	2011	Datenbanksuchmaschine zur Peptididentifikation mittels MS/MS	C#	webbasiert, Lokalinstallation: Windows
MaxQuant ^[117]	Max Planck Institute of Biochemistry	2008	Quantitative Proteomik mittels MS	C#	Windows
Proteios SE ^[118]	University Lund	2005	Plattform zur Speicherung, Organisation, Analyse und Annotation von Protein- Experimenten	Java	unabhängig
MyriMatch ^[119]	Vanderbilt University Medical Center	2006	Datenbanksuchmaschine zur Peptididentifikation mittels MS/MS	C++	Windows, Linux

2 Material und Methoden

2.1 Chemikalien

2.1.1 Enzyme

Sigma-Aldrich (Steinheim, Deutschland u. Zwijndrecht, Niederlande): Acetylcholinesterase aus *Electrophorus electricus* Typ VI-S (AChE, Enzyme Commission (EC) number 3.1.1.7, Mw ~280 kDa), intestinale alkaline Phosphatase aus der Rinder-Darmschleimhaut (iAP, EC 3.1.3.1., Mw ~160 kDa), Proteinkinase A aus dem Rinderherz (PKA, E.C. 2.7.1.37, 80 % protein; Phosphorylierungsaktivität: 1.1 Units per μg Protein), Proteinkinase A katalytische Untereinheiten aus dem Schweineherzen (PKA subunit; phosphorylation activity: 30-65 units per μg Protein), Lysozym aus dem Hühnereiweiß (HEWL, Mw 14.3 kDa, ~95 %, ~50 units per μg Protein).

2.1.2 Substrate

Sigma-Aldrich (Steinheim, Deutschland u. Zwijndrecht, Niederlande): Substrat für AChE, Acetylcholinchlorid (AChCl, Mw 181.7 Da), Substrate für iAP, Adenosin 5'-monophosphat (AMP, Mw 347.2 Da), Adenosin 5'-diphosphat (ADP, Mw 427.2 Da) und Adenosin 5'-triphosphat (ATP, Mw 507.2 Da), Substrate für PKA, Malantide (Mal, H-Arg-Thr-Lys-Arg-Ser-Gly-Ser-Val-Tyr-Glu-Pro-Leu-Lys-Ile-OH; >97 %; Mw 1633.9 Da), Kemptide (Kem, H-Leu-Arg-Arg-Ala-Ser-Leu-Gly-OH; >97 %; Mw 771.9 Da), Substrat für Lysozym, Hexa-N-acetylchitohexaose ((GlcNAc)₆, Mw 1237.2 Da, ≥ 95 %).

2.1.3 Inhibitoren und Extrakte

Sigma-Aldrich (Steinheim, Deutschland u. Zwijndrecht, Niederlande): Inhibitor für AChE, Galanthaminhydrobromid aus *Lycoris* sp. (Mw 368.3 Da), Inhibitor für PKA: Kaninchen PKA Inhibitor (5-24) Amid (P300, H-Thr-Thr-Tyr-Ala-Asp-Phe-Ile-Ala-Ser-Gly-Arg-Thr-Gly-Arg-Arg-Asn-Ala-Ile-His-Asp-OH; Mw 2222.4 Da), Inhibitor für Lysozym: N, N', N''-triacetylchitotriose (GlcNAc)₃, Mw 627.6 Da, ≥ 95 %)

Für das Screening-Experiment wurde eine Hausstaubprobe verwendet. Die Proben wurden am Institut für Energie- und Umwelttechnik e.V. (IUTA, Duisburg) vorbereitet und analysiert: 0,5 g Hausstaub wurde gesiebt und Fraktionen mit $\leq 63 \mu\text{m}$ mit 5 mL Methanol/H₂O (84:16) extrahiert. 1 mL wurde verdampft, wieder in Methanol/H₂O (80:20) gelöst, um den Anteil an organischem Lösungsmittel zu reduzieren, und mit einem 45 μm RC-Filter gefiltert.

2.1.4 Interne Standards

Bachem (Bubendorf, Schweiz): Interne Standards für PKA-Assay, Ovokinin (IS1, H-Phe-Arg-Ala-Asp-His-Pro-Phe-Leu-OH; >96 %; Mw 1002.2 Da) und PP60 c-src (phosphoryliert) (IS2, H-Thr-Ser-Thr-Glu-Pro-Gln-Tyr(PO₃H₂)-Gln-Pro-Gly-Glu-Asn-Leu-OH ; >99 %; Mw 1543.5 Da).

Merck (Darmstadt, Deutschland): Interner Standard für AChE-Assay, L-histidine (Mw 155.16 Da).

Sigma-Aldrich (Steinheim, Deutschland u. Zwijndrecht, Niederlande): Interner Standard für Lysozym-Assay, Malantide (Mal, H-Arg-Thr-Lys-Arg-Ser-Gly-Ser-Val-Tyr-Glu-Pro-Leu-Lys-Ile-OH; >97 %; Mw 1633.9 Da).

2.1.5 Weitere Chemikalien

Acros Organics (Geel, Belgien): Magnesiumsulfat (MgSO₄, >99 %).

J.-T. Baker (Deventer, Niederlande): Methanol (MeOH, 99.8 %) und Ammoniumbicarbonat (>99 %).

Millipore (Eschborn, Deutschland): Hochreines Wasser (Milli-Q-Wassersystem).

Riedel-de Haën (Seelze, Deutschland): Essigsäure (>99 %), Acetonitril (>99.9 %) und Methanol.

Merck (Darmstadt, Deutschland): Ammoniumacetat (NH₄Ac, >98 %) und Essigsäure.

Sigma-Aldrich (Steinheim, Deutschland u. Zwijndrecht, Niederlande): LC-MS Reinstwasser, Methanol und Ammoniumacetat (NH₄Ac, Mw 77.1 Da, >98 %) in LC-MS Qualität; für den PKA-Assay: Adenosin 3':5'-zyklisches Monophosphat (cAMP, freie Säure, >99 %), Adenosin 5'-Triphosphat (MgATP, Magnesiumsalz, 95-98 %).

2.2 Kontinuierliche Enzym-Assays mittels (LC-)MS

Der Schwerpunkt dieser Arbeit liegt in der Auswertung der mittels kontinuierlicher (LC)-MS Enzym-Assays gewonnenen Daten und der Entwicklung einer modularen, jederzeit erweiterbaren Software zur Auswertung dieser Daten. Nachfolgend werden die eingesetzten Versuchsaufbauten der einzelnen Applikationen beschrieben.

2.2.1 PKA-Assay: On-line Monitoring der Enzymaktivität und der Stabilität des Messsystems bei sich ändernden physikalischen Eigenschaften

Das On-line Monitoring System (Abbildung 13) bestand aus zwei HPLC Pumpen (Shicoh Engineering, Yamato, Japan), die über PEEK (Polyetheretherketon) Schlauchmaterial (Länge 0,5 m, Innendurchmesser 0,064 mm) jeweils mit dem Einlass eines SuperLoop™ (Pharmacia Biotech AB, Uppsala, Schweden) verbunden wurden. An beiden SuperLoops™ wurde eine Fließgeschwindigkeit von $10,0 \mu\text{L min}^{-1}$ angelegt. Die Enden der SuperLoops™ waren mit einer invertierten Y-Mischeinheit auf der einen Seite und auf der anderen Seite mit der maschigen ^[120] Poly(tetrafluorethylen)-Reaktionskapillare verbunden. Es wurden Reaktionskapillare mit einem Volumen von $5 \mu\text{L}$ (Länge 0,070 m, Innendurchmesser 0,3 mm), $100 \mu\text{L}$ (Länge 1,4 m, Innendurchmesser 0,3 mm) und $200 \mu\text{L}$ (Länge 2,8 m, Innendurchmesser 0,3 mm) verwendet. Der Auslass der Reaktionskapillare war wiederum an eine invertierte Y-Mischeinheit zusammen mit einer Spritzenpumpe (Harvard Apparatus 22, Harvard, MA, USA) und einer Injektionsspritze (Hamilton, Reno, NV, USA) angeschlossen. Über diese wurde vor dem Eintrag in die massenspektrometrische Detektionseinheit eine Trägerlösung (Methanol mit 1 % Essigsäure) mit einer Flussrate von $2,0 \mu\text{L min}^{-1}$ zu dosiert. Der Analyt wurde mit einer Gesamtflussrate von $22 \mu\text{L min}^{-1}$ über eine Elektrospray-Ionisierung in das Q-TOF 2 Massenspektrometer gebracht.

Zur Detektion wurde ein Q-TOF 2 Massenspektrometer von Waters-Micromass (Manchester, UK) gekoppelt mit einer Elektrospray-Ionisierung verwendet. Die Parameter für die massenspektrometrische Detektion wurden mit Malantide und IS2 (PP60 c-src) optimiert. Die Messungen wurden im positiven Ionisierungsmodus bei 353 K Kathodentemperatur und 423 K Desolvatisierungstemperatur, 250 L h^{-1} Desolvatisierungsgasflussrate, 100 L h^{-1} Kegel-Gasflussrate, 16 psi Gaszellendruck, 2500 V Kapillarspannung und 33 V Kegelspannung. Der Massenbereich wurde auf 300-1100 m/z eingestellt und alle 5 Sekunden ein Scan durchgeführt, bei einer Verweilzeit von 0,1 Sekunden. Es wurde im kontinuierlichen TOF-MS Modus gemessen, ohne die Option "MS-Profile". Als Desolvatisierungs-, Kegel- und Kollisions-Gas wurden Stickstoff und Argon verwendet. Beide Gase hatten einen Reinheitsgrad 5.0 und wurden von Praxair bezogen.

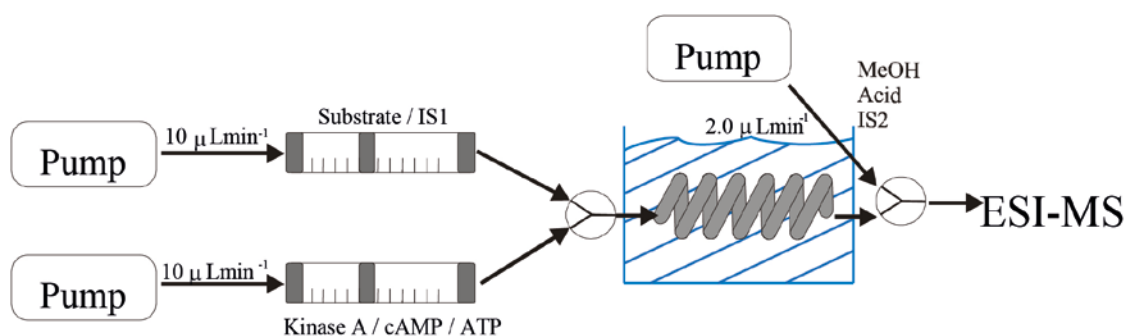


Abbildung 13: On-line Monitoring System von Enzymaktivitäten mittels ESI-MS

Ein SuperLoop™ wurde mit 2 mL wässriger Enzym-Lösung befüllt: Proteinkinase A (224-448 Units), 100 µM MgSO₄, 50 nM cAMP, 50 µM MgATP. Der andere SuperLoop™ wurde mit 2 mL wässrigen Substrat-Lösung befüllt: 2 µM Substrat (Malantide oder Kemptide), 1 µM Ovokin (IS1), 20 mM Ammoniumbicarbonat, 100 µM MgSO₄. Die Injektionsspritze für die Trägerlösung wurde mit MeOH mit 1 % Essigsäure und 10 µM IS2 befüllt. Für die Durchführung von Inhibitionsexperimenten wurde zusätzlich zur Substrat-Lösung 1 µM Inhibitor in den SuperLoop™ zugegeben.

Ein typischer Versuchsansatz für das Monitoring der Enzymaktivität ist nachfolgend gelistet:

- Wässrige Enzymlösung: Proteinkinase A (bovine oder porcine), 100 µM MgSO₄, 50 nM cAMP, 50 µM MgATP
- Wässrige Substratlösung: 2 µM Substrat (Malantide oder Kemptide), 1 µM Ovokin (IS1), 20 mM Ammoniumbicarbonat, 100 µM MgSO₄
- Inhibitorexperimente: Zugabe von 1 µM PKA-Inhibitor (P 0300)
- IS2: MeOH mit 1% Essigsäure, 10 µM IS2
- A) Temperaturabhängigkeit:
 - Temperaturstufen: 278 K, 288 K, 298 K, 308 K, 318 K, 328 K
 - Verweilzeiten bei einer Temperaturstufe 10 Minuten
 - Temperaturerhöhung alle 10 Minuten um 10 K
- B) Abhängigkeit vom Reaktionszeit im Reaktor:
 - 15, 300, 600 Sekunden
 - Die Reaktionszeiten wurden durch unterschiedliche Reaktionskapillar-Volumina (5, 100, 200 µL) bei einer Flussrate 20 µL min⁻¹ erreicht.

2.2.2 HEWL-Assay: Nachweis der Bildung von nichtkovalenten Komplexen und simultane Reaktionsinhibition

Zum Nachweis der Inhibition von Lysozym und der Bildung von Enzym-Inhibitor-Komplexen wurde ein ähnlicher Versuchsaufbau verwendet, wie er in Abbildung 7 dargestellt ist und von de Boer et al. 2004 ^[60] veröffentlicht wurde. Im Unterschied zu dem in Abbildung 7 dargestellten System wurde der Inhibitor über einen Autosampler in das System eingebracht

und nicht von einer Säule eluiert. Als Eintragungssystem wurden die Pumpen der HPLC (1100 Serie, Agilent Technologies, Waldbronn, Deutschland) genutzt. Die HPLC verfügte über eine binäre und eine isokratische Pumpe. Die binäre Pumpe wurde genutzt, um kontinuierlich Lysozym und Substrat GlcNAc₆ in das System zu pumpen. Hierzu wurde der Kopf der Pumpe vom Gradientenmischer entfernt und direkt mit der Reaktionskapillare 2 verbunden, um so das Substrat in das System einzubringen. Das Enzym wurde über den zweiten Fluss durch die HPLC in die Reaktionskapillare 1 geleitet. Die Flussrate für beide betrug 20 $\mu\text{L min}^{-1}$. Enzym und Substrat wurden jeweils in Superloops (Volumen 10 mL, Amersham Biosciences, Pittsburgh, USA) vorgehalten. Die PEEK Reaktionskapillare 1 hatte einen Innendurchmesser von 0,17 mm und eine Länge von 350 mm, die Reaktionskapillare 2 einen Innendurchmesser von 0,25 mm und eine Länge von 830 mm. Die beiden Reaktionskapillaren wurden von der Firma Chromatographiehandel Müller (Fridolfing, Deutschland) bezogen. Das Injektionsvolumen des Inhibitors betrug pro Injektion 10 μL und wurde mittels Spritzenpumpe (11 Plus, Harvard Apparatus, Hugo Sachs Elektronik, Hugstetten, Deutschland) und Autosampler mit einer Injektionsrate von 4 $\mu\text{L min}^{-1}$ und Ejektionsrate von 10 $\mu\text{L min}^{-1}$ in die Reaktionskapillare eingebracht. Die Flussrate der Pumpe betrug 20 $\mu\text{L min}^{-1}$. Der Inhibitor wurde in verschiedenen Konzentrationen (Tabelle 3) injiziert. Ein Injektionsvorgang dauerte 4 Minuten, wobei die eigentliche Inhibitorinjektion selbst ca. 2,5-3,5 Minuten betrug und somit vor und nach der Injektion noch ein Zeitraum ohne Inhibitor in Reaktionskapillare 1 innerhalb des Injektionsvorgangs zu messen ist. Dies diente zur eindeutigen Abgrenzung zwischen den einzelnen Inhibitorinjektionen. In Reaktionskapillare 1 interagieren Enzym und Inhibitor. Die Interaktionszeit von Enzym und Inhibitor beträgt 13 Sekunden, bevor Reaktionskapillare 2 erreicht und das Substrat zugeführt wurde. In Reaktionskapillare 2 konkurrierten Inhibitor und Substrat um das Enzym und es konnten sich entweder Enzym-Inhibitor und/oder Enzym-Substrat Komplexe und somit Produkt(e) bilden. Die Reaktionszeit für Reaktionen mit dem Substrat in Reaktionskapillare 2 betrug 64 Sekunden bevor der Analyt das ESI-MS (Q-TOF Ultima API) zur Detektion erreichte.

Die Detektion erfolgte mittels eines Q-TOF Ultima API Massenspektrometers (Waters-Micromass, Manchester, UK) gekoppelt mit einer Elektrospray-Ionisierung. Die Geräteparameter betragen: Kathodentemperatur 353 K, Desolvatisierungstemperatur 423 K, Desolvatisierungsgasflussrate 300 L h^{-1} , Kegel-Gasflussrate 50 L h^{-1} , 3000 V Kapillarspannung und 40 V Kegel-Spannung, Kollisionsenergie 10. Der Massenbereich wurde auf m/z 50-3000 eingestellt und jede Sekunde ein Scan durchgeführt. Als Desolvatisierungs- und Kegel-Gas wurde Stickstoff (Reinheitsgrad 4.0) und als Kollisions-Gas Argon (Reinheitsgrad 5.0) verwendet. Der Stickstoff wurde mittels Generator (Domnick Hunter GmbH, Krefeld, Deutschland) hergestellt. Zur Massenkalisierung des Geräts wurde Natriumjodid- und Cäsiumjodid-Lösung verwendet.

Ein typischer Versuchsansatz zum Nachweis der Komplexbildung ist nachfolgend gelistet:

- Wässrige Assaylösung (pH 5,2): 90% 10 mM NH₄Ac, 10% Methanol
- Substratlösung (pH 5,2): 50 µM GlcNAc₆ in 10 mM NH₄Ac-Lösung
- Enzymlösung (pH 5,2): 20 µM HEWL, 4 µM Malantide (IS) in 10 mM NH₄Ac-Lösung
- Inhibitorlösung (pH 5,2): 1-40 nM N, N', N'-triacetylchitotriose (GlcNAc)₃ in 10 mM NH₄Ac-Lösung (Tabelle 3)
- Stopp-Reagenz: 90% Acetonitril

Tabelle 3: Mengenangaben der Inhibitorinjektionen und Verhältnis der Menge an Inhibitor zu Substrat und Enzym.

Inhibitorkonzentration (µmol L ⁻¹)	Menge an Inhibitor (nmol)	Verhältnis Inhibitor : Substrat : Enzym
0,1	1	2,5 : 5 : 1
0,2	2	5 : 5 : 1
0,5	5	12,5 : 5 : 1
1,0	10	25 : 5 : 1
2,0	20	50 : 5 : 1
4,0	40	100 : 5 : 1

2.2.3 iAP und AChE/Hausstaub-Assay: Monitoring der Enzymaktivität und Screening nach unbekanntem Inhibitoren in komplexen Proben

Der Versuchsaufbau war ähnlich dem in 2.2.2: Probeneintrag über einen Injektor und/oder HPLC, 2 Reaktionskapillare für den Enzym-Assay und eine ESI-MS Detektionseinheit. Zum einen wurde die Enzymaktivität von iAP untersucht. Hierfür wurde über den Injektor (Rheodyne, California, Probenvolumen 2 µL) und einer quarternären Pumpe (1100 Series, Agilent Technologies, Waldbronn, Deutschland) ADP oder ATP^[121] in steigenden Konzentrationen dem System mit einer Flussrate von 25 µL min⁻¹ zugeführt (Abbildung 14 Mitte: graue gepunktete Linie). Über eine Spritzenpumpe (Model 11 Plus, Harvard Apparatus, Hugo Sachs Elektronik, Hugstetten, Germany) und einer Spritze (2,5 mL, Hamilton-Bonaduz, Switzerland) (Abbildung 14 oben) wurde Enzym (iAP, AChE; siehe Versuchsansatz nachfolgend) eingebracht. Der interne Standard, L-Histidin, wurde mit einer Konzentration von 80 µM zusammen mit dem Enzym in das System gegeben. In beiden Versuchsansätzen wurde eine Flussrate von 25 µL min⁻¹ verwendet. Der Substrateintrag (Abbildung 14 unten) erfolgte über einen SuperLoopTM (Volumen: 10 mL, Amersham Biosciences, Uppsala, Schweden) und über eine isokratische Pumpe (1260 Serie, Agilent Technologies, Waldbronn, Deutschland) mit einer Flussrate von 50 µL min⁻¹. Die Pumpen wurden mittels der ChemStation Software (Version B.04.03, Agilent Technologies,

Waldbronn, Deutschland) überwacht und kontrolliert. Die beiden Reaktionskapillaren (Teflon) haben einen Innendurchmesser von 0,25 mm und waren, wie in der Literatur ^[120] beschrieben, verwoben, um eine ausreichende Verweildauer zur Interaktion von Enzym und Inhibitor in Reaktionskapillare 1 (ca. 1,3 min.) zu gewährleisten. Die Verweildauer zur Interaktion von Enzym, Inhibitor und Substrat in Reaktionskapillare 2 betrug ca. 2,9 min, bevor der Analyt über eine Elektrospray-Ionisierung in das Quadrupol-MS (MSQ Plus, Wissenschaftliche Gerätebau Dr. Ing. Herbert Knauer GmbH, Berlin, Deutschland) gelangte. Das verwendete Schlauchmaterial war aus PEEK.

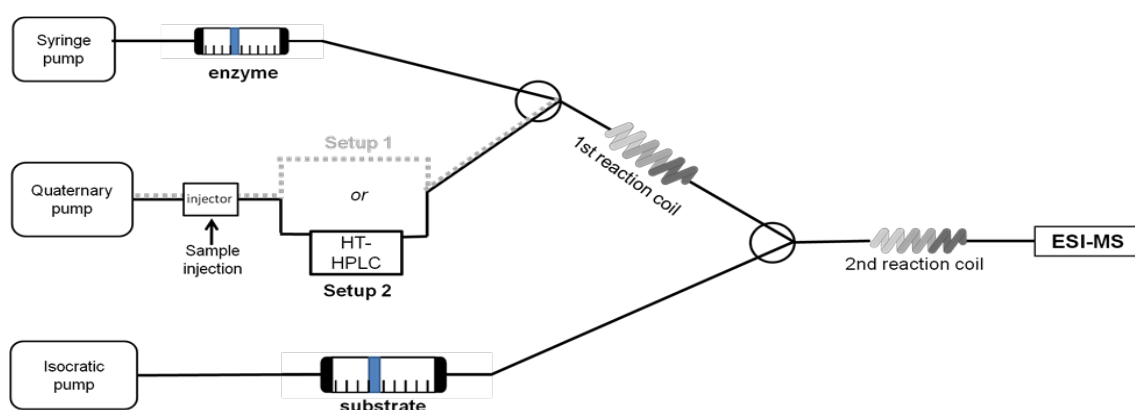


Abbildung 14: Schematische Darstellung des kontinuierlichen Flusssystems mittels (LC-)MS zum Nachweis der Enzymaktivität von iAP und zum Inhibitor-Screening in Hausstaubproben.

Für den Screening-Versuch mit AChE wurde zum Probeneintrag zusätzlich eine HT-HPLC (HT HPLC 2000, Scientific Instruments Manufacturer GmbH, Oberhausen, Deutschland) verwendet (Abbildung 14 Mitte: schwarze Linie). Zur Trennung der Probenbestandteile wurde eine hochtemperaturfähige Zirchrom PBD Säule (100 x 1 mm, 3 µm) (ZirChrom Separations, Inc., Anoka, MN, USA) in den Säulenofen eingesetzt. Die Konzentration des organischen Solvenz wurde in vorhergehenden Studien bereits getestet ^[122]. Zur Verbesserung der chromatographischen Trennung wurde ein Gradient mit einer initialen Temperatur von 30 °C eingestellt, der linear die Temperatur innerhalb von 20 Minuten auf 120 °C erhöhte. Dort wurde die Temperatur für 10 min gehalten, bevor eine Rekonditionierung der Säule automatisch durchgeführt wurde und wieder bei 30 °C begonnen werden konnte.

Die Detektion erfolgte mittels Quadrupol-MS mit einer Elektrospray-Ionisierung im positiven Ionisierungsmodus bei folgenden Parametern: Probentemperatur 300 °C, Kapillarspannung 3,5 kV, Kegel-Spannung 75 V. Der Massenbereich betrug für den iAP-Assay m/z 100-1000 und für den AChE-Assay/Hausstaub-Screening m/z 100-800.

Versuchsansatz zum Nachweis der Enzymaktivität von iAP:

Der Nachweis der Enzymaktivität von iAP erfolgte mit einer initialen Konzentration von Enzym und Substrat (AMP) von 2,4 U/mL und 80 μ M, wobei die AMP Konzentration schrittweise erhöht wurde: 160 μ M, 320 μ M, 1000 μ M, 2000 μ M. Beide wurden in 10 mM Ammoniumacetat (pH 7,4) gelöst und in das System eingebracht. ADP und ATP wurde in weiteren Versuchen in obigen Konzentrationen kontinuierlich in einer 10 mM Ammoniumacetat-Lösung (pH 7,4) über die quarternäre Pumpe zu dosiert.

Versuchsansatz zum Nachweis von Inhibitoren in einer Hausstaubprobe:

Zum Nachweis von Inhibitoren in einer Hausstaubprobe wurden jeweils 0,5 mM AChE und 10 μ M Acetylcholin (Substrat) in 10 mM Ammoniumacetat (pH 7,4) gelöst und zusammen mit dem internen Standard zugegeben. Zur Auftrennung der Hausstaub-Probe wurde diese in einer Lösung aus 90 % 10 mM Ammoniumacetat (pH 7,4) und 10 % Methanol gelöst und über die HPLC-Säule getrennt.

2.3 Datenanalyse

Alle Messungen im MS erfolgten im positiven Ionisierungsmodus. Abhängig vom Enzym Assay und dem Massenspektrometer wurden verschiedene Adduktformationen und Protonierungsgrade detektiert und gegebenenfalls aufsummiert und als extrahiertes Ionenchromatogramm (EIC) gezeigt. In Tabelle 4 werden alle m/z Werte der verwendeten Produkte, Inhibitoren und Substrate gezeigt.

Tabelle 4: Detektierte m/z Verhältnisse für alle Substrate, Produkte und Inhibitoren der untersuchten Enzym Assays

Enzym	Substrat, Produkt, Inhibitor	Mono- isotopische Masse [Da]	m/z Protonierungsgrad				
			[M+H] ⁺	[M+Na] ⁺	[M+2H] ²⁺	[M+3H] ³⁺	[M] ⁺
AChE	ACh	146,1176					146,1
	Cholin	104,1070					104,1
	Galanthamin	287,3535	288,3				
HEWL	(GlcNAc) ₆	1237,1704	1237,5	1259,5	619,2		
	(GlcNAc) ₄	830,7854	831,3	853,3			
	(GlcNAc) ₃	627,5928	628,2	650,2			
iAP	AMP	347,0631	348,1	370,1			
	ADO	267,2413	268,1	290,1			
	ADP	427,0294	428,0	450,0			
PKA	Malantide	1633,8898	1634,9	1662,9		545,3	
	Kemptide	771,9084	772,9	800,9	386,9		
	p-Mal	1712,8975			857,4	571,9	
	p-Kem	851,4379			426,7		
	P0300	2222,3767				741,3	

Die Rohdaten der massenspektrometrischen Detektion wurden mittels der herstellerspezifischen Softwares gewonnen und anschließend in das textbasierte Achroma Software (Hochschule Weihenstephan-Triesdorf, Freising, Deutschland) ASCII Datenformat konvertiert. Alle relevanten Daten der Experimente wurden mit der Achroma Software, die Teil dieser Arbeit ist (siehe auch Kapitel 3.1), ausgewertet. Die Funktionsweise der einzelnen Module von Achroma, wie z.B. die Signalerkennung, die Signalfächenberechnung und die Glättung der Chromatogramme mittels eines gleitenden Durchschnitts sind in den Kapiteln 2.4 und 3.1 ausführlich erläutert. Die Software ist Freeware und kann frei verwendet werden: <http://openmasp.hswt.de/pages/project/achroma.php>.

Wie bereits in der Einleitung dieser Arbeit erwähnt, bieten viele Hersteller die Möglichkeit an, die Rohdaten in andere Formate zu konvertieren. Die Daten für diese Arbeit wurden mittels der DataBridge (Teil der MassLynx-Software, Waters, Milford, Massachusetts, USA), dem File Converter (Teil der XCalibur-Software, Thermo Scientific, Waltham, USA) oder über einen weiteren Schritt des Microsoft Excel Exports in ein für Achroma lesbares Format transferiert, z.B. CSV. Die konvertierten Daten aus der DataBridge von Waters können direkt eingelesen werden, wenn der Modus ASCII und "No Header" ausgewählt wird. Für alle weiteren Konvertierungen wurde ein weiteres Achroma-Werkzeug (Achroma Konverter) entwickelt. Dieser kann Textdateien einlesen, die mit dem Fileconverter von Thermo Scientific oder Microsoft Excel umgewandelt wurden bzw. im CSV-Format vorliegen. Die Exportdaten anderer Hersteller können so auch verfügbar gemacht werden. Der Achroma Konverter wandelt die bereits textbasierten Daten in Achroma lesbare Daten um.

Die Achroma Module *Chromatogram comparison*, *Signal recognition* und *Spectra comparison* wurden im iAP und AChE/Hausstaub Assay genutzt, um die Enzymaktivität zu zeigen und um schnelle Ergebnisse im Inhibitoren-Screening zu bekommen (siehe auch Kapitel 3.2).

Für den PKA-Assay wurden zusätzliche Auswertungen mit Microsoft Excel 2010 durchgeführt, da für die mit Achroma gewonnen Flächenberechnungen der Signale zusätzlich eine Fehlerrechnung durchgeführt wurde (siehe auch Kapitel 3.3). Ebenso wurden über die Auswertung mit Achroma hinaus für die Konzentrationen zum Zeitpunkt der Komplexbildung im HEWL-Assay weiterführende Auswertungen mit Microsoft Excel 2010 durchgeführt. Unter der Annahme, dass ein optimales Signal der Gauß-Verteilung folgt, wurden die Zeitpunkte eines Signaleinbruchs vor und nach dem Erreichen des Signalmaximums ins Verhältnis gesetzt und über den theoretischen Wert des Maximums der Gaußkurve errechnet. Anschließend wurde auch hier eine Abschätzung des Fehlers der errechneten Werte durchgeführt und eine Inhibitorkonzentration für den Zeitpunkt des Übergangs eines Enzym-Inhibitor Komplex zum nächsthöheren Enzym-Inhibitor Komplex errechnet (siehe auch Kapitel 3.4).

2.4 Softwareimplementierung

Die Achroma-Software ist in der Programmiersprache C++ entwickelt. Zur Programmierung wurde das Microsoft Visual Studio 2003 Professional (Visual C++) genutzt. Mit Hilfe des Microsoft Visual Studios können bereits Projekte angelegt werden, die eine Architektur zur Verfügung stellen, um die Programmierung einer graphischen Benutzeroberfläche zu ermöglichen. Diese Architektur nennt sich Document-View Architektur und ist eine Variation der Model-View-Controller Architektur, die eine modulare Softwareprogrammierung ermöglicht. Bei der Model-View-Controller Architektur wird die Datenabstraktionsschicht von der Benutzerschnittstelle getrennt (Abbildung 15). Zur Umsetzung der graphischen Benutzeroberfläche wurde die *Microsoft Foundation Class* (MFC) Bibliothek genutzt, die bereits eine Vielzahl von Optionen und Klassen für die graphische Programmierung zur Verfügung stellt. Bei der Document-View Architektur ist das View-Objekt des MFC-Projekts dafür zuständig, dass die Daten vom Document-Objekt angezeigt und das Benutzereingaben an das Document-Objekt übertragen werden. Diese werden dann auf der Datenebene verarbeitet.

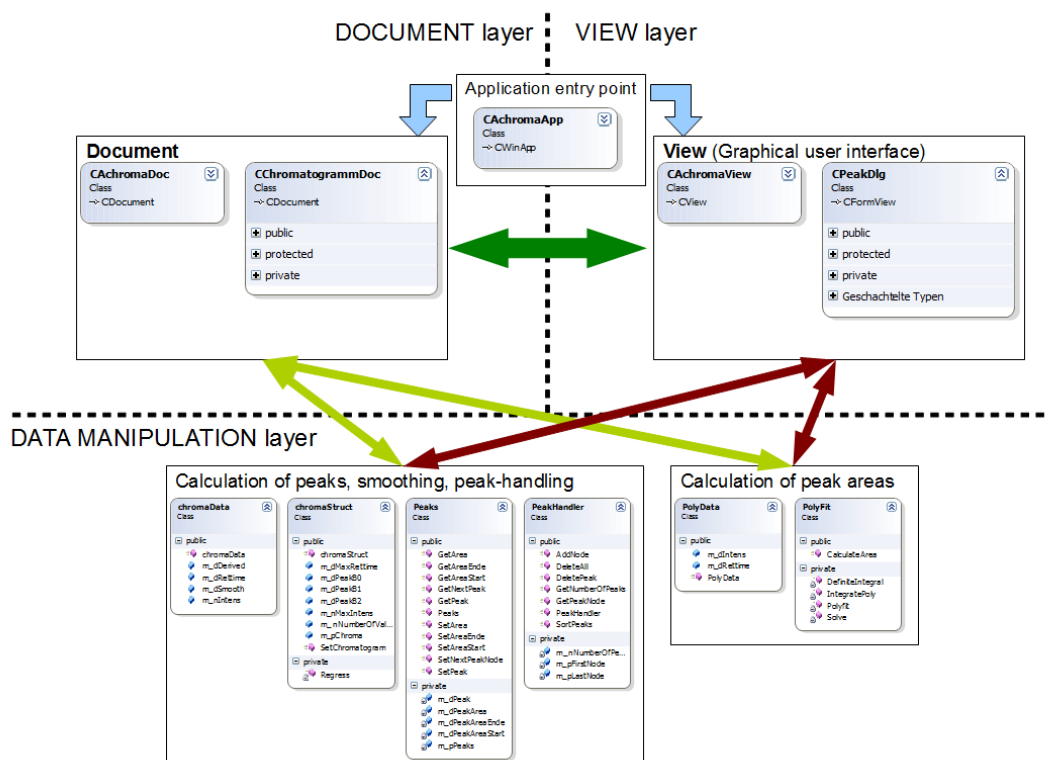


Abbildung 15: Schematische Darstellung der Document-View Architektur von Achroma.

Als Datenformat wurde ein von jedem Texteditor lesbares textbasiertes ASCII-Format gewählt. Um aber ebenfalls große Datenmengen mit Dateien von mehreren Gigabyte schnell einlesen zu können, wird bei der ersten Verwendung eines Datensatzes in Achroma zusätzlich eine Index-Datei erstellt. Diese erfasst zu jeder Retentionszeit (= Scan) die

Zeilennummern zu den m/z-Verhältnissen in definierten Abständen, z.B. alle 50 Zeilen wird ein Wert in die Index-Datei gespeichert (siehe Kapitel 3.1, Abbildung 4). So wird ein schnelles Durchsuchen möglich.

Ein kompletter Datensatz, also eine ASCII-Datei, ist zeilenweise in der nachfolgenden Form aufgebaut:

Scan	1
Retention Time	25 (Minuten)
m/z-Wert 1	Intensität 1
...	
m/z-Wert x	Intensität x
.	
m/z-Wert y	Intensität y
Scan	2
Retention Time	26 (Minuten)
m/zWert 1	Intensität 1
...	
m/z-Wert x	Intensität x
....	
m/z-Wert y	Intensität y
...	
Scan	x
...	

Aus obiger Darstellung wird ersichtlich, dass Spektren zu einer gegebenen Retentionszeit (= 1 Scan) einfach darzustellen sind, da die Datendatei nur nach dem entsprechenden Retentionszeit-Wert durchsucht werden muss und dann nachfolgend alle Werte bis zum nächsten Scan (Retentionszeit) in Achroma übernommen und dort verarbeitet werden können. Über die Index-Datei kann hier auf jeden Scan sofort zugegriffen werden. Die Darstellung eines Chromatogramms zu einem gegebenen m/z-Wert gestaltetete sich deutlich komplizierter. In der Index-Datei werden die Zeilennummern zu jedem Scan wie oben beschrieben, sowie in Abständen von 50 m/z-Werten innerhalb eines Scans jeweils die m/z-Wert-Zeilenummern-Paare gespeichert. Der Suchalgorithmus von Achroma wurde entsprechend implementiert, dass z.B. in der Index-Datei in Scan 1 die entsprechende Zeilennummer eines m/z-Wertes in der Nähe des durch den Benutzer eingestellten m/z-Wertes gesucht wird, da nicht alle m/z-Wert-Zeilenummern-Paare in der Index-Datei hinterlegt sind. Die Software springt dann in die Datendatei an diese Stelle und durchsucht die Datei zeilenweise, bis der eingestellte m/z-Wert erreicht und die Intensität automatisch in Achroma übernommen wird. Anschließend sucht Achroma die Zeilennummer von Scan 2 und wieder die Zeilennummer eines in der Nähe befindlichen m/z-Wertes und springt in die Datendatei, um die zweite Intensität zu finden. So wird verfahren bis alle Scans, sprich Retentionszeiten, durchlaufen sind.

Die graphische Präsentation der Daten erfolgte über die MFC-Bibliothek. Jedes Modul präsentiert seine Daten in einem zweigeteilten Fenster (*Split Pane*) im unteren Bereich. Der obere Bereich steht dem Benutzer zur Verfügung, um Einstellungen zu machen (Dialogfenster). Die Applikation selbst startet in einem Hauptfenster (*Main Frame*) indem alle Module mit Hilfe sogenannter *Child Frames* eingebettet wurden. Alle fünf Achroma Module wurden als *Child Frames* mit *Split Panes* implementiert.

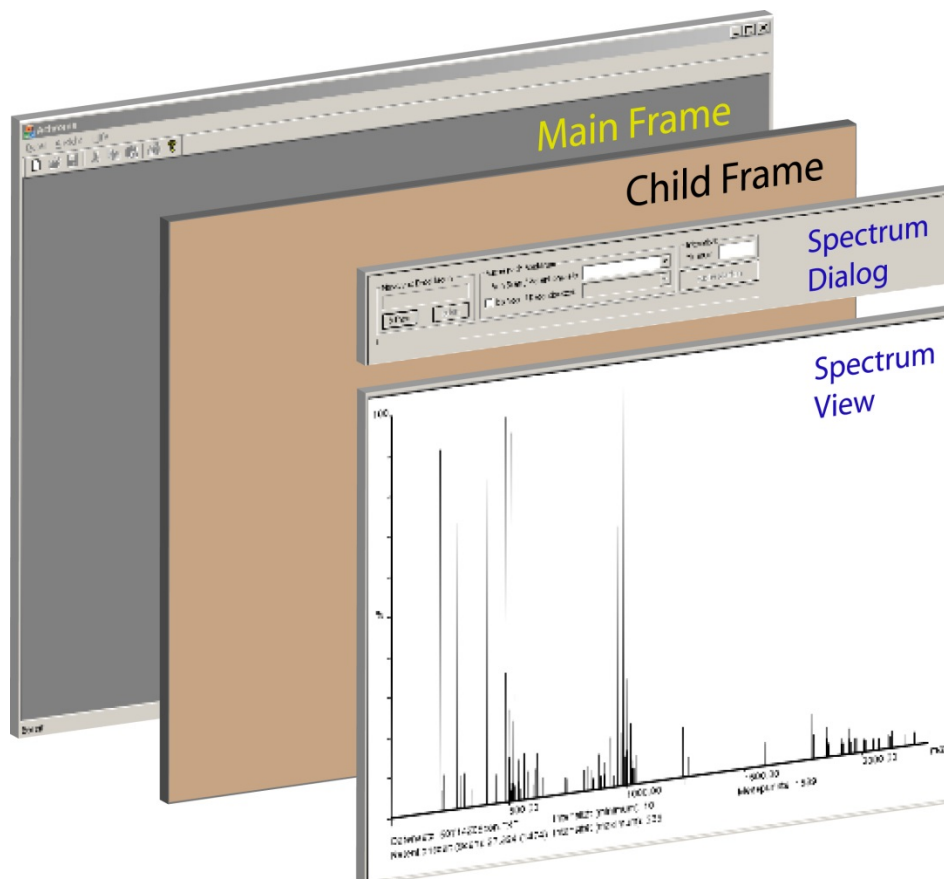


Abbildung 16: Aufbau der graphischen Oberfläche der Achroma Software.

Die Achroma Software selbst bietet fünf Module zur Auswertung von (LC-)MS Daten. Zwei Module wurden standardmäßig implementiert: "Chromatogram view" und "Spectra view". Drei weitere Module wurden speziell für die Anforderungen der Datenauswertung der in dieser Arbeit erläuterten Enzym Assays entwickelt, da manche benötigten Optionen in bereits vorhandenen Softwares noch nicht verfügbar waren. Die speziell entwickelten Module sind:

Chromatogram comparison

Dieses Modul vergleicht zwei Chromatogramme zu gegebenen m/z-Verhältnissen, berechnet daraus ein Quotientenchromatogramm und zeigt dieses an. Bei Bedarf kann auch noch eine Toleranzgrenze in % gewählt werden, dass die Intensitäts-Werte des errechneten Quotientenchromatogramms nicht über- bzw. unterschreiten dürfen.

Ein Chromatogramm wird als eindimensionales (1D) Array erzeugt, das die Länge der *Anzahl der Scans* der Messung hat. Jeder Eintrag im 1D-Array enthält ein *chromaData*-Objekt. Ein *chromaData*-Objekt besteht aus:

- Wert der Retentionszeit des Scans,
- Wert der Intensität des dazugehörigen m/z-Wertes,
- Wert der geglätteten Intensität,
- Wert der abgeleiteten Intensität (wird zum Berechnen von Anfang, Ende und Minimum/Maximum eines Signals benötigt).

Somit kann zu jedem m/z-Wert ein Chromatogramm erzeugt und über das *Chromatogram view*-Modul angezeigt werden.

Implementiert wurde die Berechnung des Quotientenchromatogramms mittels zweier *chromaData*-Objekt Arrays. Die Werte des Quotientenchromatogramms selbst werden nicht in ein drittes 1D-Array aus *chromaData*-Objekten gespeichert, sondern nur temporär in ein zweidimensionales (2D) Array aus Zahlenwerten abgelegt und zusammen mit den beiden anderen am Bildschirm dargestellt. Weitere Informationen zur Implementierung können Kapitel 3.1 entnommen werden.

Signal recognition

Dieses Modul bietet die Funktionen zur Berechnung der Signale und ihrer Flächen. Es können Optionen, wie Glättung, positives/negatives Signal oder kontinuierliche Messung eingestellt werden. Ebenso bietet dieses Modul neben der automatisierten Erkennung von Signalen auch die Möglichkeit, manuell (auch negative) Signale zu bearbeiten oder zu finden, falls die Automatik nicht die gewünschten Ergebnisse liefert.

1. Glättung des Graphen nach Savitzky und Golay^[123]:

Die Glättung des Intensitäts-Graphen über die Zeit (Chromatogramm) erfolgt mittels eines gewichteten gleitenden Durchschnitts. Es werden jeweils die Intensitäten von 5 Scans gewichtet und gemittelt (siehe Formel 1 in Kapitel 3.1). Die Glättung startet beim dritten Eintrag im 1D-Array des Chromatogramms und nimmt jeweils die beiden Intensitäts-Einträge davor und danach mit in die Berechnung des gewichteten Mittelwerts auf. So wird das ganze Array durchlaufen bis das zweitletzte *chromaData*-Objekt erreicht wird. Es wird die Glättungsprozedur insgesamt 10-mal durchlaufen und der jeweilige Wert in das *chromaData*-Objekt geschrieben.

2. Berechnung der Ableitung und finden der Maxima/Minima, sowie Start- und Endpunkte:

Nachdem die Glättung abgeschlossen ist, berechnet Achroma die erste Ableitung des dargestellten Chromatogramms. Das Finden eines Signals wurde in Achroma mittels der Ableitung des Graphen implementiert. Es kann von einem Start- oder Endpunkt eines Signal ausgegangen werden, wenn die Steigung der Ableitung nahe 0 ist und vorher oder nachher die abgeleiteten Intensitätswerte ebenfalls nahe 0 sind. Nähert sich die Steigung der Ableitung von +/- 1 der 0 und umgekehrt, dann kann an der Nullstelle (= Wendepunkt der Ableitung) ein Maximum oder Minimum angenommen werden. Achroma durchläuft das 1D-Array des ausgewählten Chromatogramms und prüft die abgeleiteten Intensitäts-Werte in den einzelnen *chromaData*-Objekten. Die gefundenen Start-, Endpunkte und Extrema werden über die Benutzeroberfläche von Achroma graphisch dargestellt (siehe Kapitel 3.1, Abbildung 6).

3. Berechnung Signalflächen:

Die Berechnung der Signalflächen der Extrema hängt davon ab, ob es sich um ein Minimum oder Maximum handelt.

A) Signalfläche eines Minimums: Berechnung der Rechtecksfläche zwischen Start- und Endzeitpunkt mit der Höhe der Intensität zum Startzeitpunkt; Subtraktion des Integrals der Intensitäten zwischen Start- und Endzeitpunkt, sowie die Berichtigung um eine Dreiecksfläche, wenn die Basislinie ab- oder zunehmen sollte.

B) Signalfläche eines Maximums: Berechnungen des Integrals der Intensitäten zwischen Start- und Endzeitpunkt; Subtraktion der Rechtecksfläche zwischen Start- und Endzeitpunkt mit der Höhe der Intensität zum Startzeitpunkt, sowie die Berichtigung um eine Dreiecksfläche, wenn die Basislinie ab- oder zunehmen sollte.

Die Berechnung für A) und B) kann in einer allgemeinen Formel (siehe auch Kapitel 3.1, Formel 2 und Abbildung 7) dargestellt werden, indem mit dem Betrag der Flächen gerechnet wird, da sonst für Minimas (A), negative Flächen berechnet würden.

Spectra comparison

Dieses Modul wird aus dem *Signal Recognition* Modul aufgerufen. Es kann im *Signal recognition* Modul ein Signalmaximum oder -minimum markiert und ein Zeitpunkt bis maximal 5 Minuten vorher auf der Basislinie des Chromatogramms gewählt werden. Anschließend ruft man das *Spectra Comparison* Modul auf und bekommt ein Differenzspektrum angezeigt, bei dem die Intensitäten der m/z Werte des Spektrums zum

Zeitpunkt bis 5 Minuten vor dem Signal von den Intensitäten des Spektrums zum Zeitpunkt des gewählten Maximums/Minimums subtrahiert wurden.

Die gesamte Implementierung, die Logik und Erstellung der Indexierung, sowie die Funktionsweise der einzelnen Module ist ausführlich im nachfolgenden Kapitel (3.1) beschrieben und dargestellt.

3 Ergebnisse

3.1 Achroma: a software strategy for analyzing (a-)typical mass spectrometric data

Michael Krappmann, Thomas Letzel
Analytical Methods, 2012, 4, 1060-1071,
DOI: 10.1039/c2ay05622d
Reproduced by permission of The Royal Society of Chemistry
<http://pubs.rsc.org/en/content/articlelanding/2012/ay/c2ay05622d>

Neue kontinuierliche (LC)-MS Enzym-Assays führen zu (un-)typischen MS-Daten. Wie man diese Daten automatisiert mit der neuprogrammierten Software Achroma auswerten kann und die zugrundeliegende Strategie wird im nachfolgenden Artikel erklärt.

Am Beispiel eines Versuchsaufbaus von De Boer et al. (2004) ^[60], mit dem Protease-Inhibitoren in einem Pflanzenextrakt gefunden wurden, wird die Strategie und die Funktionsweise der neu entwickelten Achroma Software aus der Sicht des Bioinformatikers im Kontext zur Sicht des (Bio-)Analytikers dargestellt.

Im Versuch wird zu einer kontinuierlich fließenden Enzymlyösung ein (über HPLC) getrenntes Pflanzenextrakt zugesetzt und dieses anschließend mit Substrat vermischt. Enzym und Substrat bilden in der Reaktionskapillare kontinuierlich Produkt, wobei Beide, Substrat und Produkt, im Massenspektrometer detektiert werden können. Befindet sich nun in einer eluierenden Fraktion des Extrakts ein Inhibitor, so wird die Produktbildung erniedrigt und es entsteht für den Zeitraum, ein negatives Produktsignal im MS. Eine - zur Auswertung negativer Signale - notwendige Software stand zum Zeitpunkt der Veröffentlichung jedoch nicht zur Verfügung. Mit Achroma konnte diese Lücke erstmals geschlossen werden. Als erste Software konnte Achroma solche Signale nicht nur erfassen sondern auch die erhaltenen Flächen berechnen und andere Auswertestrategien ermöglichen.

Achroma ist eine modulare Software, die auf einer Document-View Architektur basiert. Jedes einzelne Modul von Achroma selbst funktioniert ebenfalls nach dieser Architektur. Das bedeutet, dass jederzeit neue Module integriert werden können, da diese unabhängig voneinander als eigene kleine Anwendungen funktionieren. In Achroma selbst sind bisher fünf Module umgesetzt: zwei Basis-Module zur Darstellung von Chromatogrammen und Spektren, sowie drei spezialisierte Module, die auf die notwendigen Auswertungen kontinuierlicher Messungen von (LC)-MS Enzym Assays ausgerichtet sind. Die drei spezialisierten Module sind wie folgt:

- a. ein Modul zum Vergleich zweier Chromatogramme. Das aus dem Vergleich resultierende Quotientenchromatogramm zeigt dabei zum Beispiel die Systemstabilität auf, sofern im Versuchsansatz interne Standards eingesetzt wurden (siehe auch Kapitel 3.2 und 3.3). Ebenso kann dabei die Stöchiometrieänderung von nichtkovalenten Komplexen visualisiert werden (siehe auch Kapitel 3.4).
- b. ein Modul zur automatischen Berechnung der Signalmaxima/-minima und der Signalfächenberechnung sowie einer manuellen Korrekturmöglichkeit. Einzigartig ist hier die Berechnung von Signalfächen aus kontinuierlichen Versuchen, die einen kontinuierlichen Anstieg/Abfall der Intensität über die Zeit zeigen. Auch diese Flächen können mittels Achroma berechnet werden. Letztlich war Achroma - wie oben erwähnt - die erste Software, die auch negative Signale detektieren und deren Flächen berechnen konnte.
- c. ein Modul zum Vergleich zweier Spektren. Mit Hilfe eines Differenzspektrums ist bereits graphisch und mathematisch erkennbar, ob zum Zeitpunkt des Maximums bzw. Minimums neue m/z-Werte detektiert wurden. Dieses Modul wurde bereits erfolgreich zur schnellen Erkennung neuer Signale, z. B. eines unbekanntes Inhibitors in einer Probe eingesetzt (siehe auch Kapitel 3.2).

Achroma nutzt dabei textbasierte Datenformate, die von jedem Texteditor oder Tabellenkalkulationsprogramm gelesen werden können. Somit wurde hier ein Schritt in Richtung Austauschbarkeit von Daten gegangen und weg von den geschlossenen Datenformaten vieler Hersteller.

Es wurde von Anfang an mit Datenmengen in einer Größenordnung von bis zu acht Gigabyte gearbeitet. Aufgrund einer neu implementierten Indexierung konnten so selbst große Textdateien in akzeptablen Zeiten bearbeitet werden, wie folgendes Beispiel zeigt:

Datensatz 3,8 Gigabyte: Erstellung Indexierung 115 Sekunden, Erstellung Chromatogramm < 20 Sekunden auf einem 32 Bit Microsoft Windows 7 Betriebssystem mit einer Intel Core 2 Duo CPU T7500, 2.20 GHz und 2 Gigabyte RAM. Die Darstellung eines Spektrums dauerte unter 1 Sekunde.

Die Konzeptplanung, Erstellung und Implementierung der Archroma Software, sowie das Verfassen des Manuskripts wurde federführend von mir realisiert. Herr Priv.-Doz. Dr. Letzel hat die gesamte Arbeit betreut, Informationen und Daten zum experimentellen Set-up beigetragen und das Manuskript korrigiert und schließlich eingereicht.

Achroma: a software strategy for analysing (a-)typical mass-spectrometric dataMichael Krappmann^{*a} and Thomas Letzel^{*b}

Received 28th September 2011, Accepted 26th January 2012

DOI: 10.1039/c2ay05622d

Today, (bio-)analytical researchers use various software tools for improving data analysis and the evaluation of their experimental results. Bioinformaticians typically program these software tools; however, analysts and bioinformaticians have distinct views on these data. Achroma is a software tool for the analysis of spectrometric data; it is our hope that explaining the software strategy and the working modules behind Achroma may give analysts a better understanding of how (bio-)informaticians work out software solutions, thus facilitating the interaction between these two expert groups. Achroma is based on a model-view-controller (MVC) software architecture. Furthermore, Achroma has a modular structure and each module has its own MVC architecture. Typically, each module delivers just one specific function to the user. Finally, every module is embedded within Achroma as a small “stand alone” software application. Analytically, Achroma software is a tool to handle typical and untypical mass spectrometric data. Achroma was originally programmed to circumvent problems with mass spectrometric vendor software in the analysis of data from new experimental strategies. Specifically, existing software enables the data analysis from continuous-flow mixing systems monitoring enzyme–inhibitor reactions only manually and indirectly. This data-handling bottleneck is resolved in Achroma. The principles behind this implementation are described in detail for two modules—the ‘chromatogram comparison’ and ‘signal recognition’ modules. Further specific examples of data analysis are presented, such as signal recognition, chromatogram smoothing and signal area calculation. We regard such understanding between the analysts’ and programmers’ worlds as essential for future improvements in analytical software.

1 Introduction

New generation mass spectrometry (MS) in combination with liquid chromatography (LC) has become an analytical key technology for food, process monitoring, environmental, medical, pharmaceutical and biochemical research.

Particularly the rapid development of the bioanalytical methods in proteomics and metabolomics was due to high-quality identification and quantification of proteins and metabolites using LC-MS.¹ In recent years, mass spectrometry has become an excellent analytical technology in the field of ‘functional proteomics’, as for the monitoring of enzymatic activity, its inhibition and the formation of non-covalent complexes by proteins.² Mostly, high sophisticated information can be observed from chromatographic and mass spectrometric data in studying proteins or metabolites. However, samples result in

thousands of mass spectra with and without fragmentation of relevant biomolecules in the mass spectrometer, thus powerful software is needed. Chromatographic separations prior to mass spectrometric detection simplify the signals, but increase the amount of analytical data even more. Proteins (as well as peptides therefrom) and metabolites are eluted from the chromatographic system with a continuous transfer to the mass spectrometric detection system. During the analysis period, mass spectra are recorded at regular intervals, *e.g.* every second. This requires the option of an automated data handling process, the interpretation and evaluation for the large amount of newly generated data. Additionally, the rapid improvement of the analytical instruments (*e.g.* higher resolution of the mass spectrometer) enables new and even more complex experimental setups with further increasing amount of data.

Recent years have clearly shown that data handling and computational analysis of these data is an essential bottleneck. New jobs have been created and established to overcome this weakness. Today, bioinformatics and other interdisciplinary informatics deal with the need of specific software tools in the special data-producing area and program various solutions, *e.g.* for the bioanalytical chemist or the biologist. Thus,

^aInstitut für Forschung und Weiterbildung, Weihenstephan-Triesdorf University of Applied Sciences, Am Hofgarten 4, 85354 Freising, Weihenstephan, Germany. E-mail: michael.krappmann@hswt.de

^bCompetence Pool Weihenstephan, associated with Technische Universität München, Weihenstephaner Steig 23, 85354 Freising, Weihenstephan, Germany. E-mail: t.letzel@wzw.tum.de

various commercial and non-commercial software solutions have been developed for computational data from analysis techniques like the recognition and processing of chromatographic runs, mass spectra, mass spectrometric fragmentation as well as the characterisation of proteins, peptides and metabolites. Plenty of commercial software packages and open source software are available. On the other hand, commercial software packages are often part of vendor packages, *e.g.* XCalibur (Thermo Fisher), MassLynx (Waters), MassHunter (Agilent),^{3–8} mostly performing specific tasks very well. However, the data formerly were and sometimes still are stored in vendor specific data formats and the import/export into/out of open data formats is not always supported. Further disadvantages are the secret algorithms of analytical formats and therefore parameters are restricted for access, which are needed to handle the data.

Moreover, there are open source and academic software projects for mass spectrometric quantification and characterisation of proteins, peptides and metabolites. These software packages use generally open data formats and give higher control over the steps of analysis, *e.g.* OpenMS,⁹ The OpenMS Proteomics Pipeline (TOPP),¹⁰ Decon2LS,¹¹ MapQuant,¹² SpecArray,¹³ MASPECTRAS,¹⁴ msInspect,¹⁵ Superhirn¹⁶ or ATAQS.¹⁷ These tools typically support one specific step of the analysis, *e.g.* quantification, protein/peptide/metabolite identification, map alignment, or a combination of steps in a so-called step-pipeline. The applied algorithms are well described and published. Further, the software packages and/or the source code of the packages are freely available. However, all of these software packages are specialized on the main applications of proteomics and metabolomics: identification and quantification of compounds.

Finally, recently published open source software projects are more flexible in use and data handling.¹⁸ There are two solutions: mMass, cross-platform software for the analysis of MS data,^{19,20} and OpenChrom, an additional cross-platform software project for chromatographic data.²¹ mMass is a software analysing mass spectra and identifying molecules with the help of a protein database like MASCOT. OpenChrom, similarly to the presented Achroma software, can observe chromatographic peaks and calculate peak areas. In this fact there is no difference between the Achroma software tool and other software tools. On the other hand, Achroma has additional modules for the comparison of chromatograms and mass spectra, which are necessary to analyse data, obtained by typical and especially untypical experimental setups.

Concluding, there are many software tools and platforms for single needs and solutions. However, for the flexible application of analytical results obtained from chromatographic and mass spectrometric data combining and flexible software with new concepts is needed.

For instance, recently developed strategies in ‘functional proteomics’ use continuous liquid flow systems (sometimes including LC) with mass spectrometric detection. Such an experimental setup (shown in Fig. 1) can be used for the monitoring of enzyme activity, enzyme inhibition,^{22–25} and complex formation of proteins.²⁶ In several applications the chromatographic and the mass spectrometric data have to be handled simultaneously and in combination.

The software Achroma was designed as expandable software for new analytical approaches and is currently freely available *via* e-mail by the authors.

2 Analytical approach, software design and implementation

2.1 Analytical approach

Recently published,²⁵ a setup was applied with ‘untypical chromatographic results’ for the description of the software strategy and is shown in Fig. 1. The original setup was developed obtaining a continuous-flow mixing system for the screening of enzymatic inhibitors from plant extracts *via* monitoring substrate and product ion traces in MS (see Fig. 2 for mass spectrometric traces corresponding to inhibitor masses).

The first step thereby is a chromatographic separation ((RP)-HPLC, nano(RP)-HPLC or HT-(RP)-HPLC) leading to a differentiation of single molecules from a complex mixture (parts 1 and 2 in Fig. 1). The compounds eluting from the HPLC column (see part 3) are used in the forthcoming continuous-mixing flow setup. Furthermore, the elution flow out of the separation column (see part 3) containing the separated molecules is subsequently being mixed with the enzyme-containing flow (see part 4).

A separated compound that elutes and has enzyme-regulating properties then interacts with the enzyme (in coil 5) and influences the enzymatic activity, *i.e.* the enzyme is activated or inactivated. This stream is further being mixed with substrate solution in a second coil (coil 7). Subsequently, an up- or down-regulated substrate conversion is observed in comparison to the regular situation. As long as there is no regulating molecule interacting with the enzyme, a steady amount of the substrate will be converted and the chromatographic trace for the substrate as well as the product ion (as an extracted ion chromatogram, EIC) is a straight line. Inhibiting compounds lead to a negative tip and activating compounds to a positive tip in the resulting product trace. Several compounds—substrate, products and other ionic compounds—can directly be monitored and presented as EICs of the relevant *m/z* ranges (*e.g.* products *via* traces B and C in Fig. 2 and internal standards SMC1 and SMC2 *via* traces D and E).

In this setup the product formation *versus* its change marks directly the combination of a time point of an HPLC separation with biological activity, and thus leads directly to the parameter molecular mass and hydrophobicity of the regulating compound. Several similar enzymatic assays are presented in our days.²⁷

2.2 Software design

The Achroma software became necessary as there was no software available to handle various analytical data from different system sources and data like negative mass spectrometric signals as shown in Fig. 2. Consequently as described above, a modular software system was created that can be expanded at any time with new implemented modules independent of the original analytical apparatus and vendor system. Thus Achroma was programmed under our supervision by several bioinformatic students in the period of several practical courses. This successful

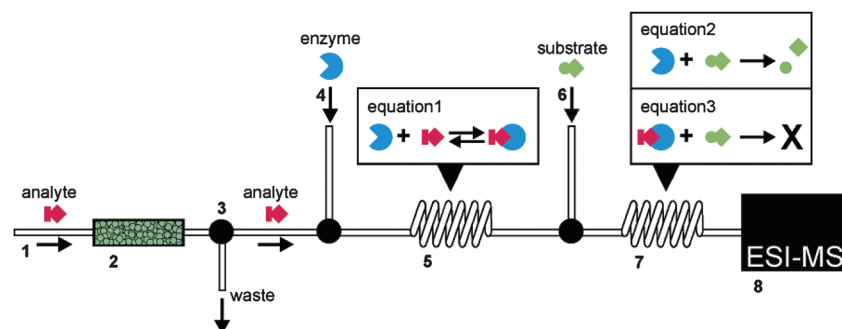


Fig. 1 Scheme of the on-line continuous flow mixing system: (1) sample introduction *via* the assay solution + 5% methanol; (2) HPLC column; (3) 1 : 3 flow splitter, 50 $\mu\text{L min}^{-1}$ bioassay/150 $\mu\text{L min}^{-1}$ waste; (4) enzyme solution; (5) reaction coil A; (6) substrate solution; (7) reaction coil B; and (8) mass spectrometer. Enzyme continuously converts the substrate into products (eqn (2)) if no inhibitor is eluting from the column. Bioactive compounds present in the eluate bind to the enzyme (eqn (1)), resulting in a decrease of the product turnover (eqn (3)). Reprinted with permission from Arjen R. de Boer *et al.*: On-line coupling of high-performance liquid chromatography to a continuous-flow enzyme assay based on electrospray ionization mass spectrometry, *Anal. Chem.*, 2004, **76**, 3155–3161. Copyright 2004 American Chemical Society.²⁵

software version shows that it can be realized to access open minded software by individual persons and during student times.

In the forthcoming sections typical procedures are shown as how to program software for the application of data from analytical instruments (including data import, data procession, and result presentation).

Several listed requirements seemed to be essential for the sufficient realization of universal software:

- Object-oriented.
- Encapsulated modules, *i.e.* every module has its own special task.
 - Smart interaction in between the encapsulated modules.
 - Easy to handle graphical user interface (GUI).
 - Expandable at any time with new modules and new features.
 - Layer-model: data layer and presentation layer of the software were separated from each other.

A model-view-controller (MVC) pattern is a software architecture, which is well known and frequently implemented (*e.g.* Java Swing, Net Framework, Microsoft Foundation Class Library, Qt, Adobe Flex, GTK+). It is a standard model for GUI programming. Thus the Achroma software was started as MVC software architecture. This architecture isolates the application logic from the user interface. The latter manages data input and data presentation.^{28,29}

The Achroma main frame which encapsulates several modules works like a MVC pattern. Furthermore, each module is implemented as a MVC pattern and therefore independent of other modules. Only the data and data manipulation layer are used by all modules (as can be seen in Fig. 3).

2.3 Implementation

The implementation of Achroma software contains several individual modules. Every module can be divided into two subunits: one unit represents the (graphical) presentation part for chromatograms (as well as mass spectra) and the other unit processes chromatographic (as well as mass spectrometric) data. The implementation of every single unit was programmed in one or more C++ classes. These classes describe the behaviour of each part of the software complex in text form as a so-called C++ programming code. C++ classes in general consist of two files. A

header file (*e.g.* Peak.h), which has a description of the used variable names and method names. In the second file, the code file (*e.g.* Peak.cpp) contains the textual description of the used methods, algorithms and variables by the object of a class definition. Objects of class definitions are generated during a compilation step. Note, using Achroma software is only possible after compilation. The internal communication of the tool runs with so-called 'object calls' and 'method calls' of objects. In forthcoming sections, objects of classes are always named by their class names and they are written in *Italian style*, for example: *txtMZStructList*. The same counts for methods: *getSpectrum()*. Here as previously mentioned, three modules were implemented to analyse the chromatographic data from the experimental design. The software contains one module for the graphical presentation of the chromatograms, one module to detect peaks and calculate peak areas, and one special module to compare two chromatograms. The latter represents the comparison results in a new quotient chromatogram.

In principle, if a new calculation step is necessary, a new module can be developed and has to be integrated into Achroma.

This type of modules is implemented in Microsoft Visual C++ using the Microsoft Windows XP operating system and the standard Microsoft Foundation Classes (MFC) of Microsoft Visual Studio. The MFC library offers a variety of methods and classes for the implementation of graphical interfaces and the data layer interfaces.

The application wizard of a Microsoft MFC project generates by default a model-view-controller model called a document/view architecture. The MFC document object manages the data handling and the coordination of data updating in multiple views. The latter can result in a spreadsheet view and a graphical presentation view. The MFC view object is responsible for the presentation of data to the user as well as for interactions entered by the user, *e.g.* selection or edition of data. Therefore the view object displays data obtained from the document object and communicates back to the document object (Fig. 3), if any data have (been) changed.³⁰

Every software module opens an own user window within the main frame of Achroma and every module is functional independently of other modules. The data structures are used by all modules jointly. In this way the software can be expanded at

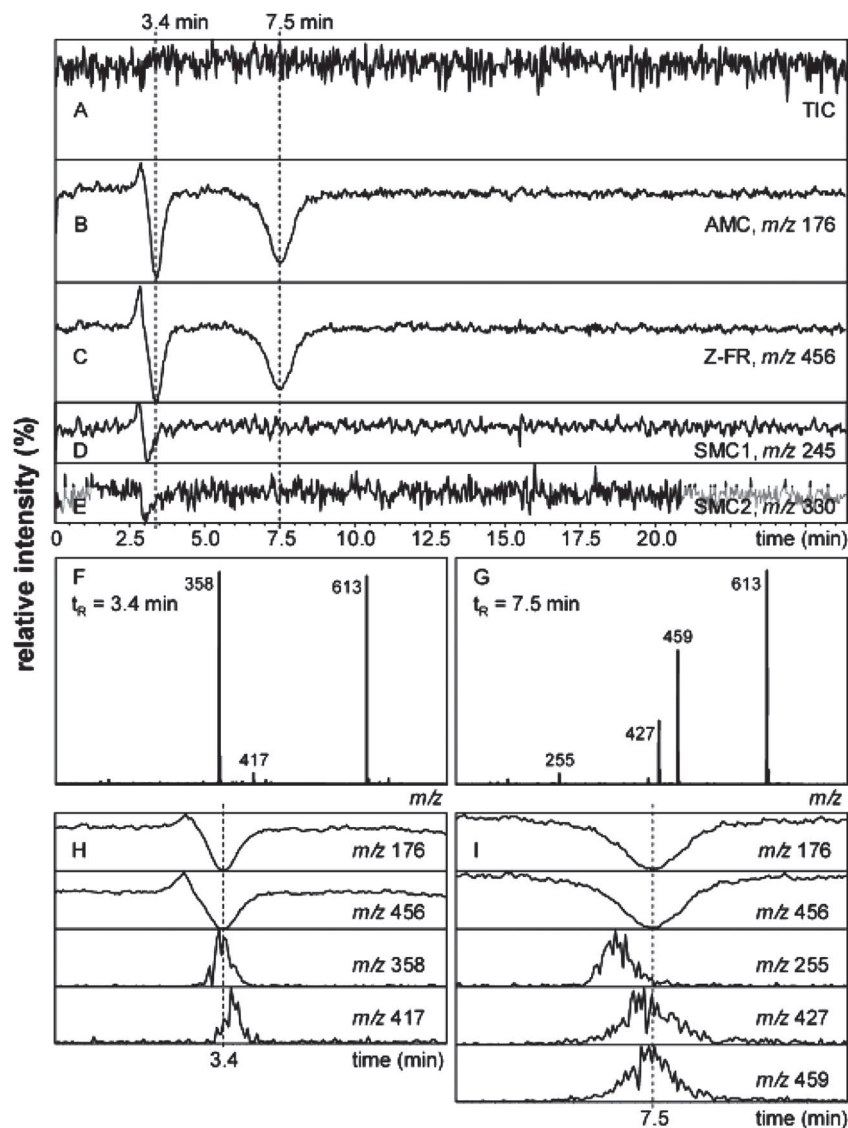


Fig. 2 On-line HPLC continuous-flow experiment of a mixture of five flavonoids spiked with two cathepsin B inhibitors, E-64 and leupeptin. (A) Total ion chromatogram (TIC) of the mixture, scan range m/z 75–750; (B) extracted ion chromatogram (EIC) of AMC (m/z 176); (C) EIC of Z-FR (m/z 456); (D) EIC of internal standard SMC1 (m/z 245); (E) EIC of internal standard SMC2 (m/z 330); (F) mass spectrum recorded at t_R 3.4 min; (G) mass spectrum recorded at t_R 7.5 min; (H) EICs of the most abundant peaks shown in the mass spectrum recorded at t_R 3.4 min; (I) EIC of the most abundant peaks shown in the mass spectrum recorded at t_R 7.5 min. Reprinted with permission from Arjen R. de Boer *et al.*: On-line coupling of high-performance liquid chromatography to a continuous-flow enzyme assay based on electrospray ionization mass spectrometry, *Anal. Chem.*, 2004, **76**, 3155–3161. Copyright 2004 American Chemical Society.²⁵

any time by additional modules without compromising the functionality of existing modules (Fig. 3).

The document/view architecture of the MFC library allows expanding the software with more features, but has the limitation that the software only works under the Microsoft Windows operating systems.

The Achroma project is currently under active development to improve the features and the compatibility over other operating systems.

2.3.1 Data transfer and data handling. Mass spectrometric data can usually be stored in a device- or company-specific data format on hard disks. Raw data stored in this way can usually

not be processed by software from another company. For this reason, mass spectrometric suppliers frequently offer some data conversion functionality. Common formats are for instance text-based ASCII, XML or binary NetCDF.³¹ There are also open formats specifically designed for mass spectrometry, such as mzDATA or mzXML,^{32,33} which are based on the open, generic XML (extensible markup language).

In an early phase of the project it has been decided that the Achroma software tool should be first used in conjunction with data in ASCII format. ASCII format is not widely applicable, but for an experimental software approach it is a good alternative. A textual based data format is very helpful showing the relations between data and software modules if new algorithms

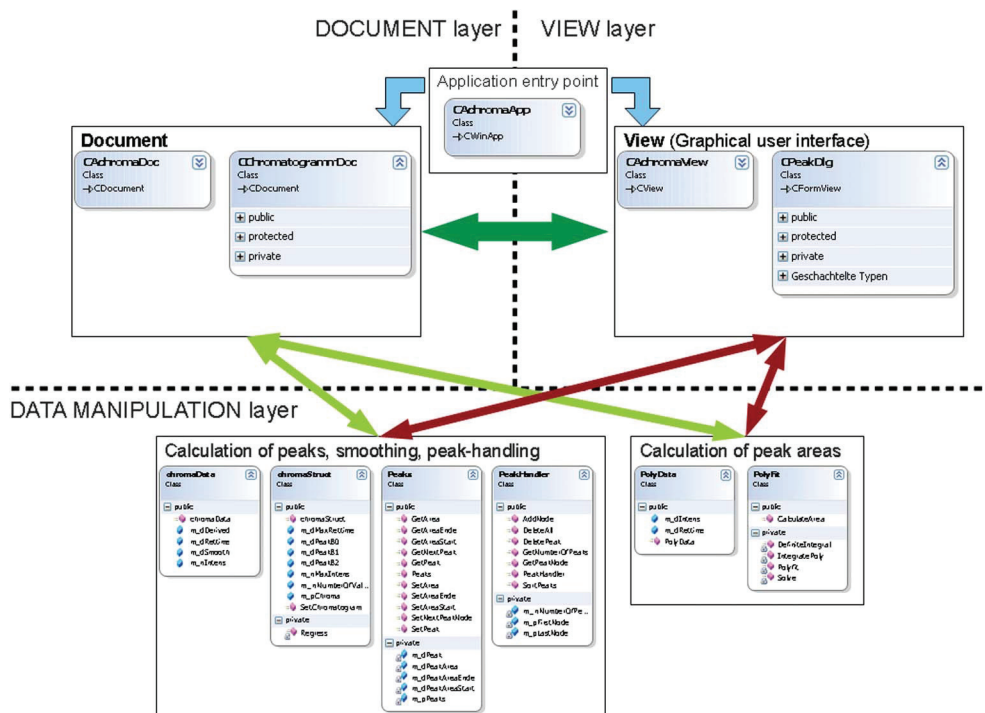


Fig. 3 Document/view architecture and the separated data layer shown with the classes for the module 'peak detection and area calculation'.

and modules are worked out with bioinformatic students. Another point is the simple way to generate specific test data (see below). In the first instance, mass spectrometric data were obtained from a mass spectrometric supplier with a conversion option into this format.

Data transfer. Exemplary continuous-flow experiments were made with an Agilent pumping and injection system, and a Waters mass spectrometer. Further systematically details can be found as published in *Protein and Peptide Analysis by LC-MS: Experimental Strategies*.²⁶ Prior to using the Achroma software a conversion step has to be performed of the original raw data format (e.g. obtained from the Waters MS) into an ASCII file format. For instance, Waters MS analysis software MassLynx (version 4.1)⁹ has a format conversion tool (DataBridge[®]) for converting the raw data files into other formats. Note, this step has to be made with the selected option in the DataBridge[®] software that no header information is transformed into the ASCII file. Thus, every generated ASCII file has the same data format (see Fig. 4: datafile.txt). Achroma data files can easily be generated with a simple text editor or with a program for spreadsheet, like Microsoft Excel or LibreOffice Calc. In forthcoming phases of the project further data formats importer, like netCDF or mzml importer, will be implemented.

Data handling. A bioinformatician has a different view to the data as an analytical chemist. The analyst looks at the obtained data and sees a three dimensional (3D) data structure of the example experiments. Behind each data point consisting of retention time with corresponding intensity there is a third dimension: a spectrum of m/z values and their corresponding intensities.

The business of the analyst is the evaluation of this amount in retention times, intensities and m/z ratios. The business of the informaticians is the handling of enormous sets of data and presenting them comprehensibly to the analyst.

The challenge is handling large data files, often performed by reduction of the data complexity and avoiding repetitions of program codes.

For example, the 3D data structure, obtained by the presented experiments, has a typical character of two 'series-connected' two dimensional (2D) data structures for the programmer:

- Retention time of chromatographic signals and corresponding intensity.
- m/z ratios of mass spectra and corresponding intensities.

Theoretically, two implementations of datasets ((a) and (b)) are needed, one for the chromatographic data and the other for the mass spectrometric data. However, here it is redundant because both datasets consist of two information sets in the form of numbers. Therefore the object-oriented programmer implements only one 2D dataset as a C++-program code description in an ordinary way. For example, two programming structures are possible: a *2D array*, if the array-size is known, and a *linked-list structure*, if the size of the dataset is unknown.

The advantage of the array structure is that every dataset has an index and allows a fast data access. On the other hand, if the array-size is reached an array extension is not available due to its static structure.

This is the advantage of a linked-list. The data-size can dynamically be expanded, but there is no possibility of indexing, like in arrays. Linked-list searches begin always at the beginning or at the end. Every element in the list has the information of the data containing the address of the first element and the prior element.

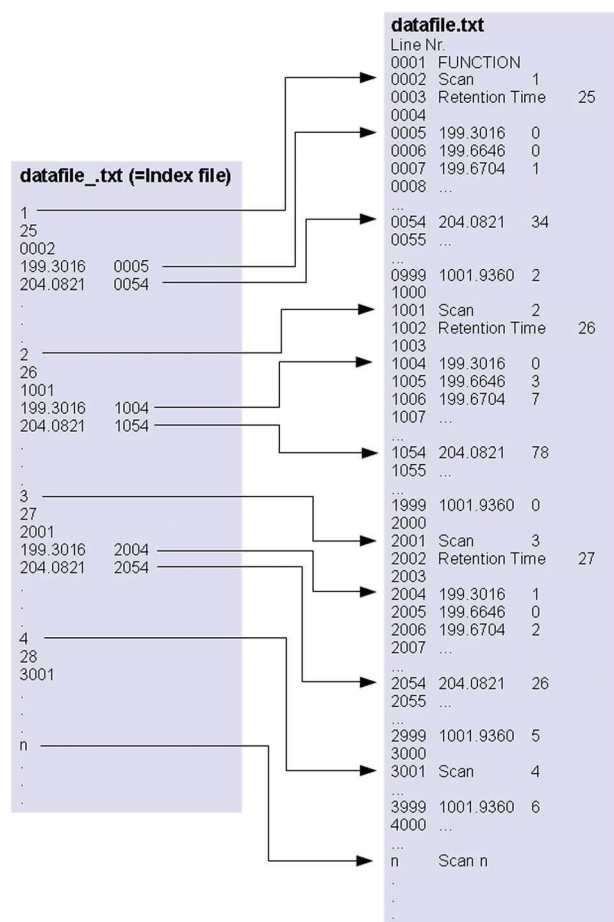


Fig. 4 Generating the index file from a data ASCII file. Left part (index file): column 1 = data from data file, column 2 = line number of the corresponding data from the data file.

Utilizing the two programming structures can reduce the complexity and can avoid a code repetition. The program code is the same, but the generated object is different, depending on the dataset which is delivered by the given experimental data and processed by an Achroma software tool.

The process described above for internal software data handling by an Achroma software tool is the base for applicable chromatographic and spectrometric data and the visualization of it for the user. Further, it is important for the generation of index files.

However, what is an index file and why is it necessary?

Mass spectrometry frequently generates a very large amount of data (e.g. file sizes up to 8 gigabytes in ASCII format). The access to specific locations in the data may consequently lead to a considerable time overhead. This problem could be overcome by generating an index file that was created by the software when the data file is loaded for the first time. This file is additionally being generated to the original ASCII data file. Furthermore it gets the same file name as the data file plus an underscore at the end of the file name and will be stored close to the corresponding data file.

The index file again consists of two parts. The first part is the index for the chromatographic data followed by the second part, the indices for the spectrometric data.

The index file contains only retention times (*i.e.* scan-numbers) and the corresponding line numbers or m/z ratios and the corresponding line numbers from the associated data file (Fig. 4). The index file has a pattern for chromatographic and spectrometric data, and is repeated for all scans: three lines are shown up for the retention time (Fig. 4 left part: 1 = scan number, 25 = retention time, 0002 = line number in the associated data file) followed by lines with two columns, one for the m/z ratio of this scan and the attached line numbers in the data file (Fig. 4 left part: 199.3016 = m/z ratio, 0005 = line number in the associated data file, ...).

The idea is not to store all line numbers for the m/z ratios of one scan, but only some information regarding the line numbers referring the m/z ratios in equal distances, for example every fiftieth line of a scan from the data file. This dramatically reduces the amount of information and improves the search performance in the data file by an associated index file.

The indexing started with the first retention time value from the original data file. Afterwards every new dataset is linked to the dataset before, until the last retention time has been reached. The index file data handling is managed by objects of the following three C++ classes: *txtStructList*, *txtStruct* and *txtStructMZPos*. The communication between the classes is not described in this publication, because it is not essential for the understanding of the basic idea behind the strategy.

Every dataset consists of one 'retention time value', several ' m/z values' and their corresponding 'intensities' (Fig. 4, right part). The index file dataset is generated similarly. Achroma software internal linked-lists will be created with the information about the line numbers of the corresponding retention times and the m/z ratios as described above.

All mass spectra can directly be found in the data file by using the 'retention time' index of the index file. The software jumps to the position of the chosen retention time and loads the m/z values and intensities into the random access memory (RAM) of the computer. The 'retention time' index accelerated the search for spectra considerably. Chromatograms, however, search all scans. In large data files, the search leads over all (several millions) lines to long loading times and thus is not user friendly. The implementation of an additional index should improve the loading time of chromatograms. The ' m/z ratio' index serves therefore as a 'search accelerator'. Additionally to the 'retention time' index, the m/z values and the position in the corresponding data file are noted in the index file to improve the performance. Not all scans are searched from the beginning to the end. The search starts instead in the adjacent area of the required m/z ratio for the first scan. The algorithm continues the check for all scans at this position and rebuilds the chromatogram graphically. Generating an index file lasted around 115 seconds for a 3,8 gigabytes ASCII data file with a 32 bit Microsoft Windows 7 operating system and a Fujitsu Siemens Amilo Xi 2528 computer with a Intel^(R) Core^(TM)2 Duo CPU T7500, 2.20 GHz and 2 gigabytes RAM. Loading a chromatogram to a given m/z ratio from the data file in a module lasted with the same computer for approximately less than 20 seconds. Visualization of spectra had no waiting time, because every spectrum belongs to a scan and all scan positions in the data file were written in the index file.

The data handling uses an index file for a faster performance, when ASCII data loading worked well. Reading only one data

format is a disadvantage, which will be broadening in future implementation cycles. However, generating data files for Achroma is very simple due to the plain structured data format. The data loading performance is good, depending on the computer hardware. Newer computers had central processor units (CPUs) with two or more cores. Generally, the performance can be improved, if the usage of more cores is implemented. Splitting the data file and calculating parts of the index in all available processors of the CPU in two or more threads is a possible solution. Also a higher RAM can increase the loading performance of data files. This is an example how to program and handle large data files in a time-spending process, as can easily be programmed from bioinformatics.

2.3.2 Graphical presentation. The resulting data can graphically be presented for each module. The retention times of the detected signals can additionally be shown in a table in the module for signal recognition. The data representation format was implemented for all modules according to the same scheme. The MFC library of Microsoft already makes a basic framework available for this.

Initially, a main-frame (*i.e.* main application graphical window) was initialised by loading Achroma. This in turn invokes the corresponding Child-frame (*i.e.* a graphical window inside the main frame) depending on the module choice. The Child-frames consist of Split Panes (*i.e.* a split graphical view inside the Child-frame) in which the upper half is a dialog area for selecting and changing calculation settings. The lower half contains the graph or table currently under inspection. Graphical representation is shown in Fig. 5.

The implementation of a split user interface ensured the easy-to-use of the software features in a dialog layer and a visualization layer with graphs and tables for all modules.

2.3.3 Processing of chromatographic data. The handling of chromatographic data on the “data manipulation” layer is implemented in *chromaData* objects (the object description can be found in the Implementation section). A *chromaData* object contains the retention time of one scan, the intensities of the corresponding *m/z* values, the smoothed and the derived value of the intensities. Each chromatogram is stored in a one dimensional (1D) array of *chromaData* objects with ‘number of values’ *chromaData* objects.

So far, Achroma software has three independent modules for processing the chromatographic data. The module is

- the visualization of single chromatograms (Fig. 5A),
- the comparison of two separate chromatograms (Fig. 5B), and
- the observation of signals and calculation of signal areas (Fig. 5C). Similar modules are available for the spectral data (Fig. 5D and 5E) as can be seen in section 2.3.4.

(a) *Visualization of single chromatograms.* The user loads a data file into the software and enters a *m/z*-ratio of interest for displaying the chromatogram after a short waiting time. The *m/z*-ratio range should be below 100 Dalton (Da). If the range is higher, the waiting time for GUI data loading will be unhandy and in worst case may lead to a software crash.

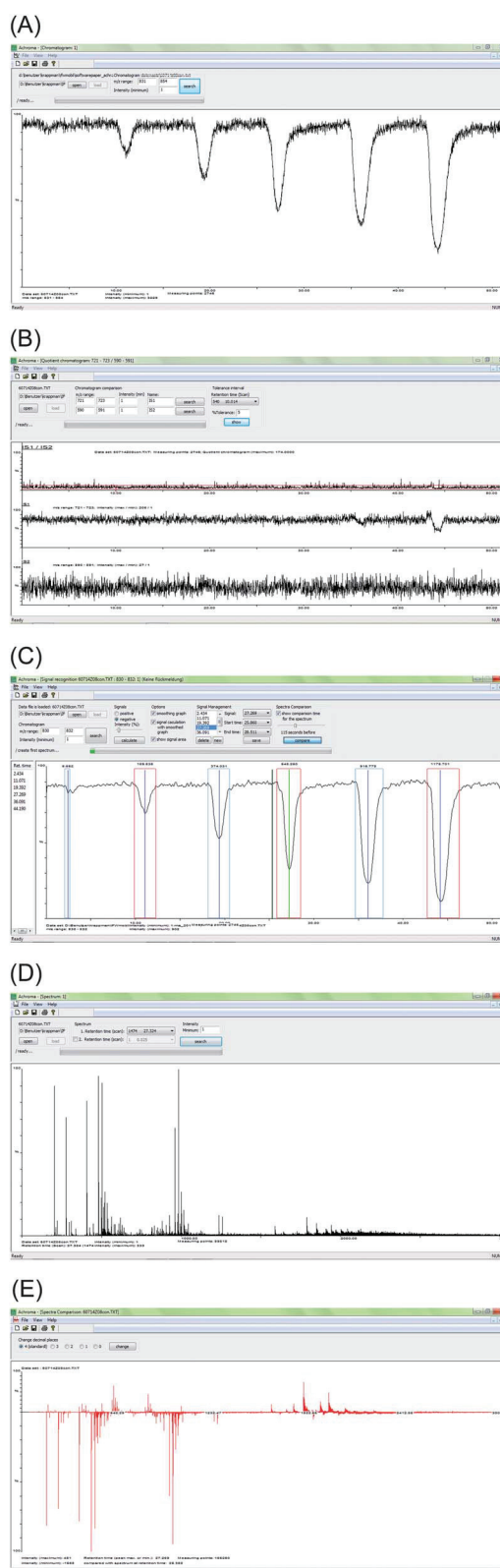


Fig. 5 (A) Visualization of a chromatogram. (B) Comparison and visualization of two chromatograms and displaying the quotient chromatogram. (C) Calculating signals and signal areas and showing them on the screen. (D) Visualization of a spectrum. (E) Generating and visualization of a difference spectrum.

(b) *Comparison of two separate chromatograms.* Achroma has an implemented module that can compare extracted ion chromatograms of two given m/z -ratios.

The corresponding retention times and signal intensities can be selected from the large data file by using the information on the two m/z traces, with the help of the previously generated index file and can be transferred into two arrays of *chromaData* objects. The quotient of the intensities is then calculated for each position in the *chromaData* object arrays, whereby the first intensity value of the first retention time in array one is divided by the first intensity of the same retention time in array two (and so on) until the last retention times in the arrays are reached. The resulting new chromatogram is constructed in temporary memory as a 2D array of numeric values and all three chromatograms were displayed graphically on the screen.

For example, after loading the data file, two m/z trace ranges are monitored (e.g. 721–723 and 590–591 in Fig. 5B) that shall be compared. Three chromatograms are displayed with the retention time on the x -axis and the intensity in percent on the y -axis. The upper chromatogram is the quotient chromatogram, which was calculated from the two chromatograms obtained from the m/z ratios for IS1 and IS2, two internal standards in Fig. 5B. The two single chromatograms are shown below.

$$pData[n].m-dSmooth = \frac{\text{intensity}(n-2) + \text{intensity}(n-1) \times 2 + \text{intensity}(n) \times 5 + \text{intensity}(n+1) \times 2 + \text{intensity}(n+2)}{11} \quad (1)$$

This module has an additional quality function. A tolerance interval for the intensity of a m/z value can be chosen. The tolerance will be shown as a red line for the positive and the negative deviation of the intensity in the graph around the quotient chromatogram. The user can see immediately, if the intensity quotient is in between the interval or not.

This function can be very useful, when an experimental setup has the shown internal standards (like in Fig. 5B) and is observed during the total experiment period like shown in the traces D (SMC1) and E (SMC2) of Fig. 2.

For example, one compound (e.g. SMC1) delivered through the complete flow system (e.g. with pump 4 in Fig. 1) can be compared with an internal standard (e.g. SMC2) injected directly prior to the mass spectrometric entry (e.g. between parts 7 and 8 in Fig. 1). Forming the quotient, a stable quotient represents then stable flow conditions in the total analysis system.

(c) *Recognition of chromatographic signal-changes.* The chromatographic signal recognition calculated by the software recognises both minima and maxima of the intensities in chromatograms. An assumption was therefore that a strong intensity increase or decrease should indicate a signal. However measurements were always subjected to noise which leads to constant intensity fluctuations and thus do not represent any signal.

The algorithms for ‘chromatographic peak detection’ and ‘peak-area calculation’ were implemented into the ‘signal recognition’ module. The user starts the peak recognition module and enters initially an m/z ratio, similarly to the comparison of the chromatograms module described above. Afterwards the method *SetChromatogram(int nNumberOfValues, chromaData*

**pData)* of the *chromaStruct* object in the ‘data manipulation’-layer is called from the view-layer (graphical user interface) through the *PeakDlg-class* object (Fig. 3). The parameter ‘int nNumberOfValues’ means that the *number of scans* is forwarded into the method and additionally as the second parameter a pointer to an 1D array of *chromaData* objects storing the chromatographic data. Now the *SetChromatogram* method recognizes the number of scans, all retention times and the corresponding intensities. This information is necessary for the three calculation steps processed by the *SetChromatogram* method and presented on screen *via* the ‘peak recognition’ module:

- signal maxima/minima calculation,
- chromatogram smoothing, and
- derivative calculation.

Smoothing the chromatograms is calculated *via* a weighted moving average³⁴ using the intensities of five scans ten times in two loops, whereby an inner loop calculates the values for the smoothed intensities and an outer loop repeats the smoothing step ten times.

Starting at the third scan in the array, the intensities of the prior two scans and the intensities of the following two scans were summed and weighted by the following scheme:

This ‘frame’ runs over the entire dataset of intensities smoothing these. The result is stored in the smoothed value variable of the corresponding *chromaData* object. The inner loop generates one smoothing cycle over all intensities in the 1D *chromaData* array (until an array size minus two is reached). This is necessary, because the algorithm (scheme above) uses values of the array beginning at the index $n + 2$ and therefore the inner loop has to stop at array size minus two to avoid errors. The outer loop repeats the smoothing cycle ten times as described above. This means, a ten times smoothing step is calculated for the graph of the chromatogram. The smoothing is needed for the derivative of the graph. The derivative of the smoothed graph is used for signal recognition. The initial and final points and also the position of a signal maximum or minimum can then be inferred from this derivative. However the smoothed curve still may have fluctuations which arise from noise. Thus also in this module of the signal recognition a tolerance limit is introduced. The tolerance limit indicates the ‘threshold value’ of intensity, above which a signal is recognised for positive signals and in the case of negative signals the intensity limit below a signal will be recognized. Each signal peak—the absolute maximum/minimum—has the same retention time as the derivative value on the point of sign changing together with a slope near $-/+1$. Finding signal starting and ending points is implemented as a comparison of the derivative values prior to and after the retention time value of the signal peak. The start or end point of a signal is closer to the retention time axis. This is also valid for the derived intensity values of the signal start and end point. The slope changes from (near) $-/+1$ to (near) 0 or *vice versa*. The implemented algorithm starts at the signal maximum/minimum, and searches in both

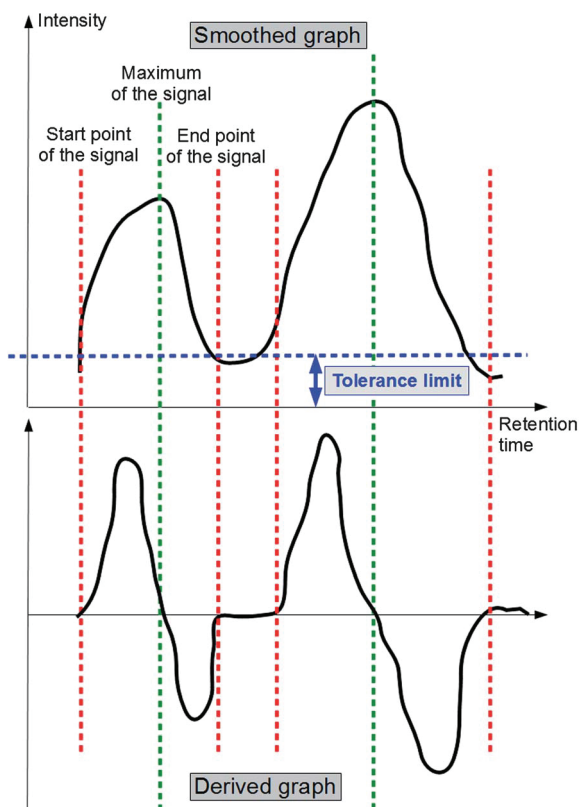


Fig. 6 Finding signal maximum/minimum, start and end retention time of a signal with the help of smoothing and derived data.

directions, obtained by the definitions from above. The points found are the starting point and ending point of the signal (Fig. 6). The calculation of the weighted moving average algorithm, the signal finding routines, finding derivative and regression were implemented in separate classes (Fig. 3).

In addition, the calculation of signal area maxima/minima is implemented, independent of whether a positive or negative signal has been detected.

The definite integral between the start and final time is calculated obtaining the area below the chromatographic signal. Any drift of the base-line is taken into account by the implementation of the area calculation.

Initially, the retention time interval is determined to be between the starting and ending point of the signal. In a next step the 1D array of *PolyData* objects is instantiated by Achroma software, with the size of the retention time interval plus one (to obtain the value of the starting point). The *PolyData* object array gets either the intensities and the corresponding retention times or the smoothed intensities and the corresponding retention times from the *chromaData* object array. This is depending on the user settings chosen through the dialog window of the software.

The complete area calculation is processed by the Achroma software. For this purpose further dedicated classes were implemented at the 'data manipulation' layer (Fig. 3). For instance a software implementation of the signal area calculation is performed like the following formula shows:

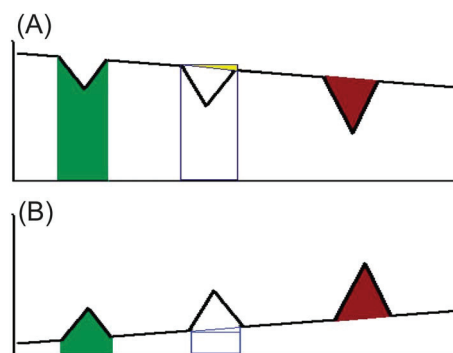


Fig. 7 Scheme of signal area calculation of negative (A) and positive (B) signals.

$$A_{\text{Signal}} = \left| \int_{R_S}^{R_E} (f(p_g)) - \frac{1}{2} \times (R_E - R_S) \times (I_S + I_E) \right| \quad (2)$$

where p_g is the smoothed peak (function for the total peak—from start to end), R_S is the retention time at the beginning of the signal, R_E is the retention time at the end of the signal, I_S is the intensity at the beginning of the signal, and I_E is the intensity at the end of the signal.

The signal area calculation starts with the instantiation of an object of the class *PolyFit*. This class has implemented the computation of the signal function definite integral. Afterwards a rectangle area is calculated by the *PolyFit* object. Thereby the object includes the following steps. The width of the rectangle is the difference value of the end and start retention time of a detected signal. The height is the intensity of the start point of a detected signal. At least a triangle area is calculated with the same width as the rectangle and the height as the difference of base line intensities between the start and end point of a signal. In case of a minimum of the signal function area the triangle area has to be subtracted from the rectangle area (Fig. 7A). In the other case (maximum) the rectangle area and the triangle area have to be subtracted from the peak function area (Fig. 7B).

Loading data works similarly to both other chromatographic processing modules. In addition, the actual or the smoothed data (shown as a graph) and a table, which presents the retention times regarding the calculated signal extrema, are displayed in the visual layer.

Various parameters can be set by observing selections in the dialog layer of the signal recognition module. The user interaction dialog area consists of four sections (Fig. 5C):

(1) *Peaks*: determines whether positive or negative signals should be presented and sets the intensity tolerance for signal recognition. Subsequently the button 'Calculate peaks' must be pressed.

(2) *Options*: indicates how the graph and the corresponding signal areas should be displayed and whether the smoothed graph should be used for the signal calculation.

(3) *Signal management*: signals can manually be added or deleted.

(4) *Spectra comparison*: the spectrum of a maximum or minimum signal can be compared with a spectrum from an earlier time point in the chromatogram. This function reflects an

additional module which is described in more detail in the following section.

The three implemented modules for chromatographic data processing worked well. The modules were tested with data of continuous flow enzyme-inhibition experiments as described above. As shown in Fig. 2 (traces B and C), the enzymatic conversion product decreases when the inhibitor concentration increases (traces not shown). The quotient chromatogram of the inhibitor m/z trace and product m/z trace showed increasing peaks during elution time. Internal standard 1 and internal standard 2 were stable during the sample runs (Fig. 2, traces D and E) thus no system fluctuation or instability may be observed. This experiment had demonstrated the function of the comparison module in a good way (Fig. 5B). In the signal detection and signal area calculation module both polarities—positive and negative—were recognised for all observable signals. The original vendor software for the example experimental setup (Fig. 2) could not calculate negative signal areas; Achroma on the other hand is able to calculate these areas and is crucial to evaluate the obtained data. The possibility to correct a signal maximum/minimum for retention time, the starting and ending retention time manually is an additional advantage.

The ten times smoothing algorithms reflected in all tests a good approximation to the raw data. A further advantage may be the implementation of other smoothing algorithms or the possibility of changing the smoothing repeating steps.

2.3.4 Processing of mass spectrometric data. The implementation of mass-spectrometric data handling was done in two parts. First, spectrum visualization was implemented and second, as part of the signal recognition module, the visualization of a difference spectrum was implemented. In the latter a spectrum from the bottom of the peak (or at the beginning of this signal) can be 'picked out' from the chromatogram and subtracted from the spectrum at the top/bottom of the chosen signal.

Every scan (*i.e.* time point in the experiment) generates a plenty of m/z values and their obtained intensities, depending on the composition of the sample (Fig. 4). This results in the fact that each scan may have a different amount of m/z values. The inconvenience was solved by implementing linked-lists for internal data handling. Therefore two new classes were designed: *SpektrNode* and *SpektrListe*, similarly to the internal handling of chromatographical datasets. *SpektrNode* represented one dataset with m/z range, the corresponding intensity and a pointer to the next *SpektrNode* object. The latter represents the next data object, which is filled with information. *SpektrListe* is responsible for managing the *SpektrNode* objects. The first index of the *SpektrNode* objects, the first quarter index, the half index, the last quarter index and the last index are stored again as search entry points in the *SpektrListe* object for a faster search through large amounts of m/z data.

The visualization of spectra to a given retention time interval was implemented in an own 'little' document/view architecture represented in objects of two classes:

- View structure = *SpektrumDLG, SpektrumView*.
- Document structure = *SpektrumDoc*.

The spectrum data can be read by the object of the *SpektrumDoc* class and visualized through the object of the *SpektrumView* class. User interactions are handled with the object of

the *SpektrumDLG* class. The spectrum module can be chosen from the main menu of Achroma software. The spectrum module shows—to a given retention time—the spectrum behind as a graphic in the visualization area of the module. First a data file had to be loaded. After the loading step is finished, a list with all scans is shown in the dialog area of the software. Afterwards a scan and a minimum intensity have to be chosen from the dialog and the spectrum is shown on the screen (Fig. 5D).

The visualization of a difference spectrum is part of the signal recognition module but also implemented in its own document/view architecture and therefore independent of other modules. The algorithm subtracts the spectrum at the chosen peak retention time from a spectrum backwards in the retention time series, depending on the amount of digits after the comma of interest (maximum is four digits).

The *Spectra Comparison* module is activated from the signal recognition module. The corresponding spectrum can be compared with a spectrum of a scan number in front of the chosen signal with a minimum or maximum in the chromatogram. Thereby the minimum of the retention time difference can be ten seconds and the maximum five minutes. The time difference selection can be realised *via* a slide control in the user interface layer of the module. This convenient function offers a quick summary of changes within the mass spectrometry sample being examined, for example due to chemical reactions within the sample.

The signals of the individual m/z ratios for reference spectra are subtracted from the signals of the spectrum at the instant of the signal extreme value, and the result is a *difference spectrum*. Signals that are more strongly represented in the reference spectrum than in the comparison spectrum are represented as positive in the difference spectrum; otherwise the representation in the difference spectrum is negative (Fig. 5E). The two modules showing spectrometric data and difference spectra worked well. A new feature of displaying difference spectra between a signal retention time and a previous retention time point is a useful and new tool analysing functional proteomics data, like in the shown example²⁵ for this strategy paper. Continuously mixing of enzyme and substrate solution into the system generates a constant product trace in the assays and is shown in the so called product chromatogram. A negative signal occurs if an inhibitor passed the system binding to the active site of an enzyme and preventing therefore the substrate conversion. Comparing the spectra of the product chromatogram at the minimum of the signal with a spectrum of a time point at the base line of the product chromatogram—before the signal occurs—shows simultaneously information about the possible inhibitor. It represents the difference spectrum obtained by subtracting the signal spectrum at the minimum from a spectrum at the base line. The resulting spectrum shows in a good way new m/z ratios, which are highly probable from the inhibitor molecule (Fig. 5E).

3 Conclusion

A software tool for analysing analytical data is not really new. However, describing the (bio-)informaticians view in relation to analysts view onto analytical data this is to our knowledge the first example. The manuscript has two primary objectives bringing the two 'worlds' together:

(A) Understanding of each other—(bio-)informaticians and analysts for better communication in future.

(B) Showing new analytical analysing features of a new software for analysing (a-)typical mass spectrometric data.

Firstly, a short description of a new experimental setup is described detecting enzymatic inhibitors in complex probes from plants. A LC-MS system was used to inject a gradient of samples in a continuous flow of enzyme and substrate. This leads to negative signals, especially if an inhibitor of the enzyme is in the injected probe. The typical vendor software had a problem to analyse the untypical negative signals, so a new solution was necessary. The need of another way to analyse the data manually or indirectly induces the idea of implementing a modular software system. It was realised together with bioinformatic students combining their teaching with practicing.

Achroma serves some standard modules, like displaying chromatograms or spectra, but has already new features that could help improving spectra analysis of atypical spectrometric data. One feature is the comparison of two chromatograms and the corresponding 'quotient-chromatogram'. The function is a valid tool to prove the stability of the experimental system, by comparing the analytical values of the two internal standards or for studying enzyme-inhibitor complex formations. Another useful module is the 'signal recognition' module, which can observe negative signals and their corresponding signal areas.

Additionally, the manuscript presents an easy description for a software strategy behind the Achroma software tool. The software strategy and the way how a software strategy will be implemented by an informatician is explained for analysts' understanding. So the benefit of reading this manuscript is: learning informaticians 'thinking' and getting knowledge about the new experimental setup and analysing methods of enzyme-inhibitor experiments.

4 Availability and requirements

Project name: Achroma

Operating system: Microsoft Windows XP or higher

Programming language: Microsoft Visual C++

Programming IDE: Microsoft Visual Studio 2003 or higher

Minimum software requirements: 1 GB RAM, 1 GHz processor

Acknowledgements

The authors thank several bioinformatic students contributing to the project as part of their practical study semester: in special these were in inverse time periods: Manuel Erhardt, Andrea Heidenreich, Patrick Schulte-Middelich, Oliver Schmidt and Wolfgang Hollweck. The authors would like to thank for financial support from the German Federal Ministry of Economics and Technology within the agenda for the promotion of industrial cooperative research and development (IGF) based on a decision of the German Bundestag. The access was organised by the AiF, Arbeitsgemeinschaft industrieller Forschungsvereinigungen, Köln.

References

- 1 T. Liu, *et al.*, Accurate mass measurements in proteomics, *Chem. Rev.*, 2007, **107**, 3621–3653.
- 2 M. Mann, C. K. Meng and J. B. Fenn, Interpreting mass spectra of multiply charged ions, *Anal. Chem.*, 1989, **61**, 1702–1708.
- 3 B. B. Reinhold and V. N. Reinhold, Electrospray ionization mass spectrometry: deconvolution by an entropy-based algorithm, *J. Am. Soc. Mass Spectrom.*, 1992, **2**, 207–215.
- 4 Z. Zhang and A. G. Marshall, A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra, *J. Am. Soc. Mass Spectrom.*, 1998, **9**(3), 225–233.
- 5 P. Kaur and P. O'Connor, Algorithms for automatic interpretation of high resolution mass spectra, *J. Am. Soc. Mass Spectrom.*, 2006, **17**, 459–468.
- 6 M. R. Hoopmann, G. L. Finney and M. J. MacCoss, High-speed data reduction, feature detection, and MS/MS spectrum quality assessment of shotgun proteomics data sets using high-resolution mass spectrometry, *Anal. Chem.*, 2007, **79**(15), 5620–5632.
- 7 P. C. Du and R. H. Angeletti, Automatic deconvolution of isotope-resolved mass spectra using variable selection and quantized peptide mass distribution, *Anal. Chem.*, 2006, **78**(10), 3385–3392.
- 8 Waters Corporation—MassLynx [<http://www.waters.com>].
- 9 M. Sturm, *et al.*, OpenMS—an open-source software framework for mass spectrometry, *BMC Bioinf.*, 2008, **9**, 163.
- 10 O. Kohlbacher, K. Reinert, C. Gröpl, E. Lange, N. Pfeifer, O. Schulz-Trieglaff and M. Sturm, TOPP—the OpenMS proteomics pipeline, *Bioinformatics*, 2007, **23**(2), e191–e197.
- 11 N. Jaitly, A. Mayampurath, K. Littlefield, J. N. Adkins, G. A. Anderson and R. D. Smith, Decon2LS: an open-source software package for automated processing and visualization of high resolution mass spectrometry data, *BMC Bioinf.*, 2009, **10**, 87.
- 12 K. C. Leptos, D. A. Sarracino, J. D. Jaffe, B. Krastins and G. M. Church, Map-quant: open-source software for large-scale protein quantification, *Proteomics*, 2006, **6**(6), 1770–1782.
- 13 X. J. Li, E. C. Yi, C. J. Kemp, H. Zhang and R. Aebersold, A software suite for the generation and comparison of peptide arrays from sets of data collected by liquid chromatography–mass spectrometry, *Mol. Cell. Proteomics*, 2005, **4**(9), 1328–1340.
- 14 J. Hartler, G. G. Thallinger, G. Stocker, A. Sturn, T. R. Burkard, E. Korner, R. Rader, A. Schmidt, K. Mechtler and Z. Trajanoski, MASPECTRAS: a platform for management and analysis of proteomics LCMS/MS data, *BMC Bioinf.*, 2007, **8**, 197.
- 15 M. Bellew, M. Coram, M. Fitzgibbon, M. Igra, T. Randolph, P. Wang, D. May, J. Eng, R. Fang, C. W. Lin, J. Chen, D. Goodlett, J. Whiteaker, A. Paulovich and M. McIntosh, A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS, *Bioinformatics*, 2006, **22**(15), 1902–1909.
- 16 L. N. N. Mueller, O. Rinner, A. Schmidt, S. Letarte, B. Bodenmiller, M. Y. Y. Brusniak, O. Vitek, R. Aebersold and M. Mller, SuperHirn—a novel tool for high resolution LC-MS-based peptide/protein profiling, *Proteomics*, 2007, **7**(19), 3470–3480.
- 17 Brusniak, *et al.*, ATAQS: a computational software tool for high throughput transition optimization and validation for selected reaction monitoring mass spectrometry, *BMC Bioinf.*, 2011, **12**, 78.
- 18 Nelson, *et al.*, LabKey server: an open source platform for scientific data integration, analysis and collaboration, *BMC Bioinf.*, 2011, **12**, 71.
- 19 M. Strohalm, M. Hassman, B. Kořata and M. Kodíček, mMass data miner: an open source alternative for mass spectrometric data analysis, *Rapid Commun. Mass Spectrom.*, 2008, **22**(6), 905–908.
- 20 M. Strohalm, D. Kavan, P. Novak, M. Volny and V. Havlicek, mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data, *Anal. Chem.*, 2010, **82**(11), 4648–4651.
- 21 P. Wenig and J. Odermatt, Openchrom: a cross-platform open source software for the mass spectrometric analysis of chromatographic data, *BMC Bioinf.*, 2010, **11**, 405.
- 22 A. R. de Boer, H. Lingeman, W. M. A. Niessen and H. Irth, Mass spectrometry-based biochemical assays for enzyme-inhibitor screening, *TrAC, Trends Anal. Chem.*, 2007, **26**(9), 867–883.
- 23 A. R. De Boer, J. M. Alcaide-Hidalgo, J. G. Krabbe, J. Kolkman, C. N. Van Emde Boas, W. M. A. Niessen, H. Lingeman and H. Irth, High-temperature liquid chromatography coupled on-line to a continuous-flow biochemical screening assay with electrospray ionization mass spectrometric detection, *Anal. Chem.*, 2005, **77**(24), 7894–7900.

- 24 A. C. Hogenboom, A. R. De Boer, R. J. E. Derks and H. Irth, Continuous-flow, on-line monitoring of biospecific interactions using electrospray mass spectrometry, *Anal. Chem.*, 2001, **73**(16), 3816–3823.
- 25 A. R. de Boer, T. Letzel, D. A. van Elswijk, H. Lingemann, W. M. A. Niessen and H. Irth, On-line coupling of high-performance liquid chromatography to a continuous-flow enzyme assay based on electrospray ionization mass spectrometry, *Anal. Chem.*, 2004, **76**(11), 3155–3161.
- 26 *Protein and Peptide Analysis by LC-MS: Experimental Strategies*, ed. T. Letzel, Royal Society of Chemistry, Cambridge, UK, DOI: 10.1039/9781849733144-00142.
- 27 J. Grassmann, R. K. Scheerle and T. Letzel, Functional proteomics: application of mass spectrometry to the study of enzymology in complex mixtures, *Anal. Bioanal. Chem.*, 2012, **2**, 625–645.
- 28 E. Gamma, R. Helm, R. Johnson and J. Vlissides, *Design Patterns*, Addison-Wesley, 12/1995. ISBN 0-201-63361-2.
- 29 G. E. Krasner and T. Stephen, Pope: a cookbook for using the model view controller user interface paradigm in Smalltalk-80, *J. Object-Oriented Programming*, 1988, **1**(3), 26–49.
- 30 <http://msdn.microsoft.com/de-de/library/4x1xy43a%28v=VS.100%29.aspx>, Stand: 10.09.2011.
- 31 <http://www.unidata.ucar.edu/software/netcdf/>, Stand: 10.09.2011.
- 32 P. G. A. Pedrioli, J. K. Eng, R. Hubley, M. Vogelzang, E. W. Deutsch, B. Raught, B. Pratt, E. Nilsson, R. H. Angeletti, R. Apweiler, K. Cheung, C. E. Costello, H. Hermjakob, S. Huang, R. K. Julian, E. Kapp, M. E. McComb, S. G. Oliver, G. Omenn, N. W. Paton, R. Simpson, R. Smith, C. F. Taylor, W. Zhu and R. Aebersold, A common open representation of mass spectrometry data and its application to proteomics research, *Nat. Biotechnol.*, 2004, **22**(11), 1459–1466.
- 33 S. Orchard, L. Montechi-Palazzi, E. W. Deutsch, P. A. Binz, A. R. Jones, N. Paton, A. Pizarro, D. M. Creasy, J. Wojcik and H. Hermjakob, Five years of progress in the standardization of proteomics data 4(th) annual spring workshop of the HUPO-proteomics standards initiative April 23–25, 2007 Ecole Nationale Supérieure (ENS), Lyon, France, *Proteomics*, 2007, **7**(19), 3436–3440.
- 34 A. Savitzky and M. J. E. Golay, Smoothing + differentiation of data by simplified least squares procedures, *Anal. Chem.*, 1964, **36**(8), 1627–1639.

3.2 Achroma Software - High-Quality Policy in (a-)Typical Mass Spectrometric Data Handling and Applied Functional Proteomics

Michael Krappmann, Christine Kaufmann, Romy Scheerle, Johanna Grassmann und Thomas Letzel

Journal of Proteomics & Bioinformatics, 2014, 7:264-271,

DOI: 10.1039/c2ay05622d

OPEN-ACCESS Artikel - Copyright: 2014 © Krappmann M, et al.

<http://omicsonline.org/open-access/achroma-software-for-functional-proteomics-jpb.1000328.php?aid=29977>

LC-MS basierte Enzym-Assays, aber auch die Proteomik im Allgemeinen, produzieren immense Mengen an Rohdaten. Zur Interpretation dieser bedarf es der spezialisierten Module der Software Achroma.

In weitergehenden kontinuierlichen LC-MS Enzym-Assays wurde die Änderung der enzymatischen Aktivität untersucht, sobald eine Interaktion des eingesetzten Enzyms mit regulatorischen Molekülen wie kompetitiven Inhibitoren oder alternativen Substraten in der Reaktionskapillare möglich war. Die Auswertung erfolgte mit den einzelnen Modulen von Achroma. Neben der Ermittlung der Ergebnisse, sowie deren Validierung wurde ebenso die Software Validität, Funktionalität und Benutzerfreundlichkeit geprüft.

In einem ersten Ansatz wurde als Enzym intestinale alkaline Phosphatase (iAP) aus der Rinder-Darmschleimhaut eingesetzt. Es wurde a) untersucht, wie sich die Produktbildung verändert, wenn zusätzlich zum Substrat AMP noch die alternativen Substrate ADP und ATP in der Reaktionskapillare zur Verfügung standen und b) die Messsystemstabilität beobachtet. Hierfür wurden die beiden Achroma Module *Chromatogram comparison* und *Signal recognition* von Achroma zur Visualisierung und Auswertung der Ergebnisse eingesetzt.

1. Es konnte gezeigt werden, dass der Versuchsaufbau stabil ist und es zu keinen Interaktionen der eingesetzten Enzyme, Substrate und des internen Standards mit dem Messsystem kommt. Mittels interner Standards wurden die Daten validiert. Beispielsweise konnte durch die Berechnung des Quotientenchromatogramms mit Achroma ausgeschlossen werden, dass es sich bei den erhaltenen Signalen um Systemfehler handelte. Die Standardabweichungen lagen hierbei jeweils unter +/- 10%.

2. Standen dem Enzym iAP neben dem Substrat AMP weitere konkurrierende Substrate (ADP, ATP) zur Verfügung, so konnte gezeigt werden, dass die Produktbildung von primären Produkt (ADO) beeinflusst war. Die Visualisierung der einzelnen Signale, die automatische Signalerkennung, sowie die Möglichkeit der Datenevaluation mit Achroma durchzuführen, war erfolgreich und qualitativ nutzbar.

Im zweiten Versuchsansatz wurde Acetylcholinesterase (AChE) als Enzym verwendet. Ziel war es dabei Achroma als speziell programmiertes Werkzeug für ein schnelles Inhibitorscreening zu nutzen. Hierzu wurde eine Hausstaubprobe mit einem bekannten Inhibitor (Galanthaminhydrobromid) versetzt und nach HPLC-Trennung in das System eingebracht. Die kontinuierlich gemessene Produk Spur zeigte nach Auswertung mit Achroma ein negatives Signal nach dem Eintrag der über eine LC-Säule getrennten Hausstaubprobe. Anhand des berechneten Differenzspektrums wurde ein neues m/z Signal zum Zeitpunkt des negativen Produktsignals beobachtet. Dieses entsprach dem m/z-Wert von 288 (M+H)⁺ und somit dem protonierten Molekülion des bekannten Inhibitors. Eine Kontrolle der Inhibitorspur von m/z 288 zeigte genau zum Zeitpunkt des Einbruchs der Produk Spur ein positives Signal und bestätigte die Annahme, dass der Einbruch der Produk Spur auf eine Interaktion Enzym mit Inhibitor zurückzuführen war. Achroma eignet sich als Werkzeug für die Aussage, ob in einer komplexen Probe, wie z.B. Hausstaub, Inhibitoren bzw. Regulatoren detektiert werden können.

Das Manuskript wurde federführend zu gleichen Teilen von Frau Kaufmann und mir verfasst. Die experimentellen Versuche wurden durch die Co-AutorInnen durchgeführt. Die Datenauswertung erfolgte gleichwertig sowohl von Frau Kaufmann als auch von mir. Frau Scheerle lieferte Beiträge zum Kapitel Acetylcholinesterase Assay und Frau Grassmann zur Diskussion der Enzym Assays. Priv.-Doz. Dr. Letzel begleitete die komplette Arbeit, lieferte Beiträge, korrigierte das Manuskript und reichte es ein.

Achroma Software-High-Quality Policy in (a-)Typical Mass Spectrometric Data Handling and Applied Functional Proteomics

Michael Krappmann^{1#}, Christine M Kaufmann^{2#}, Romy K Scheerle², Johanna Grassmann² and Thomas Letzel^{2*}

¹University of Applied Sciences, Am Hofgarten 4, 85354 Freising, Weihenstephan, Germany

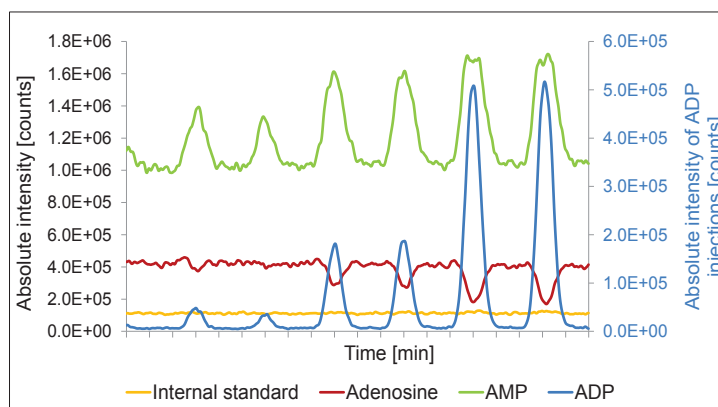
²Chair of Urban Water Systems Engineering, TU München, Am Coulombwall, 85748 Garching, Germany

[#]Equal contribution of the authors

Abstract

Data evaluation of mass spectrometric raw data is an essential step to obtain high-quality results. In our days an exorbitant amount of raw data are produced in analytical (bio)chemistry due to the utilization of sophisticated experimental setups. The recently published free software Achroma has been developed to overcome increasing data processing challenges by providing the possibility of a comprehensive data analysis (<http://openmasp.hswt.de/pages/project/achroma.php>). To illustrate (a-)typical data evaluation, an online coupled continuous flow system hyphenated with mass spectrometric detection was applied to investigate enzymatic activity changes in the presence of regulatory molecules and alternative substrates. The extended software strategy, the processing as well as evaluation of data is presented in detail based on enzymatic assays of intestinal alkaline phosphatase (iAP) and acetylcholine esterase (AChE). Different Achroma data evaluation modules enabled a high quality analysis. This included the elucidation of enzymatic substrate preferences by means of the calculation of negative and positive peak areas as well as the identification of an inhibitory molecule by comparing different mass spectra with regard to their overall composition. The possibility of an automatically performed validity control to monitor the systems robustness furthermore emphasized the usefulness of Achroma software regarding its applicability in the area of 'functional proteomics' data handling.

Graphical Abstract



Challenges regarding (a-)typical data processing and their overcoming by applying Achroma Software are outlined. Data evaluation includes e.g. the assessment of data quality and the calculation of peak area.

Keywords: Functional proteomics; Enzyme; Data evaluation software; Online-coupled continuous flow system; Chromatography; Mass spectrometry

Abbreviations: LC: Liquid Chromatography; MS: Mass Spectrometry; EIC: Extracted Ion Chromatograms; AChE: Acetylcholine Esterase; SRM/MRM: Selective or Multiple Reaction Monitoring; IS: Internal Standards; openMASP: Open Source Modular Analytical Software Platform

Introduction

Mass spectrometry (MS) combined with liquid chromatography (LC) is nowadays a key technology in the field of analytical chemistry. Research areas like 'small compound screening' in food and environmental chemistry and also the investigation of large molecules like proteins may lead to a plethora of data, which need to be processed, evaluated and interpreted. Beyond that, (LC-) MS is a key technology in

the field of 'functional proteomics', as part of experimental setups used to monitor enzymatic reactions in real-time or to study the formation of non-covalent complexes [1-6].

***Corresponding author:** Thomas Letzel, Chair of Urban Water Systems Engineering, TU München, Am Coulombwall, 85748 Garching, Germany, Tel: +49 (0)89 2891 3780; Fax: +49 (0)89-289-13718; E-mail: T.Letzel@tum.de

Received July 03, 2014; **Accepted** August 25, 2014; **Published** August 28, 2014

Citation: Krappmann M, Kaufmann CM, Scheerle RK, Grassmann J, Letzel T (2014) Achroma Software-High-Quality Policy in (a-)Typical Mass Spectrometric Data Handling and Applied Functional Proteomics. J Proteomics Bioinform 7: 264-271. doi:[10.4172/jpb.1000328](https://doi.org/10.4172/jpb.1000328)

Copyright: © 2014 Krappmann M, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

The rapid improvement of analytical instruments and thereby the emergence of new and complex experimental setups further contribute to an increasing amount of data and in this manner to data evaluation challenges. One single experiment may result in thousands of mass spectra and extracted ion chromatograms (EIC), depending on the complexity of the analyzed mixtures, the measurement time and the number of substrate and product traces, especially in multiplexing approaches.

In this regard, the application of an online coupled continuous flow system to detect changes in enzymatic activity in the presence of regulatory molecules may result in the emergence of complex data sets. This well-established technique has already been applied for the detection of interactions between proteins and compounds of interest [7], for the detailed investigation of enzyme inhibitors [8,9] or even on the screening of complex mixtures for unknown enzyme regulatory compounds [8-10]. The applicability of this systems has already been proven for the investigation of several enzymes, i.e. acetylcholine esterase (AChE) [9,8], phosphodiesterase [10], glutathione S-transferase [11] and cathepsin B [8]. In the presented work the setup was applied for the investigation of AChE and iAP enzymatic assays, either to find regulatory molecules from complex mixtures or to detect the enzymatic response to alternative substrates. Both applications were found to result in unusual data sets, which enclose the presence of positive and negative peaks within the assay EICs. The introduction of complex mixtures with unknown composition furthermore necessitated a fast and easy identification of potential enzyme regulatory compounds. Standard software provided by LC/MS hardware vendors is usually expensive and often does not comprise all functions required for entire data analysis. Researchers may have to conduct additional procedures like the export of data and further handling with common spreadsheet software. This software type however in most cases lacks certain requirements for a detailed data evaluation. Thus, the development of new software allowing automated data handling and rapid as well as easy data interpretation and data evaluation is essential. Consequently, a special free software—Achroma—was developed and applied to clearly visualize, process and analyze a large amount of chromatographic and mass spectrometric data. So far, the Achroma software includes several independent data handling modules, like a) 'chromatogram comparison' for the calculation of a quotient of particular EICs to track signal instancies, b) 'signal recognition' of positive and negative peak signals (incl. signal area calculation) and c) 'signal comparison' tool. Latter enables the discovery of differences between two spectra at two different defined time points, e.g. comparing the EICs baseline range with a signal of enzymatic regulation. In contrast to common data evaluation software the possibility of a data quality examination was implemented into Achroma software to obtain helpful information about the robustness of the applied experimental setup. Automatically performed signal recognition and signal comparison furthermore provide a distinct reduction of elaborate manual data processing time.

In contrast to other available software tools for selective or multiple reaction monitoring (SRM/MRM) [12], e.g. like Skyline [13] or ATAQS [14], Achroma provides the calculation of negative peak areas, which may occur in continuous flow experiments of enzymatic activity determination in the presence of regulatory compounds. By adding the possibility of a difference spectrum calculation, Achroma software tool further stands out from classical SRM/MRM software regarding helpful features for functional proteomics data evaluation. Achroma software and its features were exemplarily employed on AChE and iAP enzymatic assays to evaluate data quality features and to illustrate application-oriented data evaluation. The enzymatic assay data were

processed with respect to a) performing data quality control and signal robustness assessment via EIC correlation of an internal standard (IS) and the EIC of an assay component, b) monitoring signal changes of iAP enzymatic assay induced by the injection of competitive substrates and c) screening of single compounds contained in complex mixtures on their regulatory effect on AChE, e.g. inhibitory effects.

Materials and Methods

Reagents and chemicals

All compounds for the enzymatic assays were purchased from Sigma-Aldrich (Steinheim, Germany): AChE from *Electrophorus electricus* type VI-S (AChE, Enzyme Commission (EC) number 3.1.1.7, M_w ~280 kDa), acetylcholine chloride (AChCl, M_w 181.7 Da), galanthamine hydrobromide from *Lycoris* sp. (M_w 368.3 Da), iAP from bovine intestinal mucosa (EC 3.1.3.1., M_w ~160 kDa), Adenosine 5'-triphosphate (ATP, M_w 507.2 Da), Adenosine 5'-diphosphate (ADP, M_w 427.2 Da), Adenosine 5'-monophosphate (AMP, M_w 347.2 Da), LC-MS grade reagent water, LC-MS grade methanol and ammonium acetate (M_w 77.1 Da, >98%). L-histidine (M_w 155.16 Da) served as internal standard and was obtained from Merck Chemicals (Darmstadt, Germany).

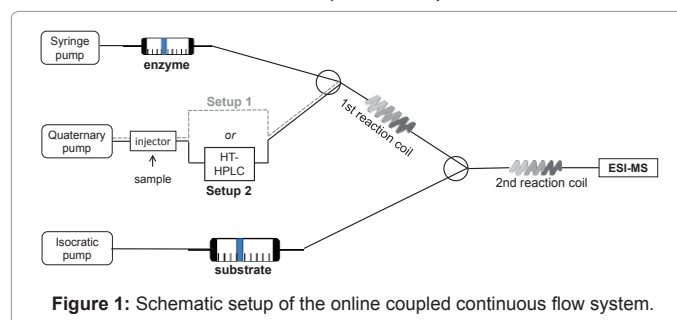
A house dust extract sample was used for screening experiments. The sample was taken from a household with visible mould infestation and was prepared as follows: 0.5 g house dust was sieved and the particle fraction $\leq 63 \mu\text{m}$ was extracted with 5 mL methanol/ H_2O (84:15). 1 mL was evaporated to dryness, resolved in methanol/ H_2O (20:80), filtered with a $45 \mu\text{m}$ RC-filter and used for application in enzymatic assays.

Instrumentation

Online coupled continuous flow system: The experiments were performed using an online coupled continuous flow system with sample introduction part (Injector, Rheodyne, California, sample loop volume: 2 μL) as presented in Figure 1.

In setup 1 (Figure 1, middle trace: light grey dashed line), the enzymatic activity of iAP was measured towards its nucleotide substrate AMP. Activity changes in AMP substrate degradation along with adenosine (Ado) product generation, caused by the introduction of additional AMP substrate as well as increasing concentrations of the competitive nucleotide substrates ADP and ATP [15], were investigated. Concentration of the enzyme and the AMP substrate introduced to the system were 2.4 U/mL and 80 μM , respectively. Both were solved in 10 mM ammonium acetate pH 7.4.

Setup 2 (Figure 1, middle trace: black solid line) was used for the determination of the AChE activity with regard to an enzymatic regulation potentially caused by a chromatographically separated house dust mixture of unknown molecules. Initial concentrations were set to 0.05 nM of AChE and 10 μM of acetylcholine, either solved in



10 mM ammonium acetate pH 7.4. Internal standard L-histidine was introduced to the systems along with the enzyme solutions and was set to an initial concentration of 80 μM . In both setups the enzyme was provided in a syringe (2.5 mL, Hamilton-Bonaduz, Switzerland) located in a syringe pump with a flow rate set to 25 $\mu\text{L}/\text{min}$ (Figure 1, upper trace) (Model 11 Plus, Harvard Apparatus, Hugo Sachs Elektronik, Hugstetten, Germany). Substrate solution was filled in a superloop (volume 10 mL, Amersham Biosciences, Uppsala, Sweden) and introduced to the system with a flow rate of 50 $\mu\text{L}/\text{min}$ (Figure 1, bottom trace). Samples were injected into the continuous flow delivered by the quaternary pump via an injector (Rheodyne, California, sample loop volume: 2 μL). The quaternary pump flow rate (1100 series, Agilent Technologies, Waldbronn, Germany) was set to 25 $\mu\text{L}/\text{min}$, taking the sample either directly to the first reaction coil (setup 1) or along to the chromatographic column (setup 2). The isocratic pump (1260 series, Agilent Technologies, Waldbronn, Germany) was connected to the superloop for pumping the substrate solution to the system (Figure 1, bottom trace). Pumps were controlled by ChemStation software (version B.04.03, Agilent Technologies, Waldbronn, Germany). Reaction coils 1 and 2 (Teflon, 0.25 mm ID) were knitted as described in literature [16] to provide a sufficient mixing of enzyme and injected sample (reaction coil 1, flow-through time of ~ 1.3 min) and of the enzyme-sample mixture with the substrate (reaction coil 2, flow-through time of ~ 2.9 min), respectively. Apart from reaction coil 1 and 2 all further modules in both setups were connected with PEEK tubing. In setup 1, 10 mM ammonium acetate pH 7.4 was delivered continuously by the quaternary pump (Figure 1, middle trace). In setup 2 an isocratic eluent flow with 90% 10 mM ammonium acetate pH 7.4 with 10% methanol was used to separate the sample on a high temperature-compatible Zirchrom PBD column (100x1 mm, 3 μm) (ZirChrom Separations, Inc., Anoka, MN, USA), which was integrated into a column oven (HT HPLC 2000, Scientific Instruments Manufacturer GmbH, Oberhausen, Germany) (Figure 1, middle trace). The organic solvent concentration tolerable for AChE was preliminarily tested [17], whereupon 10% methanol was found to be suitable to maintain sufficient enzymatic activity. A temperature gradient was applied to improve the chromatographic separation. Initial temperature was set 30°C, followed by a linear temperature increase up to 120°C within a time range of 20 min, whereupon 120°C was held for 10 min. After completion of the gradient, the column was automatically reconditioned to the starting temperature of 30°C.

Mass spectrometric assay detection: Samples were analyzed using a single quadrupole mass spectrometer (MSQ Plus, Wissenschaftliche Gerätebau Dr. Ing. Herbert Knauer GmbH, Berlin, Germany) equipped with an electrospray ionization source in positive ionization mode. Mass spectrometric parameters were as follows: 300°C probe temperature, 3.5 kV needle voltage and 75 V cone voltage. In setup 1, the mass range for the iAP assay component detection was 100-1000 m/z , whereas in setup 2, the mass range for experiments performed with AChE was 100-800 m/z .

The following signals were analyzed in the AChE assay with the substrate acetylcholine: substrate signal m/z 146 (M)⁺ and product signal m/z 104 ($\text{M}-\text{CH}_3\text{COOH}$)⁺.

For data evaluation of iAP assays, related m/z signals were summarized, using m/z 268.1 and 290.1, *i.e.* ($\text{Ado}+\text{H}$)⁺ and ($\text{Ado}+\text{Na}$)⁺; m/z 348.1, 370.1, 386.0, 392.0 and 408.0, *i.e.* ($\text{AMP}+\text{H}$)⁺, ($\text{AMP}+\text{Na}$)⁺, ($\text{AMP}+\text{K}$)⁺, ($\text{AMP}-\text{H}+2\text{Na}$)⁺ and ($\text{AMP}-\text{H}+\text{Na}+\text{K}$)⁺ and m/z 428.0, 450.0, 466.0, 472.0 and 488.0, *i.e.* ($\text{ADP}+\text{H}$)⁺, ($\text{ADP}+\text{Na}$)⁺, ($\text{ADP}+\text{K}$)⁺, ($\text{ADP}-\text{H}+2\text{Na}$)⁺ and ($\text{ADP}-\text{H}+\text{Na}+\text{K}$)⁺.

Data evaluation

Mass spectrometric data was acquired using Xcalibur software 2.1.0.1139 (Thermo Fisher Scientific Inc., Waltham, MA, USA). For data evaluation with Achroma software either an entire Xcalibur data file was processed directly with an Xcalibur-implemented file converter tool, followed by a further conversion step with Achroma converter software or individual EICs of interest were transferred to Office Excel, whereupon further processing was again performed with Achroma converter (Figure 2).

The latter procedure leads to a minor amount of information, only containing data of particular EICs, which is sufficient in case of known m/z values of interest and advantageous regarding the reduction of the final data volume and file sizes. In contrast by converting the entire Xcalibur data file the mass spectrometric information recorded is entirely preserved. A more detailed explanation regarding the conversion of raw data is provided on the Achroma web page [18].

Data interpretation of enzymatic assays was accomplished by applying several Achroma modules after smoothing the data: ‘Chromatogram comparison’ tool allows the calculation of a quotient of two independent EICs, e.g. the traces of an internal standard and an assay component. Hence, the ratio obtained enables the correction of experimental data for intensity irregularities.

Moreover, enzymatic regulation manifested as negative or positive peaks present in the EIC traces was detected with ‘signal recognition’ module to assess peak areas, thereby obtaining numerical values that correspond to an alteration of enzymatic activity, whether it is inhibitory or activating. Further evaluation of data was carried out using ‘spectra comparison’. By means of this module, regulatory molecules affecting enzymatic activity can be easily identified by comparing the mass spectrum of a time point of altered enzymatic activity with the mass spectrum of a non-regulated time range. In doing so, differences

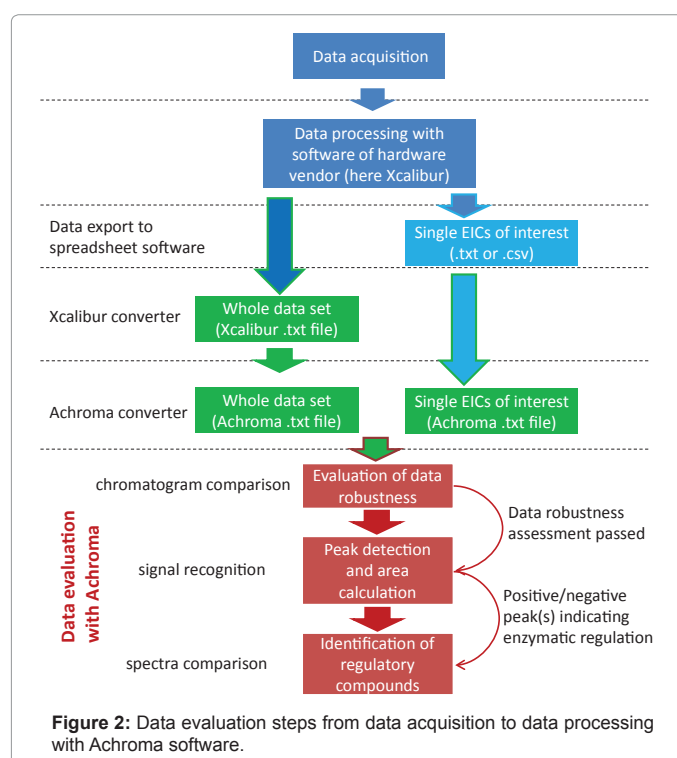


Figure 2: Data evaluation steps from data acquisition to data processing with Achroma software.

regarding the total composition of both spectra or merely the alteration of single m/z signal intensities respectively can be revealed. Temporarily emerging m/z values or m/z values with a distinct increase of intensity are presented as negative signals and vice versa.

Results and Discussion

A continuous flow system coupled with MS detection (Figure 1) was applied to study enzymatic activity of iAP and AChE, respectively. Temporary increases or decreases of mass spectrometric substrate or product traces in the presence of inhibiting or activating molecules can be used to identify regulatory molecules from complex mixtures. In the absence of regulatory molecules, a constant trace of substrate and product is continuously detected, which reflects an unchanged enzymatic activity.

The presence of *e.g.* an inhibiting molecule is reflected by a temporary increase of substrate intensity (positive peak) and a simultaneously occurring decrease of product intensity (negative peak). In this regard, the evaluation of positive as well as negative peak signals to assess enzymatic activity changes is inevitable. To address this need a former Achroma software version was developed and recently published [19]. Data quality and also the overall data evaluation procedure was considered applying three modules of Achroma software-‘signal recognition’, ‘spectra comparison’ and ‘chromatogram comparison’.

The handling and processing of raw data with a new Achroma version (incl. additional functionality like an extended file converter tool) is demonstrated by means of iAP enzymatic activity measured with the online coupled continuous flow system in section 3.1 (Figure 1 and Setup 1). A further verification of proper data analysis is shown by the evaluation of AChE enzymatic assay measured in the presence of a chromatographically separated complex real house dust extract, which was introduced to the system (Figure 1 and Setup 2).

Intestinal alkaline phosphatase assay data evaluation

iAP is able to dephosphorylate the nucleotide AMP, thereby generating the nucleoside product adenosine. iAP (Figure 1, upper trace) and AMP (Figure 1, bottom trace) were introduced to the online coupled continuous flow system, resulting in a constant AMP substrate trace and an enzymatically dephosphorylated Ado product trace. Enzyme and substrate concentrations were chosen to obtain an enzymatic substrate conversion rate that results in a sufficient mass spectrometric substrate as well as product trace intensity. The injection of different concentrations of substrate AMP or alternative substrates ADP or ATP led to the generation of intermediates and an altered Ado product generation, which is due to the enzyme’s ability to dephosphorylate its substrates in a stepwise manner [15].

As previously investigated, iAP activity measurements towards AMP compared to ADP and ATP revealed the enzyme’s preference to dephosphorylate the latter two with a rate about twice and thrice as high, respectively [15].

ADP and ATP injections (Figure 1, middle trace) consequently caused a reduced generation of the final product Ado, which is solely formed through AMP dephosphorylation. Additionally, injected ADP or ATP are largely degraded to AMP or AMP and ADP by the removal of one or two phosphate groups, respectively [20].

Assessment of signal stability and overall data quality: After successfully adapting iAP enzymatic assay to the online coupled continuous flow system, first experiments showed recurring drops within assay component traces. To investigate whether these

inconstancies are due to mass spectrometric detection or related to changes of enzymatic activity, data was assessed with Achroma software module ‘chromatogram comparison’. The extent of signal inconstancies within the course of internal standard and assay component traces (substrate, intermediates or product) can easily be displayed, compared and corrected (Figure 3). By doing so, actual enzymatic activity changes due to the presence of regulatory molecules and events of system-related intensity loss can be distinguished. This data correction step is therefore particularly important, since signal inconstancies might lead to a feigning of enzyme regulation. ‘Chromatogram comparison’ tool hence provides the possibility to monitor and correct assay EICs for signal detection irregularities and potentially assign them either to signal suppression effects due to injected molecules, pumping inconsistencies, system leakage or others.

Assay traces of the internal standard (IS) and the substrate AMP (Figure 3) are selected by entering the respective masses into the provided input field of Achroma software (Figure 3, solid black line, exemplary m/z range 155 to 156 and 347 to 348). By means of chromatogram comparison the data were edited by calculating the quotient of those molecule EICs, resulting in a constant and consistent ratio which is corrected for signal inconstancies (Figure 3C).

A tolerance interval can be additionally selected (Figure 3, dashed line box and Figure 3C) to control the resulting EIC after correction. The tolerance interval is shown as additional lines below and above the calculated quotient chromatogram (Figure 3C). The correction procedure applied to the data might therefore also reveal further inconstancies beyond the events of obvious signal losses, which would be clearly revealed as signals ranging out of the tolerance interval set. Those exceeding signals might therefore be accounted to enzymatic regulation events, which were previously concealed by the more dominant signal inconstancies. In case of the data set exemplarily used, no further irregularity is apparent, visible by the trace’s course lying within the tolerance limit of 10%.

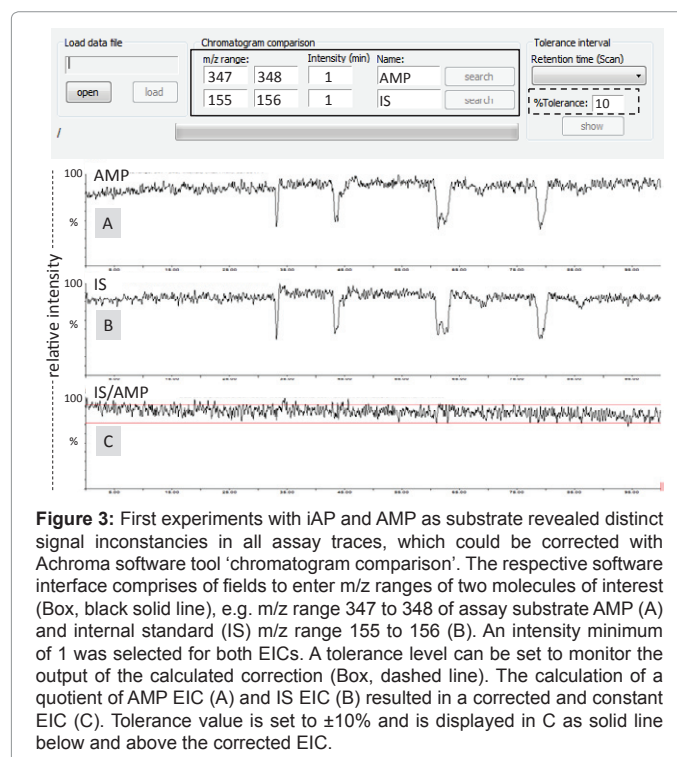


Figure 3: First experiments with iAP and AMP as substrate revealed distinct signal inconstancies in all assay traces, which could be corrected with Achroma software tool ‘chromatogram comparison’. The respective software interface comprises of fields to enter m/z ranges of two molecules of interest (Box, black solid line), *e.g.* m/z range 347 to 348 of assay substrate AMP (A) and internal standard (IS) m/z range 155 to 156 (B). An intensity minimum of 1 was selected for both EICs. A tolerance level can be set to monitor the output of the calculated correction (Box, dashed line). The calculation of a quotient of AMP EIC (A) and IS EIC (B) resulted in a corrected and constant EIC (C). Tolerance value is set to $\pm 10\%$ and is displayed in C as solid line below and above the corrected EIC.

After successfully examining the quality of the data set, evaluation regarding the effects of an alternative substrate introduced to iAP assay as well as the finding of regulatory molecules were accomplished by applying additional Achroma software modules.

Enzymatic regulation in the presence of alternative substrates: The activity of iAP towards its substrate AMP was investigated in the presence of different concentrations of injected AMP and alternative substrates ADP and ATP. Changes in Ado product and intermediate generation as well as enzymatic preferences regarding nucleotide substrates ought to be detected and analyzed. Data processing is exemplarily described by means of iAP enzymatic assay with AMP substrate and ADP injections to the online coupled continuous flow system (Figure 4). Evaluation was performed by entering the *m/z* of an assay component of interest into the respective input field (Figure 4, I) to display the corresponding EIC traces of the product (Ado), the substrate (AMP) and the competitive substrate (ADP) (Figure 4A-C). EIC smoothing was conducted with a moving average algorithm, similar to the Savitzky-Golay algorithm (Figure 4, III) [18].

Subsequently, peak recognition was automatically performed with Achroma software by selecting the provided option “negative signals” (Figure 4, II) in case of data showing mainly negative peaks (Figure 4A) and vice versa (Figure 4B and C). If necessary, the software provides the possibility to manually adjust the starting and end point of individual peaks (Figure 4, IV). Alterations of trace intensities of product Ado (Figure 4A), substrate AMP (Figure 4B) and competing substrate ADP (Figure 4C) were further evaluated using the peak areas obtained from Achroma Software (Figure 4D). A peak area average value of four

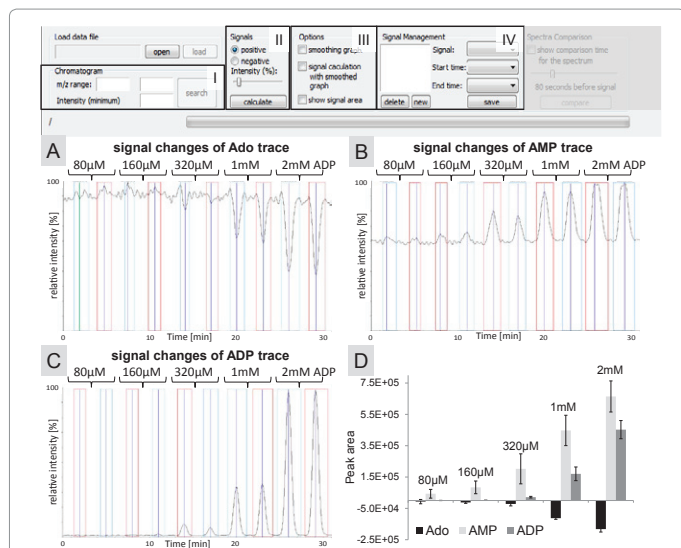


Figure 4: Data evaluation of positive and negative peak signals with Achroma ‘signal recognition’ module is exemplarily demonstrated by iAP assay with AMP substrate and injections of different ADP concentrations. EICs of Ado (A), AMP (B) and ADP (C) are presented. Each ADP concentration was injected twice, resulting in trace responses of all assay components (framed areas). Positive peak areas (AMP, ADP) and negative peak areas (Ado) are furthermore displayed as positive and negative bars with their respective standard deviations out of n=4 (D). The data evaluation interface is divided into functional subdivisions: selection of *m/z* range of interest with specified intensity minimum (I), automatic detection of positive or negative peak signals respectively (II), further options to edit data like smoothing of EICs, signal calculation with smoothed graph and displaying of automatically calculated peak areas (III), manual adjustment of peak starting and end points and possibility to include or exclude non-detected or false detected peaks respectively (IV).

injections of ADP was calculated for Ado, AMP as well as ADP. The decrease of trace intensities is presented as negative bars, whereas the intensity increases are displayed as positive bars (Figure 4D).

Injection of 80 µM and 160 µM ADP solutions to the iAP assay with AMP substrate led to a preferred and complete degradation of ADP in reaction coil 1 (Figure 4C) and in this manner to an associated slight increase of AMP (Figure 4B) and of the final product Ado (Figure

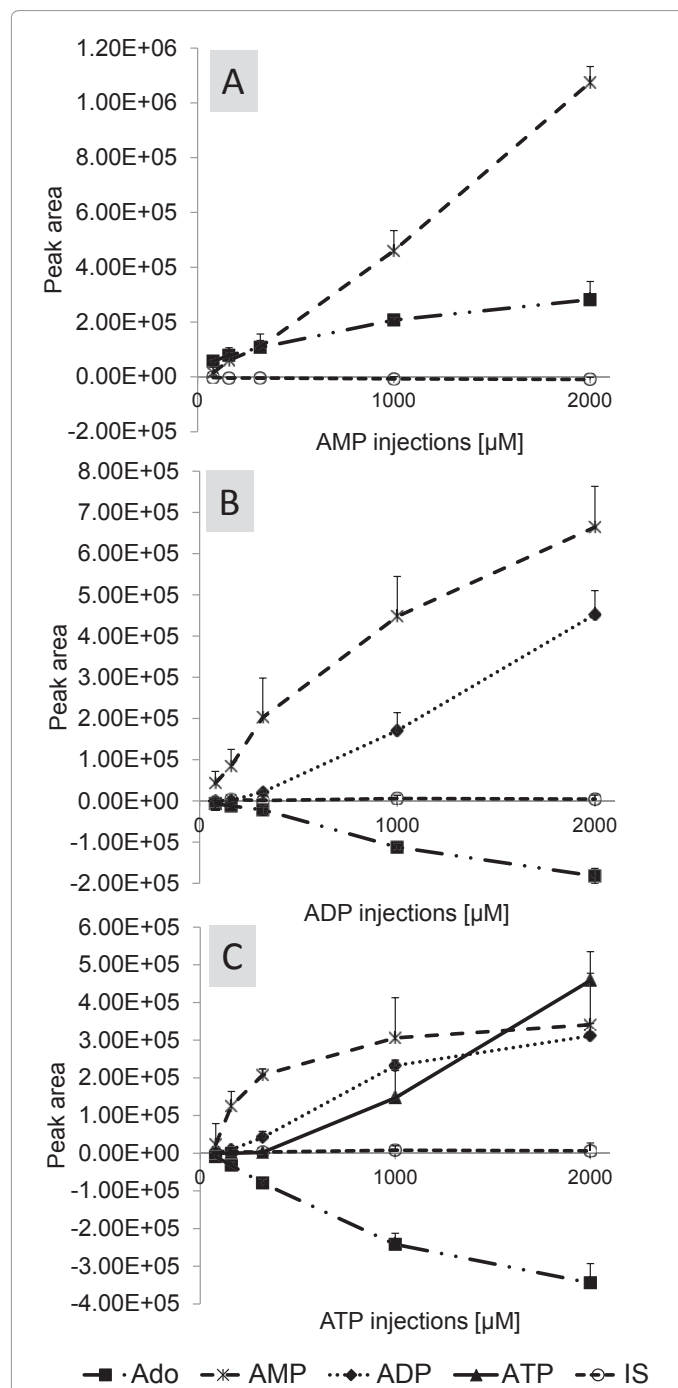


Figure 5: Dependency between calculated peak areas of Ado (■), AMP (*), ADP (◆) and ATP (▲) and the injection of AMP, ADP or ATP in different concentrations (80 µM, 160 µM, 320 µM, 1 mM and 2 mM); internal standard (IS) peak areas are presented as well (○). Standard deviations shown are calculated out of n=4.

4A). ADP concentration above 160 μM caused a more pronounced and gradual increase of AMP intensities and a progressive decrease of the Ado level, which can be explained by the availability of non-degraded, enzymatically favored substrate ADP. High levels of competing ADP therefore resulted in the generation of mainly AMP and a further reduced Ado generation. Detailed data evaluation with AMP, ADP and ATP injections was performed in the same manner as described and is comparatively displayed in Figure 5.

The injection of additional AMP to the system led to a distinct increase of AMP peak areas, which is likely due to the inability of iAP to degrade the majority of AMP injected to the system. The rise of Ado product level was consequently found to be minor and only slightly increased for all AMP concentrations (Figure 5A).

In contrast, data examination with ADP injections revealed a more pronounced increase of AMP intermediate product peak areas (Figure 5B), which reflects the enzymatic substrate preference towards ADP compared to AMP. The simultaneous decrease of Ado furthermore indicates the preferred degradation of ADP to AMP rather than AMP to Ado (Figure 5B). The non-linearity of AMP peak areas with increasing ADP concentrations injected might be accounted to a substrate saturation occurring above a certain level of injected ADP (Figure 5B). This results in an increasing amount of non-degraded ADP and concurrently less generated AMP, due to the systems limited reaction time span (Figure 1, 2nd reaction coil). To verify the thesis of actual substrate saturation and to exclude the possibility of mass spectrometric signal suppression effects, controls containing no enzyme were measured in the online coupled continuous flow system. 80, 160, 320, 1000 and 2000 μM AMP was injected respectively, investigating the peak area progression with increasing concentrations. Data evaluation of AMP peak areas with Achroma showed a linear increase up to 2 mM of injected AMP (data not shown), which in turn supports the assumption of the non-linear AMP increase visible in Figure 5B to be due to substrate saturation.

The injection of ATP to the system furthermore reduces the generation of Ado, which is again due to the enzymes substrate preference towards ATP and ADP. Minor concentrations of injected ATP resulted in a distinct increase of AMP and only a slight rise of ADP levels, which is accounted to the enzyme's ability to rapidly degrade ATP as well as a major proportion of the thereby generated ADP to AMP within the reaction time span provided (Figure 1, 2nd reaction coil) (Figure 5C). ATP injections of 1mM and beyond led to a flattening of AMP, ADP and Ado peak area progression, which is again attributed to substrate saturation.

Within the time ranges of assay component peak calculations, also the peak areas of the internal standard were assessed (Figure 5A-C). The absence of distinct intensity changes in the course of the internal standard EIC furthermore confirms the finding of negative peaks within the EIC of Ado to be an actual alteration in enzyme activity rather than mere signal suppression.

In this manner the signal recognition tool of Achroma software supports a more detailed data interpretation by providing the possibility to calculate positive and negative peak areas as well as regulation effects.

Identification of regulatory molecules and differences in spectra composition and intensities: Injection of alternative substrates to iAP enzymatic assay resulted in the emergence of negative and positive peaks within the assay component EICs. To be able to investigate the spectra composition, which corresponds to the time point of altered substrate degradation, 'spectra comparison' module was exemplarily

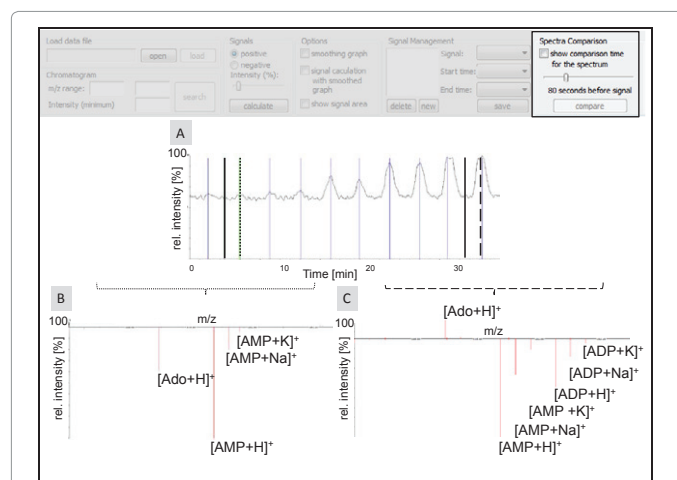


Figure 6: 'Spectra comparison' interface of the Achroma software and data evaluation of signal peaks: Injection of 80 μM (A, left dashed line) and 2 mM ADP (A, right dashed line) with no injection baseline time points (A, dark solid lines); difference spectrum of 80 μM ADP injection (B); difference spectrum of 2 mM ADP injection (C). AMP and ADP appeared as positively charged ions, either with proton, sodium or potassium. Ado was solely detected as protonated ion.

applied for the data evaluation of ADP injections (compare Figure 4 and 5B). The module is convenient in terms of investigating differences between two different spectra or even provides the possibility to identify signal intensity changes, e.g. caused by the injection of a competitive molecule like ADP. To illustrate the procedure of data evaluation with spectra comparison, two independent signals present in the AMP substrate trace were exemplarily selected (Figure 6A, dashed lines, corresponding to the injection of 80 μM ADP and 2 mM ADP). In both cases the time point of the respective assay baseline spectrum is marked as solid black line (manually set to 80 seconds prior to the respective peak maximum; see Figure 6, top right).

A spectra comparison was automatically performed by subtracting the spectrum corresponding to the selected peak maximum (Figure 6A, dashed lines) from the reference baseline spectrum corresponding to the selected minimum (Figure 6A, solid lines). Molecules possessing higher intensities within the selected peak spectrum compared to the reference peak spectrum are therefore displayed as negative signal and vice versa (Figure 6B and 6C) [19]. The injection of 80 μM ADP (Figure 6A, left dashed line) results in an increase of Ado and AMP signal and no detectable ADP signal (Figure 6B). As already discussed and displayed in Figure 4 at low ADP concentrations, the total amount of ADP can be enzymatically degraded to AMP and Ado, therefore causing negative signals for both molecules in the spectra comparison data evaluation.

Injection of 2 mM ADP (Figure 6A, right dashed line) however reveals a decrease of Ado and an increase of AMP as well as the emergence of ADP signals and co-occurring increases of ADP and AMP Na- and K adducts (Figure 6C).

With high ADP concentrations introduced to the system, ADP can be only partially degraded to AMP, resulting in a negative signal of remaining ADP and a distinctively higher negative signal for AMP. The positive Ado signal visible in the spectrum reflects the reduced amount of the final enzymatic product Ado, which is generated in lower quantity due to the presence of the favored ADP substrate.

Acetylcholine esterase assay data evaluation

The identification of regulatory substances from complex mixtures

is an emerging research area, which becomes more and more important within the scope of pharmaceutical, environmental or nutritional issues [1,8,21]. In a current project different house dust extracts were used for the screening of their regulatory influence on enzymatic systems. The investigation of the extract and especially their partially unknown molecular composition led to a large amount of data that had to be evaluated.

Enzymatic activity response and identification of regulatory molecules from complex mixtures: The identification of negative and/or positive EIC peak areas, indicating enzymatic regulation, and simultaneously the evaluation of the related mass spectra detected at time points of regulation, serve as a basis for the finding of enzyme regulatory compounds. In this regard, the effects of house dust extracts were investigated on their influence on AChE activity.

A positive control was included by spiking the house dust extract with the known AChE inhibitor galanthamine. The spiked extract sample was injected to the system and chromatographically separated by HT-HPLC (Figure 1, setup 2, middle trace) before it was introduced to the continuous flow containing AChE (Figure 1, upper trace). Data evaluation with the Achroma 'signal recognition' module revealed two negative peaks, which were automatically identified and tagged in the continuous trace of the assay product choline (Figure 7A, light and dark grey area). The negative peaks may therefore indicate the presence of an AChE regulatory compound contained in the house dust sample. The peak areas in the product trace can be easily calculated with Achroma and may also provide further information about the strength of an enzymatic regulation. After calculating and analyzing the recognized negative signal peaks, software module 'spectra comparison' was applied. By comparing the spectrum underlying a negative peak signal time range with a constant signal time range spectrum (baseline), information about differences between their compositions was obtained (Figure 7B and 7C). Subsequently regulatory extract compounds can be easily identified by means of their

m/z . The spectrum comparison of the negative signal 1 (Figure 7A, 7B and 7D, light grey area) with the constant non-regulated baseline spectrum (Figure 7A, dashed line) revealed no significant characteristic difference of m/z signals. In contrast, the comparison of the negative peak 2 spectrum (Figure 7A and 7C, dark grey area) with the baseline spectrum (Figure 7A, dashed line) showed a distinct difference, which is displayed as negative spectrum signal with a m/z of 288 (Figure 7C). The signal is only apparent in the spectrum of the negative peak 2 area and seems to affect the activity of AChE. To verify this finding, the EIC of m/z 288 was visualized (Figure 7E), hereby revealing a retention time, which is identical with the time point of the negative EIC peak signal within the (choline) product trace. Since the house dust extract was spiked with a known AChE inhibitor, m/z 288 could be assigned to galanthamine (M+H)⁺, therewith clearly showing the applicability of the experimental setup as well as the benefits of Achroma software for a comprehensive data analysis.

Conclusion

Achroma software features were presented in detail, in this manner demonstrating the software to be a powerful and beneficial tool. It supports a fast and wide-ranging interpretation of (a-)typical mass spectrometrically detected enzymatic assay data, e.g. data which are obtained with an online coupled continuous flow system. The three modules 'signal recognition', 'signal comparison' and 'chromatogram comparison' discussed in this work provide several possibilities regarding data evaluation. Negative and positive signal peak areas within the trace of a substrate, intermediate and/or product EIC could be assessed for iAP enzymatic assay with 'signal recognition' module. Detailed examination of the recorded mass spectrometric data thus revealed the enzyme's substrate preference towards ADP and ATP in comparison to AMP. Besides the injection of alternative substrates, negative and positive peaks are often related to events of enzymatic regulation, e.g. caused by an inhibitor. In this context AChE activity was investigated in the presence of chromatographically separated house dust extracts, spiked with a known AChE inhibitor as positive control. An enzymatic assay response related to the inhibitor added could be detected and evaluated with Achroma modules 'signal recognition' and 'spectra comparison'. The latter provided the comparative assessment of mass spectrometric spectra, detected at different time points to identify the compound responsible for AChE inhibition. Besides the finding of newly emerging molecule signals, e.g. like enzyme inhibiting or activating compounds from complex mixtures, also intensity changes could be easily identified.

The third module, 'chromatogram comparison', provides an easy means to control and correct mass spectrometric data for signal loss or signal inconsistencies. It was therefore employed to assess signal inconsistencies within iAP enzymatic assay traces in comparison to the course of an internal standard also taken along. The calculation of a quotient and in this manner the correction of assay EICs revealed the nature of signal loss events as related to mass spectrometric detection rather than enzymatic regulation.

Achroma software was shown to be efficient for the analysis of mass spectrometric data, whose evaluation has been too complex, time consuming or even impossible heretofore. However, further adaptations are necessary to fully exploit the improvement potential. For this purpose the functionality of Achroma software will be transferred into an open source modular analytical software platform (openMASP) to revise and expand the modules explained in this work. By switching to the open source platform the implemented code of Achroma will

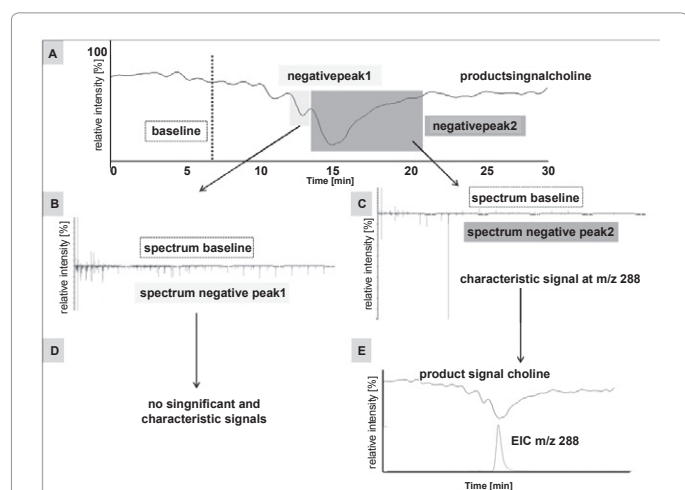


Figure 7: Screening of a house dust extract on its regulatory influence on AChE activity. Detection of negative signals and calculation of signal areas present in the course of the product trace was performed with the 'signal recognition' software tool (A). Comparison of the product EIC baseline spectrum with the spectra corresponding to the negative peak areas was done with 'spectra comparison' tool (B,C). No significant difference of baseline spectrum to negative peak 1 spectrum was detected (B,D). A potentially enzyme regulatory substance was identified, reflected by a characteristic difference between baseline spectrum and negative area spectrum (m/z 288, C). EIC of regulatory substance revealed the concurrence of the compound's retention time and the time point of regulation (D).

be free, open and downloadable for users, in this manner enabling the participation of external software developers/analysts in improving and adjusting the software functions or even adding new modules to the analytical software platform. Until then every potential user can download the Achroma software for free and may use it as helpful as it is. Further information regarding software and hardware requirements can be found in a previous publication [19] or on the Achroma web page [18].

Acknowledgment

The authors would like to thank for financial support the German Federal Ministry of Economics and technology within the agenda for the promotion of industrial cooperative research and development (IGF) on the basis of a decision by the German Bundestag. The access was opened by the member organization DECHEMA and organized by the AiF, Arbeitsgemeinschaft industrieller Forschungsvereinigungen, Cologne, Germany (IGF Project No. 450 ZN). We also thank Knauer Wissenschaftliche Gerätebau GmbH (Berlin, Germany) for the Quadrupole mass spectrometer as a loan and the Institut für Energie- und Umwelttechnik e.V., Duisburg (especially Dr. Thorsten Teutenberg) for the house dust samples.

References

- Grassmann J, Scheerle RK, Letzel T (2012) Functional proteomics: application of mass spectrometry to the study of enzymology in complex mixtures. *Anal Bioanal Chem* 402: 625-645.
- Greis KD (2007) Mass spectrometry for enzyme assays and inhibitor screening: an emerging application in pharmaceutical research. *Mass Spectrom Rev* 26: 324-339.
- Letzel T (2008) Real-time mass spectrometry in enzymology. *Anal Bioanal Chem* 390: 257-261.
- Ruotolo BT, Verbeck GF, Thomson LM, Gillig KJ, Russell DH (2002) Observation of conserved solution-phase secondary structure in gas-phase tryptic peptides. *J Am Chem Soc* 124: 4214-4215.
- Dennhart N, Letzel T (2006) Mass spectrometric real-time monitoring of enzymatic glycosidic hydrolysis, enzymatic inhibition and enzyme complexes. *Anal Bioanal Chem* 386: 689-698.
- Kaltashov IA, Fenselau C (1997) Stability of secondary structural elements in a solvent-free environment: the alpha helix. *Proteins* 27: 165-170.
- Hogenboom AC, de Boer AR, Derks RJ, Irth H (2001) Continuous-flow, on-line monitoring of biospecific interactions using electrospray mass spectrometry. *Anal Chem* 73: 3816-3823.
- Kool J, Giera M, Irth H, Niessen WM (2011) Advances in mass spectrometry-based post-column bioaffinity profiling of mixtures. *Anal Bioanal Chem* 399: 2655-2668.
- de Jong CF, Derks RJE, Bruyneel B, Niessen W, Irth H (2006) High-performance liquid chromatography-mass spectrometry-based acetylcholinesterase assay for the screening of inhibitors in natural extracts. *J Chromatogr A* 1112: 303-310.
- Schenk T, Breeel GJ, Koevoets P, van den Berg S, Hogenboom AC, et al. (2003) Screening of natural products extracts for the presence of phosphodiesterase inhibitors using liquid chromatography coupled online to parallel biochemical detection and chemical characterization. *J Biomol Screen* 8: 421-429.
- Kool J, Eggink M, van Rossum H, van Liempd SM, van Elswijk DA, et al. (2007) Online biochemical detection of glutathione-S-transferase P1-specific inhibitors in complex mixtures. *J Biomol Screen* 12: 396-405.
- Colangelo CM, Chung L, Bruce C, Cheung KH (2013) Review of software tools for design and analysis of large scale MRM proteomic datasets. *Methods* 61: 287-298.
- MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, et al. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26: 966-968.
- Brusniak MY, Kwok ST, Christiansen M, Campbell D, Reiter L, et al. (2011) ATAQS: A computational software tool for high throughput transition optimization and validation for selected reaction monitoring mass spectrometry. *BMC bioinformatics* 12: 78.
- Kaufmann CM, Grassmann J, Treutter D, Letzel T (2014) Utilization of real-time electrospray ionization mass spectrometry to gain further insight into the course of nucleotide degradation by intestinal alkaline phosphatase. *Rapid Commun Mass Spectrom* 28: 869-878.
- Schebb NH, Vielhaber T, Jousset A, Karst U (2009) Development of a liquid chromatography-based screening methodology for proteolytic enzyme activity. *J Chromatogr A* 1216: 4407-4415.
- Scheerle RK, Grassmann J, Letzel T (2012) Real-time ESI-MS of enzymatic conversion: impact of organic solvents and multiplexing. *Anal Sci* 28: 607-612.
- Krappmann (2012) open MASP.
- Krappmann M, Letzel T (2012) Achroma: a software strategy for analysing (a-) typical mass-spectrometric data. *Analytical Methods* 4: 1060-1071.
- Moss DW, Walli AK (1969) Intermediates in the hydrolysis of ATP by human alkaline phosphatase. *Biochim Biophys Acta* 191: 476-477.
- Shi SY, Zhou HH, Zhang YP, Jiang XY, Chen XQ, et al. (2009) Coupling HPLC to on-line, post-column (bio)chemical assays for high-resolution screening of bioactive compounds from complex mixtures. *TrAC-Trends Anal Chem* 28: 865-877.

Citation: Krappmann M, Kaufmann CM, Scheerle RK, Grassmann J, Letzel T (2014) Achroma Software-High-Quality Policy in (a-)Typical Mass Spectrometric Data Handling and Applied Functional Proteomics. *J Proteomics Bioinform* 7: 264-271. doi:10.4172/jpb.1000328

Submit your next manuscript and get advantages of OMICS Group submissions

Unique features:

- User friendly/feasible website-translation of your paper to 50 world's leading languages
- Audio Version of published paper
- Digital articles to share and explore

Special features:

- 350 Open Access Journals
- 30,000 editorial team
- 21 days rapid review process
- Quality and quick editorial, review and publication processing
- Indexing at PubMed (partial), Scopus, EBSCO, Index Copernicus and Google Scholar etc
- Sharing Option: Social Networking Enabled
- Authors, Reviewers and Editors rewarded with online Scientific Credits
- Better discount for your subsequent articles

Submit your manuscript at: <http://www.editorialmanager.com/jmbt>

3.3 Mass Spectrometric Real-Time Monitoring of an Enzymatic Phosphorylation Assay using Internal Standards and Data Handling Freeware

Michael Krappmann, Arjen R. de Boer, Daniël R.W. Kool, Hubertus Irth and Thomas Letzel, submitted to *Rapid Communications in Mass Spectrometry*

In dieser Studie wurde für die Versuchsdurchführung ein zuvor publiziertes (massenspektrometrisch kompatibles) Enzymsystem zur Peptid-Phosphorylierung mittels cAMP-abhängiger Protein Kinase A (PKA) von de Boer et al. ^[124] verwendet.

Neben der Bestimmung von Enzymaktivität bei verschiedenen Temperaturen (278 K-328 K) und verschiedenen Reaktionszeiten (15, 300 und 600 Sekunden) in der Reaktionskapillare sollte mit dem Versuchsaufbau gezeigt werden, dass eine Quantifizierungsstrategie unter Verwendung von Achroma erfolgreich ist. Der Versuchsansatz wurde kontinuierlich in das Massenspektrometer geleitet und die Produktbildung bei steigender Temperatur überprüft. Gleichzeitig wurde mit diesen Experimenten die Stabilität des Messsystems bei Anstieg der Temperatur kontrolliert.

Als Substrat kamen die phosphorylierbaren Peptide Malantide und Kempide zum Einsatz. Beide Substrate konnten jeweils in einzelnen Ansätzen mit PKA (vom Rind und Schwein) in einer Reaktionskapillare interagieren, während alle 10 Minuten die Temperatur, beginnend bei 278 K und endend bei 328 K, um 10 K-Intervalle erhöht wurde (siehe auch Kapitel 2.2.1). Die Versuche wurden dabei jeweils mit und ohne Inhibitorzugabe durchgeführt, um die Abhängigkeit der Aktivität des Enzyms von der Temperatur und Inhibitorkonzentration zu bestimmen. Gleichzeitig sollten die Auswirkungen weiterer variierender äußerer Bedingungen - wie Reaktionszeit- auf die Stabilität des Messsystems untersucht werden.

Systemstabilität bei sich ändernder Temperatur:

Es konnte gezeigt werden, dass die Abweichung von der idealisierten Linie des Quotientenchromatogramms der internen Standards zwischen 5% und 15% betrug, sodass von einem thermisch stabilem Messsystem ausgegangen werden konnte.

Einfluss der Reaktionstemperatur auf die Enzymaktivität:

Aufgrund des kontinuierlichen Betriebs entstand eine kontinuierliche Produktpur bei der Phosphorylierung des Substrats, die bei konstanter Temperatur eine Parallele zur Zeitachse darstellte, da die Intensität über diesen Zeitraum gleich blieb. Die stufenweise Temperaturerhöhung zeigte jeweils einen stufenweisen Anstieg in der Intensität während des

Erwärmungsprozesses. Über den Zeitraum gleichbleibender Temperaturen konnte eine gleichbleibende Intensität detektiert werden. Ein Intensitätseinbruch resultierend aus der vermutlichen Denaturierung des Enzyms wurde bei 328 K beobachtet.

Die Herausforderung der Auswertung bestand darin, dass keine gaußförmigen Signale entstanden. Somit konnten bereits bekannte Quantifizierungssoftwares nicht herangezogen werden. Es wurde speziell für das Achroma Modul *Signal recognition* eine zusätzliche Option zur Berechnung von kontinuierlichen Flächen implementiert. Mit Hilfe dieser Funktion konnte, aufgrund der bekannten Zeitintervalle, Start- und Endpunkt markiert und die Fläche unterhalb der Intensitätslinie berechnet werden. Die so gewonnenen Daten wurden in Microsoft Excel überführt, um eine Fehlerberechnung durchzuführen.

Das beschriebene Verfahren wurde für die Flächen von Produkt und Substrat, sowie für die beiden internen Standards durchgeführt. Aufgrund der geringen Wiederfindungsraten der eingesetzten Standards, konnte zwar ein System zur Quantifizierung untypischer Signale (keine Gauß-Verteilung) etabliert, aber noch nicht abschließend mit einer erfolgreichen Quantifizierung belegt werden.

Die Flächenberechnung der internen Standards bei den Zeitintervallen der Temperaturstufen hingegen zeigte zusätzlich zu der bereits visuell dargestellten Stabilität, dass das System stabil funktioniert. Es wurde ein relativer Fehler von maximal 6% ermittelt.

Die Temperaturabhängigkeit der Enzymaktivität und deren Inhibition konnte mit dem eingesetzten Versuchsaufbau in Verbindung mit Achroma ausgewertet werden.

Die Implementierung der Auswertung von kontinuierlichen Signalfächen des *Signal recognition* Moduls von Achroma wurde von mir federführend konzipiert und programmiert. Die automatisierte Datenauswertung mit Achroma wurde von mir durchgeführt, ebenso die weiterführende Auswertung der kontinuierlichen Signalfächen. Das Manuskript wurde federführend von mir erstellt. Die experimentellen Versuche wurden durch die Co-Autoren ausgeführt. Die Co-Autoren lieferten auch einzelne Beiträge zum Manuskript. Priv.-Doz. Dr. Letzel betreute die gesamte Arbeit, korrigierte das Manuskript und reichte es ein.

Mass Spectrometric Real-Time Monitoring of an Enzymatic Phosphorylation Assay using Internal Standards and Data Handling Freeware

Michael Krappmann¹, Arjen R. de Boer^{o3}, Daniël R.W. Kool⁺³, Hubertus Irth³ and Thomas Letzel^{*2,3}

¹*Zentrum für Forschung und Weiterbildung, Weihenstephan-Triesdorf – University of Applied Science, Am Staudengarten 9, D-85354 Freising-Weihenstephan, Germany.*

²*Analytical Research Group, Chair of Urban Water Systems Engineering, Technische Universität München, Am Coulombwall 8, D-85748 Garching, Germany.*

³*Faculty of Sciences, Division of Chemistry, Department of Analytical Chemistry and Applied Spectroscopy, Free University of Amsterdam, De Boelelaan 1083, NL-1081 HV Amsterdam, The Netherlands.*

* Corresponding author. Fax: +49-89-2891 3780

e-mail address: T.Letzel@ tum.de

+ Present address: Waters Chromatography B.V., 4870 AJ Ettenleur, The Netherlands

° Present address: Centraal Bureau voor de Statistiek, 2490 Den Haag/Heerlen, The Netherlands

Abstract

An application previously developed for the mass spectrometric detection of peptide phosphorylation by cAMP-dependent protein kinase A (PKA) [De Boer *et al.*, *Anal. Bioanal. Chem.* **2005**, 381, 647-655], was transmitted to a continuous-flow reaction detection system. This enzymatic reaction, involving the enzyme activation as well as the transfer of a phosphate group from ATP to a peptide substrate, was used to prove the compatibility of a quantitative enzymatic assay in a continuous-flow real-time system (connected to MS).

Therefore, two independent internal standards (IS) were implemented, on the one hand to monitor the online system stability and on the other hand to obtain quantitative values of conversion substrate and product as well as high mass spectrometric accuracy. One internal standard passed the complete reaction line similar to the substrate and a second internal standard was continuously added to the reaction flow prior entering the mass spectrometric interface. A ratio of both internal standards permits to distinguish signal intensity changes caused by different enzymatic activities, fluctuations in the pumping system or ion source contaminations. The data workout could be observed with the newly available freeware platform Achroma.

Moreover, (using the internal standards) critical parameter, such as reaction temperature (with the solution density variations depending on temperature), were studied in the continuous-flow mixing system. Furthermore, two substrates (malantide and kemptide), and two enzyme types (catalytic subunit of PKA and complete PKA) are tested to determine system robustness and long-term availability. Even spraying solutions -that contain significant amount of mass spectrometric contaminants (e.g. the polluted catalytic subunit)- resulted in significant MS signal intensities. Subsequent recalculations using the internal standards lead to results representing the power of internal standards. The on-line monitoring system was also applied to study inhibitor influence on the enzyme activity performing a competitive assay in continuous flow experiments.

Concluding, the present methodology and the data evaluation with the Achroma platform enables the direct coupling of different biochemical assays with mass spectrometric detection. Monitoring changes as temperature, reaction time, inhibition, or compound concentrations can be observed quantitatively and thus enzymatic activity can be calculated.

INTRODUCTION

The finalization of the Human Genome Project has led to the identification of several thousands of potential new drug targets, most of them being proteins. Among these targets, protein kinases play a key role since they are involved in a wide range of physiological processes. Kinases catalyse the phosphorylation of proteins and the importance of this process is reflected by the fact that the human genome contains as many as 2000 protein kinases. The screening of protein kinases in order to find suitable inhibitors for both in-vitro/in-vivo studies and lead discovery represents a major challenge, and preferably, generic screening formats are still required to minimize assay development times^{1,2,3}.

In the past, a multitude of screening assays for protein kinases have been developed⁴, including ³²P-radioisotope labeling⁵⁻¹¹, fluorescence based assays¹², affinity assays against phosphorylated¹³ or labelled phosphorylated¹⁴ peptides and proteins. Since a decade, mass spectrometry has become an important tool in the investigation of enzyme kinetics and inhibition. Various approaches¹⁵ have been chosen to detect conversion products of enzyme-substrate reactions, either directly^{16,17}, after gel separation and digestion^{18,19} or after liquid chromatographic separation²⁰⁻²⁵.

In the case of protein kinases, the detection of phosphorylated peptides offers the possibility to develop a more or less generic assay format since both native and phosphorylated peptides can be detected with excellent sensitivity by mass spectrometry.

Most approaches described above, as well as earlier published enzymatic kinetic²⁶ and inhibition²⁷⁻²⁹ studies are based on off-line (batch) incubation. However, several authors have described continuous-flow^{23,24} approaches for enzyme assays, like on-line digestion by immobilized trypsin³⁰⁻³² or other immobilized enzyme reactors, like protein tyrosine phosphatase³³, glucose oxidase³⁴, and alkaline phosphatase³⁵⁻³⁷. As utilized in this study, similar on-line reaction detection systems were used in earlier publications to monitor acetylcholinesterase inhibition³⁸⁻⁴⁰, enzymatic hydrolysis⁴¹⁻⁴³, solvolysis⁴³, disulfide reduction⁴³ and protein unfolding⁴⁴⁻⁴⁶. However, the data obtained could not be calculated to quantitative values, due to the lack of using internal standards. That qualitative data were not comparable with results after system modifications or with results obtained from other research groups. Currently, the mass spectrometric determination of kinetic data is performed with different variants of internal standards (IS)^{26-28,47-50}. However, these quantification methods were of limited usefulness for on-line monitoring setups.

To address these problems, a study was published describing a continuous-flow biochemical assay format in which two internal standards were applied for the first time. Both standards were pumped through the total mixing system monitoring fluctuations that occur in the

pumping parts²³. However, if the pumping parts were not the source of problems, the search for other possible reasons was difficult using that IS arrangement.

In the present paper an optimized internal standard approach is described performing the peptide phosphorylation by protein kinase A²⁸ using the real-time monitoring of electrospray mass spectrometry and a special modular software platform (Achroma)^{40,51,52}. The new constellation of the two IS allows to distinguish changing conditions resulted from both, the reaction and the detection part.

EXPERIMENTAL SECTION

Chemicals and Reagents.

Protein kinase A (E.C. 2.7.1.37; from bovine heart; 80 % protein; phosphorylation activity: 1.1 units per μg protein), protein kinase A catalytic subunits (from porcine heart; phosphorylation activity: 30-65 units per μg protein), malantide (H-Arg-Thr-Lys-Arg-Ser-Gly-Ser-Val-Tyr-Glu-Pro-Leu-Lys-Ile-OH; >97 %; MW 1633.9; Mal), kemptide (H-Leu-Arg-Arg-Ala-Ser-Leu-Gly-OH; >97 %; MW 771.9; Kem), adenosine 3':5'-cyclic monophosphate (free acid; >99 %; cAMP), adenosine 5'-triphosphate (magnesium salt; 95-98 %; MgATP), and the protein kinase inhibitor from rabbit phosphorylation PKA inhibitor (5-24) amide (*P* 0300, H-Thr-Thr-Tyr-Ala-Asp-Phe-Ile-Ala-Ser-Gly-Arg-Thr-Gly-Arg-Arg-Asn-Ala-Ile-His-Asp-OH; MW 2222.4; Inh) were obtained from Sigma (Zwijndrecht, The Netherlands). The internal standards ovokinin (H-Phe-Arg-Ala-Asp-His-Pro-Phe-Leu-OH; >96 %; MW 1002.2; IS1) and PP60 c-src (phosphorylated) (H-Thr-Ser-Thr-Glu-Pro-Gln-Tyr(PO₃H₂)-Gln-Pro-Gly-Glu-Asn-Leu-OH; >99 %; MW 1543.5; IS2) were supplied by Bachem (Bubendorf, Switzerland). Methanol (99.8 %; MeOH) and ammonium bicarbonate (>99 %) were purchased from J.-T. Baker (Deventer, The Netherlands), acetic acid (>99.8 %) from Riedel-de Haën (Seelze, Germany), and magnesium sulfate (>99 %; MgSO₄) from Acros Organics (Geel, Belgium). High purity water was taken from a Milli-Q water system (Millipore, Eschborn, Germany).

Salt Test (Batch) Experiments.

1 μM malantide and 10 μM MgATP were solved for comparison measurements in 50 mM phosphate buffer (pH 7.0) containing 10 mM MgSO₄, 10 mM dithiothreitol, and 0.5 mg BSA as described in the original study used for radio assay conditions⁵.

In further samples, different volatile salts were developed for mass spectrometric detection. Thus 1 μM malantide and 10 μM MgATP were solved in water, containing 100 μM MgSO₄ and 20 mM ammonium bicarbonate, ammonium acetate or ammonium formate, respectively.

For better ionization properties of the analytes and good comparison of experiments, 0.1 % acetic acid and 10 % MeOH was added before infusion via a so called make –up flow. Several samples were measured by continuous infusion into the mass spectrometer at a flow rate of $20 \mu\text{L min}^{-1}$ and the mass spectrometric signal intensity was obtained by area integration of one minute.

Further assay optimization procedures are described in [De Boer *et al.*, Anal. Bioanal. Chem. **2005**, 381, 647-655]²⁸ and optimal solvent conditions are used as specified therein.

Reaction System for the On-line Monitoring of Enzymatic Reactions.

The on-line monitoring system is shown in *Figure 1*. Reagent delivery was performed using 10 mL SuperLoopsTM (Pharmacia Biotech AB, Uppsala, Sweden). The input flow of the SuperLoopsTM was generated using Shicoh Engineering (Yamato, Japan) HPLC pumps connected to the inlet of SuperLoopsTM by Peek tubing of 0.5 m length and 0.064 mm inner diameter. Both SuperLoopsTM were operated at a flow rate of $10.0 \mu\text{L min}^{-1}$. The outlet of the SuperLoopsTM were connected to a reaction detection system comprised of inverted Y-type mixing unions (inhouse) and knitted poly(tetrafluoroethylene) reaction coils with 5 μL , 100 μL , or 200 μL internal volume, respectively (ID 0.3 mm, length 0.070, 1.4, and 2.8 m, respectively). Prior to infusion into the mass spectrometer, a make-up solution containing MeOH with 1 % acetic acid was added via an inverted Y-type mixing union at a flow rate of $2.0 \mu\text{L min}^{-1}$. The make-up solution was pumped by a syringe pump (Harvard Apparatus 22, Harvard, MA, USA) via a syringe (Hamilton, Reno, NV, USA). The total flow of $22 \mu\text{L min}^{-1}$ was directed via an electrospray interface into a Q-TOF 2 mass spectrometer (Micromass, Manchester, UK).

In a typical set-up one SuperLoopTM was filled with 2 mL aqueous solution containing 20 mM ammonium bicarbonate, 100 μM MgSO_4 , 50 nM cAMP, 50 μM MgATP, and protein kinase A (224-448 units). The other SuperLoopTM was filled with 2 mL aqueous solution containing 20 mM ammonium bicarbonate, 100 μM MgSO_4 , 2 μM peptide (malantide or kemptide) and 1 μM IS1. The syringe for the make-up flow was filled with MeOH containing 1 % acetic acid and 10 μM IS2. For inhibition measurements additionally 2 μM of inhibitor²³ was added to the substrate solution in the SuperLoopTM containing the peptide and IS1.

Electrospray Ionization-Mass Spectrometry Conditions.

The parameters for mass spectrometric detection were optimized with malantide and IS2 using a Q-TOF 2 mass spectrometer coupled to an electrospray interface. The measurements were performed in the positive ionization mode with 353 K source temperature, 423 K desolvation temperature, 250 L h⁻¹ desolvation gas flow, 100 L h⁻¹ cone gas flow, 16 psi gas cell pressure, 2500 V capillary voltage, 33 V cone voltage. The mass-range was set to 300 - 1100 m/z and data acquisition parameters were 5.0 sec/scan, 0.1 sec dwell time, full TOF MS continuous scan mode, and without using the option 'MS profile'. Nitrogen (purity 5.0; Praxair) and argon (purity 5.0; Praxair) were used as desolvation / cone gas and collision gas, respectively.

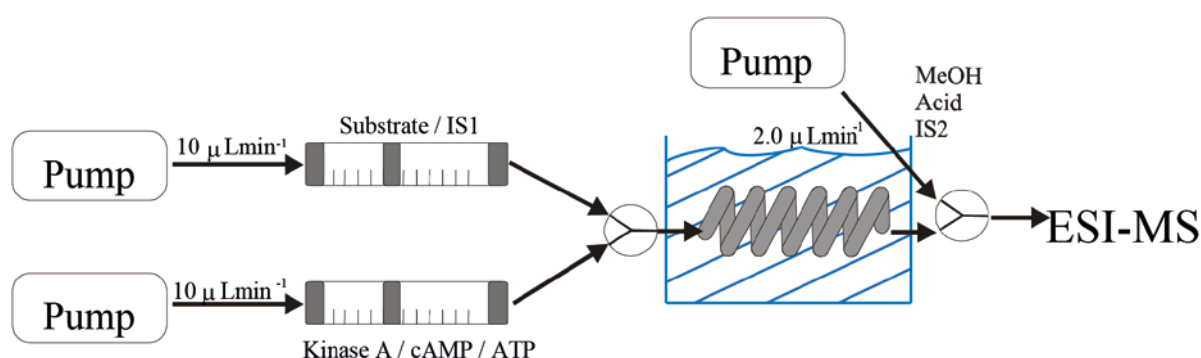


Figure 1: Scheme of the continuous-flow enzymatic reaction system with pumps, solvent container (SuperLoop™), reaction coil, water bath for temperature regulation, make-up flow line, and online electrospray mass spectrometric detection by a Q-TOF mass spectrometer.

Data Evaluation.

Continuous flow enzymatic assays deliver data that cannot be evaluated with standard software due to the continuously observed MS intensity instead of chromatographic signals. Figure 2 sketches the work flow of the data evaluation. The freeware Achroma can realize this data workout using the steps 'chromatogram comparison', 'signal recognition' and 'signal area calculation'. However, Achroma can read text based formats like ASCII. Therefore the raw data had to be converted using a data conversion step. The mass spectrometry software from Waters delivers a DataBridge-Software for raw data conversion into an ASCII file format. Subsequently data conversion, the specialized modules from Achroma⁵¹ can be applied for the data evaluation.

Chromatogram comparison module

This module compares the intensities of two preferred chromatograms regarding the time to a given m/z ratio and shows the result in a new quotient chromatogram⁵¹. For example, the m/z

ratios of IS1 and IS2 were compared to monitor the stability of the experimental set up. Thereby the masses were set to m/z 501-502 (IS1) and m/z 772-773 (IS2), respectively. After pressing the search button for each chromatogram, both chromatograms were shown in the display window. Automatically a quotient chromatogram IS1/IS2 was calculated and additionally presented. The signal intensity deviation has to be chosen to a given time point and is marked in the quotient chromatogram. A deviation of $\pm 10\%$ of the intensity graph was chosen from the intensity value at 40 min in all experiments (for the inhibition experiments with substrate kemptide the deviation was set to 15%).

Signal recognition module

This module has the option of switching between signal area evaluation of peaks and continuous flow experiments. Normally a chromatogram has chromatographic peaks and vendor software typically calculates the signal area below this peaks. In continuous flow experiments a chromatogram has continuously changing intensities without chromatographic peaks if parameters change, e.g. temperature in defined time intervals. The signal area of a typical chromatographic peak is calculated as the integral between start and end point of a signal and subsequently subtracted by the integral of the base line between start and end point of the signal⁵¹. The calculation of continuous flow experiments (without signals) follows an easier process. In this case the integral of the changing base line is calculated. For example, the intensity of the time plot will increase with every increasing temperature step and may show constant intensities during the stable periods.

The continuous flow chromatograms in the temperature dependent experiments were evaluated for the m/z ratio of the reaction products (p-Mal, p-Kem), the m/z ratio of the substrates (malantide, kemptide), and the m/z ratio of IS1 and IS2 using the signal recognition module. The time settings of temperature changes were performed manually in the Achroma signal recognition module (see signal management section). The areas below the intensity graph were calculated at each temperature.

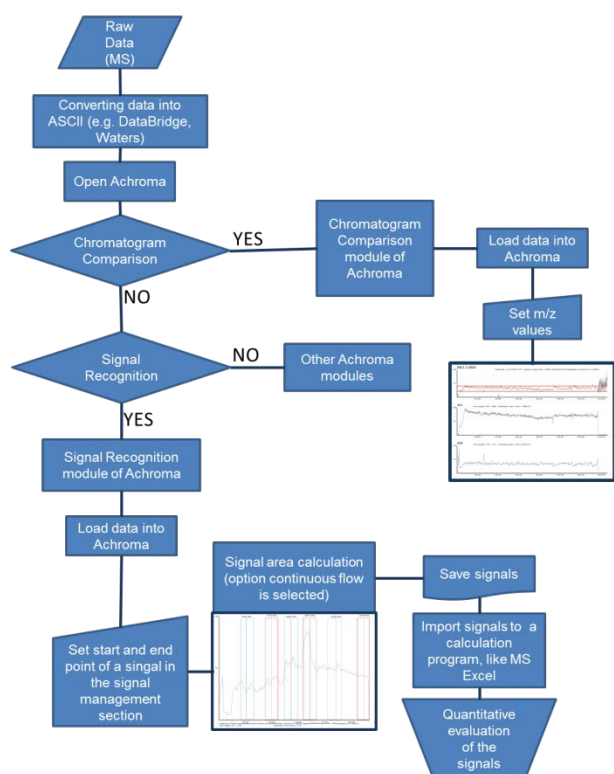


Figure 2: Work flow of the applied data evaluation with Achroma software.

Data preparation with spreadsheet software

The initial assay temperature was set to 278 K and every 10 minutes the temperature was increased by 10 K. The intensities between start and end scan number at each temperature was inserted into a Microsoft Excel table and the ratio was always set to 12 scans/minute, i.e. 5 seconds per scan. So the exact time of a temperature interval could be calculated by dividing: $(\text{end count} - \text{start count}) / 12$. The result is a prepared table with columns for the signal areas of substrate(s), product(s) and Internal Standard(s) and rows for the temperature intervals. In a next step Achroma software can be started with the 'signal recognition module' and the signal areas can be calculated.

So every temperature interval was set manually in the 'signal management' section (start and end time for each temperature), like the prepared Excel table above, and the signal area of each temperature step was calculated automatically with Achroma. The list of the signal areas were stored in a text file and imported into the Excel table. This step was done with all substrates, conversion products, IS1 and IS2.

System stability using two internal standards (with spreadsheet software)

The stability of the experimental system can be presented with the 'Chromatogram comparison module' of Achroma - see above. Another way for presenting the system stability is a standard deviation of both internal standards. It should be constant under 10% over all

temperature steps. All signal areas of the IS1 and IS2 were imported into the Excel table form Achroma. The next step is, calculating the mean value of the signal areas for each temperature step of IS1 and IS2. Afterwards the standard deviation from the mean value of every internal standard has to be calculated (calculation of errors). A stable system may have an error below +/- 10% from the mean value of the used internal standards.

Quantifying conversion products (with spreadsheet software)

IS1 is a similar molecule like the substrate and IS2 is similar to the phosphorylated product. Since the experimental procedure is performed with known recoveries of the conversion product, a quantification calculation is possible.

With the known recovery rates of IS1 and IS2 and the calculated signal areas of substrate (similar to IS1) and the corresponding conversion product (similar to IS2) the 100% recovery and therefore the 100% signal area of both may be calculated. In a continuous experiment the substrate is introduced into the system in a constant rate (for 2 μM malantide and kemptide, respectively). Is no enzyme in the system the 100% signal area must be 2 μM for the substrate, if no losses are assumed. The introduction of enzyme causes a phosphorylation reaction and the concentration of the conversion product could be calculated via the recovery rate of IS2 and the signal area of the conversion product, caused by the precondition, that IS2 is comparable to the conversion product. So the mass balance of calculated substrate concentration and calculated conversion product concentration has nearly to be 2 μM .

RESULTS AND DISCUSSION

Comparison Experiments and Assay Setup.

Biochemical studies of phosphorylation reactions with cAMP-dependent kinases (published in biological research) are mainly performed in aqueous solutions containing millimolar salt and buffer concentrations⁵⁻¹¹ (e.g. substrate in phosphate buffer). As could be seen earlier⁵², the molecules of interest cannot be detected using mass spectrometric detection if phosphate buffer and additives are used in mM-range. In general this may have following reasons:

- a) cluster forming with neutral salt molecules and an occurring mass shift of the analyte,
- b) the neutralisation of a positive charged analyte ion by negative salt molecules,
- c) signal suppression by non-volatile salts in the nebulized droplets due to changed drying conditions, and/or

d) mass spectrometric inlet contamination by sprayed non-volatile salts.

Volatile salts like ammonium bicarbonate, ammonium acetate or ammonium formate are applied performing mass spectrometric compatible enzymatic reactions to avoid ion and/or signal suppression in mass spectrometric detection. A typical assay represents a spectrum containing all essential compounds (excluding the enzyme) in aqueous ammonium bicarbonate solution clearly showing the mass spectrometric signals of interest⁵². However measurements in ammonium acetate and formate gave similar mass spectrometric signal intensities thus they all can be used for quantitative analysis. Spectra prior enzymatic reaction (including the enzyme) and after 45 min reaction time were published earlier²⁸ and are identical to spectra obtained in this study (data not shown).

In the earlier published study we optimized an enzymatic assay with PKA in a systematic way leading to an active biological phosphorylation assay compatible with sensitive mass spectrometric detection and clearly showing the use of 'non physiological solutions' for enzymatic kinetic studies²⁸. The most compatible conditions obtained there could be transferred to a continuous flow system directly coupled with MS. With that 'time and labour saving' online technique, various experiments speed up receiving information by MS. However, in this study ammonium bicarbonate was applied as volatile salt proving the use of a slightly stable salt system (easy pH change by carbon dioxide degassing) as reaction system salt (at pH 7.0 with high MS response as published previously)²⁸.

The enzyme reaction proceeds in a closed, open-tubular reactor composed of a knitted reaction coil. The reactants in this study, i.e., the enzyme, substrate and co-factors, are added via a pumping system that is optimized for low dead-volumes and pulsation (see Experimental). The flow rates of the reactant pumps and the volume of the reaction coil and the mass spectrometric interface determine the reaction time for the enzyme assay. Prior to infusion into the electrospray interface, a make-up solution (methanol and acetic acid) is added in order to improve the detection of substrate and/or product. It should be noted that the addition of methanol and acetic acid terminates the enzyme reaction, i.e. the reaction stopped before entering the mass spectrometric interface.²⁸

Some mass spectrometric assumptions have to be made for this study.

First, the phosphorylation was proven, by a fragmentation of the protonated phosphorylated malantide (p-Mal; $[M+3H]^{3+}$ and $[M+2H]^{2+}$) vs. protonated phosphorylated kemptide (p-Kem; $[M+2H]^{2+}$) in collision-induced dissociation (CID) experiments. In each experiment the specific difference of 98 Da was monitored, the typical neutral loss of H_3PO_4 ⁵³⁻⁵⁷ (data not shown). In further online experiments the phosphorylated molecules were detected without fragmentation.

A second assumption is that protonated and Na-adduct ions although with significant recovery have to be neglected in this study. Thus resulting in an easier calculation and a 'similar' behaviour in each experiment using Achroma modules. However, besides protonated molecules, especially the following Na-adduct ions could be detected in experiments: m/z 531.2 ($[\text{ATP}+\text{Na}]^+$), m/z 553.1 ($[\text{Mal}+\text{Na}+2\text{H}]^{3+}$) and m/z 579.2 ($[\text{p-Mal}+\text{Na}+2\text{H}]^{3+}$ in Fig. 2). Note that, in this study these adducts as well as the doubly charged protonated molecules of malantide and phosphorylated malantide ($[\text{M}+2\text{H}]^{2+}$) will not be considered. But someone has to keep in mind that these traces subsequently have to be taken into account.

Since the enzymatic phosphorylation reaction was developed systematically to be compatible with MS detection²⁸, further experiments were performed in a continuous flow mixing system without further sample enrichment prior to ESI-MS analysis as mentioned there.

In a first set of experiments, various enzymatic reaction times were studied for comparison of the conversion rate with batch experiments and check the effective transfer to an online system. The total reaction time of the enzymatic assay in this study was controlled by reactors with different volumes, i.e., 5, 100 and 200 μL . At a total reaction flow rate of 20 $\mu\text{L min}^{-1}$ the corresponding reaction times were 15, 300 and 600 seconds. Figure 3 presents mass spectra for kinase assays performed at the three different reaction times. The mass range representing the triply charged protonated molecule of phosphorylated malantide is highlighted (m/z 571.9 $[\text{p-Mal}+3\text{H}]^{3+}$). As can be expected, the maximum response of the phosphorylated malantide was obtained at the longest reaction time. At the reaction times of 15 and 300 seconds, approximately 18% and 70%, respectively, of the total intensity of phosphorylated malantide measured at 600 seconds were detected. These contents were obtained by comparison of the signal intensities with the ion at m/z 567.9 in Fig.3 (similar to a one-point normalization factor)^{26,28}.

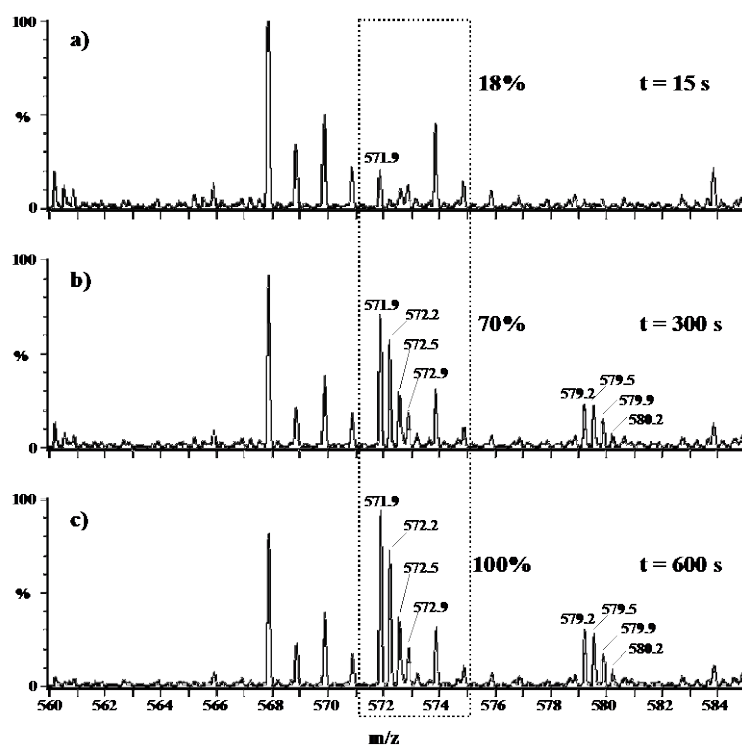


Figure 3: Non-corrected MS spectra of the triply charged phosphorylated malantide at reaction times of a) 15 sec, b) 300 sec, and c) 600 sec with mass spectrometric intensity factors [%]; zoomed into the mass range m/z 560-585 (total range m/z 300 – 1100). The mass spectrometric intensity factor mirrors the mass spectrometric response of m/z 571.9 in relation to m/z 567.9 and the relation at 600 sec reaction time as 100 % content.

Complementary Internal Standards (IS).

The need of an one-point normalization factor²⁸ in mass spectrometry and the other studies performed with internal standards^{24,26,27,47-50} illustrates an information lack using continuous flow experiments. A new approach with two independent internal standards will fill this lack for quantitative measurements. Changes caused by the enzymatic reaction, the reaction system and the mass spectrometric detection can be distinguishing with this IS constellation. In the following the conditions and needs of the two IS are described in general and in special:

IS 1. This standard is flowing through the total pumping system, is very similar to the original substrate, but is inert to enzymatic conversion. The reservoir filled with substrate solution (in Fig. 1, contains ovokinin as IS1. Ovokinin contains no amino acid-residue (serine, tyrosine, and threonine) that possibly can be phosphorylated by protein kinase A. On the other hand, the hydrophobicity of ovokinin is comparable with malantide and kemptide; thus similar physico-chemical behaviour is expected.

For this study, IS1 was used in the on-line system at a final concentration of 500 nM. A flow calibration was performed for IS1 in the range of 5 - 1000 nM (Table 1). As can be seen, the IS1 is more sensitive in concentration-dependent signal intensities compared to the substrate

malantide and kemptide. Thus further corrections of mass spectrometric signal intensities for malantide and kemptide will be sensitive enough.

IS 2 A second internal standard (PP-60 (phos.), a phosphorylated thirteen residue peptide) was introduced with the make-up flow through an inverted Y-type mixing union positioned directly in front of the mass spectrometer. Due to the contained phosphate group the standard is similar to the reaction products and simultaneously a sensitive control of the mass spectrometric performance. If the molecule shows fragmentation with the loss of phosphoric acid the detection was chosen inadequate. Due to the lack of a serine phosphorylated standard (at the project time) a commercial available tyrosine phosphorylated standard was used.

IS2 was applied in the on-line system at a final concentration of 1.0 μM and calibration was performed in the range of 0.010 - 2.0 μM (Table 1).

The mass spectrometric calibration method was identical with the solution and detection conditions in the on-line experiments. Table 1 presents the calibration equation, the R-squared value and the between-day repeatability for each standard and analyte. A between-day repeatability standard deviation of about 20 % was observed for both standards calibration measurements in fifteen days using the same spray capillary before starting the enzymatic experiments. The standard deviation of the ratio IS1 / IS2 over the same period was 6 % (data not shown), which corresponds with a change of the spray efficiency affected both internal standards similarly (see below). The calculated average ratio of IS1 / IS2 (out of the calibration equations in Table 1) was 1.74, compared to the average ratio of 1.78 for the standards directly observed from the calibration solutions (data not shown). Thus the calculated ratio fits well with the direct observed ratio.

Table 1: Studied compounds (and abbreviations) with calibration equation (x = concentration [μM] and y = MS signal intensity [A.U.]), R-squared value and other calibration parameter.

Compound	MS signal [m/z] of $[\text{M}+\alpha\text{H}]^{\alpha+}$	Calibration equation	R^2 (n=5)	Calibration range [μM]	day repeatability [dev.]; (period)
Ovokinin (IS1)	501.6 ; $\alpha = 2$	$y = 8031.6x + 101.3$	0.9978	0.005 - 1.0	22 % (15days)
PP60-phos. (IS2)	772.2 ; $\alpha = 2$	$y = 2396.9x - 36.2$	0.9982	0.01 - 2.0	20 % (15days)
Malantide (Mal)	545.6 ; $\alpha = 3$	$y = 3331.9x - 69.9$	0.9997	0.05 - 1.0	16 % (7days)
Kemptide (Kem)	386.6 ; $\alpha = 2$	$y = 4388.1x - 26.2$	0.9962	0.05 - 1.0	25 % (7days)

Changes in the enzyme reaction section and in the mass spectrometric part (interface and mass spectrometer) can be distinguished by calculating the IS1 / IS2 ratio. Subsequently, some propositions are made about potential trends of this internal standard ratio if the enzymatic assay is not changed in reactivity, but MS intensity changes occur. The topics represent not a complete set of possibilities, but demonstrate clearly how to use the ratio tendency:

a) Compound A ↓ and IS1 ↓ and IS1/IS2 const. => MS inlet contamination, experiment is o.k.

If the MS signal intensity of the detected analyte A is changing during unchanged reaction conditions and IS1 shows the same trend while the ratio of IS1 / IS2 remains constant, it can be stated that the spray conditions are changed. Due to an absolute change in intensity, but a stable ratio of the internal standards, the ionization or detection was affected for all compounds in the same way. Normally, the interface spray and the inlet conditions may change by time due to contamination, whereas the mass spectrometric properties do not change, as long as parameter settings are unaltered.

b) Compound A ↓ and IS1 ↓ and IS1/IS2 ↓or↑ => Mixing system causes problems, experiment is not o.k.

If the MS response of the detected analyte is changing during unchanged reaction conditions, if the IS1 shows the same trend and if the ratio of IS1 / IS2 decreases or increases, it can be stated that the conditions in the reaction part are changed. If the standard ratio decreases, it is possible that less amount of the reaction mixture is flowing through the reaction coil (due to a lack in the system) or an increased adsorption of molecules takes place in tubing. If the standard ratio increases, it is possible that the non-specific binding of compounds to the tubing is reduced (e.g., at decreased pH-values).

c) Compound A ↓ and IS1 const. and IS1/IS2 const. => Compound stability causes problems, experiment is not o.k.

If the intensity of the analyte MS response is decreasing during unchanged reaction conditions, if the IS1 as well as the ratio of IS1 / IS2 are still stable, it can be stated that the analyte is degraded during storage in the container.

Concluding the combinations above it can be stated, just if the MS response of IS1 shows the same trend as the analyte response at unchanged reaction conditions and the ratio of IS1 / IS2 remains stable (like in a)), the accuracy for quantitative measurements is given. In the other cases the experiments have to be repeated as a consequence of irreproducible results.

The temperature experiments with different enzymatic assays show acceptable results with the used internal standard ratios (Figure 4 A-E). The chromatogram comparison with Achroma and the resulting deviation of quotient chromatograms were in all experiments between 5% (Figure 4 C) and 15% (Figure 4 E). The experiment using malantide as substrate, PKA (bov.) as enzyme and inhibitor introduction shows intensity fluctuation of IS1 after 60 minutes (Figure 4 B). This causes system problems which disallow quantitative calculations after this time point. For all experiments the deviation (random error) of the mean value of the signal areas of IS1 and IS2 was between 0% and 6% (data not shown).

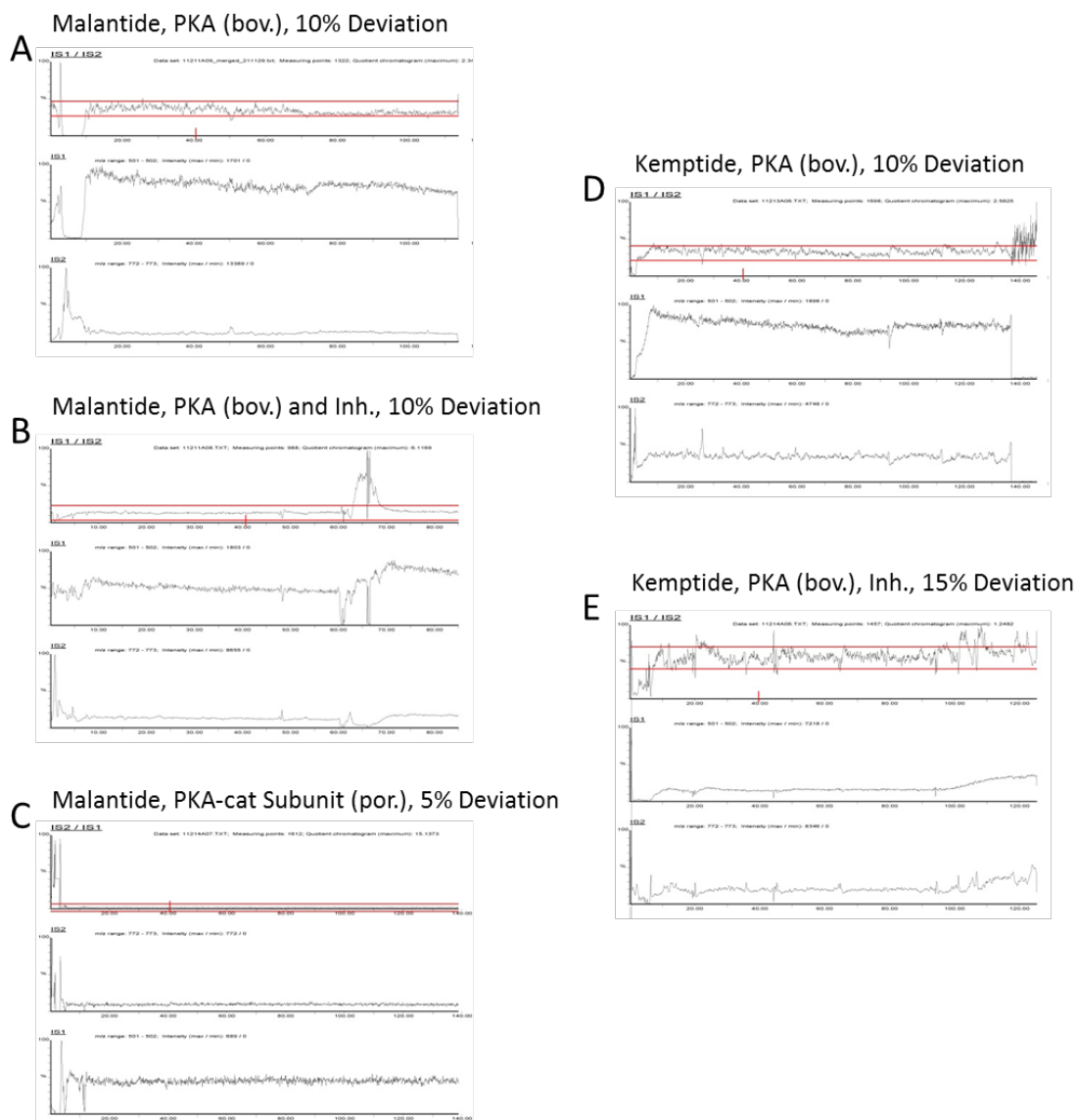


Figure 4 A-E: Mass traces of the quotient chromatogram of IS1 and IS2 (upper traces), chromatogram of IS1 (middle traces) and IS2 (lower traces) and the deviations (red lines). **A + B)** Mass trace ratio of IS1 / IS2 (m/z 501.6 ([IS 1+2H]²⁺) / and m/z 772.2 ([IS 2+2H]²⁺)) with an deviation interval of +/- 10% (red line) for the temperature dependent monitoring of phosphorylation conversion of the peptide malantide by protein kinase A (bov.) and with inhibition (B); **C)** Monitoring of phosphorylation conversion of the peptide malantide by protein kinase A cat subunit (por.), Deviation +/- 5%; **D + E)** Monitoring of phosphorylation conversion of the peptide kemptide by protein kinase A (bov.), showing additional the mass trace of m/z 426.7 ([phos. Kem+3H]²⁺) and m/z 501.6 ([IS 1+2H]²⁺) and in presence of a strong inhibitor (E).

Application Experiments (Reaction Temperature).

The temperature experiments were carried out with enzyme PKA from bovine and porcine, malantide and kemptide as substrates and PKA inhibitor for inhibition of the enzyme substrate reaction. Therefore five experiments were performed: PKA (bov.)/malantide in absence and presence of an inhibitor, PKA (bov.)/kemptide in absence and presence of an inhibitor and PKA (porc.)/malantide. The reaction coil was placed in a water bath, see figure 1. The starting temperature for every experiment was 278 K and every 10 minutes the temperature was increased by 10 K. The experiments were stopped at 328 K. The solution conditions were:

- aqueous enzyme solution: PKA (bov./porc.), 100 μM MgSO_4 , 50 μM MgATP , 50 nM cAMP
- aqueous substrate solution: 2 μM malantide or kemptide, 1 μM Ovokin (IS1), 20 mM Ammoniumbicarbonat, 100 μM MgSO_4
- for inhibition experiments additional 2 μM PKA-inhibitor (P 0300)
- IS2: MeOH mit 1% Essigsäure, 10 μM IS2

The mass spectrometric response of internal standards in on-line experiments was decreased differently by non-specific adsorption to tubing and enzyme⁵⁸ as well as signal suppression effects⁵⁹ as can be seen in Figure 4. The mass spectrometric responses shown in Table 2 for each standard in each run (details below) were obtained by averaging ten times integrating one minute in each temperature section. The standard deviation for the standards is shown in brackets subsequent to the averaged response.

Due to the higher recovery of IS1 compared to IS2 in all on-line experiments, the resulting ratio of IS1 / IS2 was significant higher compared to the ratio of the unaffected standards (1.74 and 1.78). The real ratio in the different temperature periods was between 1.84 and 2.81 for runs with protein kinase A. The resulting ratio in experiments with the protein kinase A catalytic subunit was significantly higher (4.00 - 4.39) but stable. The ratio was generally stable for all experiments during the total measurement period of longest 120 min (see average deviations in Table 2). The standard deviations of the mass spectrometric responses were for all runs below 10 %, in two cases even below 5 %. However, these deviations permit a quantitative treatment of the obtained signal responses of peptide substrate and conversion product.

The recovery for each internal standard (Table 2; column 'Rec.') was calculated, calculating the ratio of mass spectrometric response obtained in a run (Table 2; Columns 'IS1' and 'IS2') with the theoretical mass spectrometric response obtained from the calibration equation in Table 1 (e.g., 4117 counts for 0.50 μM IS1 and 2361 counts for 1.0 μM IS2). The resulting

recoveries were in this study between 34 – 46 % for IS1 and 26 – 34 % for IS2 in measurements with protein kinase A (Table 2a and 2d) and 27 – 31 % for IS1 and 22 – 27 % for IS2 with protein kinase A in presence of inhibitor (Table 2b and 2e). In the on-line run with protein kinase A catalytic subunit (Table 2c) the recoveries were 4 % and 2 % for IS1 and IS2, respectively. The recoveries for both standards were very low due to the high contamination potential of subunit additives which were also leading to a fast blockage of the mass spectrometric inlet.

The higher recovery of IS1 in all experiments leads to the assumption that binding of the internal standards to tubing seems in this study to be negligible compared to the mass spectrometric signal suppression effects.

Table 2: Mass spectrometric intensities for the two internal standards with standard deviations and recoveries in online reaction detection experiments for the peptides malantide (a-c) and kemptide (d,e) with protein kinase A (a,b,d,e) and protein kinase A catalytic subunit (c), in presence (b,e) or in absence of a strong inhibitor (a,c,d). The ratio of both internal standards is listed in each last column and the time stability (see deviations) in each last row).

a) Malantide with PKA (bov.)

Temp. (300 s)	IS 1 Int. [A.U.]	Rec.	IS 2 Int. [A.U.]	Rec.	IS 1 / IS 2
278 K	1671 (3%)	41%	759 (7%)	32%	2.20 (6%)
288 K	1601 (4%)	39%	752 (8%)	32%	2.13 (8%)
298 K	1400 (4%)	34%	742 (7%)	31%	1.89 (7%)
308 K	1485 (5%)	36%	808 (6%)	34%	1.84 (4%)
318 K	1391 (2%)	34%	682 (6%)	29%	2.04 (7%)
328 K	1400 (2%)	34%	667 (3%)	28%	2.10 (3%)
Average (dev.)	1491 (8%)	36%	735 (7%)	31%	2.03(7%)

d) Kemptide with PKA (bov.)

Temp. (300 s)	IS 1 Int. [A.U.]	Rec.	IS 2 Int. [A.U.]	Rec.	IS 1 / IS 2
278 K	1906 (1%)	46%	705 (6%)	30%	2.70 (6%)
288 K	1770 (2%)	43%	631 (7%)	27%	2.81 (7%)
298 K	1619 (2%)	39%	639 (7%)	27%	2.53 (8%)
308 K	1446 (2%)	35%	664 (6%)	28%	2.18 (5%)
318 K	1619 (2%)	39%	664 (9%)	28%	2.44 (12%)
328 K	1619 (5%)	39%	622 (10%)	26%	2.60 (9%)
Average (dev.)	1663 (9%)	40%	654 (5%)	28%	2.54 (9%)

b) Malantide with PKA (bov.) and Inh

Temp. (300 s)	IS 1 Int. [A.U.]	Rec.	IS 2 Int. [A.U.]	Rec.	IS 1 / IS 2
298 K	1094 (3%)	27%	519 (4%)	22%	2.11 (4%)
308 K	1144 (4%)	28%	556 (4%)	24%	2.06 (6%)
318 K	1250 (3%)	30%	590 (5%)	25%	2.12 (4%)
Average (dev.)	1163 (7%)	28%	555 (6%)	24%	2.09 (2 %)

e) Kemptide with PKA (bov.) and Inh

Temp. (300 s)	IS 1 Int. [A.U.]	Rec.	IS 2 Int. [A.U.]	Rec.	IS 1 / IS 2
298 K	1283 (1%)	31%	595 (3%)	25%	2.16 (3%)
308 K	1215 (6%)	30%	648 (7%)	27%	1.88 (9%)
318 K	1277 (6%)	31%	574 (13%)	24%	2.23 (16%)
Average (dev.)	1258 (3%)	31%	606 (6%)	26%	2.09 (9%)

c) Malantide with PKA cat. subunit (por.)

Temp. (300 s)	IS 1 Int. [A.U.]	Rec.	IS 2 Int. [A.U.]	Rec.	IS 1 / IS 2
278 K	164 (11%)	4%	41 (10%)	2%	4.03 (15%)
288 K	165 (5%)	4%	41 (8%)	2%	4.00 (9%)
298 K	178 (4%)	4%	41 (7%)	2%	4.39 (8%)
308 K	175 (8%)	4%	43 (6%)	2%	4.12 (11%)
318 K	162 (6%)	4%	40 (6%)	2%	4.04 (11%)
328 K	168 (3%)	4%	39 (6%)	2%	4.33 (7%)
Average (dev.)	169 (4%)	4%	41 (3%)	2%	4.15 (4%)

The obtained and calculated amounts of non-converted peptides malantide and kemptide, respectively, are listed in Table 3. Assuming equal relative losses of the analyte and IS1 during the flow process, IS1 was used to correct the remaining substrate (malantide and kemptide) to 100% recovery. Therefore the mass spectrometric response of substrate obtained (e.g., 1487 A.U. for malantide at 278 K Table 3a) was divided by the mass spectrometric recovery of IS1 (e.g., 0.41 for IS1 at 278 K in Table 2a). The calculated 100% recovery signal response of non reacted substrate (e.g., 3664 A.U. for malantide at 278 K in Table 3a) was transformed with its calibration equation (Table 1) into concentrations (e.g., 1.12 μM for malantide at 278 K in Table 3a). The calibration for malantide and kemptide was performed with a series of standard solutions (Table 1).

Table 3: Mass spectrometric intensities and corrected data (details see text) for the non converted substrates malantide (a-c) and kemptide (d,e) after reaction with protein kinase A (a,b,d,e) and protein kinase A catalytic subunit (c), in presence (b,e) or in absence of a strong inhibitor (a,c,d). The values in brackets represent the standard deviations of ten times one minute integrations for each temperature

a) Malantide with PKA (bov.)**d) Kemptide with PKA (bov.)**

Temp. (300 s)	Mal Int. [A.U.]	Mal(cor) Int. [A.U.]	Mal(cor) [μM]	Cor & norm	Temp. (300 s)	Kem Int. [A.U.]	Kem(cor) Int. [A.U.]	Kem(cor) [μM]	Cor & norm
278 K	1487 (4%)	3664 (3%)	1.12	1.00	278 K	2758 (2%)	5959 (3%)	1.37	1.00
288 K	1386 (3%)	3579 (5%)	1.10	0.98	288 K	2533 (4%)	5871 (4%)	1.35	0.99
298 K	1161 (5%)	3380 (3%)	1.04	0.92	298 K	2128 (2%)	5407 (3%)	1.24	0.91
308 K	1069 (5%)	2916 (4%)	0.89	0.80	308 K	1665 (2%)	4742 (2%)	1.09	0.80
318 K	899 (3%)	2658 (4%)	0.81	0.73	318 K	1869 (3%)	4738 (2%)	1.09	0.80
328 K	1055 (2%)	3094 (2%)	0.95	0.84	328 K	2029 (1%)	5204 (4%)	1.19	0.87

b) Malantide with PKA (bov.) and Inh

Temp. (300 s)	Mal Int. [A.U.]	Mal(cor) Int. [A.U.]	Mal(cor) [μ M]	Cor & norm
298 K	769 (3%)	2923 (3%)	0.90	1.00
308 K	806 (4%)	2902 (4%)	0.89	0.99
318 K	772 (5%)	2554 (4%)	0.78	0.87

e) Kemptide with PKA (bov.) and Inh

Temp. (300 s)	Kem Int. [A.U.]	Kem(cor) Int. [A.U.]	Kem(cor) [μ M]	Cor & norm
298 K	1844 (2%)	5917 (2%)	1.36	1.00
308 K	1716 (5%)	5725 (4%)	1.31	0.97
318 K	1670 (6%)	5263 (3%)	1.21	0.89

c) Malantide with PKA cat. subunit (por.)

Temp. (300 s)	Mal Int. [A.U.]	Mal(cor) Int. [A.U.]	Mal(cor) [μ M]	Cor & norm
278 K	148 (12%)	3748 (18%)	1.15	1.00
288 K	134 (6%)	3363 (6%)	1.03	0.90
298 K	78 (12%)	1827 (13%)	0.56	0.49
308 K	58 (9%)	1406 (9%)	0.43	0.38
318 K	118 (9%)	3027 (10%)	0.93	0.81
328 K	201 (10%)	4930 (8%)	1.51	1.32

To overcome this overcharges and for the better comparison of all on-line reactions with several kinases and substrates a normalization factor was required. The substrate concentration corrected at each temperature was divided by the substrate concentration corrected at 278 K (for measurements with inhibitor at 298 K) to obtain data normalized to 278 K (or 298 K).

The mass spectrometric responses for the products could not be recalculated and normalized due to a lack of phosphorylated malantide and phosphorylated kemptide as reference materials for calibration. Further on these results are handled in absolute or corrected mass signal responses.

The evaluation with Achroma shows, that all signal areas could be calculated: malantide, kemptide, phosphorylated malantide and kemptide, and the areas of IS1 and IS2 for all experiments (data not shown). If the internal standards were chosen like described above, further calculations, e.g. quantification could be done with the recovery rates of IS1 and IS2. The recovery rate of substrate and conversion product is not important for calculation, only the signal areas.

The influence of the reaction temperature on the turnover of the phosphorylation reaction has been investigated to prove the usability of the internal standard ratio for quantitative measurements even in changing reaction conditions.

As a first application, the phosphorylation of kemptide by protein kinase A was performed by using the 100 μL reaction coil at a reaction flow rate of 20 $\mu\text{L min}^{-1}$, and varied temperature (from 278 K to 328 K by 10 K intervals). *Figure 5A* shows the mass trace of phosphorylated kemptide (m/z 426.7 [p-Kem+2H] $^{2+}$) and the ratio of the two internal standards (IS1 / IS2) for phosphorylation measurements in absence (*Fig. 4D*) and in presence (*Fig. 4E*) of a strong protein kinase inhibitor.

An increase of phosphorylated kemptide signal response was observed in absence of an inhibitor by raising temperature with a maximum conversion rate at 318 K. At further temperature increase, the amount of phosphorylated kemptide decreased indicating a deactivation of the enzyme.

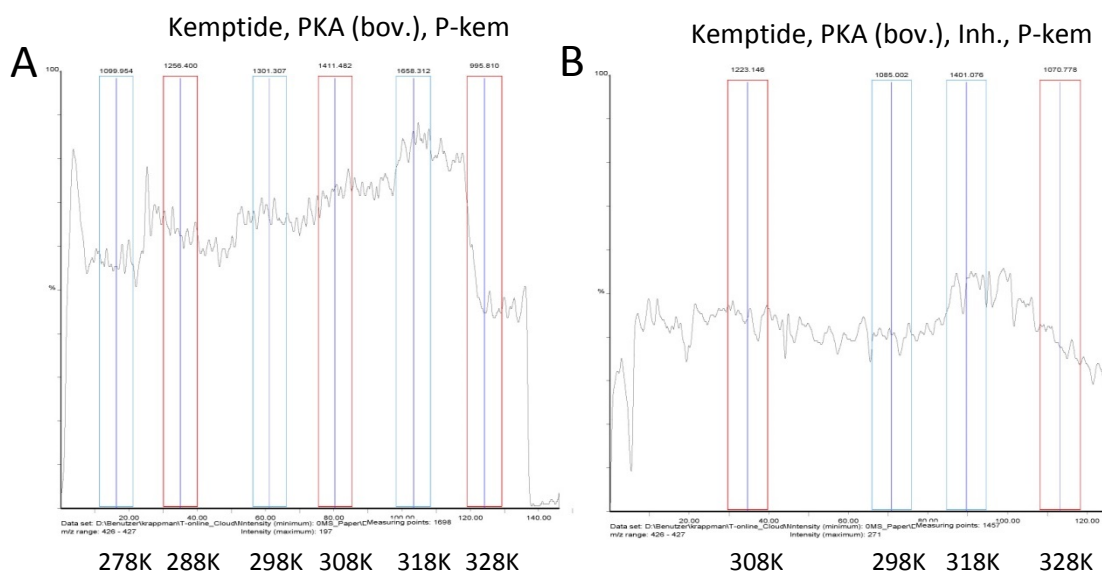


Figure 5: A) Monitoring of phosphorylation conversion of the peptide kemptide by protein kinase A (bov.) and the dependence of the reaction temperature; B) in presence of a strong inhibitor.

The signals of IS1 and IS2 exhibited only minor variations during the entire measurement sequence and the calculated ratio IS1 / IS2 remained stable (*Fig. 4D*) indicating virtually constant operation conditions and ionization behaviour of the ESI interface. The drop of all mass traces indicates clearly the importance of acidic methanol for a sensitive detection. Please note, as described in the last chapter, only sections with stable IS1 / IS2 ratio can be used for the quantification procedure. However, for all sections stable plateaus could be obtained and for each temperature ten times one-minute integrations could be averaged. The

visual impression of the stable ratio in Figure 4D and 4E was confirmed by the data presented in Table 2. The intensity increase of phosphorylated kemptide as a function of temperature is therefore only based on an increase of the substrate conversion activity of the kinase.

The phosphorylation behaviour was additionally proven by blocking the kinase with a suitable inhibitor repeating the temperature dependency experiment. The effect of inhibitors on the kinase activity was investigated by adding a kinase inhibitor with a K_i value of 2.3 nM to the substrate solution at a final concentration of 1 μ M. The presence of this inhibitor leads to a significant decrease of the phosphorylated kemptide production (Figure 5B; trace m/z 426.7 [p-Kem+2H]²⁺). Again, the ratio IS1 / IS2 is constant, clearly proving that ESI-MS can be successfully applied for monitoring kinase reactions and their inhibition.

As a next application, the phosphorylation of the peptide malantide by protein kinase A was performed by using the 100 μ L reaction coil at a reaction flow rate of 20 μ Lmin⁻¹, and at varied temperature (from 278 K to 328 K by 10 K intervals). In this experiment malantide exhibited a comparable phosphorylation behaviour as kemptide and its phosphorylation can significant be blocked by the strong inhibitor. The phosphorylation rate of the substrate malantide was at all temperatures similar compared to the rate of the substrate kemptide with a conversion maximum at 318 K. The quantitative treatment for the temperature dependency for malantide as substrate in the absence and presence of 1 μ M inhibitor is presented in Table 3a and 3b, respectively.

As a third application, the phosphorylation of malantide by protein kinase A catalytic subunit from porcine⁵⁹ was performed by using the 100 μ L reaction coil at a reaction flow rate of 20 μ L min⁻¹, and varied temperature (from 278 K to 328 K by 10 K intervals). Compared to the protein kinase A from bovine similar conditions were chosen, however, no cAMP had to be added. Ion suppression was more pronounced using the catalytic subunit when compared with the intact enzyme (see Recoveries in Table 2c). The results in Table 3c reveal that the maximum phosphorylation efficiency for the catalytic subunit was at 308 K and above the phosphorylation is decreasing due to denaturation of the enzyme subunit. All three measurements confirm the statement to achieve quantitative enzymatic conversion data with a stable ratio of the applied internal standards. As described in the quantification theory section, the recoveries of IS1 and IS2 and the signal areas of the substrate (kemptide, malantide) and the conversion products (P-Kem, P-Mal) got from the freeware Achroma could be used for quantification.

Conclusions

The development of mass spectrometric assays was addressed for the on-line determination of enzyme kinetics and inhibition. The results have shown that electrospray MS can be used to monitor enzyme-substrate reactions by detecting specific enzyme reaction products. Unlike fluorescence- and radioactivity based enzyme assay, no labelled substrate is required for this purpose. The present method may find application both in the study of enzyme kinetics and in high-throughput screening of, for example, combinatorial chemistry libraries or natural products.

This kind of continuous-flow mixing system can easily be modified to the FIA-continuous-flow mixing system as published previously for a protease reaction. Therefore an automated injector has to be implemented in the substrate flow (Fig. 1; upper pumping line between Super-loop and mixing Y-piece) and regulating substances can be identifier via injections.

Correct adjustment of assay conditions is critical for a successful on-line coupling of enzyme assays to mass spectrometry. Special attention was paid to the optimization of buffer conditions which have to be compatible with the enzyme substrate reaction and mass spectrometry. Clearly, the present method can only be applied without alteration to those enzymes that exhibit activity in buffer systems containing volatile salts such as ammonium acetate, formate or bicarbonate. If MS compatible buffer conditions can be establish the approach presented allows a rapid assay development, consisting essentially of determining the reaction products that can be detected at the highest sensitivity and selectivity by electrospray MS. The use of two internal standards is an essential tool to control the stability of the on-line system and efficiently detect changes both in the enzymatic reaction and electrospray interface. Monitoring a ratio of internal standards permits to measure quantitative data of enzymatic reactions and inhibitions even for long time periods, like presented for the protein kinase A. The new constellation opens the possibility to perform on-line reaction detection systems quantitative. Additionally, Achroma gives the possibility of evaluating continuous flow 'signals' and therefore a simple method for quantification of the conversion products, if the internal standards chemically similar to the substrate (for IS1) and/or the conversion product (IS2).

Software Availability and Computer Requirements

Project name: Achroma

Download Address (latest 01.06.2015): <http://openmasp.hswt.de/pages/downloads.php>

Operating system: Microsoft Windows XP or higher

Programming language: Microsoft Visual C++

Programming IDE: Microsoft Visual Studio 2003 or higher

Minimum software requirements: 1 GB RAM, 1 GHz processor Acknowledgement

Acknowledgements

T.L. thanks Günter Mayer and Marion Lackhoff for very helpful biochemical information and discussions.

References

- (1) Johnson DA, Akamine P, Radzio-Andzelm E, Madhusudan, Taylor SS. *Chem. Rev.* 2001; **101**: 2243.
- (2) Adams JA. *Chem. Rev.* 2001; **101**: 2271.
- (3) Scheerle RK, Graßmann J, Letzel T, *Anal. Methods* 2011, **3**: 822-830.
- (4) Yan JX, Packer NH, Gooley AA, Williams KL. *J. Chromatogr. A* 1998; **808**: 23.
- (5) Rosen OM, Erlichman J. *J. Biol. Chem.* 1975; **250**: 7788.
- (6) Hofmann F, Beavo JA, Bechtel PJ, Krebs EG. *J. Biol. Chem.* 1975; **250**: 7795.
- (7) Usui H, Inoue R, Tanabe O, Nishito Y, Shimizu M, Hayashi H, Kagamiyama H, Takeda M. *FEBS Letters* 1998; **430**: 312.
- (8) Manganello JM, Djellas Y, Borg C, Antonakis K, Breton GCL. *J. Biol. Chem.* 1999; **274**: 28003.
- (9) Pei L. *J. Biol. Chem.* 2000; **275**: 1176.
- (10) Anderson AE, Adams JP, Qian Y, Cook RG, Pfaffinger PJ, Sweatt JD. *J. Biol. Chem.* 2000; **275**: 5337.
- (11) Tokuo H, Yunoue S, Feng L, Kimoto M, Tsuji H, Ono T, Saya H, Araki N. *FEBS Letters* 2001; **494**: 48.
- (12) Negash S, Yao Q, Sun H, Li J, Bigelow DJ, Squier TC, *Biochem. J.* 2000; **351**: 195.
- (13) Cao P, Stults JT. *J. Chromatogr. A* 1999; **853**: 225.
- (14) Goshe MB, Conrads TP, Panisko EA, Angell NH, Veenstra TD, Smith RD. *Anal. Chem.* 2001; **73**: 2578.
- (15) Grassmann J, Scheerle RK, Letzel T, *Anal. Bioanal. Chem.* 2012, **402**: 625.

- (16) Wilm M, Neubauer G, Mann M. *Anal. Chem.* 1996; **68**: 527.
- (17) Asara JM, Allison J. *J. Am. Soc. Mass Spectrom.* 1999; **10**: 35.
- (18) Neubauer G, Mann M. *Anal. Chem.* 1999; **71**: 235.
- (19) Steen H, Küster B, Fernandez M, Pandey A, Mann M. *Anal. Chem.* 2001; **73**: 1440.
- (20) Huddleston MJ, Annan RS, Bean MF, Carr SA. *J. Am. Soc. Mass Spectrom.* 1993; **4**: 710.
- (21) Jedrzejewski PT, Lehmann WD. *Anal. Chem.* 1997; **69**: 294.
- (22) Letzel T, *Anal. Bioanal. Chem.* 2008, **390**(1): 257.
- (23) de Boer A, Letzel T, Lingeman H, Irth H. *Anal. Chem.* 2004; **76** (11): 3155.
- (24) de Boer AR, Alcaide-Hidalgo JM, Krabbe JG, Kolkman J, Boas CNV, Niessen WMA, Lingeman H, Irth H, *Anal. Chem.* 2005, **77**(24):7894..
- (25) de Boer AR, Bruyneel B, Krabbe JG, Lingeman H, Niessen WMA, Irth H, *Lab. Chip* 2005, **5**(11): 1286.
- (26) Ge X, Sirich TL, Beyer MK, Desaire H, Leary JA. *Anal. Chem.* 2001; **73**: 5078.
- (27) Gao H, Leary JA. *J. Am. Soc. Mass Spectrom.* 2003; **14**: 173.
- (28) De Boer A R, Letzel T, Lingeman H, Irth H. *Anal. Bioanal. Chem.* 2005, **381**, 647.
- (29) Norris AJ, Whitelegge JP, Faull KF, Toyokuni T. *Anal. Chem.* 2001; **73**: 6024.
- (30) Jiang Y, Lee CS. *J. Chromatogr. A* 2001; **924**: 315.
- (31) Gao J, Xu J, Locascio LE, Lee CS. *Anal. Chem.* 2001; **73**: 2648.
- (32) Sakai-Kato K, Kato M, Toyo'oka T. *Anal. Chem.* 2002; **74**: 2943.
- (33) Amankwa LN, Harder K, Jirik F, Aebersold R. *Protein Sci.* 1995; **4**: 113.
- (34) Richter P, Lopez-Ruiz B, Sanchez-Cabezudo M, Mottola HA. *Anal. Chem.* 1996; **68**: 1701.
- (35) Mao H, Yang T, Cremer PS. *Anal. Chem.* 2002; **74**: 379.
- (36) Malherbe CJ, de Beer D, Joubert E, *Int. J. Mol. Sci.* 2012, **13**: 3101.
- (37) Kaufmann CM, Graßmann J, Treutter D, Letzel T, *Rapid Commun. Mass Spectrom.* 2014, **28** (8): 869.
- (38) Ingkaninan K, de Best CM, van der Heijden R, Hofte AJP, Karabatak B, Irth H, Tjaden UR, van der Greef J, Verpoorte R. *J. Chromatogr. A* 2000; **872**: 61.
- (39) de Jong CF, Derks RJE, Bruyneel B, Niessen W, Irth H. *J. Chromatogr. A* 2006, **1112**(1–2): 303.
- (40) Krappmann M, Kaufmann CM, Scheerle RK, Grassmann J, Letzel T. *J. Proteomics Bioinform* 2014, **07**.
- (41) Zechel DL, Konermann L, Withers SG, Douglas DJ. *Biochemistry* 1998; **37**: 7664.

- (42) Kolakowski BM, Simmons DA, Konermann L. *Rapid Commun. Mass Spectrom.* 2000; **14**: 772.
- (43) Lee ED, Mueck W, Henion JD, Covey TR. *J. Am. Chem. Soc.* 1989; **111**: 4600.
- (44) Konermann L, Collings BA, Douglas DJ. *Biochemistry* 1997; **36**: 5554.
- (45) Konermann L, Rosell FI, Mauk AG, Douglas DJ. *Biochemistry* 1997; **36**: 6448.
- (46) Simmons DA, Konermann L. *Biochemistry* 2002; **41**:1906.
- (47) Tao WA, Clark RL, Cooks RG. *Anal. Chem.* 2002; **74**: 3783.
- (48) Patton WF. *J. Chromatogr. B* 2002; **771**: 3.
- (49) Mo W, Karger BL. *Curr. Opin. Chem. Biol.* 2002; **6**: 666.
- (50) Sechi S, Oda Y. *Curr. Opin. Chem. Biol.* 2003; **7**: 70.
- (51) Krappmann M, Letzel T. *Analytical Methods* 2012; **4**: 1060.
- (52) Letzel T. *RSC Chromatography Monographs* 2011, **15**: Chapter 11; ISBN: 978-1-84973-182-9.
- (53) Cheng HC, Kemp BE, Pearson RB, Smith AJ, Misconi L, van Patten SM, Walsh DA. *J. Biol. Chem.* 1986; **261**: 989.
- (54) McLachlin DT, Chait BT. *Curr. Opin. Chem. Biol.* 2001, **5**, 591.
- (55) Ogueta S, Rogado R, Marina A, Moreno F, Redondo JM, Vázquez J. *J. Mass Spectrom.* 2000; **35**: 556.
- (56) Schlosser A, Pipkorn R, Bossemeyer D, Lehmann WD. *Anal. Chem.* 2001; **73**: 170.
- (57) Moyer SC, Cotter RJ, Woods AS. *J. Am. Soc. Mass Spectrom.* 2002; **13**: 274.
- (58) Castelletti L, Verzola B, Gelfi C, Stoyanov A, Righetti PG. *J. Chromatogr. A* 2000; **894**: 281.
- (59) Kebarle P, Verkerk UH. *Mass. Spectrom Rev.* 2009; **28(6)**: 898.

3.4 Functional Analysis of Proteins, Including LC-MS and Special Freeware

Michael Krappmann and Thomas Letzel

Protein and Peptide Analysis by LC-MS: Experimental Strategies, Chapter 11, 2011, 142-155,

DOI: 10.1039/9781849733144-00142

Reproduced by permission of The Royal Society of Chemistry

<http://pubs.rsc.org/en/content/chapter/bk9781849731829-00142/978-1-84973-182-9>

Mit einem Versuchsaufbau, ähnlich wie er in Abbildung 7 in Kapitel 1.1.3 beschrieben wurde, und dem Einsatz von Achroma wurde mit einem LC-MS Enzym-Assay die Bildung nichtkovalenter Inhibitor-Enzym-Komplexe bei der Interaktion von Enzym und Inhibitor massenspektrometrisch nachgewiesen. Es wurde gezeigt, dass abhängig von der Inhibitorkonzentration bevorzugt unterschiedliche Komplexe gebildet werden.

Hierzu wurde ein System mit Lysozym aus dem Hühnereiweiß als Enzym, Hexa-N-acetylchitohexaose ((GlcNAc)₆) als Substrat und N,N',N''-triacetylchitotriose ((GlcNAc)₃) als Inhibitor verwendet.

In einem Versuch mit verschiedenen Inhibitormengen (1 nmol, 2 nmol, 5 nmol, 10 nmol, 20 nmol und 40 nmol) konnte Lysozym in der ersten Reaktionskapillare mit dem Inhibitor interagieren. In der zweiten Reaktionskapillare wurde kontinuierlich eine konstante Konzentration Substrat beigegeben. Das Verhältnis Inhibitor zu Substrat zu Enzym war: X : 5 : 1 (X=2,5; 5; 12,5; 25; 50; 100). Als interner Standard wurde das Peptid Malantide verwendet.

Folgende m/z Verhältnisse der möglichen Komplexe wurden mit Achroma näher untersucht:

- 1 Enzym-Molekül und 1 Inhibitor-Molekül
- 1 Enzym-Molekül und 2 Inhibitor-Moleküle
- 1 Enzym-Molekül und 3 Inhibitor-Moleküle
- 1 Enzym-Molekül und 4 Inhibitor-Moleküle

Es konnte gezeigt werden, dass ein Einbruch der Intensität im Chromatogramm eines kleineren Komplexes den Übergang zum nächsthöheren Komplex darstellt, indem der Quotient zweier Komplex-Chromatogramme gebildet und das daraus resultierende Quotientenchromatogramm (siehe auch Kapitel 2.4 und 3.1) analysiert wurde. Solange ein Komplex bevorzugt gebildet wurde, zeigte das Quotientenchromatogramm ein Plateau. An der Stelle des Übergangs zum nächsthöheren Enzym-Inhibitor Komplex brach das Plateau ein.

Aus obiger Beobachtung konnte geschlossen werden, dass der Komplex, bestehend aus einem Enzym-Molekül und einem Inhibitor-Molekül bevorzugt bei niedrigen Inhibitormengen (d.h. bis 5 nmol) gebildet wurde und höhere Komplexe mit mehr als einem Inhibitor-Molekül bei Inhibitormengen größer als 5 nmol. Bei niedrigen Inhibitormengen ist die Wahrscheinlichkeit höher, dass spezifische Interaktionen im aktiven Zentrum des Enzyms mit einem Inhibitor Monomer stattfinden. Mit steigender Inhibitormenge kommt es vermehrt auch zur Bildung von Inhibitor-Homodimeren, die offensichtlich zusätzlich zu den Monomeren mit der Oberfläche des Enzyms unspezifisch interagieren (siehe Abbildung 11.3 des nachfolgenden Kapitels). Es zeigt auch, dass bei unterschiedlicher Inhibitormenge nichtkovalente Komplexe mit unterschiedlicher Anzahl von Inhibitor-Molekülen gebildet werden, je nach Menge mit verschobenen Häufigkeiten in der Stöchiometrie. Mit weiterführenden Auswertungen, die über die Ausführungen im nachfolgenden Buchkapitel hinausgehen, konnte gezeigt werden, dass sich der Übergang zum nächsthöheren Enzym-Inhibitor Komplex bei den durchgeführten Versuchen wie folgt verhält:

Tabelle 5: Inhibitormenge am Übergang zum nächsthöheren Enzym-Inhibitor Komplex.

Enzym : Inhibitor-Komplex: 1 : X	Inhibitormenge des Übergangs (gemittelt)	Standardabweichung
X = 1	3,3 nmol	20 %
X = 2	6,3 nmol	6 %
X = 3	10,2 nmol	3 %
X = 4	12,0 nmol	4 %

Die Daten wurden aus der Achroma Software exportiert und mittels Gaußscher Verteilung die Inhibitormengen zum Zeitpunkt der Signaleinbrüche und anschließend die Standardabweichung berechnet. Die Gauß-Verteilung wurde den vorliegenden Signalformen genähert, da aufgrund des Singaleinbruchs beim Übergang zum nächsthöheren Komplex kein Signalmaximum bestimmt werden konnte. Damit können die ermittelten Übergänge als gute Näherung beschrieben und als Richtwert gesehen werden, bei welcher vorliegenden Inhibitormenge bevorzugt mit welchen Komplexen zu rechnen ist. Allerdings zeigt die hohe Standardabweichung bei X=1, dass das Näherungsverfahren stark von den gewählten Parametern für die Gauß-Verteilung und den zur Verfügung stehenden Daten abhängt. Zur besseren Vergleichbarkeit der Daten wurden die Parameter der Gauß-Verteilung für alle eingesetzten Inhibitormengen nicht verändert.

Das Manuskript wurde federführend von mir verfasst. Die experimentellen Versuche wurden in der Arbeitsgruppe von Herrn Priv.-Doz. Dr. Letzel gemacht. Die neukonzipierte Datenauswertung im *Chromatogram comparison* Modul von Achroma wurde von mir ein- und durchgeführt, ebenso die weiterführende Auswertung zur Berechnung der Übergänge. Herr Priv.-Doz. Dr. Letzel hat die Informationen zum experimentellen Set-up beigetragen, die Arbeit betreut und das Manuskript korrigiert und eingereicht.

CHAPTER 11

Functional Analysis of Proteins, Including LC-MS and Special Freeware

MICHAEL KRAPPMANN¹ AND THOMAS LETZEL²

¹ Institut für Forschung und Weiterbildung, University of Applied Sciences, Hochschule Weihenstephan – Triesdorf, Am Hofgarten 4, 85350 Freising, Germany; ² Competence Pool Weihenstephan (CPW), Technische Universität München, Weihenstephaner Steig 23, 85350 Freising, Germany

11.1 Introduction

Proteins have a function in organisms when they are acting as an enzyme (catalyst), when they are regulating as complexes (covalent/non-covalent) and/or when they are transporting compounds *via* non-covalent complexes. For this reason analytical methods often are applied that need labelled molecules to obtain signals in the detection systems. Mass spectrometry does not need a labelling group, if the analyte can be transferred in the mass spectrometer as an ion (see also Chapter 2).^{1,2} Thus, mass spectrometry can be used to detect protein function indirectly *via* post-translational modifications (see also Chapter 9) or directly by monitoring enzymatic conversion reactions (see Chapter 10) or complex formation (see this chapter).

New strategies in coupling liquid flow techniques with mass spectrometry result in protein reaction or interaction in continuous-flow or continuous-flow mixing assays. Chapter 10 gives an example of a continuous flow–mass spectrometry setup for the detection of enzyme kinetics using hen egg white lysozyme (HEWL) and

substrate. The same setup can be used for the monitoring of noncovalent complexes of HEWL with inhibitors. However, such studies favour continuous-flow mixing assays, as presented in this chapter. Injections of molecules that can inhibit enzymes, especially by forming specific non-covalent protein–ligand complexes, can sometimes be detected by increasing mass to charge ratios of enzyme complexes and/or negative mass to charge (m/z) ratios of product. Such an experiment, and special free software to work out the data, are presented in this chapter.

11.2 Materials and Systems

11.2.1 Chemicals

- Enzyme: Hen egg white lysozyme (HEWL, M_r 14.3 kDa, ~95%, ~50 000 units protein mg)
- Substrate: Hexa-*N*-acetylchitohexaose ((GlcNAc)₆), M_r 1237.2, ≥95%)
- Inhibitor: N, N', N'-triacetylchitotriose ((GlcNAc)₃), M_r 627.6, ≥95%)
- Internal standard: Malantide (IS, M_r 1633.9, ≥97%)
 - Enzyme, substrate, inhibitor and internal standard were obtained from Sigma-Aldrich (Steinheim, Germany)
- Further chemicals: Ammonium acetate (NH₄Ac, >98%) and acetic acid were purchased from Merck (Darmstadt, Germany), acetonitrile (>99.9%) and methanol (MS grade) from Riedel-de Haën (Seelze, Germany). High purity water was from a Milli-Q system (Millipore, Eschborn, Germany)
 - Ⓢ The higher the purity of the applied compounds, the lower the impurity signals in the obtained mass spectra

11.2.2 The Example Enzymatic Assay

- Enzyme HEWL and substrate ((GlcNAc)₆): see also Chapter 10
- Inhibitor: N, N', N'-triacetylchitotriose ((GlcNAc)₃)
 - Ⓢ An inhibitor is a regulator (ligand, molecule) that decreases the rate of an enzyme-catalysed reaction if it is present in the reaction mixture^{3,4}

11.2.3 HPLC (Sample Introduction)

- HPLC: 1100 Series, Agilent Technologies, Waldbronn, Germany with binary pump, isocratic pump, degasser, well-plate-sampler, column oven and multiwavelength detector
- Software: ChemStation (Rev.B.01.01), Agilent Technologies, Waldbronn, Germany

11.2.4 Mass Spectrometer

- Mass spectrometer: Q-TOF Ultima API, Waters-Micromass, Manchester, UK

- Desolvation gas and cone gas was Nitrogen (purity 4.0) obtained from a nitrogen generator (Domnick Hunter GmbH, Krefeld, Germany) and the collision gas was argon (purity 5.0)

The setting parameters were optimized for the ESI interface (see Chapter 2 for a similar optimization) and final parameters are like the following:

- The measurements were carried out with 80 °C source temperature, 150 °C desolvation temperature, 300 L h⁻¹ desolvation gas flow and 50 L h⁻¹ cone gas flow
- The mass spectrometric detection was performed in positive detection mode. The capillary voltage was 3000 V, the cone voltage 40 V and the collision energy 10. The mass-range was set to *m/z* range of 50–3000 with 1 s scan time
 - ① The Q-TOF calibration was done by using NaI and CsI solution. The salt cluster can very well be used for mass spectrometric calibrations in the region of higher mass to charge ratio and to cover a large mass range

11.2.5 Complex Study with Flow Injection Analysis (FIA)

- Assay solution: 90% 10 mM NH₄Ac containing aqueous solution (pH 5.2) with 10% methanol
- Substrate solution: 50 μM (GlcNAc)₆ in 10 mM NH₄Ac solution (pH 5.2)
- Enzyme solution: 20 μM HEWL + 4 μM malantide in 10 mM NH₄Ac solution (pH 5.2)
- Inhibitor solution: 1–40 nM in 10 mM NH₄Ac solution (pH 5.2)
- Stopping reagent: 90% acetonitrile
- Syringe pump: 11 Plus, Harvard Apparatus, Hugo Sachs Elektronik, Hugstetten, Germany
- Superloops: Volume: 10 mL, Amersham Biosciences, Pittsburgh, USA
- Tubing: PEEK (polyetheretherketone), 0.17 mm × 350 mm and 0.25 mm × 830 mm, Chromatographie Handel Müller, Fridolfing, Germany
- Mixer: 1/16" PEEK mixing-T, 10 μm UHMWPE frit, Chromatographie Handel Müller, Fridolfing, Germany

11.2.6 Analysis Software

- MassLynx Software, Version 4.0, Waters-Micromass, Manchester, UK
- Achroma Software Tool, University of Applied Science Weihenstephan-Triesdorf, Freising, Germany⁵

11.3 Methods

A typical setup for an optimal performance of complex formation experiments is shown in Figure 11.1. The original setup was developed to obtain a

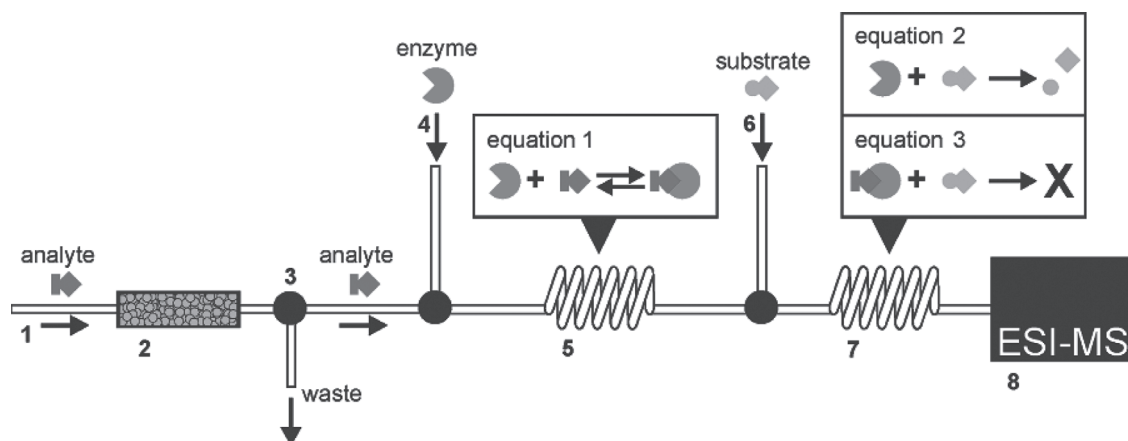


Figure 11.1 Scheme of the on-line continuous-flow mixing system: (1) sample introduction *via* the assay solution +5% methanol; (2) HPLC column; (3) 1:3 flow splitter, 50 $\mu\text{L}/\text{min}$ bioassay/150 $\mu\text{L}/\text{min}$ waste; (4) enzyme solution; (5) reaction coil A; (6) substrate solution; (7) reaction coil B; (8) mass spectrometer. Enzyme continuously converts the substrate into products (equation 2) if no inhibitor is eluting from the column. Bioactive compounds present in the eluate bind to the enzyme (equation 1), resulting in a decrease of product turnover (equation 3).⁶ Reprinted with permission from A.R. de Boer *et al.*: On-line coupling of high-performance liquid chromatography to a continuous-flow enzyme assay based on electrospray ionization mass spectrometry. *Anal. Chem.*, **76**, 2004; 3155–161. Copyright 2004 American Chemical Society.

continuous-flow mixing system for the monitoring of enzymatic inhibitors in complex solvents extracted from plants *via* substrate and product ion traces in MS.⁶

The first step is a chromatographic separation leading to a differentiation of single molecules from a complex mixture (parts 1 and 2 in Figure 11.1). The HPLC column eluting compounds (see part 3 in Figure 11.1) is used further in the continuous-mixing flow setup.

- ❶ Take care that the HPLC mobile phase contains as little organic solvent as possible, because the higher the content the lower the enzymatic activity (due to protein denaturation and compound dissociation)
- ❷ On the other hand, organic solvents protect proteins from being adsorbed on capillary surfaces and support an efficient ion transfer out of the interface spray into the mass spectrometer (see also Chapter 2). Thus effective detection often requires the addition of a ‘make-up’ flow after eluting the reaction system prior to entering the mass spectrometer (*e.g.* behind coil 7 and prior to MS 8 in Figure 11.1)

Classical reversed-phase (RP)-HPLC can be applied if the organic content is low enough and no protein denaturation occurs. This is contrary to the requirements of a typical HPLC separation, whereby non-polar compounds elute at high organic content. Techniques like high-temperature (HT)-HPLC are often used to overcome this weakness. The effect of decreasing polarity by increasing the temperature of water and water/organic solvent mixtures thereby

plays an extensive role. A detailed description of HT-HPLC can be found in a recent RSC monograph.⁷ As a consequence, solvent gradients can partially be substituted by temperature gradients, with the result that lower quantities of organic solvents are needed.

Back to the continuous-mixing flow system; the elution flow out of the separation column (see part 3 in Figure 11.1) containing the separated molecules is mixed with the enzyme-containing flow (see part 4 in Figure 11.1).

A separated compound that elutes and has enzyme-regulating properties is interacting with the enzyme (in coil 5 in Figure 11.1) and influences the enzymatic activity, *i.e.* the enzyme is activated or inactivated. In a second coil (coil 7 in Figure 11.1) this stream is mixed with the substrate solution. Now, an up- or down-regulated substrate conversion compared to the regular situation can be monitored. As long as there is no regulating molecule interacting with the enzyme, a steady amount of substrate will be converted and the chromatographic result for the substrate as well as the product ion trace (as an extracted ion chromatogram, EIC) is a straight line. Inhibiting compounds lead to a negative tip and activating compounds to a positive tip in the resulting product trace. Several compounds—substrate, products and other ionic compounds—can be directly monitored and presented as EICs of the relevant m/z ranges.⁸

In this setup the product formation *versus* its change directly marks the combination of a time point of an HPLC separation with biological activity, and thus leads directly to molecular masses and to hydrophobicity of the regulating compounds. The same is true for the monitoring of non-covalent complexes. The experiment explained below shows the development of non-covalent enzyme–inhibitor complexes through the EIC of the relevant m/z ranges and the use of the Achroma software tool.

11.3.1 Studying Complex Formation (and Simultaneous Reaction Inhibition) with Continuous-Flow Mixing Analysis

The experimental setup in this chapter is a similar on-line system as described above (from de Boer *et al.* in 2004).⁶ The continuous-flow mixing system consists of a sample introduction part (see parts 1–3 in Figure 11.1), an enzymatic assay (see parts 4–7 in Figure 11.1) and an ESI-MS (see part 8 in Figure 11.1). The HPLC sample introduction part is substituted by an auto injector for studying the complex formation.

11.3.1.1 Enzymatic Assay and Inhibition (Complex Formation) Study

- First, the inhibitor (complex-forming compound) is injected into the system by an autosampler at periodic time intervals (sample introduction part). In the reaction coil (see part 5 in Figure 11.1) inhibitor and enzyme form complexes (see equation 1 in Figure 11.1) for a defined interaction time

- In this way enzyme (see part 4 in Figure 11.1) and substrate (see part 6 in Figure 11.1) are pumped continuously from the two containing superloops into the flow system
- In the reaction coil (see part 7 in Figure 11.1) enzyme–substrate complexes can be formed at times when no inhibitor is injected. During the inhibitor injection periods, both inhibitor–enzyme complexes and substrate–enzyme complexes can be formed in competition. The equilibrium therefore depends thereby on the inhibitor concentration and affinity (see equations 2 and 3 in Figure 11.1)
- The continuous-flow system ends in an ESI-MS, which detects the relevant m/z ranges by time for enzyme, inhibitor, substrate, product, and possible complexes
 - Example assay: Injection volume of inhibitor is 10 μL (injection rate 4 $\mu\text{L}/\text{min}$, ejection rate 10 $\mu\text{L}/\text{min}$) *via* autosampler. The flow rate of the pump is 20 $\mu\text{L}/\text{min}$. 10 μL of various inhibitor concentrations (see Table 11.1) are injected with 4 min runtime before the next injection. Each injection takes 2.5–3.5 min. The first tubing is used as interaction coil (see part 5 in Figure 11.1) for enzyme–inhibitor interaction with a length of 350 mm and a diameter of 0.17 mm. Thus the interaction time between enzyme and inhibitor is 13 s (0.21 min)
 - ① The reaction time ($r^2\pi*L/fr$) of the mixtures is determined by the tubing length L , inner diameter ($2r$) of the tubing and the flow rate (fr) of the solutions and can be adapted by varying these parameters
- Enzyme (HEWL) and substrate (GlcNAc_6) are continuously pumped by the HPLC binary pump or two single pumps with a flow rate of 20 $\mu\text{L}/\text{min}$ each from a superloop container into the flow system
 - ① The tubing of one pump head of the binary pump is unscrewed from the gradient mixer and is directly connected to one superloop (if possible, use an external pulse damper). The other flow is guided through the HPLC system and then connected with the second superloop
 - ① Cool the superloops down if you store large volumes of enzyme/substrate for an extended continuous-flow enzymatic assay (with ice, if

Table 11.1 Injection concentrations and absolute amount injected of the inhibitor ($(\text{GlcNAc})_3$) into the enzyme containing assay with final stoichiometry at each flow-injection analysis

Concentration of inhibitor ($\mu\text{mol L}^{-1}$)	Abundance of ligand (inhibitor) (nmol)	Ratio $(\text{GlcNAc})_3:(\text{GlcNAc})_6:\text{HEWL}$ (inhibitor:substrate:enzyme)
0.1	1	2.5:5:1
0.2	2	5:5:1
0.5	5	12.5:5:1
1.0	10	25:5:1
2.0	20	50:5:1
4.0	40	100:5:1

compounds tolerate it). On the other hand, take care that the solutions are at a reaction temperature for a stable interaction equilibrium

- The concentration ratio of substrate to enzyme in this experiment is 5:1 (see Table 11.1). The interaction time of enzyme (respectively, enzyme–inhibitor complex) with substrate in the second tubing (see part 7 in Figure 11.1, length = 830 mm, diameter = 0.25 mm) is 64 s (1.03 min) before the solution reaches the ESI-MS (Q-TOF Ultima API)

11.3.1.2 Mass Spectrometric Detection

- Example assay: Monitor the reaction for 120 min. After this time all inhibitor injections are finished (Table 11.1). Note, every concentration is injected twice

11.3.2 Data Analysis

Typically, the manufacturer's analysis software works well for standardized experimental setups but lacks some special features for the evaluation of experimental data from the setup presented here. For instance, studying complex formation with the current experimental setup generates negative chromatographic signals (see NAG_4+H^+ trace in Figure 11.2), which in general can be shown but cannot be used for calculation by manufacturer's software. Thus, we developed and applied a specialized freeware program to allow specialized features and calculations with 'atypical' analytical datasets.

The injection of an inhibitor ($(\text{GlcNAc})_3$) into the flow of the enzyme (HEWL) resulted in an increased mass trace of the inhibitor itself (NAG_3+H^+ trace in Figure 11.2) and a decrease of the product ion trace ($(\text{GlcNAc})_4$; NAG_4+H^+ trace in Figure 11.2). Also the IS [Mal+3H] ion trace is decreasing beginning with 5 nmol injection of NAG_3 . Thus the results cannot be directly used quantitatively without correction procedures. However, because the focus in this chapter is elsewhere, the uncorrected data is used, employing a qualitative processing.

Proceed as follows in the manufacturers' software:

- Extract the EIC of substrate, product(s) and inhibitor (ligand) from the mass spectrum. Plot the intensity according to the time points
- Compare the EIC of the product and inhibitor. When the inhibitor plot shows a positive signal, a negative signal should be observed in the product EIC plot and no signal change in the IS EIC plot at the same time point
 - Example assay: The EIC of substrate signal with m/z 1237.4 [$(\text{GlcNAc})_6$], inhibitor signal with m/z 628.3 [$(\text{GlcNAc})_3$], and the product signals with m/z 831.3 [$(\text{GlcNAc})_4$] as well as the signal of the internal standard malantide with m/z 545.3 are plotted in Figure 11.2. The continuous-flow mixing assay shows the expected decreasing product trace while the inhibitor trace is increasing

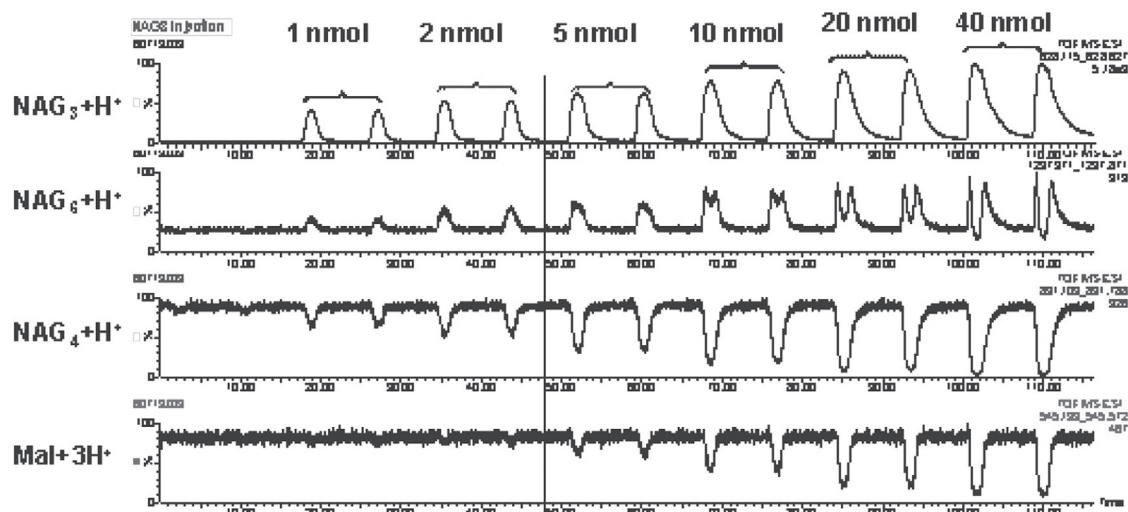


Figure 11.2 EICs of inhibitor ((GlcNAc)₃ or NAG₃), substrate ((GlcNAc)₆ or NAG₆), product ((GlcNAc)₄ or NAG₄) and IS (Mal+3H⁺) in HEWL assay with (GlcNAc)₃ injections, presented with MassLynx software from Waters-Micromass.

- ① Note: The abundances of equal or different signal intensities of the various ion traces do not reflect quantitative data, and signal intensities are not directly comparable between the different ions. The negative tip of IS above 5 nmol inhibitor leads to further need for data processing prior quantitative statements. However, as stated above, here the uncorrected data is applied for the proof of the software tool
- ① The different EIC can be presented by using MassLynx software from Waters for data acquisition (see Figure 11.2). The chromatograms can also be plotted with our Achroma software. As well as the typical positive signals, the negative signals applied in this study can also be integrated and the peak areas calculated
- ① Before using the Achroma software, a data transformation step is necessary. MassLynx, for instance, includes the so-called DataBridge software for converting data files into other formats
- Open the DataBridge software and convert the raw data file into a (MassLynx) ASCII file without header
 - ① An ASCII file without header has the format seen in Table 11.2 and the file extension ‘txt’
- The module applied in processing and calculating the enzyme–inhibitor complex EICs is an ion trace comparison. Therefore, open the Achroma software and choose the tool: ‘Chromatogram Comparison’. Then load the ASCII data file into the software
 - ① Achroma generates an index file, for faster internal work. The first loading of a data file, *e.g.* with a size of 8 GB, takes approximately 2 min, depending on the computer hardware, because the index file is generated simultaneously during the loading process. Later loadings of this data file do not take so long if the index file still exists.

Table 11.2 Format of an ASCII data file (left) and the presentation in an example file (right)

	<i>Data file format</i>	<i>Example file (60711Z02.txt)</i>
Line number		FUNCTION
001	FUNCTION	Scan 1
002	[Empty line]	Retention Time 25
003	Scan[tabulator][tabulator][Number]	199.3016 0
004	Retention Time[tabulator][time]	199.6646 0
005	[Empty line]	199.6704 0
006	[mz value][tabulator][intensity]	...
007	[mz value][tabulator][intensity]	3002.8877 1
...		Scan 2
100	[Empty line]	Retention Time 44
101	Scan[tabulator][tabulator][Number]	199.3016 0

The index file has the same file name as the data file plus an underscore sign at the end of the name (Example file in Table 11.2: 60711Z02.txt, index file name: 60711Z02_.txt) and will be saved on the same place where the data file is saved

- Next, load the first complex EIC into the software and then load the second complex EIC that has to be compared. The software automatically calculates a quotient chromatogram with the data of the first-loaded EIC divided by data of the second-loaded EIC
 - ① Such a quotient chromatogram (see the upper traces in Figure 11.3 a–c) is a useful software function for data correction (*e.g.* product with IS; data not shown) or the interpretation of the higher-order complex formation as shown in Figure 11.3
 - ① Another task for the calculation of a quotient chromatogram is the following: if the experimental setup has two internal standards which flow with the samples through the system and/or are mixed into the flow before MS, then the quotient of the data of the internal standards should be constant over the complete runtime. The experimental setup works without any influence of system errors, if the quotient chromatogram is constant, whereas fluctuations occur when the flow system shows instabilities
- Example data analysis: The results of the study of HEWL–inhibitor (GlcNAc)₃–complex (LIC) formation by inhibitor flow injections from 1 nmol to 40 nmol (Table 11.1) are shown in Figure 11.3. In each section (A–C) there are always three plots. The upper plot represents a quotient chromatogram calculated from the middle chromatogram plot divided by the lower chromatogram plot as described above. The middle LIC EIC plot in each section has a higher order (n) than the lower LIC EIC plot in the same section. The complexes studied are eightfold charged ions and the ratio enzyme:inhibitor is 1:n, so n = 1–4. The following shortcuts are used for the complexes:
 - LIC1 ([HEWL((GlcNAc)₃)⁸⁺]) for a complex with n = 1
 - LIC2 ([HEWL((GlcNAc)₃)₂⁸⁺]) for n = 2

- LIC3 ([HEWL((GlcNAc)₃)₃]⁸⁺) for n = 3
- LIC4 ([HEWL((GlcNAc)₃)₄]⁸⁺) for n = 4

In Figure 11.3a, the comparison of LIC2 with LIC1 shows that LIC1 is formed above 1 nmol inhibitor injection and LIC2 is formed above 2 nmol. Peak breaks are observed for LIC1 above 5 nmol injections and for LIC2 above 10 nmol injections. The signal of LIC1 has generally a higher intensity than LIC2. However, as stated above, this is not a direct quantitative information, because the ionization efficiency of similar molecules can differ immensely.⁹ On the other hand, it can be assumed that LIC1 is formed preferably, because one inhibitor molecule could bind stably at the active site of the enzyme at the lowest inhibitor concentration.

- ① Note, the results in this form give no clear picture of interaction specificities. Specific and unspecific interactions can be identified by further experiments with known ligands or tandem mass spectrometry

The LIC2 is probably the specific interaction of an inhibitor monomer and the active centre of one enzyme molecule and a second unspecific interaction of one inhibitor molecule with the surface of the enzyme (Figure 11.3a). The LIC3 is probably the specific interaction of an inhibitor monomer and the active centre of one enzyme molecule and a second unspecific interaction of an inhibitor homodimer with the surface of the enzyme (Figure 11.3b). Finally, LIC4 is probably the specific interaction of an inhibitor monomer and the active centre of one enzyme molecule and a second unspecific interaction of an inhibitor homodimer and a monomer with the surface of the enzyme (Figure 11.3c). The formation of (GlcNAc)₃ homodimers in the gas phase was shown in a previous publication with the general mass spectrometric detection applied here.¹⁰

The quotient chromatogram always shows negative signals, resulting from the fact that the data of the EIC of LIC1 are the divisor of the quotients with higher intensities and therefore higher abundance. Other statements can be made from the quotient chromatogram: a signal in the quotient chromatogram shows that both LICs at the corresponding abundance have a signal; a plateau on the top/bottom of a quotient chromatogram signal indicates that one LIC chromatogram is near or in the stadium of a signal break; the occurrence of signal break in the quotient chromatogram is a sign for signal breaks in both LICs, thus a higher order complex can occur in this inhibitor concentrations or signal suppression (see also Chapter 2). There are analogous considerations for all other cases (see B and C in Figure 11.3).

Another option of the Achroma software is the discovery and calculation of negative chromatographic peaks.

- In the next section, as an example, negative signal areas of the product trace are calculated at the time points of inhibitor injections. Open a new tool of Achroma software, 'Signal Recognition', and load the data file

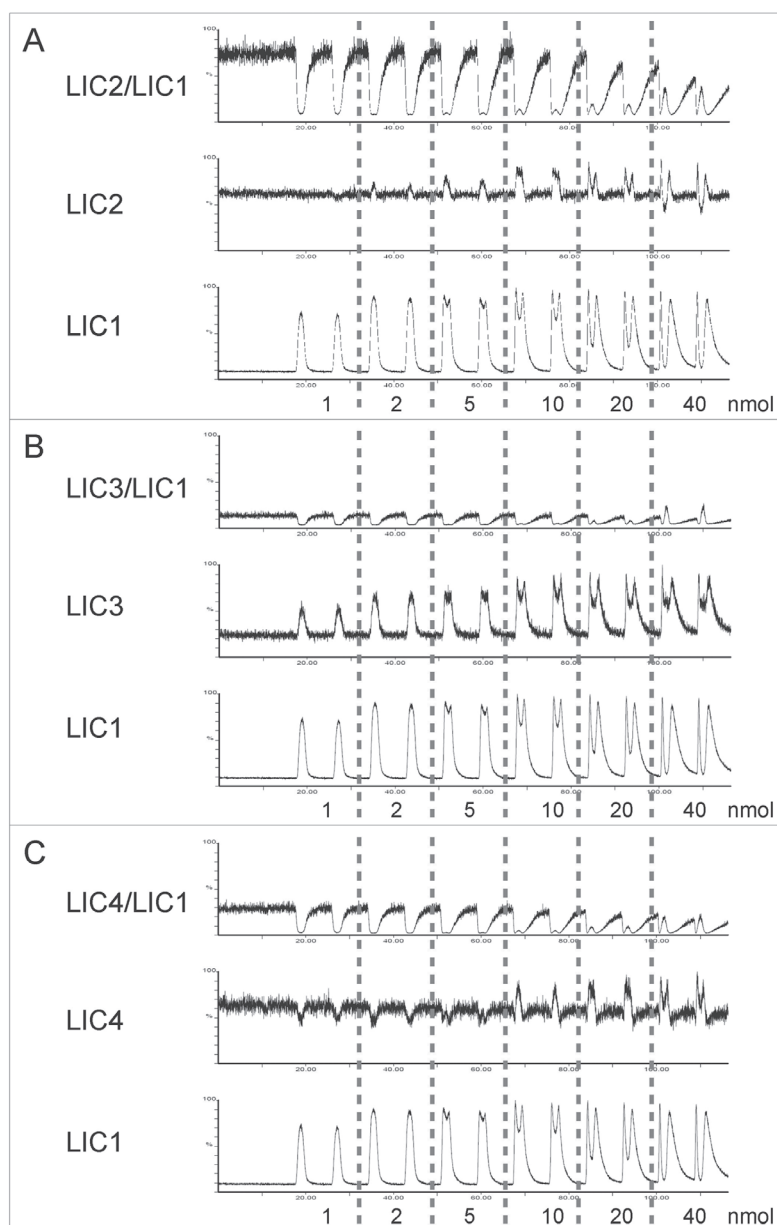


Figure 11.3 Results of data acquisition with Achroma software. Scheme of the plots from A to C: upper plot = quotient chromatogram (middle plot/lower plot), middle plot = complex EIC with higher order than the lower EIC complex plot. (A) quotient chromatogram LIC2/LIC1, LIC2, LIC1; (B) quotient chromatogram LIC3/LIC1, LIC3, LIC1; (C) quotient chromatogram LIC4/LIC1, LIC4, LIC1.

again. After setting the m/z range and choosing the calculation of negative signals, the peak integration can be carried out

- Example assay: Enter m/z range of 831–832, select ‘negative signals’ into Achroma and then press the ‘search’ button. The product chromatogram will be shown. After this step press the ‘calculate’ button and the signal areas will be calculated and be shown in the plot and additionally in a table (Figure 11.4)



Figure 11.4 Results of the signal recognition with Achroma software tool. Signal areas of the product ((GlcNAc)₄) at different inhibitor concentrations (see Table 11.1) are shown.

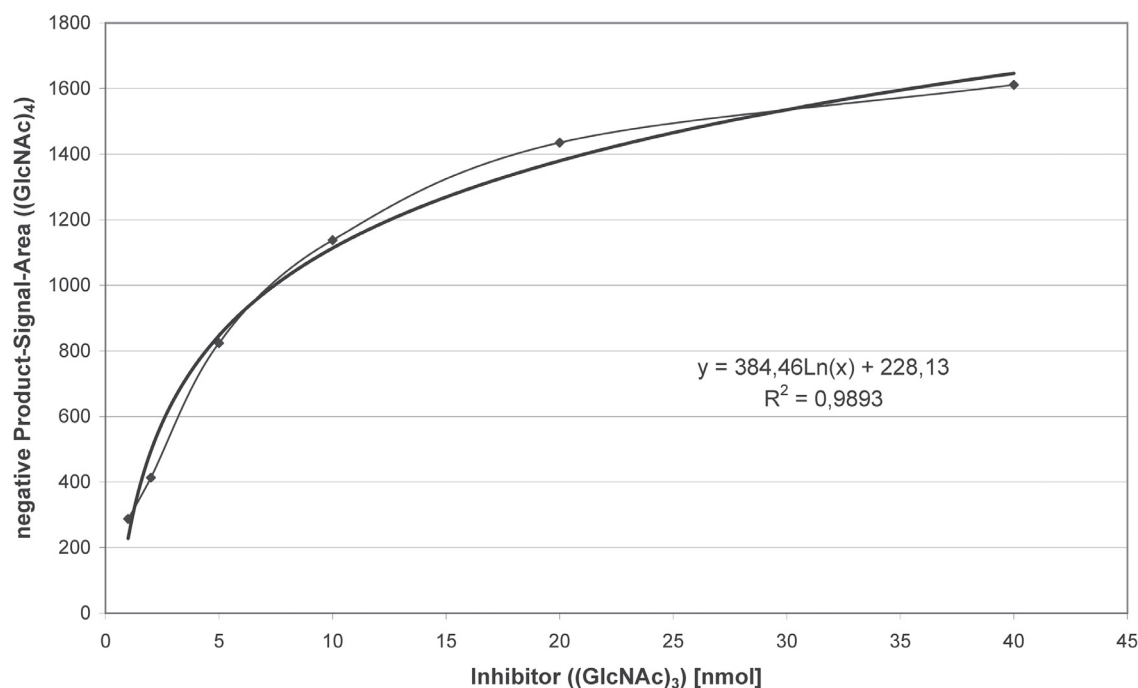


Figure 11.5 Plot of inhibitor against negative product signal areas with trend curve.

- Present a plot: inhibitor concentrations are plotted according to the calculated signal areas of the product
 - Example assay: Product signal areas (GlcNAc)₄ *versus* inhibitor concentrations from Table 11.1 are shown in Figure 11.5. It can be shown that the increase of inhibitor abundance does not lead to a linear increase of a 1:1 enzyme–inhibitor complex formation in the mass spectrometer. Several reasons can be ascribed: an equation not established at a specific concentration, a mass spectrometric dissociation, and formation of enzyme–inhibitor complexes with higher inhibitor content (data as shown in Figure 11.3)

11.4 Notes

11.4.1 Working with Enzymes and MS-Based Enzymatic Assays

- See Chapter 10

11.4.2 Working with Achroma External Analytical Software

- The data must have an appropriate format (*e.g.* MassLynx ASCII file without header) to be able to work with Achroma software. If the data file is in a raw data format the Achroma software will not be able to handle it. However, nowadays several raw data formats should be convertible by manufacturer's software into ASCII files or similar data format (txt)
- If the data is in the appropriate format, the analytical results from all commercial mass spectrometers can be used with the software and be combined
- Loading two chromatograms into the module 'Chromatogram comparison' should be done sequentially, because opening/processing two chromatograms simultaneously may lead to software problems, depending on the hardware (processor-intensive calculations)
- For the quotient calculation of two chromatograms the data file may not contain zero values, since division by zero is not defined and will therefore lead to an error
- 'Signal recognition' has an additional feature. It is possible to generate a *difference spectrum* between two time coordinates of the graph (shown in Figure 11.4). For example, subtract the spectrum at the maximum/minimum of a signal and the beginning of a signal and the new *difference spectrum* shows immediately which molecules increase and which molecules decrease in intensity.⁵

At present the Achroma software can be received from MK by e-mail. In future it will be downloadable from an Achroma website and is free for non-commercial use.

References

1. T. Niwa, U. Doi, Y. Kato and T. Osawa, *FEBS Lett.*, 1999, **459**, 43.
2. X. Xiao, X. Luo, B. Chen and S. Yao, *J. Chromatogr.*, 2006, **834**, 48.
3. D. Koley, *Enzymology*, World Scientific Publishing, NJ, 2007.
4. A. Cornish-Bowden, *Fundamentals of Enzyme Kinetics*, Portland Press, London, 2004.
5. M. Krappmann, T. Letzel, Achroma—a software for analysing ‘functional proteomics’ research data, like non covalent complex bonding (e.g. enzymatic inhibition) got via a new LC-MS experimental design, *BMC Bioinformatics*, to be submitted.
6. A. R. deBoer, *Anal. Chem.*, 2004, **76**, 3155.
7. T. Teutenberg, *High-Temperature High-Performance Liquid Chromatography—A User’s Guide for Method Development*, Royal Society of Chemistry, Cambridge, 2010.
8. T. Letzel, *Anal. Bioanal. Chem.*, 2008, **390**, 257.
9. N. Dennhart, T. Fukamizo, R. Brzezinski, M. E. Lacombe-Harvey and T. Letzel, *J. Biotechnol.*, 2008, **134**, 253.
10. N. Dennhart and T. Letzel, *Anal. Bioanal. Chem.*, 2006, **386**, 689.

4 Diskussion

Kontinuierliche LC-MS Enzym-Assays haben sich in den letzten Jahren zu einer etablierten Methode im Bereich der "funktionellen" Proteomik ^[40] entwickelt. Dies zeigt die stetig steigende Anzahl von Veröffentlichungen (siehe Abbildung 6). Mit der Etablierung neuer Messmethoden nimmt ebenso der Anspruch und Bedarf an Informationstechnologien zur Auswertung und Interpretation der gewonnenen Datenmengen zu. Gerade im Bereich der experimentellen Forschung, mit nicht etablierten Versuchsaufbauten, ist eine schnelle Softwareadaption zur Auswertung der Daten von Vorteil. Ziel dieser Arbeit war es eine solche Software-Strategie (Achroma) zu entwickeln, mit der auf wechselnde Versuchsanforderungen reagiert und die Software bei entsprechenden Programmierkenntnissen schnell erweitert werden kann. Die neu entwickelten Module wurden anhand von konkreten Anwendungen im Rahmen dieser Arbeit auf Validität der ermittelten Ergebnisse und auf Funktionalität sowie Benutzerfreundlichkeit getestet.

Im Bereich der (Prote-)Omik-Forschung gibt es bereits eine Vielzahl von Softwarelösungen, die aus historischen Gründen, aber auch aus den verschiedensten Ansprüchen an die Datenevaluation und -austauschbarkeit entstanden sind. In Kapitel 1.4 dieser Arbeit wurde ausführlich und themenübergreifend auf den aktuellen Stand der Technologie eingegangen. Auch zeigt die nach wie vor steigende Zahl an Veröffentlichungen zu Softwarelösungen in Verbindung mit Massenspektrometrie und Enzym-Assays den steten Bedarf nach neuen Softwarestrategien (siehe Abbildung 10). Trotz der mittlerweile bestehenden Vielzahl an Softwarelösungen gab es zu Beginn dieser Arbeit keine Software, die den Ansprüchen der Datenauswertung der neu entwickelten Enzym-Inhibitor Versuche, sowie den Versuchen zu nicht-kovalenten Komplexen genügte.

Mit Achroma wurde ein modulares Softwaresystem entwickelt. Achroma ist in C++ geschrieben und für Windowssysteme entwickelt. Zur schnellen Verarbeitung großer Datenmengen (mehrere Gigabyte) wurde ein Indexierungssystem konzipiert. Die Daten werden in einem textbasiertem Format gespeichert, damit diese möglichst von vielen anderen Softwarelösungen verarbeitet werden können bzw. ohne Einsatz von Datenkonvertern lesbar bleiben. Alle Achroma-Module sind in sich gekapselt, sodass jederzeit weitere Module implementiert werden können. Es wurden dabei bisher fünf Module umgesetzt (siehe auch Kapitel 2.4 und Kapitel 3.1):

1. *Chromatogram view* Modul (Basis-Modul)
2. *Spectrum view* Modul (Basis-Modul)
3. *Chromatogram comparison* Modul (erstmalige Umsetzung in Achroma)

4. *Signal recognition* Modul (modifizierte Signalfächenberechnung im Vergleich zu bestehenden Softwarelösungen, sowie erstmals kontinuierliche Flächenberechnung)
5. *Spectra comparison* Modul (erstmalige Umsetzung in Achroma)

Die beiden Basis-Module zur Visualisierung von Spektren und Chromatogrammen (1. und 2.) werden nicht näher diskutiert, da diese Funktionalität die Grundvoraussetzung darstellt um die chromatographischen und massenspektrometrischen Daten handhaben zu können.

Chromatogram comparison Modul:

Dieses Modul vergleicht zwei Chromatogramme zu gegebenen m/z Verhältnissen und stellt das Ergebnis in einem dritten Chromatogramm am Bildschirm dar. Der Vergleich erfolgt in Form eines Quotienten. Es wird zu identischen Zeitpunkten der beiden Chromatogramme ein Quotient aus den Intensitäten gebildet. Die Achroma Software war die erste Software die diesen Vergleich zweier Chromatogramme in dieser Form implementiert hat. Das *Chromatogram comparison* Modul wurde ursprünglich dazu entwickelt zu zeigen, dass es bei Versuchen mit LC-MS Enzym-Flusssystemen nicht zu Interaktionen der Analyten mit den Komponenten des Systems kommt. Hierzu werden zwei interne Standards, einer zu Beginn und einer kurz vor dem MS, in das System eingebracht - siehe auch Kapitel 3.2 und 3.3. Mit der Bildung des Quotientenchromatogramms der m/z Spuren der internen Standards wird jede Abweichung sofort sichtbar. Jegliche Abweichung gibt dem Experimentator somit Hinweise auf mögliche Reaktionen der Analyten mit dem System.

In Kapitel 3.3 wird ein LC-MS Enzym-Assay (PKA) beschrieben indem zwei komplementäre interne Standards verwendet wurden. Der erste interne Standard (IS1), der das komplette System durchlief, hatte eine ähnliche Hydrophobizität, wie die eingesetzten Substrate, so dass auch ein ähnliches physikalisches/chemisches Verhalten erwartet werden konnte. Der zweite interne Standard (IS2), der direkt vor dem Eintrag ins MS zugegeben wurde, war in seinen Eigenschaften den Reaktionsprodukten ähnlicher. Diese Strategie sollte zum einen eine Quantifizierung ermöglichen (siehe unten: *Signal recognition* Modul) und zum anderen zeigen, dass das System wie erwähnt stabil betrieben werden kann. Gerade über längere Versuchsreihen, die sich teilweise über Stunden bis Tage erstrecken, ist eine Kontrolle über die Zuverlässigkeit des Messsystems von großer Bedeutung. Mit dem *Chromatogram comparison* Modul kann die Stabilität über das Quotientenchromatogramm schnell und einfach gezeigt werden:

- Bleibt der Quotient konstant, aber die Intensitäten von z.B. dem Substrat und IS1 sinken über die Zeit oder wiederholter Experimente, dann kann von einer Kontamination im Einlass zum MS ausgegangen werden. Die Ionisation/Detektion

wird für alle Analyten in gleicherweise beeinträchtigt. Die Messwerte sind valide und können ausgewertet werden.

- Ebenso schnell können Unregelmäßigkeiten im Messsystem erkannt werden, wenn sich der Quotient der internen Standards über die Zeit in der Intensität verändert. Ändern sich gleichzeitig auch noch die Intensitäten von IS1 und den Analyten, dann kann angenommen werden, dass das System aufgrund von veränderten Systemvoraussetzungen, z.B. zu geringen Flussraten (Leck im System), oder veränderten unspezifischen Interaktionen der Analyten/IS1 mit der Reaktionskapillare. Die Messergebnisse können dann nicht zur Auswertung herangezogen werden.
- Ein drittes Beispiel ist die Wiederholbarkeit von Experimenten. Bleiben Quotientenchromatogramm und die Intensität von IS1 konstant, aber die Intensität eines Analyten sinkt, dann kann davon ausgegangen werden, dass der Analyt während seiner Verweilzeit im Vorratsbehälter zum Teil abgebaut wurde. Die Reproduzierbarkeit der Ergebnisse ist nicht gegeben und das Experiment muss wiederholt werden.

Neben der Systemstabilität konnte gezeigt werden, dass mit dem *Chromatogram comparison* Modul die Komplexbildung zwischen Enzym und Inhibitor visualisiert werden kann - siehe auch Kapitel 3.4. Diese neue Auswertemethode wurde während dieser Arbeit entwickelt und erstmalig dargestellt:

Neben der spezifischen Interaktion von Enzym und Inhibitor im aktiven Zentrum des Enzyms kommt es bei steigenden Inhibitorkonzentrationen in der Reaktionskapillare vermehrt zu unspezifischen Interaktionen von Inhibitor Monomeren oder Homodimeren außerhalb des aktiven Zentrums. Setzt man die Intensitäten der Chromatogramme der höheren Komplexe mit dem Komplex aus einem Inhibitor-Molekül und einem Enzym-Molekül ins Verhältnis, so kann beobachtet werden, dass bestimmte Komplexe erst bei Überschreiten einer Konzentration an Inhibitor in der Reaktionskapillare bevorzugt gebildet werden. Die Darstellung der Verhältnisse führt zu einem Quotientenchromatogramm mit Signalen, die ein Plateau an der Stelle des Maximums oder Minimums zeigen. Am Übergang von einem Komplex zum nächsthöheren Komplex bricht dieses Plateau ein. Diese Darstellung ist sehr hilfreich für eine schnelle massenspektrometrische Vorhersage in welchem Intervall an Inhibitorkonzentration bevorzugt mit welchem Komplex zu rechnen ist. Allerdings kann zu diesem Zeitpunkt noch keine Aussage über die Zusammensetzung des Komplexes (Anzahl Inhibitor-Monomere, Inhibitor-Homodimere) gemacht werden. In weiterführenden Auswertungen konnte mit Hilfe des *Signal recognition* Moduls näherungsweise die Inhibitormenge berechnet werden, wo der Übergang zum nächsthöheren Enzym-Inhibitor

Komplex stattfindet. Generell ist der Bedarf nach weiterführenden Auswertungen schwer abzuschätzen. Ein guter Indikator ist die Anzahl von Publikationen zu einem Thema. So ist es sinnvoll die weiterführende Auswertung zur Komplexbildung in Achroma zu implementieren, wenn der Bedarf entsprechend groß wird.

Signal recognition Modul:

In diesem Modul sind zum einen Standards zur Signalerkennung und Flächenberechnung implementiert, aber auch neue Funktionen, die mit Achroma erstmalig im Bereich der Software für funktionelle Proteomik, respektive der Enzymfunktion und -regulation umgesetzt wurden. So erfolgt die Signalerkennung über den Wechsel in der Steigung der ersten Ableitung des Chromatogramms. Die Signal-Extrema werden von Achroma automatisch erkannt, können aber manuell nachbearbeitet werden, sodass bei nichtoptimaler automatischer Erkennung von Extrema justiert werden kann. Der Graph kann geglättet werden. Die Glättung erfolgt bisher mittels einer voreingestellten 10-fachen Glättung nach Savitzky und Golay ^[123]. Die fest vorgegebene Glättung des Graphen sollte bei einer Überarbeitung des Moduls als frei durch den Benutzer einstellbar implementiert werden. Das würde die Benutzerfreundlichkeit zusätzlich erhöhen. Die Flächenberechnung erfolgt über das Integral unterhalb des Graphen von Start- bis Endzeitpunkt des Signals und eine etwaige Drift der Basislinie wird automatisch bei der Berechnung mit berücksichtigt. Einzigartig in Achroma, war zum Zeitpunkt der Entwicklung des Moduls, die Detektion negativer Signale, sowie deren Flächenberechnung (siehe auch Kapitel 3.2). Das *Signal recognition* Modul stellt zusätzlich die Möglichkeit einer Flächenberechnung kontinuierlicher Signale zur Verfügung. Bei kontinuierlichen Signalen handelt es sich nicht um eine klassische Signalform, der Gauß-Verteilung, sondern um eine Veränderung der Basislinie nach oben oder unten, parallel zur Zeitachse. Diese Funktion ist bisher in der aktuellen Literatur noch für keine Software beschrieben und stellt für Achroma ein Alleinstellungsmerkmal dar.

Das *Signal recognition* Modul ist mit seinen Funktionen ein zentrales Modul von Achroma und ist Bestandteil jeder Auswertung von (kontinuierlichen) Signalen und sehr vielseitig einsetzbar. So wurde das Modul zur automatisierten Detektion von negativen, positiven (siehe auch Kapitel 3.2 und 3.4) und kontinuierlichen (siehe auch Kapitel 3.3) Signalen in allen durchgeführten Versuchen eingesetzt.

Aufgrund der regulatorischen Funktion von Enzymen in biologischen Systemen ist die Aktivität eines Enzyms und deren Beeinflussung durch Veränderung der physikalischen Eigenschaften, wie pH-Wert oder Temperatur, sowie durch regulatorische Moleküle, wie Inhibitoren oder konkurrierender Substrate von wissenschaftlicher Bedeutung. Diese

Einflüsse können sehr gut mit kontinuierlichen LC-MS Enzym-Assays erfasst werden, indem kontinuierlich die Produk Spur über die Zeit beobachtet wird und die Zeitpunkte an denen sich die im MS detektierte Intensität des Produktes ändert, näher untersucht werden. Achroma wurde speziell für diese Änderungen entwickelt. So konnte mit Hilfe von Achroma semi-quantitativ gezeigt werden, dass die Enzymaktivität mit steigender Temperatur zunimmt und ein Maximum erreicht, bis aufgrund der starken thermischen Belastung der Enzym-Moleküle vermutlich eine Denaturierung eintritt. Neben der erstmaligen Auswertung kontinuierlicher Signalfächen, wurde mit dem verwendeten System gezeigt, dass die Temperatur keinen Einfluss auf die Spezifität der Interaktion hat (siehe auch Kapitel 3.3). Der Einsatz unterschiedlicher Substrate aus unterschiedlichen Spezies, sowie die Inhibition mit unterschiedlichen Inhibitoren bestätigte die zuvor gewonnenen Ergebnisse. Somit kann Achroma Hilfestellung bei der Wahl der richtigen physikalischen Parameter für Versuchsreihen leisten.

Prinzipiell kann das oben beschriebene System auch zur Quantifizierung herangezogen werden (siehe ebenfalls Kapitel 3.3). Der in den durchgeführten Versuchen eingesetzte interne Standard war in seinen physikochemischen Eigenschaften zur quantitativen Bestimmung des Produkts dem Enzym-Assay noch nicht ähnlich genug, sodass sich bei der Erfassung mit Achroma und einer nachgelagerten Auswertung mit Microsoft Excel keine validen Ergebnisse ergaben. Allerdings ist die beschriebene Strategie für die Quantifizierung kontinuierlicher Signale (nicht Gauß-Verteilung) mit Hilfe von Achroma grundsätzlich nutzbar und zurzeit die einzig in der Literatur über eine Software beschriebene. In wie weit quantifizierbare Enzym-Assays mit passenden internen Standards zur Verfügung stehen, müssen weitere Versuche zeigen.

In Anwesenheit von konkurrierenden Substraten (siehe auch Kapitel 3.2) konnte mit Achroma gezeigt werden, dass die Produktbildung verändert wird. Die Intensität der Produk Spur nimmt temporär ab, bei einem gleichzeitigen Intensitätsanstieg des jeweiligen Substrates. Je höher die Konzentration und Affinität des konkurrierenden Substrates im Vergleich zum ursprünglichen Substrat ist, desto weniger Produkt entsteht von diesem. Generell ist der LC-MS Enzym-Assay in Verbindung mit Achroma eine hervorragende Methode, um Inhibition und Aktivierung der Produktbildung oder den Einfluss physikalischer und/oder chemischer Eigenschaften in einem frühen Stadium zu untersuchen.

Gerade in Verbindung mit dem *Spectra comparison* Modul stellt das *Signal recognition* Modul eine schnelle Auswertemethode auf neue m/z-Verhältnisse und somit neue Ionen zu einem bestimmten Zeitpunkt im Chromatogramm im Vergleich zu einem vorherigen Zeitpunkt dar. Das ist eine komfortable Funktion z.B. für das Inhibitorscreening in komplexen Proben (siehe auch Kapitel 3.2). So wurde einer komplexen Hausstaubprobe ein bekannter Inhibitor

zugegeben und über eine HPLC Trennung die einzelnen Komponenten aufgetrennt, dem Enzymsystem zugeführt und nach Interaktion mittels MS detektiert. Mit Achroma wurde dann das Chromatogramm der Produktpur des Enzym-Assays (AChE) auf Unregelmäßigkeiten untersucht. Eine Darstellung des Peak-Minimums im Vergleich zur Basislinie der Produktpur als Differenzspektrum zeigte ein neues m/z Verhältnis. Bei der Visualisierung dieses m/z-Verhältnisses als Chromatogramm zeigte sich ein positiver Peak genau zum Zeitpunkt des Signaleinbruchs. Das gefundene neue Signal im Differenzspektrum war dem m/z-Verhältnis des bekannten Inhibitors zuordenbar. Die Kombination *Signal recognition* Modul mit dem *Spectra comparison* Modul (siehe nachfolgend) ist für das schnelle Screening nach unbekanntem Komponenten einer komplexen biologischen Probe eine Neuentwicklung, die in der themenübergreifenden Literatur für die Auswertung (un-)typischer Signale bisher nicht beschrieben ist.

***Spectra comparison* Modul:**

Das *Spectra comparison* Modul ist ein eigenständiges Modul innerhalb Achroma, kann aber nur aus dem *Signal recognition* Modul heraus aufgerufen werden. Dieses Modul vergleicht zwei Spektren zu gegebenen Zeitpunkten eines Chromatogramms miteinander und bildet daraus ein Differenzspektrum. Anhand des Differenzspektrums lässt sich sehr schnell erkennen, welche m/z Verhältnisse zu den unterschiedlichen Zeitpunkten im MS neu hinzukommen bzw. nicht mehr vorhanden waren. Ist zum Zeitpunkt eines Signaleinbruchs in der Produktpur ein neues m/z Verhältnis im Spektrenvergleich sichtbar, so kann in weiteren Auswerteschritten eingegrenzt werden, ob es sich tatsächlich um einen Inhibitor handelt, z.B. durch Darstellung des Inhibitor-Chromatogramms mit Achroma. Hier bietet Achroma neue Möglichkeiten für schnelle Aussagen. Weiterführende Untersuchungen, ob es sich um einen bekannten oder neuen Inhibitor handelt können mit Achroma nicht gemacht werden. Dafür gibt es bereits eine Vielzahl von Software- und Datenbanklösungen die genutzt werden können - siehe auch Kapitel 1.4.

Innerhalb dieser Arbeit sollte neben der richtigen Strategie ebenso die Frage nach dem notwendigen Spezialisierungsgrad des/der Entwickler/-in beantwortet werden. Hierzu wurden kleine Teile der Achroma Software zusammen mit Studierenden umgesetzt, die ihr praktisches Semester unter meiner Anleitung absolvierten. Inklusive Einarbeitungszeit war es möglich, innerhalb von 4-6 Monaten ein kleines funktionsfähiges Modul in Achroma zu verankern. Somit ist diese Softwarestrategie auch für Forschergruppen interessant, die sich mit sehr neuen wissenschaftlichen Mess- und Auswertemethoden beschäftigen und u. U. selbst Teile der Software weiterentwickeln müssen. Es können so in sehr kurzer Zeit neue Module entweder durch Eigeninitiative oder Zusammenarbeit mit spezialisierten Einrichtungen entwickelt werden.

Die Softwareentwicklung ist einem rasanten Fortschritt, vielen Trends und neuen Technologien unterworfen. Aus diesem Grund wurde während der Durchführungsphase dieser Arbeit über einen Paradigmenwechsel nachgedacht; weg von C++ und der Model-View-Controller Architektur hin zu Java als Programmiersprache in Verbindung mit der Eclipse-Plattform und deren Plug-In basierter Architektur. So wurden mit zwei studentischen Projektgruppen die folgenden Fragen erörtert:

1. Wie lässt sich ein Umzug der Achroma Software nach Eclipse gestalten?
2. wie muss die Architektur und das Datensystem gestaltet werden, damit diese Plattform möglichst kompatibel ist?

Es konnte eine lauffähige Testversion erstellt werden, die in vereinfachter Form die Funktionalität von Achroma hatte. Die neue Projektidee, für Analytiker/-innen eine allesumfassende frei nutzbare Plattform zu konzipieren und implementieren, die alle möglichen Bereiche der Bioanalytik abdeckt und offen ist für jegliche Art von Erweiterung, würde in weiterführenden Arbeiten, zunächst auf Basis der Achroma Module umgesetzt werden (siehe auch openMASP-Projektidee, <http://openmasp.hswt.de>), Abbildung 12).

Aus heutiger Sicht sollte, neben der Entwicklung einer neuen Plattform, eine andere, dem *Open Source*-Gedanken dienende, Strategie zielführend sein. Dabei könnten einzelne Module von Achroma auch in Java umgesetzt werden und diese dann als Eclipse-Plug-Ins über den *Eclipse Market Place* der "Community" zur Verfügung gestellt werden. Der Vorteil dieser Strategie ist, dass keine Ressourcen für eine neu zu entwickelnde Plattform vorhanden sein müssen. Der Nachteil allerdings ist, dass die neuen Plug-Ins so konzipiert sein müssen, dass die Schnittstellen mit bereits vorhandenen Lösungen kompatibel sind und somit entsprechend Ressourcen in das Schnittstellenkonzept fließen müssen.

Die Gründung einer Science Working Group der Eclipse Foundation ^[125] in diesem Bereich war ein erster Schritt. In Kapitel 1.4 dieser Arbeit wird die Vielzahl der Lösungen dargestellt, die aktuell in der Literatur und im World Wide Web beschrieben sind. Die Science Working Group allerdings spezialisiert sich hauptsächlich auf mit Java entwickelte Eclipse RCP Projekte. Die Realisierbarkeit einer weitaus größeren Science Working Group, tatsächlich plattform- und programmiersprachenunabhängig zu untersuchen, wäre ein weiterer konsequenter Schritt in Richtung Interoperabilität und Austauschbarkeit von Daten.

Im Rahmen dieser Arbeit hat sich ebenso herausgestellt, dass die Entwicklung einer webbasierte Plattform sehr sinnvoll wäre (siehe auch Kapitel 1.4) die Bioanalytikern/-innen hilft, die richtige Software für die Auswertung ihrer Daten zu finden. So könnte z.B. über eine Stichworteingabe das World Wide Web automatisiert von der Plattform durchsucht werden,

welche Softwarelösungen am besten zum beschriebenen Versuch passen und die Plattform macht dann entsprechend Vorschläge zu geeigneten Softwarelösungen zur Auswertung der Daten.

Neben den vielen Plattformen rückt langsam die Cloud-Technologie in den Fokus der "omik"-Forschungsfelder, besonders da damit Rechnerleistungen gebündelt bzw. erhöht werden können, was für komplexe Berechnungen einen erheblichen Geschwindigkeitsvorteil bringt. Ebenso können mit dieser Technologie theoretisch sofort mit der weltweiten "Forschergemeinde" Ergebnisse ausgetauscht werden, da die Daten nicht mehr lokal vorliegen, sondern zentral auf einem Server oder in einem Rechenzentrum. Allerdings bestehen noch viele Bedenken hinsichtlich der rechtlichen Situation im Bereich der Datenspeicherung. "Analytik in der Cloud" können sich viele Wissenschaftler/-innen (noch) nicht vorstellen. Die Erforschung der Möglichkeiten, der Machbarkeit und des Potenzials der Cloud-Technologie für die "omik"-Forschungsfelder ist ein weites Feld für zukünftige Softwarestrategien.

Abschließend ist mit Achroma ein Software-Werkzeug entstanden, das im Forschungsfeld der Enzymfunktion und -regulation neue Möglichkeiten der Auswertung bietet, gerade für untypische massenspektrometrische Signale. In der themenübergreifenden Reflexion aktueller Literatur hat sich gezeigt, dass Achroma gerade für das Inhibitorscreening in komplexen biologischen Proben einige neue innovative Ansätze implementiert hat, die schnell zu Aussagen über die eingesetzte Probe führen. Achroma bietet ein Funktionsspektrum beginnend bei der Auswertung des klassischen LC-MS Enzym-Assays mit der Auswertung des Chromatogramms der Produktpur bis hin zu Signalen, die nicht der Gauß-Verteilung folgen. Ebenso ist die parallele Beobachtung der Intensitäten der eingesetzten internen Standards über das Quotientenchromatogramm eine neue und schnelle Methode, um Aussagen über die Systemstabilität machen zu können. Zu guter Letzt bietet Achroma aufgrund seiner Implementierung noch Raum zur Programmierung weiterer Module zur Auswertung untypischer Signale, die aus den Daten von neuen, innovativen experimentellen Setups gewonnen werden.

5 Ausblick

Die Adaption von LC-MS Systemen für Enzym-Assays hat die Qualität der Rohdaten eindeutig verbessert. Aufgrund der massenspektrometrischen Detektion können die an den Interaktionen beteiligten Moleküle simultan gemessen werden. Achroma ist ein Software-Werkzeug das gut in Bereichen wie der Biotechnologie, Pharmazie oder der Umwelttechnologie eingesetzt werden kann, da es mit seinen spezialisierten Modulen gerade beim Screening nach regulatorischen Molekülen von Enzymen in komplexen Proben eine Arbeitserleichterung darstellt.

Achroma ist aufgrund der gewählten Architektur jederzeit erweiterbar und kann gerade für neue Versuchsansätze eine gute Basis darstellen, wenn es um die Auswertung neuer, untypischer Daten im Bereich funktionelle Proteomik geht. Aktuell wird Achroma beim Inhibitorscreening realer Proben auf unbekannte Inhibitoren zusammen mit neuester Micro Array Chip Technologie (nano ESI Chips mit automatisierter MS-Detektion) und weiteren Anwendungen mit verschiedenen Enzymen eingesetzt.

Die Softwareentwicklung unterliegt allerdings einem rasanten Fortschritt. Daher ist es eine sinnvolle Überlegung, die einzelnen Module von Achroma auf neue Architekturen zu übertragen, wie z.B. ganz neue webbasierte Technologien oder die Cloud-Technologie. Gerade die noch junge Cloud-Technologie ist für rechenintensive Forschungsbereiche wie die (Prote)omik sehr interessant und sollte zukünftig mehr in den Fokus für neue Softwarestrategien rücken.

6 Referenzen

- [1] M. Chitty, "Biopharmaceutical -Omes & -omics glossary taxonomy", zu finden unter <http://www.genomicglossaries.com/content/omes.asp>, 01.06.2015.
- [2] "Alphabetically ordered list of omes and omics - Omics.org", zu finden unter http://omics.org/index.php/Alphabetically_ordered_list_of_omes_and_omics, 01.06.2015.
- [3] pubmeddev, "proteomic - PubMed - NCBI", zu finden unter <http://www.ncbi.nlm.nih.gov/pubmed/?term=proteomic>, 01.06.2015.
- [4] M. R. Wilkins, J. C. Sanchez, A. A. Gooley, R. D. Appel, I. Humphery-Smith, D. F. Hochstrasser, K. L. Williams, *Biotechnol Genet Eng Rev* **1996**, *13*, 19–50.
- [5] S. D. Patterson, R. H. Aebersold, *Nat Genet* **2003**, *33*, 311–323.
- [6] M.-S. Kim, S. M. Pinto, D. Getnet, R. S. Nirujogi, S. S. Manda, R. Chaerkady, A. K. Madugundu, D. S. Kelkar, R. Isserlin, S. Jain et al., *Nature* **2014**, *509*, 575–581.
- [7] M. Wilhelm, J. Schlegl, H. Hahne, A. Moghaddas Gholami, M. Lieberenz, M. M. Savitski, E. Ziegler, L. Butzmann, S. Gessulat, H. Marx et al., *Nature* **2014**, *509*, 582–587.
- [8] S. Naylor, R. Kumar, *Adv Protein Chem* **2003**, *65*, 217–248.
- [9] J. Godovac-Zimmermann, L. R. Brown, *Mass Spectrom Rev* **2001**, *20*, 1–57.
- [10] M. Yanagida, *J Chromatogr B Analyt Technol Biomed Life Sci* **2002**, *771*, 89–106.
- [11] T. C. Hunter, N. L. Andon, A. Koller, Yates III, John R., P. A. Haynes, *J Chromatogr B* **2002**, *782*, 165–181.
- [12] P. R. Graves, Haystead, T. A. J., *Microbiol Mol Biol Rev* **2002**, *66*, 39–63.
- [13] M. J. Hubbard, *Proteomics* **2002**, *2*, 1069–1078.
- [14] T. Kocher, G. Superti-Furga, *Nat Methods* **2007**, *4*, 807–815.
- [15] A. Sali, R. Glaeser, T. Earnest, W. Baumeister, *Nature* **2003**, *422*, 216–225.
- [16] S. Yokoyama, *Curr Opin Chem Bio* **2003**, *7*, 39–43.
- [17] M. J. Niphakis, B. F. Cravatt, *Annu Rev Biochem* **2014**, *83*, 341–377.
- [18] K. A. Stubbs, *Carbohydr Res* **2014**, *390*, 9–19.
- [19] L. E. Sanman, M. Bogyo, *Annu Rev Biochem* **2014**, *83*, 249–273.
- [20] H. Schmidinger, A. Hermetter, R. Birner-Gruenberger, *Amino Acids* **2006**, *30*, 333–350.
- [21] T. Yamamoto, T. Mori, Y. Katayama, *Curr Med Chem* **2014**, *21*, 2542–2552.
- [22] D.-H. Kim, T. Sim, *BMB Rep* **2010**, *43*, 711–719.
- [23] S. A. Johnson, T. Hunter, *Nat Methods* **2005**, *2*, 17–25.
- [24] D. Martins-de-Souza, *Dialogues Clin Neurosci* **2014**, *16*, 63–73.
- [25] V. Collura, G. Boissy, *Subcell Biochem* **2007**, *43*, 135–183.

- [26] pubmeddev, "enzyme - PubMed - NCBI", zu finden unter <http://www.ncbi.nlm.nih.gov/pubmed?term=enzyme>, 01.06.2015.
- [27] W. Stirling, *J Anat Physiol* **1884**, 18, 426–435.
- [28] *Enzyme nomenclature. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes*, Acad Press, San Diego [u.a.], **1992**.
- [29] "Historical Introduction", zu finden unter <http://www.chem.qmul.ac.uk/iubmb/enzyme/history.html>, 01.06.2015.
- [30] "Enzyme Database - BRENDA", zu finden unter http://www.brenda-enzymes.org/index.php4?page=information/all_enzymes.php4, 01.06.2015.
- [31] P. H. O'Farrell, *J Biol Chem* **1975**, 250, 4007–4021.
- [32] J. Klose, *Humangenetik* **1975**, 26, 231–243.
- [33] S. P. Gygi, G. L. Corthals, Y. Zhang, Y. Rochon, R. Aebersold, *Proc Natl Acad Sci U.S.A.* **2000**, 97, 9390–9395.
- [34] E. Appella, E. A. Padlan, D. F. Hunt, *EXS* **1995**, 73, 105–119.
- [35] A. J. Link, J. Eng, D. M. Schieltz, E. Carmack, G. J. Mize, D. R. Morris, B. M. Garvik, Yates, J R 3rd, *Nat Biotechnol* **1999**, 17, 676–682.
- [36] P. Mallick, B. Kuster, *Nat Biotechnol* **2010**, 28, 695–709.
- [37] R. Eisenthal, M. J. Danson, *Enzyme assays. A practical approach*, 2. Aufl., Oxford University Press, Oxford, OX, New York, **2002**.
- [38] H. Bisswanger, *Practical enzymology*, Wiley-VCH, Weinheim, **2004**.
- [39] T. Letzel, E. Sahmel-Schneider, K. Skriver, T. Ohnuma, T. Fukamizo, *Carbohydr Res* **2011**, 346, 863–866.
- [40] J. Grassmann, R. K. Scheerle, T. Letzel, *Anal Bioanal Chem* **2012**, 402, 625–645.
- [41] R. Smith, A. D. Mathis, D. Ventura, J. T. Prince, *BMC Bioinformatics* **2014**, 15, 9–23.
- [42] Benesch, Justin L P, C. V. Robinson, *Curr Opin Struct Biol* **2006**, 16, 245–251.
- [43] R. L. Beardsley, C. M. Jones, A. S. Galhena, V. H. Wysocki, *Anal Chem* **2009**, 81, 1347–1356.
- [44] T. Letzel, *Anal Bioanal Chem* **2008**, 390, 257–261.
- [45] K. D. Greis, *Mass Spectrom Rev* **2007**, 26, 324–339.
- [46] C. J. Malherbe, D. de Beer, E. Joubert, *Int J Mol Sci* **2012**, 13, 3101–3133.
- [47] K. F. Geoghegan, M. A. Kelly, *Mass Spectrom Rev* **2005**, 24, 347–366.
- [48] A. Liesener, U. Karst, *Anal Bioanal Chem* **2005**, 382, 1451–1464.
- [49] X. Ge, T. L. Sirich, M. K. Beyer, H. Desaire, J. A. Leary, *Anal Chem* **2001**, 73, 5078–5082.
- [50] D. J. Wilson, L. Konermann, *Anal Chem* **2004**, 76, 2537–2543.

- [51] E. D. Lee, W. Mueck, J. D. Henion, T. R. Covey, *J Am Chem Soc* **1989**, *111*, 4600–4604.
- [52] B. Ganem, Y. T. Li, J. D. Henion, *J Am Chem Soc* **1991**, *113*, 7818–7819.
- [53] S. M. Clark, L. Konermann, *Anal Chem* **2004**, *76*, 1257–1263.
- [54] N. Denhart, T. Letzel, *Anal Bioanal Chem* **2006**, *386*, 689–698.
- [55] N. Denhart, T. Fukamizo, R. Brzezinski, M.-E. Lacombe-Harvey, T. Letzel, *J Biotechnol* **2008**, *134*, 253–260.
- [56] N. Denhart, Weigang, Linda M. M., M. Fujiwara, T. Fukamizo, K. Skriver, T. Letzel, *J Biotec* **2009**, *143*, 274–283.
- [57] T. Taira, M. Fujiwara, N. Denhart, H. Hayashi, S. Onaga, T. Ohnuma, T. Letzel, S. Sakuda, T. Fukamizo, *Biochim Biophys Acta* **2010**, *1804*, 668–675.
- [58] R. K. Scheerle, J. Graßmann, T. Letzel, *Anal Methods* **2011**, *3*, 822–830.
- [59] Z. Li, F. Song, Z. Zhuang, D. Dunaway-Mariano, K. S. Anderson, *Anal Biochem* **2009**, *394*, 209–216.
- [60] de Boer, Arjen R, T. Letzel, van Elswijk, Danny A, H. Lingeman, Niessen, Wilfried M A, H. Irth, *Anal Chem* **2004**, *76*, 3155–3161.
- [61] B. Anandkumar, S. W. Haga, H.-F. Wu, *Comb Chem High Throughput Screen* **2014**, *17*, 173–182.
- [62] D. N. Perkins, D. J. Pappin, D. M. Creasy, J. S. Cottrell, *Electrophoresis* **1999**, *20*, 3551–3567.
- [63] M. Mann, P. Hojrup, P. Roepstorff, *Biol Mass Spectrom* **1993**, *22*, 338–345.
- [64] J. K. Eng, A. L. McCormack, J. R. Yates, *J Am Soc Mass Spectrom* **1994**, *5*, 976–989.
- [65] S. M. Bridges, G. B. Magee, N. Wang, W. P. Williams, S. C. Burgess, B. Nanduri, *BMC Bioinformatics* **2007**, *8 Suppl 7*, 24-33.
- [66] "ExpASY: SIB Bioinformatics Resource Portal - Categories", zu finden unter <http://www.expasy.org/proteomics>, 15.05.2015.
- [67] D. Kessner, M. Chambers, R. Burke, D. Agus, P. Mallick, *Bioinformatics* **2008**, *24*, 2534–2536.
- [68] P. Wenig, J. Odermatt, *BMC Bioinformatics* **2010**, *11*, 405-414.
- [69] A. Keller, J. Eng, N. Zhang, X.-j. Li, R. Aebersold, *Mol. Syst. Biol.* **2005**, *1*, 2005.0017.
- [70] O. Kohlbacher, K. Reinert, C. Gröpl, E. Lange, N. Pfeifer, O. Schulz-Trieglaff, M. Sturm, *Bioinformatics* **2007**, *23*, e191–e197.
- [71] C. H. Johnson, J. Ivanisevic, H. P. Benton, G. Siuzdak, *Anal Chem* **2015**, *87 (1)*, 147–156.
- [72] C. Guindon, "Eclipse Luna", zu finden unter <https://www.eclipse.org/>, 01.06.2015.
- [73] "Welcome to NetBeans", zu finden unter <https://netbeans.org/>, 01.06.2015.

- [74] "HUPO-PSI Working Groups and Outputs | HUPO Proteomics Standards Initiative", zu finden unter <http://www.psidev.info/overview>, 01.06.**2015**.
- [75] Pedrioli, Patrick G A, J. K. Eng, R. Hubley, M. Vogelzang, E. W. Deutsch, B. Raught, B. Pratt, E. Nilsson, R. H. Angeletti, R. Apweiler et al., *Nat Biotechnol* **2004**, *22*, 1459–1466.
- [76] M. Wilhelm, M. Kirchner, Steen, Judith A J, H. Steen, *Mol Cell Proteomics* **2012**, *11*, O111.011379.
- [77] "Unidata | NetCDF", zu finden unter <http://www.unidata.ucar.edu/software/netcdf/>, **2013**, 01.06.**2015**.
- [78] "Mass Spectrometry Software | Analyst Software | Triple Quadrupole TOF System Software", zu finden unter <http://www.absciex.com/products/software>, 01.06.**2015**.
- [79] "Software & Informatics | Agilent", zu finden unter <http://www.chem.agilent.com/en-US/products-services/Software-Informatics/Pages/default.aspx#>, 01.06.**2015**.
- [80] "Proteome Discoverer™ Software", zu finden unter <http://www.thermoscientific.de/product/proteome-discoverer-software.html>, 01.06.**2015**.
- [81] waters, "Mass Lynx Massenspektrometrie Software : Waters", zu finden unter http://www.waters.com/waters/de_DE/MassLynx-Mass-Spectrometry-Software-/nav.htm?cid=513164&locale=de_DE, 01.06.**2015**.
- [82] "GeneBio : Phenyx", zu finden unter <http://www.genebio.com/products/phenyx/index.html>, 01.06.**2015**.
- [83] "Overview - Proteomics Mass Spectrometry Software | PEAKS", zu finden unter <http://www.bioinform.com/peaks/features/overview.html>, 01.06.**2015**.
- [84] B. Ma, K. Zhang, C. Hendrie, C. Liang, M. Li, A. Doherty-Kirby, G. Lajoie, *Rapid Commun Mass Spectrom* **2003**, *17*, 2337–2342.
- [85] M. Krappmann, T. Letzel, *Anal Methods* **2012**, *4*, 1060–1071.
- [86] M. Krappmann, C. M. Kaufmann, R. K. Scheerle, J. Grassmann, T. Letzel, *J Proteomics Bioinform* **2014**, *07*, 264–271.
- [87] N. Hoffmann, M. Keck, H. Neuweger, M. Wilhelm, P. Högy, K. Niehaus, J. Stoye, *BMC Bioinformatics* **2012**, *13*, 214–234.
- [88] F. Rusconi, *Bioinformatics* **2009**, *25*, 2741–2742.
- [89] N. Jaitly, A. Mayampurath, K. Littlefield, J. N. Adkins, G. A. Anderson, R. D. Smith, *BMC Bioinformatics* **2009**, *10*, 87–102.
- [90] O. Spjuth, T. Helmus, E. L. Willighagen, S. Kuhn, M. Eklund, J. Wagener, P. Murray-Rust, C. Steinbeck, Wikberg, Jarl E S, *BMC Bioinformatics* **2007**, *8*, 59–69.
- [91] O. Spjuth, J. Alvarsson, A. Berg, M. Eklund, S. Kuhn, C. Mäsak, G. Torrance, J. Wagener, E. L. Willighagen, C. Steinbeck et al., *BMC Bioinformatics* **2009**, *10*, 397–402.

- [92] Y. Cao, N. Wang, X. Ying, A. Li, H. Wang, X. Zhang, W. Li, *BMC Med Inform Decis Mak* **2009**, 9, 13–22.
- [93] M. Sturm, A. Bertsch, C. Gröpl, A. Hildebrandt, R. Hussong, E. Lange, N. Pfeifer, O. Schulz-Trieglaff, A. Zerck, K. Reinert et al., *BMC Bioinformatics* **2008**, 9, 163-174.
- [94] B. Bunk, M. Kucklick, R. Jonas, R. Munch, M. Schobert, D. Jahn, K. Hiller, *Bioinformatics* **2006**, 22, 2962–2965.
- [95] M.-S. Kwon, H.-J. Lee, S.-K. Jeong, E.-Y. Lee, S. Y. Cho, Y.-K. Paik, *Proteomics* **2008**, 8, 4910–4913.
- [96] M. Strohal, M. Hassman, B. Kosata, M. Kodíček, *Rapid Commun Mass Spectrom* **2008**, 22, 905–908.
- [97] T. Pluskal, S. Castillo, A. Villar-Briones, M. Oresic, *BMC Bioinformatics* **2010**, 11, 395–406.
- [98] T. Bald, J. Barth, A. Niehues, M. Specht, M. Hippler, C. Fufezan, *Bioinformatics* **2012**, 28, 1052–1053.
- [99] V. Bessels, "MS-Spectre - Mass Spectrometry analysis software", zu finden unter <http://ms-spectre.sourceforge.net/news.html>, **2007**, 01.06.2015.
- [100] D. T. Duncan, R. Craig, A. J. Link, *J Proteome Res* **2005**, 4, 1842–1847.
- [101] M. Hirosawa, M. Hoshida, M. Ishikawa, T. Toya, *Comput Appl Biosci* **1993**, 9, 161–167.
- [102] L. Martens, H. Hermjakob, P. Jones, M. Adamski, C. Taylor, D. States, K. Gevaert, J. Vandekerckhove, R. Apweiler, *Proteomics* **2005**, 5, 3537–3545.
- [103] K. C. Leptos, D. A. Sarracino, J. D. Jaffe, B. Krastins, G. M. Church, *Proteomics* **2006**, 6, 1770–1782.
- [104] X.-j. Li, E. C. Yi, C. J. Kemp, H. Zhang, R. Aebersold, *Mol Cell Proteomics* **2005**, 4, 1328–1340.
- [105] M.-Y. K. Brusniak, S.-T. Kwok, M. Christiansen, D. Campbell, L. Reiter, P. Picotti, U. Kusebauch, H. Ramos, E. W. Deutsch, J. Chen et al., *BMC Bioinformatics* **2011**, 12, 78–93.
- [106] B. Maclean, D. M. Tomazela, N. Shulman, M. Chambers, G. L. Finney, B. Frewen, R. Kern, D. L. Tabb, D. C. Liebler, M. J. Maccoss, *Bioinformatics* **2010**, 26, 966–968.
- [107] M. Bellew, M. Coram, M. Fitzgibbon, M. Igra, T. Randolph, P. Wang, D. May, J. Eng, R. Fang, C. Lin et al., *Bioinformatics* **2006**, 22, 1902–1909.
- [108] L. N. Mueller, O. Rinner, A. Schmidt, S. Letarte, B. Bodenmiller, M.-Y. Brusniak, O. Vitek, R. Aebersold, M. Müller, *Proteomics* **2007**, 7, 3470–3480.
- [109] H. Barsnes, S.-O. Mikalsen, I. Eidhammer, *BMC Bioinformatics* **2006**, 7, 42-51.
- [110] T. Farrah, E. W. Deutsch, R. Kreisberg, Z. Sun, D. S. Campbell, L. Mendoza, U. Kusebauch, M.-Y. Brusniak, R. Huttenhain, R. Schiess et al., *Proteomics* **2012**, 12, 1170–1175.

- [111] K. Helsens, M.-Y. Brusniak, E. Deutsch, R. L. Moritz, L. Martens, *J Proteome Res* **2011**, *10*, 5260–5263.
- [112] S. Liu, L. Liu, U. Uzuner, X. Zhou, M. Gu, W. Shi, Y. Zhang, S. Y. Dai, J. S. Yuan, *BMC Bioinformatics* **2011**, *12 Suppl 1*, 43–53.
- [113] Y. Mohammed, D. Domanski, A. M. Jackson, D. S. Smith, A. M. Deelder, M. Palmblad, C. H. Borchers, *J Proteomics* **2014**, *106C*, 151–161.
- [114] H. Lam, E. W. Deutsch, J. S. Eddes, J. K. Eng, N. King, S. E. Stein, R. Aebersold, *Proteomics* **2007**, *7*, 655–667.
- [115] L. Y. Geer, S. P. Markey, J. A. Kowalak, L. Wagner, M. Xu, D. M. Maynard, X. Yang, W. Shi, S. H. Bryant, *J Proteome Res* **2004**, *3*, 958–964.
- [116] J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, M. Mann, *J Proteome Res* **2011**, *10*, 1794–1805.
- [117] J. Cox, M. Mann, *Nat Biotechnol* **2008**, *26*, 1367–1372.
- [118] P. Gärdén, R. Alm, J. Häkkinen, *Bioinformatics* **2005**, *21*, 2085–2087.
- [119] D. L. Tabb, C. G. Fernando, M. C. Chambers, *J Proteome Res* **2007**, *6*, 654–661.
- [120] N. H. Schebb, T. Vielhaber, A. Jousset, U. Karst, *J Chrom A* **2009**, *1216*, 4407–4415.
- [121] C. M. Kaufmann, J. Graßmann, D. Treutter, T. Letzel, *Rapid Commun Mass Spectrom* **2014**, *28*, 869–878.
- [122] R. K. Scheerle, J. Grassmann, T. Letzel, *Anal Sci* **2012**, *28*, 607–612.
- [123] A. Savitzky, Golay, M. J. E., *Anal Chem* **1964**, *36*, 1627–1639.
- [124] A. R. de Boer, T. Letzel, H. Lingeman, H. Irth, *Anal Bioanal Chem* **2005**, *381*, 647–655.
- [125] "Science WG - Eclipsepedia", zu finden unter https://wiki.eclipse.org/Science_WG, 01.06.2015.

Danksagung

Ich bedanke mich bei allen die mich bei dieser Arbeit unterstützt haben, ganz besonders bei meinem Betreuer Priv.-Doz. Dr. Thomas Letzel und seiner Arbeitsgruppe und vor allem bei meinen beiden Mit-Doktorandinnen Romy Scheerle und Christine Kaufmann für die Zusammenarbeit und die gegenseitige Unterstützung.

Herrn Prof. Dr. Küster danke ich für die Zweitbegutachtung und die hilfreiche Kritik, die wichtig war, um die Arbeit abzurunden und Herrn Prof. Dr. Langosch danke ich für die Übernahme des Prüfungsvorsitzes.

Ich bedanke mich ebenfalls bei Prof. Dr. Frank Leßke, der mich seitens der Hochschule Weihenstephan-Triesdorf bei dieser Arbeit betreut hat und immer ein guter Ratgeber war und ist.

Ebenso bedanke ich mich bei Marco Luthardt, mit dem ich so manche Diskussion über neue Softwaretechnologien führen durfte.

Ein herzliches Dankeschön gilt auch allen Studierenden, die während ihrer praktischen Studiensemester sowie in den Projektgruppen einen kleinen aber wichtigen Teil zum Gelingen der Software und zu den Erkenntnissen über die Eclipse Plattform beigetragen haben: Wolfgang Hollweck, Oliver Schmidt, Andrea Heidenreich, Patrick Schulte-Middelich, Manuel Erhardt, Yassine Azyrit, Tim Codanda, Daniel Fink, Gabriele Härtinger, Juri Ivanov, Maria Kreitmayr, Mathias Lehmann, Andreas Nürnberg, Martin Scholz, Marcel Suttner, Reinhard Wahl und Ludwig Wildner.

Vor allem bedanke ich mich bei meiner Frau für die Geduld und den Rückhalt, besonders in den "schwierigen Zeiten" dieser Arbeit. Bei meinem kleinen Sohn möchte ich mich für die Freude bedanken, die er mir jeden Tag schenkt. Es gibt nichts schöneres, als sein Lachen und seine strahlenden Augen, wenn wir gemeinsam spielen.

Ich danke meinen Eltern, die mich immer unterstützt haben, unabhängig davon, ob sie meine Entscheidungen richtig oder falsch fanden. Schlussendlich hatte ich den Freiraum meinen Weg zu gehen und dafür danke ich Euch von ganzem Herzen.

Liste der Publikationen, Vorträge und Posterbeiträge

Begutachtete Artikel in Fachzeitschriften

- **Krappmann M**, Luthardt M, Lesske F, Letzel T (2015) The Software-Landscape in (Prote)Omic Research. *J Proteomics Bioinform* 8:164-175. doi: 10.4172/jpb.1000365
- **Krappmann M**, Kaufmann CM, Scheerle RK, Grassmann J, Letzel T (2014) Achroma Software-High-Quality Policy in (a-)Typical Mass Spectrometric Data Handling and Applied Functional Proteomics. *J Proteomics Bioinform* 7:264-271. doi: 10.4172/jpb.1000328.
- **Michael Krappmann** and Thomas Letzel: "Achroma: a software strategy for analysing (a-)typical mass-spectrometric data", *Journal Article Anal. Methods*, 2012, 4, 1060-1071, DOI: 10.1039/C2AY05622D.

Eingereichte Artikel in Fachzeitschriften:

- **Michael Krappmann**, Arjen R. de Boer, Daniël R.W. Kool, Hubertus Irth and Thomas Letzel, (2015) *Mass Spectrometric Real-Time Monitoring of an Enzymatic Phosphorylation Assay using Internal Standards and Data Handling Freeware*, *Rapid Communication in Mass Spectrometry*.

Buchkapitel

- **Michael Krappmann** and Thomas Letzel, Chapter 11: "Functional Analysis of Proteins, Including LC-MS and Special Freeware", Book Chapter in *Protein and Peptide Analysis by LC-MS: Experimental Strategies*, 2011, 142-155, DOI:10.1039/9781849733144-00142

Posterbeiträge:

- Luthardt M., **Krappmann M.**, Hoffmann N., Wenig P, 2013: CDSclipse - Proposal for an open Chromatography Data System, *EclipseCon Europe*, Ludwigsburg - Deutschland.
- **Krappmann M.**, Leßke F., Letzel T., 2013: Achroma - a software strategy for analysing (a -)typical mass spectrometric data, *Anakon-Tagung*, Essen - Deutschland.
- Luthardt M., **Krappmann M.**, Leßke F., Letzel T., 2012: openMASP – high performance software technology for analytical chemistry, *Umweltsymposium*, Leipzig - Deutschland.
- **Krappmann M.**, Luthardt M., Wenig P., Leßke F., Letzel T., 2011: Cloudy Analytics – High Performance Software Technology for Analytical Chemistry, *Langenauer Wassertage*, Langenau - Deutschland.

- **Krappmann M.**, Letzel T., 2011: NEED NEW EXPERIMENTS NEW SOFTWARE SOLUTIONS?, *Anakon-Tagung*, Zürich - Schweiz.

Vorträge:

- **Krappmann M.**, Letzel T., 2014: Achroma → openMASP, (a)typical LC-MS data → Open Source Analytical Software Platform, *Doktorandenseminar - TUM - Lehrstuhl für Siedlungswasserwirtschaft*, Garching - Deutschland.
- **Krappmann M.**, Letzel T., 2013: openMASP, was bisher geschah..., *Workshop: RiskIdent*, Freising - Deutschland.
- **Krappmann M.**, Letzel T., 2012: openMASP oder die Geschichte einer Idee, *Akademischer Nachmittag der Fakultät Biotechnologie und Bioinformatik der Hochschule Weihenstephan-Triesdorf*, Freising - Deutschland.
- Scheerle R. K., Kaufmann C., **Krappmann M.**, Graßmann J., Letzel T., 2012: How to handle huge amount of mass spectrometric data? A new strategy with the Achroma software, *Joint Conference of the Polish Mass Spectrometry Society and the German Mass Spectrometry Society (DGMS)*, Poznan - Polen.
- **Krappmann M.**, 2012: openMASP – the history of openMASP, *1. Treffen der Science Working Group der Eclipse Foundation*, Hamburg - Deutschland.
- **Krappmann M.**, Letzel T., 2012: Functional Analysis of Proteins – Monitoring of non covalent Complexes, *Doktorandenseminar – TUM – Lehrstuhl für chemisch technische Analyse*, Freising - Deutschland.
- **Krappmann M.**, Letzel T., 2011: Vom Problem – Zur Idee – Zu openMASP, *Erster nationaler openMASP-Workshop*, Freising - Deutschland.
- **Krappmann M.**, Letzel T., 2011: openMASP - a Rich Client Platform (RCP) application, *Doktorandenseminar – TUM – Lehrstuhl für chemisch technische Analyse*, Freising - Deutschland.
- **Krappmann M.**, Letzel T., 2010: Development of a modular software platform for universal analysis and presentation of data in modern analytical laboratories, *Doktorandenseminar – TUM – Lehrstuhl für chemisch technische Analyse*, Freising - Deutschland.

Appendix

Krappmann M, Luthardt M, Lesske F, Letzel T (2015) The Software-Landscape in (Prote)Omic Research. J Proteomics Bioinform 8:164-175. doi: 10.4172/jpb.1000365

OPEN-ACCESS Artikel - Copyright: 2015 © Krappmann M, et al.

<http://www.omicsonline.org/open-access/the-softwarelandscape-in-proteomic-research-jpb-1000365.php?aid=56242>

Die Inhalte und das Konzept des Reviews wurden federführend von mir verfasst. Herr Prof. Dr. Leßke und Herr Luthardt haben im Kapitel zu den Programmiersprachen Inhalte zur Veröffentlichung beigetragen und Herr PD Dr. Letzel im analytischen Bereich, sowie die finale Korrektur und Einreichung des Manuskripts übernommen.

The Software-Landscape in (Prote)Omic Research

Michael Krappmann¹, Marco Luthardt¹, Frank Lesske¹ and Thomas Letzel^{2*}

¹University of Applied Sciences, Am Hofgarten 4, 85354 Freising, Weihenstephan, Germany

²Analytical Research Group, Chair of Urban Water Systems Engineering, TU München, Am Coulombwall, 85748 Garching, Germany

Abstract

(Bio)Informatics plays a major role in (prote)omic research experiments and applications. Analysis of an entire proteome including protein identification, protein quantification, detecting biological pathways, metabolite identification and others is not possible without software solutions for analyzing the resulting huge data sets. In the last decade plenty of software-tools, -platforms and databases have been developed by vendors of analytical hardware, as well as by freeware developers and the open source software community. Some of these software packages are very much specialized for one (omic) topic, as for example genomics, proteomics, interactomics or metabolomics. Other software tools and platforms can be applied in a more general manner, e.g. for generating workflows, or performing data conversion and data management, or statistics. Nowadays the main problem is not to find out a way, how to analyze the experimental data, but to identify the most suitable software for this purpose in the vast software-landscape.

This review focuses on the following issue: How complex is the link between biology, analysis and (bio)informatics, and how complex is the variety of software tools to be used for scientific investigations, starting from microorganisms up to the detection of a proteome. Thereby the main emphasis is on the variety in software for (LC) MS(/MS) proteomics. In the World Wide Web sites like ExPASy show extensive lists of proteomics software, leaving it to the user to identify which software actually serves their purposes.

First we consider the huge variability of software in the field of proteomics research. Then we take a closer look on the variability of MS data and the incompatibilities of software tools with respect to that. We give an overview over commonly used software technologies and finally end up with the question, whether open source software would not add more value to this field.

Keywords: Proteomics; Informatics; Language; Java; C++; Pipeline; Database

Understanding the complex mechanisms in (micro)organisms has been developed to a big challenge, called system biology. The included “-omic” research fields received a wide focus during the recent years. New terms were formed, like Genome, Transcriptome, Proteome, Metabolome; in total more than 140 [1] and 250 [2] terms ending on the suffix “-ome” or “-omics”- have been counted so far. The progress of analytical technologies therein and the possibility of getting more sensitive accurate and robust results was a prerequisite to gain an in-depth insight into organisms.

Not only organisms are complex, but also is the analytical hardware to obtain data just as the software landscape that helps in data interpretation. Figure 1 exemplarily represents, how complex the possibilities of analysing the “-omes” and their data workout can be [3]. To analyse the proteome of a microorganism for example, proteins can be separated with liquid chromatography (LC) or 2D electrophoresis (2DE). Using software, like Melanie or 2DHunt, the results can be interpreted and further conclusions about the proteome can be drawn. An alternative approach employs the fragmentation of proteins (i.e. their tryptic peptides) in mass spectrometry (i.e. LC-MS), using software like MASCOT, PEPSEA or Peptidmass for getting information about the proteome of the microorganism (as shown in Figure 1).

A good example for the integration of new informatics and technologies in “-ome” research fields was shown in a recently published review [4]. The review describes the dependency of metabolomics research from bioinformatics such as: streamlining data acquisition (e.g., data alignment, automated metabolomics, and cloud based metabolomics), feature analysis (e.g. mass spectral annotations, statistical analysis, and targeted validation), pathway analysis and the

biological context. It appears that bioinformatics help researchers to identify metabolite features from LC-MS data and to describe their biological roles by identifying their involvement in chemical pathways.

Using the example of mass spectrometric technologies associated with various software tools in proteomics, we want to demonstrate the general complexity of the software landscape in “-omics” research up to system biology. The term proteome occurred in 1994 and since that time the amount of publications rose extremely. In the year 2014 almost 7.400 publications with the keyword “proteomic” were listed in Pubmed (Figure 2a), about 2.800 publications with the keywords “proteomic” and “mass spectrometry”, about 350 with “proteomic” and “software” and about 200 including all three terms “proteomic”, “mass spectrometry” and “software”. In all categories the publications have been doubled since 2004. This clearly shows that the combination of analytical methods for proteomics and software development meanwhile evolved into an important field of research, resulting in a large number of available software tools.

Bioinformatics has a wide range of application in proteomics-

***Corresponding author:** Thomas Letzel, Analytical Research Group, Chair of Urban Water Systems Engineering, TU München, Am Coulombwall, 85748 Garching, Germany, Tel: +49 (0)89 2891 3780; Fax: +49 (0)89 - 289-13718; E-mail: T.Letzel@tum.de

Received May 28, 2015; Accepted July 13, 2015; Published July 17, 2015

Citation: Krappmann M, Luthardt M, Lesske F, Letzel T (2015) The Software-Landscape in (Prote)Omic Research. J Proteomics Bioinform 8: 164-175. doi:[10.4172/jpb.1000365](http://dx.doi.org/10.4172/jpb.1000365)

Copyright: © 2015 Krappmann M, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

In recent years, the Proteomic Standard Initiative (PSI) of the Human Proteome Organization (HUPO) has developed various standards (e.g. mzdata, see Table 1 “1 data formats”) based on the “Extensible Markup Language” (XML), a text-based language for describing hierarchical data. XML-formats are very well suited for data exchange due to the structured composition of these text documents (i.e. clearly defined elements with markers for beginning and end, respectively).

Nowadays, several mass spectrometer manufacturers offer software solutions for evaluation of proteomic experiments in combination with their devices (Table 1 “2.commercial software”). Commercial software solutions generally work with proprietary file formats. But due to the large “proteomics community” and its pressure, many commercial manufacturers and distributors offer an exchange format based on XML (see above) and/or deliver their software together with a so called data converter. This necessity arose from the fact that many laboratories and institutes kept developing new methods for analysing MS- or MS/MS- data, that, however, the standard software could not yet support. Thereby new algorithms and specific software tools were developed. The proprietary data formats had to be decoded in order to be able to test the new algorithms with existing data. Through constant development, free data converters were created, such as ReAdW (.raw-converter), mzWiff (.wiff-converter), MassWolf (.raw-Konverter) and Trapper (.d directories data-converter). These were all transferred to the software msconvert, which is part of the ProteoWizart libraries [5] (Table 1). Still there are other open-source solutions, like OpenChrom [6], that offer more formats specific to certain manufacturers.

The demand for new software is often caused by the needs of users from individual research groups. They work on very specific topics and though search for automated solutions to evaluate their data. Usually these users do not wish to evaluate the data manually with a table calculation program like Microsoft Excel. Therefore purpose-built software tools and algorithms are developed. An example for very specialized software is Achroma, which was developed for the evaluation of continuous flow LC-MS enzymatic assays in the field of functional proteomics [7,8]. The combination of well-known software/algorithms and these specific ones could however be extremely valuable. This is a big challenge for the future, especially to enable the interoperability between these software platforms and tools.

To some extent, interoperability is already provided by data exchange formats (Table 1 “1. Data formats”). But how can this applied to the software tools and platforms? For this purpose it makes sense to examine the programs from the point of software development view. Due to the large number of different solutions it is not possible to unite all solution in one platform. On the other hand there are of course many solutions which use the same or similar approaches for evaluating data. They resemble another in terms of their algorithms or their overlap in solution approaches. Amongst this multitude of solutions some can be found with a function as “pipeline”, like the Trans-Proteomic Pipeline (TPP)-software [9] (Figure 3) or TOPP [10]. These “pipelines” are specialised on connecting previously existing tools into a processing chain automatically executed in sequence.

Software	Developer/Vendor	Publication Date	Relation to Proteomics	Programming language	Platform
1. Data Formats					
mzData [23]	HUPO PSI	2007	Standardized data-exchange format	XML	independent
mzXML [24]	Seattle Proteome Center; Institute for Systems Biology	2004	Standardized data-exchange format	XML	independent
mzML [23] (replacement of mzXML u. mzData)	HUPO PSI	2008	Standardized data-exchange format	XML	independent
TraML [23]	HUPO PSI	2010	Standardized data-exchange format	XML	independent
mzIdentML [23]	HUPO PSI	2009	Standardized data-exchange format	XML	independent
mzQuantML [23]	HUPO PSI	2011	Standardized data-exchange format	XML	independent
mz5 [25]	Boston Children's Hospital; Steen & Steen Laboratory	2012	Storage-optimized data-exchange format	XML, HDF5	independent
netCDF [26]	UCAR	1997	Standardized data-exchange format	binary	independent
2. Commercial Software					
Analyst	ABSciex [27]	-	Automated MS a. MS/MS Detection, MRM Monitoring, Database-search possible	C, C++	Windows
SWATH Acquisition	ABSciex	-	Data-independent data capture, MRM, High resolution, MS/MS, label-free Quantification	C, C++	Windows
MetabolitePilot	ABSciex	-	Detection and identification of metabolites (drug design)	C, C++	Windows
ProteinPilot	ABSciex	-	Protein-identification a. -quantification	C, C++	Windows
MultiQuant	ABSciex	-	MRM, Quantification	C, C++	Windows
PeakView	ABSciex	-	TripleTOF, Mass interpretation, Accurate mass, Protein structure	C, C++	Windows
Masshunter	Agilent [28]	-	MS-Detection platform of Agilent, Peak detection, -integration	C, C++	Windows
LCMS ChemStation	Agilent	-	High-throughput quantification, Peak detection, NIST-search	C, C++	Windows
OpenLAB	Agilent	-	Data center, Crosslinkage between analytical hardware and software of Agilent	C, C++	Windows
Proteome Discoverer	Thermo [29]	-	Extensible platform for qualitative a. quantitative Proteomics	C, C++	Windows
Xcalibur	Thermo Fisher Scientific Inc.	-	Data system for device control and data analysis	C, C++	Windows
ProMass	Thermo Fisher Scientific Inc.	-	Automated biomolecule deconvolution	C, C++	Windows
MassLynx	Waters [30]	-	Platform for MS-detection of Waters	C, C++	Windows

OpenLynx	Waters	-	Qualitative screening a. identification	C, C++	Windows
QuanLynx	Waters	-	Automated Quantification	C, C++	Windows
ProteinLynx Global SERVER	Waters	-	Platform for qualitative a. quantitative proteomics	C, C++	Windows
Phenyx [31]	GeneBio	2006	MS/MS Protein identification, -quantification	-	web based
Seaquest [32] [33]	Yates Lab	1993/1994	Protein identification algorithm throug MS-data	-	Windwos, Linux
PEAKS [34] [35]	Bioinformatics Solutions Inc.	2003	De Novo sequencing, Data base search, PTMs, SPIDER-search, Quatnification, Protein/Peptid identification	C, C++	Windows
3. Free/Open Source Platforms and Software-Tools					
Achroma [7] [8]	Weihenstephan-Triesdorf - University of Applied Sciences	2012	LC-MS, Functional proteomics	C++, C#, Microsoft NET Framework a. Visual Studio 2003	Windows
OpenChrom (free Community Edition) [9]	University of Hamburg, Lablicate	2010	GC/MS, LC/MS, HPLC-MS, ICP-MS, MALDI-MS	Java, Eclipse	independent
Maltcms/Maui [20]	University of Bielefeld	2012	LC-MS-, GS-MS- u. GC-MS/MS-high-throughput metabolomics	Java, NetBeans	independent
polyXmass / massXpert [36]	Filippo Rusconi	2001 / 2009	Simulation of biological a. chemical reactions, Fragmentation	C++, Qt-Bibliotheken	independent
DeconLS [37]	Pacific Northwest National Laboratory	2009	Detection of molecule properties with MS-data and isotopes signatures, Deconvolution	C++, C#, Microsoft NET Framework a. Visual Studio 2003	Windows
BioClipse [38][39]	Dept. of Pharmaceutical Biosciences, Uppsala University; European Bioinformatics Institute	2007	Tool box for bio- a. chemoinformatics	Java, Eclipse	independent
BioSunMS [40]	Center of Computational Biology, Beijing Institute of Basic Medical Sciences	2009	Peptide profiles via MS	Java, R, SQL	Windows, Linux
openMS [41]	University of Tübingen, Free University of Berlin, ETH Zürich	2008	LC-MS data management a. -analytics, Label-free quantification, SILAC, iTRAQ, SRM, SWATH, De-novo identification, Data base search	C++	Windows, Linux
MetaQuant [42]	Bioinformatics Organization	2006	Automated quantification of GC/MS based Metabolome-data	Java	independent
intelliMS [43]	Yonsei Proteome Research Center	2008	Management a. Visualization of tandem-MS data	PHP, MySQL	web based
mMass [44]	Martin Strohm	2008	Protein- a lipid-identification, Protein modification, Deconvolution, Data base search, Standard-MS evaluations, e.g. peak detection, Spectrum-Viewer,...	Python, wxPython	independent
MZMine2 [45]	Okinawa Institute of Science and Technology	2010	Peak-identification of MS-data, Peak deconvolution, 3D-Visualization, Statistics, Data base search	Java, Java3D	independent
pymzML [46]	University of Münster	2012	High-throughput bioinformatics with MS-data	Python	independent
MS-Spectre [47]	Netherlands Proteomics Centre	2007	Spectra visualization, Filter, Peak detection, Peak matching	Java, Eclipse	independent
ProteoWizard [8]	ProteoWizard Software Foundation	2008	Proteomics a. MS software toolbox (msconvert converting tool for vendor formats into mzML)	C++	independent
X!Tandem / X!!Tandem [48]	The Global Proteome Machine Organization	2004	Data base search (algorithm) for matching tandem-MS sequences with peptide sequences for protein identification	C	Linux
Mascot (Server) [49] [50] (free search for 1200 spectra and limited functions)	Matrix Science	1993	Data base search (algorithm) for Protein identification, characterization a. quantification via MS-data	-	Windows, Linux
PRIDE [51]	EMBL-EBI	2005	Data base search (algorithm) for Protein identification	Java, HTML; SQL, XML	Windows, Linux a. web based
TOPP [11]	University of Tübingen	2007	OpenMS Proteomik Pipeline for batch processing of analytical tasks	C++	Windows, Linux
MapQuant [52]	Havard Molecular Technologies	2006	Large-scale Protein quantification	C++	Windows, Linux
SpecArray [53]	Seattle Proteome Center	2005	Comparison of peptide-arrays with LC-MS data	C	Linux
ATAQS [54]	Seattle Proteome Center	2011	Automated a. targeted analysis with quantitative SRM a. MS	Java, R, mysql	Linux
Skyline [55]	MacCoss Lab Software	2010	Document editor for designing a. analysing targeted proteomic experiments	C#, Windwos Forms	Windows
msInspect [56]	Fred Hutchinson Cancer Research Center	2006	Analysing complex protein mixtures with LC-MS	Java	independent

SuperHirn [57]	Seattle Proteome Center	2007	LC-MS peptide-/protein-profiling	C++	Linux, MacOS X
MassSorter [58]	FUGE Bioinformatics Platform	2006	Analysing of MS protein experiments mit known amino acids	Java, NetBeans, Java3D	independent
PASSEL [59]	Institute of System Biology, Seattle	2012	PeptideAtlas SRM experiment library	-	web based
jTraML [60]	CompOmics	2011	Java API of TraML-Standard of the Proteome Standards Initiative, SRM experiments	Java	independent
HDX-analyzer [61]	Texas A&M University	2011	Package for statistical analysis of protein structure dynamics	Python, R	independent
PeptidePicker [62]	Genome British Columbia Proteomics Center	2014	Toolbox for selection of possible peptides for targeted proteomics	-	web based
TPP (Trans-Proteomic Pipeline) [10]	Seattle Proteome Center	2005	MS/MS based "shotgun"-proteomics, Identification, Quantification	C++, Python	independent
SpectraST [63]	Institute of System Biology, Seattle	2007	Search algorithm for spectra libraries for identification of peptides via MS/MS	a part of TPP in Python	independent
OMSSA (Open Mass Spectrometry Search Algorithm) [64]	National Center of Biotechnology Information	2004	Data base search algorithm for peptide identification via MS/MS	-	independent
Andromeda [65] (Part of MaxQuant)	Max Planck Institute of Biochemistry	2011	Data base search engine for peptide identification via MS/MS	C#	web based, local installation: Windows
MaxQuant [66]	Max Planck Institute of Biochemistry	2008	Quantitative proteomics via MS	C#	Windows
Proteios SE [67]	University Lund	2005	Platform for storing, organizing, analyzing a. annotation of protein experiments	Java	independent
MyriMatch [68]	Vanderbilt University Medical Center	2006	Data base search engine for peptide identification via MS/MS	C++	Windows, Linux
PAGIT [69]	Sanger Institute	2012	Tools to generate automatically high quality sequence by ordering contigs, closing gaps, correcting sequence errors and transferring annotation	C++	Linux, virtual machine (for other operating systems)
MUMmer [70]	J. Craig Venter Institute	1999,2004	Ultra-fast alignment of large-scale DNA and protein sequences	C++, Python	Linux
VeSPA [71]	Computational Biology and Bioinformatics, Pacific Northwest National Laboratory	2012	Facilitation of genomic annotation of prokaryotic organisms through integration of proteomic and transcriptomic data	Python	independent
MINT [72]	University of Rome	2006	Molecular interaction database, focus on experimentally verified protein-protein interactions	Java	web based
Mass Bank [73]	Institute for Advanced Biosciences, Keio University	2010	Data base for mass spectra of small chemical compounds for life sciences (<3000 Da)	Java	web based
MSEA [74]	University of Alberta	2010	Tool to identify biological patterns in quantitative metabolomic data	Java, R	web based
XANNpred [75]	University of Dundee	2010	Crystallisation propensity predictor	Perl	web based

Table 1: List of software tools and platforms for the evaluation of proteomic data got from mass spectrometer detection.

Large Number and High Variability of Software Tools for Data Evaluation

In relation to the tools depicted in Figure 1, the distinction must be made that entire software systems and "small" software tools exist for particular tasks, such as individual specific algorithms, for example search algorithms for protein databases. Many companies offer complete software packages which employ an easy-to-use graphic user interface and enable the execution of several steps within a single software system. In order to avoid blurring only a few tools are depicted in this figure. In reality these relations are far more complex.

The proteomics website of ExPASy [11] gives a good overview of the diversity of the software landscape in proteomics. Under the heading of proteomics, it lists 31 databases and 240 software tools in eight categories: 1) protein sequence identification, 2) mass spectrometry and 2-DE data, 3) protein characterisation and function, 4) families, patterns and profiles, 5) post-translational modifications,

6) protein structure, 7) protein-protein interaction, and 8) similarity search/alignment. But the whole list at ExPASy, does not represent the complete range. For example, the proteomic tools site [12] of the Seattle Proteome Center contains 32 software tools and the software site [13] of the Pacific Northwest National Laboratory 52 softwares. Beside the software systems belonging to manufacturers of appliances such as mass spectrometers, there is by now a large open source community and many freely accessible software tools (freeware) which often have their origin in research institutes or universities. These tools are mostly not listed by ExPASy, since they are highly specific or have not yet reached the level of popularity required for a respective linkage. In addition, there are many specific categories in the free online-encyclopaedia Wikipedia list of proteomics software. An example for this is the "List of mass spectrometry software" [14]. It contains 76 commercial and free software systems, categorized into three main groups ("Proteomics software", "MS/MS peptide quantification" and "Other software"). Another Wikipedia Site is "ms-utils" [15] (229 softwares listed). The listed software in Wikipedia is not a reliable scientific source, but it

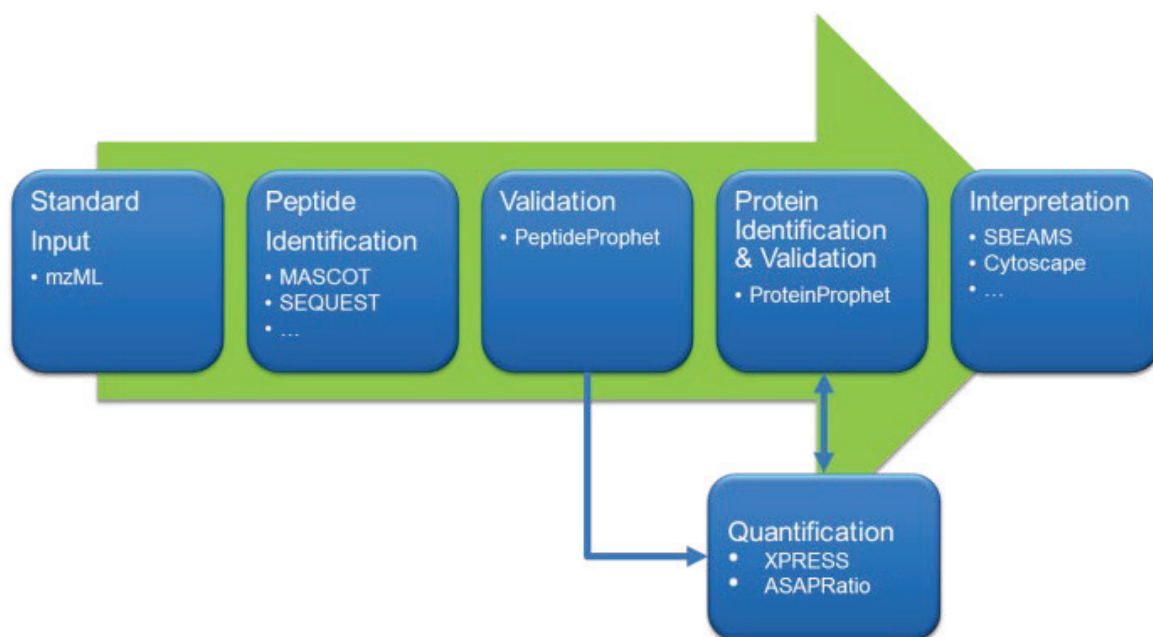


Figure 3: Sequence of an automated processing workflow of various software solutions (bullet points) using the example of Trans Proteomic Pipeline (TPP) [9].

Software	Link
mzData	http://www.psivdev.info/mzdata-1_0_5-docs
mzXML	http://sashimi.sourceforge.net/schema_revision/
mzML (replacement of mzXML u. mzData)	http://www.psivdev.info/mzml_1_0_0%20
TraML	http://www.psivdev.info/traml
mzIdentML	http://www.psivdev.info/mzidentml
mzQuantML	http://www.psivdev.info/mzquantml
mz5	http://software.steenlab.org/mz5/
netCDF	http://www.unidata.ucar.edu/software/netcdf/
Analyst	http://sciex.com/products/software/analyst-software
SWATH Acquisition	http://sciex.com/swath-acquisition
MetabolitePilot	http://sciex.com/products/software/metabolitepilot-software
ProteinPilot	http://sciex.com/products/software/proteinpilot-software
MultiQuant	http://sciex.com/products/software/multiquant-software
PeakView	http://sciex.com/products/software/peakview-software
Masshunter	http://www.chem.agilent.com/en-US/products-services/Software-Informatics/MassHunter-Workstation-Software/Pages/default.aspx
LCMS ChemStation	http://www.chem.agilent.com/en-US/products-services/Software-Informatics/LC-MS-Chemstation-Software/Pages/default.aspx
OpenLAB	http://www.chem.agilent.com/en-US/products-services/Software-Informatics/OpenLAB-Chromatography-Data-System-(CDS)/Pages/default.aspx
Proteome Discoverer	http://www.thermoscientific.com/content/tfs/en/product/proteome-discoverer-software.html
Xcalibur	http://www.thermoscientific.com/en/product/xcalibur-software.html
ProMass	http://www.thermoscientific.com/en/product/promass-deconvolution-2-8-software.html
MassLynx	http://www.waters.com/waters/en_GB/MassLynx-Mass-Spectrometry-Software-/nav.htm?cid=513164&lset=1&locale=en_GB
OpenLynx	http://www.waters.com/waters/en_GB/OpenLynx-Open-Access/nav.htm?cid=10008851&locale=en_GB
QuanLynx	http://www.waters.com/waters/library.htm?locale=en_US&lid=1545661
ProteinLynx Global SERVER	http://www.waters.com/waters/en_US/ProteinLynx-Global-SERVER-(PLGS)/nav.htm?cid=513821&locale=en_US
Phenyx	http://www.genebio.com/products/phenyx/solutions.html
Seaquest	http://fields.scripps.edu/researchtools.php
PEAKS	http://www.bioinform.com
Achroma	http://openmasp.hswt.de/pages/project/achroma.php
OpenChrom (free Community Edition)	https://www.openchrom.net/
Maltcms/Maui [17]	http://maltcms.sourceforge.net/

polyXmass / massXpert	polyXmass: http://www.gnu.org/software/polyxmass/ massXpert: http://www.massxpert.org/wiki/
DeconLS	http://omics.pnl.gov/software/decontools-decon2ls
BioClipse	http://www.bioclipse.net/
BioSunMS	http://biosunms.sourceforge.net/
openMS	http://open-ms.sourceforge.net/
MetaQuant	http://bioinformatics.org/metaquant/
intelliMS	http://intellims.proteomix.org/
mMass	http://www.mmass.org/
MZMine2	http://mzmine.sourceforge.net/
pymzML	http://pymzml.github.io/
MS-Spectre	http://ms-spectre.sourceforge.net/
ProteoWizard	http://proteowizard.sourceforge.net/
X!Tandem / X!!Tandem	X!Tandem: http://www.thegpm.org/tandem/ X!!Tandem: http://wiki.thegpm.org/wiki/X!!Tandem
Mascot (Server) (free search for 1200 spectra and limited functions)	http://www.matrixscience.com/search_form_select.html
PRIDE	http://www.ebi.ac.uk/pride/archive/
TOPP	http://open-ms.sourceforge.net/
MapQuant	http://arep.med.harvard.edu/MapQuant/
SpecArray	http://tools.proteomecenter.org/wiki/index.php?title=Software:SpecArray
ATAQS	http://tools.proteomecenter.org/ATAQS/ATAQS.html
Skyline	https://skyline.gs.washington.edu/labkey/project/home/software/Skyline/begin.view
msInspect	http://proteomics.fhrc.org/CPL/msinspect/
SuperHirn	http://tools.proteomecenter.org/wiki/index.php?title=Software:SuperHirn
MassSorter	http://services.cbu.uib.no/software/massSorter
PASSEL	http://www.peptideatlas.org/passel/
jTraML	https://github.com/compomics/jtraml
HDX-analyzer	http://people.tamu.edu/~syuan/hdxanalyzer/
PeptidePicker	http://mrmpeptidepicker.proteincentre.com/peptidepicker9/
TPP (Trans-Proteomic Pipeline)	http://tools.proteomecenter.org/wiki/index.php?title=Software:TPP
SpectraST	http://tools.proteomecenter.org/wiki/index.php?title=Software:SpectraST
OMSSA (Open Mass Spectrometry Search Algorithm)	No longer available, but some artifacts can be found here: http://www.ncbi.nlm.nih.gov/Web/Newsitr/V14N2/
Andromeda (Part of MaxQuant)	http://www.andromeda-search.org/
MaxQuant	http://www.maxquant.org/downloads.htm
Proteios SE	http://www.proteios.org/
MyriMatch	No longer available, but a description how to obtain the source code can be found here: http://www.mc.vanderbilt.edu/root/vumc.php?site=msrc/bioinformatics&doc=27122
PAGIT	http://www.sanger.ac.uk/resources/software/pagit/
MUMmer	http://mummer.sourceforge.net/
VeSPA	https://scion.duhs.duke.edu/vespa/project/wiki
MINT	http://mint.bio.uniroma2.it/mint/Welcome.do
Mass Bank	http://www.massbank.jp/
MSEA	http://www.msea.ca/MSEA/
XANNpred	http://www.compbio.dundee.ac.uk/xtal/cgi-bin/xannpred_in.pl

Table 2: List of web-links (last accessed on 2015.07.09) to the softwares presented in table 1.

displays the variety of the software landscape. Experience shows that many searches by users are starting from *Google* and *Wikipedia*.

Many software tools have been developed in the area of protein/peptide identification and quantification, both main topics in proteomics. Enumerating all of these would go beyond the dimension of this paper. However, there is an extensive, but not exhaustive, list of referring commercial and free software platforms and tools in Table 1. The software listed there can only partially be found in ExpASY and thus extends the diversity of software solutions for evaluating proteomic experiments in a very complementary way. The main focus of this list

is on the proteomics field as well as on the programming language and platform.

The demand for software to evaluate specific, possibly untypical MS data is still high. Particularly in functional proteomics for the data output resulting from batch or continuous flow enzyme assays [16], there is still a big need for optimising the evaluation. A search using PubMed with key words: “enzyme” and “software” and “mass spectrometry” for example delivered 56 results overall since 2002 (Figure 2b). The highest level has been reached in 2014 with 13 publications and 39 of the 56 publications have been registered since 2009. However this search also

shows that the development of software for specialized topics like “MS based enzyme assays” is still a very new field of research, which is not in the focus of main proteomics research.

Finally, in the World Wide Web one can find plenty lists for application specific software and also platforms for listing software tools and systems, but no platform that helps the researcher and analytical scientists to identify the most suitable software for their problems (without searching for hours in the internet). An idea for solving this problem could be a web based search platform, in which researcher describe (in headwords) the analytical evaluation tasks and, based on this description, the search platform will result in a set of currently relevant software tools, collected from several sites in the World Wide Web.

The Applied Software Languages and Development Platforms

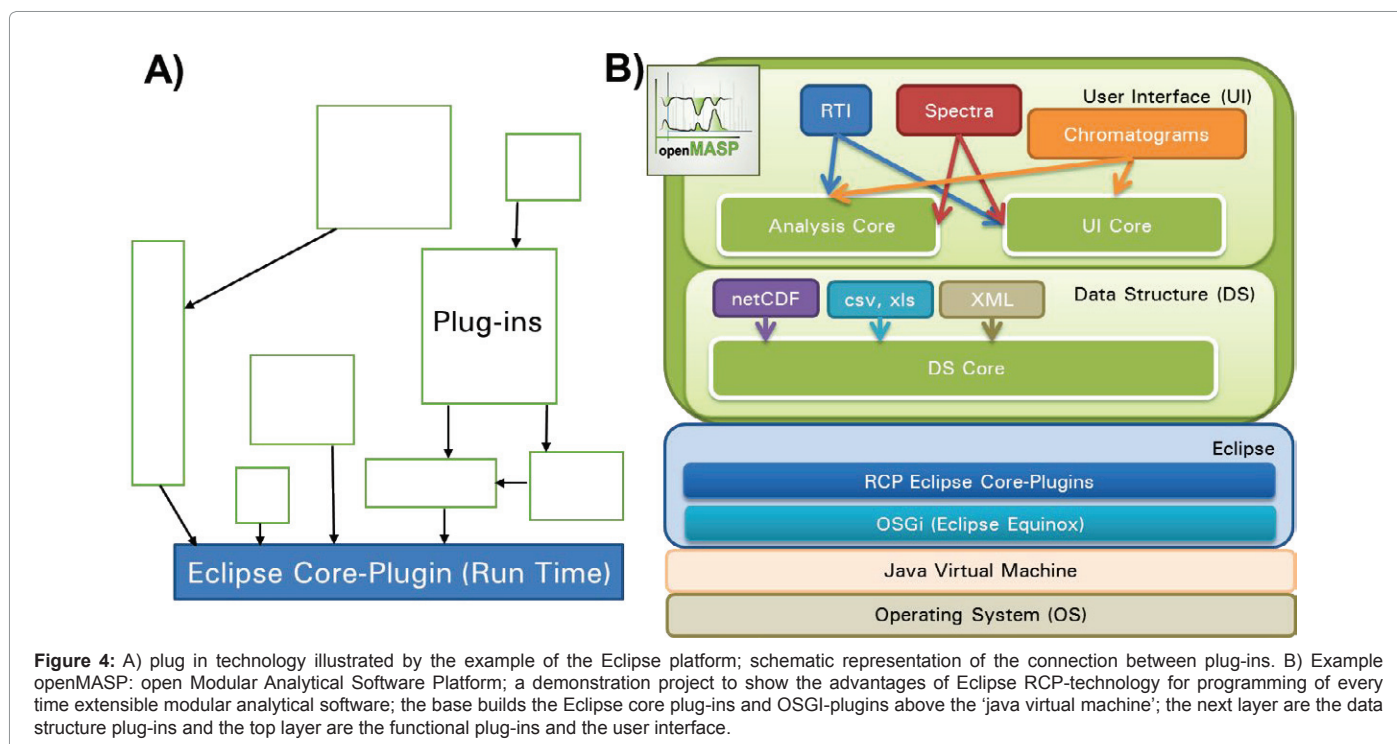
The previous explanations obviously indicate that there is already a large diversity of software solutions concerning proteomics. They vary from exchange data formats, over appropriate converters and solutions for the management of proteomic data, to solutions developed specially for proteomics and software combining several software solutions to a processing sequence (Table 1). But users cannot easily combine these programs, since in many cases they are not compatible due to the different technologies used for development. To understand this, it is necessary to take a closer look on software architecture and the methods used for software development.

First of all there is to consider the underlying programming language. Two main programming languages are popular for the development of such software solutions: on the one hand the languages of the C-family: C, C++ and C#, where C++ is most widespread, and on the other hand the programming language Java.

For historical reasons the majority of commercial systems is programmed in C and C++, since these generally originate from

the manufacturers of appliances and in the classical study of mass spectrometers. C and C++ were for a long time the standard languages for device oriented application development and are still used in that area nowadays. The source code is compiled into machine code, which means that the programmes are very much dependent on specific hardware features. Therefore many of these applications have the disadvantage that they run exclusively on specific platforms, i.e. hardware combined with its operating system (like Microsoft Windows or (exclusively) Linux).

Next, one has to consider the specific approach to software development. A large group of free or even open source software was developed with the emergence of the “Proteomics Community”. In the case of open source the entire source code is public available. Thus many different people can participate and help in development and support. These freely accessible software solutions are frequently platform independent or at least run with both, Linux and Windows operating systems. This leads to the fact that in many cases these tools are used for development and research. Table 1 “3 Free and open source platforms and software tools” shows clearly that in this area the programming language Java and its auxiliaries (sixteen software solutions are listed) is becoming standard, although C++ (ten entries) is still widespread. However, particularly for new software solutions developed since 2010 the majority use Java. The largest advantage of using Java is that -thanks to its technology using a virtual machine- the software can be run independently of the underlying specific operating system. Python (seven entries) for example is as well a programming language that is not compiled into machine code but interpreted and in the meantime widely used in the area of natural sciences, since Python is very well adapted for batch processing and therefore is very suitable for the programming of pipelines. Table 1 (third part) also lists web based solutions with likewise nine entries, which may gain more interest from the bioinformatics and proteomics community in the future, particularly with the continued worldwide development of web based programming (Web 2.0) and Cloud solutions.



Furthermore, when we consider extendability and compatibility of software we have to take the underlying technologies for building a software architecture into account. Many so called “Integrated Development Environments” (IDE), like e.g. Eclipse [17] or NetBeans [18], provide an elegant mechanism for extending an application. The main software package then contains just the core functionality. Optional features can be made available using a so called plug-in technology. A plug-in is an encapsulated part of software with a well-defined interface that can be plugged into an existing software application in order to enhance its functionality (Figure 4A). This enables customization of applications, without needing to update the full application. A further benefit of the plug-in technology is the possibility of re-using the plug-ins in other software projects. This fits perfectly, if the underlying technology is the same, as e.g. when using eclipse and the eclipse market place. In this way the programming environment already provides most parts necessary to generate a modular project. Such tools that allow programmers to easily integrate further software components, where most of the processing is executed on the client side, is called Rich Client Platform (RCP). Figure 4 illustrates by means of an example of the Eclipse IDE platform. The left hand side (Figure 4A) schematically depicts the fact that the entire application is composed of plug-ins. Even the Eclipse core, which constitutes the starting basis for the application development is itself a plug-in, the so called entry-point of the application. Figure 4B) shows the concept of an analytical software platform, called openMASP (open Modular Analytical Software Platform) [19]: the green area represents the layer that interacts with the users the freely programmable functional plug-ins in connection

with the graphical user interface (the UI). These functions are based on the implementation of various data structures. The blue area in the middle comprises the modules that are responsible for the connection to the core plug-ins belonging to Eclipse (or another RCP-platform) and to the “Java Virtual Machine” that executes the interpreted Java code. The “Java Virtual Machine” provides an execution environment, independent of the underlying operating system. By this way we achieve so called “Cross-Platform” applications.

A good example for such an extendable application is OpenChrom [6], a software for chromatography and mass spectrometry. It is built with Eclipse RCP technology. The main focus is on handling mass spectrometry files. For example, OpenChrom can handle many venture formats natively and can be used to analyse GC/MS, and LC/MS data. It contains a growing number of processing and analysing procedures. Another example is Maui that is built on top of the NetBeans RCP technology. It provides a user interface to handle Maltcms (Modular Application Toolkit for Chromatography Mass-Spectrometry) [20], a framework mainly developed for developers in the domain of bioinformatics for metabolomics and proteomics. It offers integrated functions to handle different file formats, like mzXML, mzData and netcdf. The data are processed using a free configurable processing queue. Maui gives the users a visual interface to control the Maltcms framework and display its processing results.

Both, OpenChrom and Maui with Maltcms, open up new possibilities, since they are Open Source projects. The user can use it

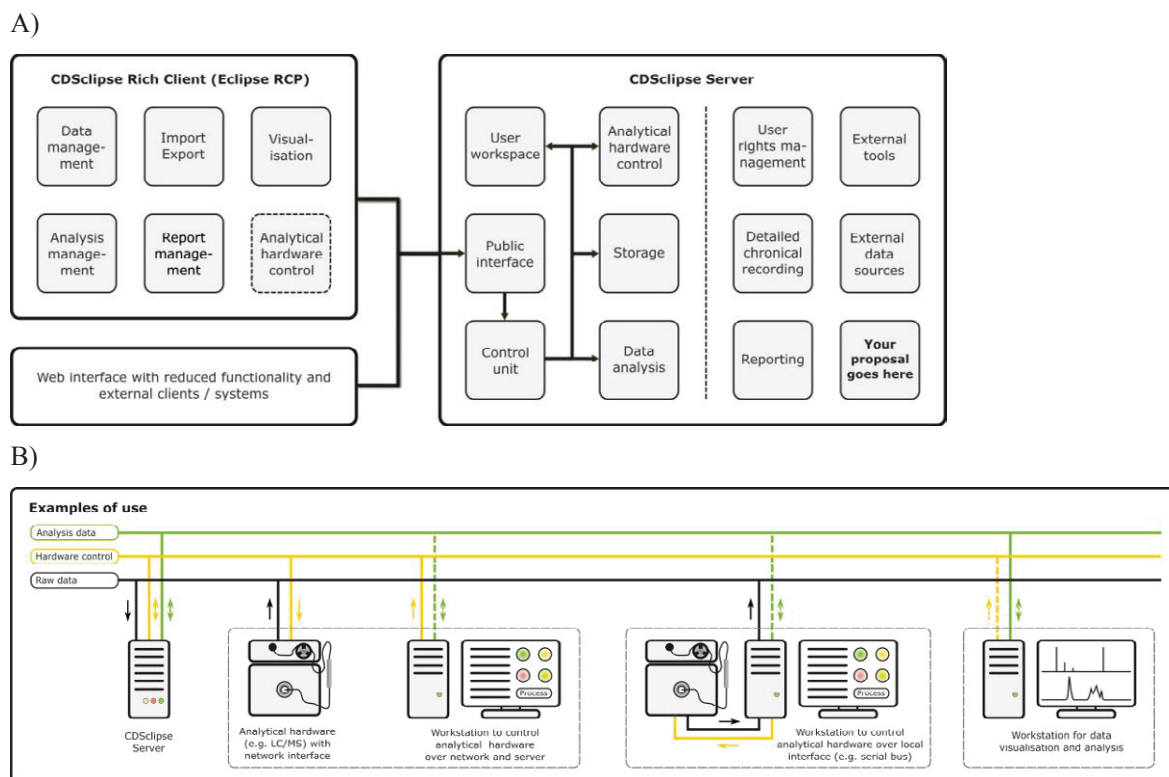


Figure 5: CDSclipse: Chromatography Data System based on Eclipse – a software project study for a SaaS chromatography data system based on Eclipse RCP-technology. The server delivers all functionality to the clients (workstations, LC/MS systems or Tablets). A) the left side shows the Eclipse Rich Client and its functions (data import/export, visualization) and its independence from any system (Webclient, Eclipse plug-in, local client, external software); the right side shows the server and the necessary/possible modules. B) visualisation of the analytical data flow (green line) and the hardware control flow (yellow line) and some examples of possible hardware setups under control of CDSclipse.

for free and adopt it to for personal needs, assumed that the technical skills are at hand. It is possible to handle files and data using the projects infrastructure and integrate further software modules for analysis easily. Therefore the standard analysis related modules of the applications can be enriched by this way of customization, giving the user potential in data analysis and allowing individual customized solutions. Both projects are part of the Eclipse Science Working Group [21].

Beside local and server based platforms, online databases and tools are commonly used in proteomics. These are available via the internet and most of them are free to use. Some of them provide only access with a browser, as e.g. www.chemicalize.org. Others offer a so called Web Service. This is an interface designed for machine to machine communication. The service provider defines the interface and decides on the protocol for communication. Any computer can connect to such a Web Service via the protocol and request the service, as for example a search in a database. The result will be resent to the computer which must be capable of interpreting the result. So it is possible to use data from online services in desktop applications. When using web-based applications or databases one has to consider license issues. Many free data sets are provided under a certain license, like e.g. the Creative Commons, that might restrict the data usage to non-commercial applications. These issues sometimes lack adequate awareness from the users.

In the last years Cloud Computing has come into the focus of the public. The definition of Cloud Computing by the National Institute of Standards and Technology (NIST) [22] is: "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." Using a cloud in modern analytical research may provide the advantage that data can be processed in powerful external processing centres and can be easily shared with other researcher in research institutions worldwide.

The most interesting service model for cloud computing is the so called SaaS (Software as a service, Figure 5). The resources for heavy calculations are in the cloud that is built by the servers and the network of the specific provider. Users can use the high computing power of the cloud service provider in order to execute their algorithms. The client machines would just be used for configuring and starting computing pipelines, but the execution itself would be carried out on the server machines. After processing just the results will be transferred to the client, which means that such a scenario would neither require a particular high bandwidth nor a high end client computer. Analysis processes can particularly be accelerated with the power of the cloud as soon as there are algorithms that can be scheduled in parallel. Such a service can be used through many different clients, like a web browser or a program on the client computer. The user has no control of the underlying cloud components.

A critical examination of cloud computing may easily lead to serious concerns about data security. Data shared in a public cloud may be available for other users or organisations. But on the other side, the utilisation of cloud computing can lead to an improved flow of knowledge by sharing data and can speed up the analysis processes. When using a community cloud, that means a restricted accessible cloud as opposed to a public cloud, data can be shared among trusted and cooperating laboratories and therefore a substantial benefit may be achieved.

Open Source, Freeware or Commercial Software - What is the Best Way?

In Table 1 and above it is shown, that there are many very useful software solutions available, but often are not interoperable among each other. For that reason, researchers usually have to use a lot of different software tools to evaluate proteomics data. In our opinion a big challenge for the future is to bring commercial, freeware and open source together and make them interoperable. A first step in this direction was done by the founding of the Science Working Group of the Eclipse based platforms [21]. This group is currently a consortium of 23 open source projects in the field of natural sciences, 6 proprietary projects, 4 miscellaneous projects and 10 software companies. Their aim is to make their products interoperable and interchangeable. That would add values for the users of these products within the Eclipse community. Still it remains open, how other software tools can be integrated. Our suggestion is to develop an open platform for collecting all relevant software tools in the field of proteomics or biochemistry, similar to the Science WG of Eclipse, but open for all the solutions, independent of the programming language or programming platform used. In order to achieve this, first of all the vendors of analytical hardware would have to open their own data formats for more interoperability. Furthermore standard interfaces need to be defined and pushed forward to achieve the required interchangeability between the several software solutions. At the end the research in proteomics could become easier, much more structured and much more reproducible.

It cannot be decided which of the two approaches, open-source or proprietary, is better or worse in general. One has to choose the right tool for the right purpose. In high dynamic areas where requirements change rapidly, open source software has proven to be a good decision, since existing software that is based on open source can be extended or adapted to new situations, without needing too many expensive resources (such as money or manpower). Furthermore, developing or extending an open source application, results in substantially better options, with regard to the sharing and development of new and innovative algorithms and workflows. The Open Source approach often leads to the establishment of new communities, which can further extend the software according to new perspectives and ways of working.

Conclusion

It could be shown, that the software landscape in the proteomics research field, as an example for omic-fields, is very heterogeneous and different. Plenty of tools and platforms are available and it is the user's choice, which is the best software for analyzing proteomics data. Data exchange and interoperability between the software is problematic and last but not least the software is developed with different programming languages and different development strategies. So, three conclusions can be made to improve this situation: firstly, the proteomics community should bring researcher, commercial, freeware and open source developer together to make the software solutions interoperable and interchangeable. A good step in this direction was the founding of the science working group of the Eclipse Foundation. Secondly, the definition of standards and interfaces for interoperability between the software solutions should be the next step. The definition of data exchange formats, like mzML, is not enough. Thirdly, the development of a web based platform (as well as pipelines) which help researcher finding the best software in the variety of software solutions could be helpful to improve the data evaluation.

References

- Chitty M (2015) Biopharmaceutical -Omics & -omics glossary taxonomy.
- Alphabetically ordered list of omes and omics - Omics.org.
- Anandkumar B, Haga SW, Wu HF (2014) Computer applications making rapid advances in high throughput microbial proteomics (HTMP). *Comb Chem High Throughput Screen* 17: 173-182.
- Gowda H, Ivanisevic J, Johnson CH, Kurczy ME, Benton HP, et al. (2014) Interactive XCMS Online: simplifying advanced metabolomic data processing and subsequent statistical analyses. *Analytical Chemistry* 86: 6931-6939.
- Kessner D, Chambers M, Burke R, Agus D, Mallick P (2008) ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* 24: 2534-2536.
- Wenig P, Odermatt J (2010) OpenChrom: a cross-platform open source software for the mass spectrometric analysis of chromatographic data. *BMC Bioinformatics* 11: 405.
- Krappmann M, Letzel T, *Anal. Methods* 201, 4 (4), 1060.
- Krappmann M, Kaufmann CM, Scheerle RK, Grassmann J, Letzel T (2014) Achroma Software-High-Quality Policy in (a-)Typical Mass Spectrometric Data Handling and Applied Functional Proteomics. *J Proteomics Bioinform* 7: 264-271.
- Keller A, et al., *Mol. Syst. Biol.* 2005, 2005.0017.
- Kohlbacher O, Reinert K, Gröpl C, Lange E, Pfeifer N, et al. (2007) TOPP-the OpenMS proteomics pipeline. *Bioinformatics* 23: e191-197.
- ExpASY: SIB Bioinformatics Resource Portal - Categories.
- SPC Proteomics Tools.
- Software | Pan-Omics Research.
- Wikipedia, List of mass spectrometry software - Wikipedia, the free encyclopedia.
- ms-utils.org - Software List.
- Grassmann J, Scheerle RK, Letzel T (2012) Functional proteomics: application of mass spectrometry to the study of enzymology in complex mixtures. *Anal Bioanal Chem* 402: 625-645.
- Guindon C, Eclipse Luna.
- Welcome to NetBeans.
- openMASP - Vision.
- Hoffmann N, Keck M, Neuweiger H, Wilhelm M, Högy P, et al. (2012) Combining peak- and chromatogram-based retention time alignment algorithms for multiple chromatography-mass spectrometry datasets. *BMC Bioinformatics* 13: 214.
- Science WG - Eclipsepedia.
- NIST Computer Security Division (CSD).
- HUPO-PSI Working Groups and Outputs | HUPO Proteomics Standards Initiative.
- Pedrioli PG, Eng JK, Hubley R, Vogelzang M, Deutsch EW, et al. (2004) A common open representation of mass spectrometry data and its application to proteomics research. *Nat Biotechnol* 22: 1459-1466.
- Wilhelm M, Kirchner M, Steen JA, Steen H (2012) mz5: space- and time-efficient storage of mass spectrometry data sets. *Mol Cell Proteomics* 11: O111.
- Unidata | NetCDF.
- Mass Spectrometry Software | Analyst Software | Triple Quadrupole TOF System Software.
- Software & Informatics | Agilent.
- Proteome Discoverer™ Software.
- waters, Mass Lynx Massenspektrometrie Software Waters.
- GeneBio Phenyx.
- Mann M, Højrup P, Roepstorff P (1993) Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol Mass Spectrom* 22: 338-345.
- Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 5: 976-989.
- Overview - Proteomics Mass Spectrometry Software | PEAKS.
- Ma B, Zhang K, Hendrie C, Liang C, Li M, et al. (2003) PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* 17: 2337-2342.
- Rusconi F1 (2009) massXpert 2: a cross-platform software environment for polymer chemistry modelling and simulation/analysis of mass spectrometric data. *Bioinformatics* 25: 2741-2742.
- Jaitly N, Mayampurath A, Littlefield K, Adkins JN, Anderson GA, et al. (2008) Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. *BMC Bioinformatics* 10: 87.
- Spjuth O, Alvarsson J, Berg A, Eklund M, Kuhn S, et al. (2009) Bioclipse 2: a scriptable integration platform for the life sciences. *BMC Bioinformatics* 10: 397.
- Spjuth O, Helmus T, Willighagen EL, Kuhn S, Eklund M, et al. (2007) Bioclipse: an open source workbench for chemo- and bioinformatics. *BMC Bioinformatics* 8: 59.
- Cao Y, Wang N, Ying X, Li A, Wang H, et al. (2009) BioSunMS: a plug-in-based software for the management of patients information and the analysis of peptide profiles from mass spectrometry. *BMC Med Inform Decis Mak* 9: 13.
- Sturm M, Bertsch A, Gräßl C, Hildebrandt A, Hussong R, et al. (2008) OpenMS - an open-source software framework for mass spectrometry. *BMC Bioinformatics* 9: 163.
- Bunk B, Kucklick M, Jonas R, Mänch R, Schobert M, et al. (2006) MetaQuant: a tool for the automatic quantification of GC/MS-based metabolome data. *Bioinformatics* 22: 2962-2965.
- Kwon MS, Lee HJ, Jeong SK, Lee EY, Cho SY, et al. (2008) IntelliMS: a platform to efficiently manage and visualize tandem mass spectral data. *Proteomics* 8: 4910-4913.
- Strohalm M, Hassman M, Kosata B, Kodáček M (2008) mMass data miner: an open source alternative for mass spectrometric data analysis. *Rapid Commun Mass Spectrom* 22: 905-908.
- Pluskal T, Castillo S, Villar-Briones A, Oresic M (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 11: 395.
- Bald T, Barth J, Niehues A, Specht M, Hippler M, et al. (2012) pymzML--Python module for high-throughput bioinformatics on mass spectrometry data. *Bioinformatics* 28: 1052-1053.
- V. Bessels, MS-Spectre - Mass Spectrometry analysis software.
- Duncan DT, Craig R, Link AJ (2005) Parallel tandem: a program for parallel processing of tandem mass spectra using PVM or MPI and X!Tandem. *J Proteome Res* 4: 1842-1847.
- Hirosawa M, Hoshida M, Ishikawa M, Taya T (1993) MASCOT: multiple alignment system for protein sequences based on three-way dynamic programming. *Comput Appl Biosci* 9: 161-167.
- Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20: 3551-3567.
- Martens L, Hermjakob H, Jones P, Adamski M, Taylor C, et al. (2005) PRIDE: the proteomics identifications database. *Proteomics* 5: 3537-3545.
- Leptos KC, Sarracino DA, Jaffe JD, Krastins B, Church GM (2006) MapQuant: open-source software for large-scale protein quantification. *Proteomics* 6: 1770-1782.
- Li XJ, Yi EC, Kemp CJ, Zhang H, Aebersold R (2005) A software suite for the generation and comparison of peptide arrays from sets of data collected by liquid chromatography-mass spectrometry. *Mol Cell Proteomics* 4: 1328-1340.
- Brusniak MY, Kwok ST, Christiansen M, Campbell D, Reiter L, et al. (2011) ATAQS: A computational software tool for high throughput transition optimization and validation for selected reaction monitoring mass spectrometry. *BMC Bioinformatics* 12: 78.
- MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, et al. (2010)

- Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26: 966-968.
56. Bellew M, Coram M, Fitzgibbon M, Igra M, Randolph T, et al. (2006) A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS. *Bioinformatics* 22: 1902-1909.
57. Mueller LN, Rinner O, Schmidt A, Letarte S, Bodenmiller B, et al. (2007) SuperHirn - a novel tool for high resolution LC-MS-based peptide/protein profiling. *Proteomics* 7: 3470-3480.
58. Barsnes H, Mikalsen SO, Eidhammer I (2006) MassSorter: a tool for administrating and analyzing data from mass spectrometry experiments on proteins with known amino acid sequences. *BMC Bioinformatics* 7: 42.
59. Farrah T, Deutsch EW, Kreisberg R, Sun Z, Campbell DS, et al. (2012) PASSEL: the PeptideAtlas SRMexperiment library. *Proteomics* 12: 1170-1175.
60. Helsen K, Brusniak MY, Deutsch E, Moritz RL, Martens L (2011) jTraML: an open source Java API for TraML, the PSI standard for sharing SRM transitions. *J Proteome Res* 10: 5260-5263.
61. Liu S, Liu L, Uzuner U, Zhou X, Gu M, et al. (2011) HDX-analyzer: a novel package for statistical analysis of protein structure dynamics. *BMC Bioinformatics* 12 Suppl 1: S43.
62. Mohammed Y, Domański D, Jackson AM, Smith DS, Deelder AM, et al. (2014) PeptidePicker: a scientific workflow with web interface for selecting appropriate peptides for targeted proteomics experiments. *J Proteomics* 106: 151-161.
63. Lam H, Deutsch EW, Eddes JS, Eng JK, King N, et al. (2007) Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* 7: 655-667.
64. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, et al. (2004) Open mass spectrometry search algorithm. *J Proteome Res* 3: 958-964.
65. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, et al. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* 10: 1794-1805.
66. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26: 1367-1372.
67. Gärdén P, Alm R, Häkkinen J (2005) PROTEIOS: an open source proteomics initiative. *Bioinformatics* 21: 2085-2087.
68. Tabb DL, Fernando CG, Chambers MC (2007) MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J Proteome Res* 6: 654-661.
69. Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, et al. (2012) A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. *Nat Protoc* 7: 1260-1284.
70. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, et al. (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5: R12.
71. Peterson ES, McCue LA, Schrimpe-Rutledge AC, Jensen JL, Walker H, et al. (2012) VESPA: software to facilitate genomic annotation of prokaryotic organisms through integration of proteomic and transcriptomic data. *BMC Genomics* 13: 131.
72. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, et al. (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* 40: D857-861.
73. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, et al. (2010) MassBank: a public repository for sharing mass spectral data for life sciences. *J Mass Spectrom* 45: 703-714.
74. Xia J, Wishart DS (2010) MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Res* 38: W71-77.
75. Overton IM, van Niekerk CA, Barton GJ (2011) XANNpred: neural nets that predict the propensity of a protein to yield diffraction-quality crystals. *Proteins* 79: 1027-1033.

Citation: Krappmann M, Luthardt M, Lesske F, Letzel T (2015) The Software-Landscape in (Prote)Omic Research. *J Proteomics Bioinform* 8: 164-175. doi:10.4172/jpb.1000365

Submit your next manuscript and get advantages of OMICS Group submissions

Unique features:

- User friendly/feasible website-translation of your paper to 50 world's leading languages
- Audio Version of published paper
- Digital articles to share and explore

Special features:

- 400 Open Access Journals
- 30,000 editorial team
- 21 days rapid review process
- Quality and quick editorial, review and publication processing
- Indexing at PubMed (partial), Scopus, EBSCO, Index Copernicus and Google Scholar etc
- Sharing Option: Social Networking Enabled
- Authors, Reviewers and Editors rewarded with online Scientific Credits
- Better discount for your subsequent articles

Submit your manuscript at: <http://www.editorialmanager.com/proteomics>