

**Automatisierte Identifizierung von nativen Peptiden  
aus humanen Körperflüssigkeiten mittels  
Massenspektrometrie**

Dissertation  
zur Erlangung des Doktorgrades  
der Naturwissenschaften

vorgelegt beim Fachbereich  
Chemische und Pharmazeutische Wissenschaften  
der Johann Wolfgang Goethe-Universität  
in Frankfurt am Main

von  
Markus Kellmann  
aus Bremen

Frankfurt (2004)  
(DF 1)

vom Fachbereich Chemische und Pharmazeutische Wissenschaften der  
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Harald Schwalbe

Gutachter: Prof. Dr. Michael Karas, PD Dr. Peter Schulz-Knappe

Datum der Disputation:

Für

Kerstin, Annemieke und Ilvi

# Inhaltsverzeichnis

INHALTSVERZEICHNIS.....	I
1	EINLEITUNG..... 1
2	THEORETISCHE GRUNDLAGEN..... 7
2.1	Grundlagen der massenspektrometrischen Analyse von Peptiden ..... 7
2.1.1	<i>Matrix Assisted Laser Desorption Ionization Mass Spectrometry (MALDI-TOF-MS)</i> ..... 13
2.1.1.1	<i>Ionisierung</i> ..... 13
2.1.1.2	<i>Massenanalyse</i> ..... 15
2.1.2	<i>Elektrospray Ionization Mass Spectrometry (ESI-MS)</i> ..... 17
2.1.2.1	<i>Ionisierung</i> ..... 18
2.1.2.1.1	Varianten der ESI-MS ..... 22
2.1.2.2	<i>Massenanalyse</i> ..... 24
2.1.2.2.1	Der Quadrupol Analysator..... 24
2.1.2.2.2	Die Ionenfalle ..... 26
2.1.2.2.3	Quadrupol-TOF-Hybrid-Geräte (QToF)..... 29
2.1.3	<i>Massenspektrometrische Strukturanalyse: MS<sup>n</sup>-Methoden</i> ..... 30
2.1.3.1	<i>Die Prinzipien der Tandem-Massenspektrometrie: MS<sup>2</sup> und MS<sup>n</sup></i> 32
2.1.3.2	<i>Fragmentierung von Peptiden</i> ..... 33
2.2	Der gängige Identifizierungsweg von Proteinen in der Proteomanalytik..... 36
2.3	Unterschiede der Peptidom- und Proteom-Analytik..... 37
2.4	Peptide als funktionelle Einheiten im Körper ..... 39
3	MATERIAL UND METHODEN ..... 43
3.1	Material ..... 43
3.1.1	<i>Peptidbank aus humanem Blutfiltrat</i> ..... 43
3.1.2	<i>Lösungsmittel und Chemikalien</i> ..... 45

3.1.3	<i>Peptidstandards</i> .....	45
3.1.3.1	<i>MALDI-MS</i> .....	46
3.1.3.2	<i>ESI MS</i> .....	46
3.1.3.3	<i>Massenspektrometer</i> .....	47
3.2	Methoden .....	47
3.2.1	<i>Reversed Phase Chromatographie</i> .....	47
3.2.2	<i>Resuspendierung der Proben</i> .....	48
3.2.3	<i>Massenspektrometrische Methoden</i> .....	48
3.2.3.1	<i>MALDI-MS</i> .....	48
3.2.3.2	<i>ESI-Quadrupol-Ion-Trap-MS (ESI-QIT-MS)</i> .....	50
3.2.3.3	<i>ESI-Quadrupol-Time-of-Flight-MS (ESI-QTOF-MS)</i> .....	51
3.2.3.4	<i>Darstellung der MALDI-MS Spektren mit der Software Spectromania</i> .....	52
3.2.3.5	<i>Aminosäure-Sequenzanalyse nach Edman</i> .....	52
3.2.3.6	<i>Modifizierungen am Cystein</i> .....	53
3.2.3.7	<i>Entsalzung der Proben</i> .....	54
3.2.3.8	<i>Datenbankrecherchen</i> .....	54
4	ERGEBNIS UND DISKUSSION .....	56
4.1	Problemstellung.....	56
4.2	Entwicklung eines allgemeinen Ablaufschemas zur Erstellung eines <i>Inventories</i> .....	58
4.2.1	<i>Fraktionierung der Probe</i> .....	61
4.2.2	<i>Kartierung der Molekülmassen einer Peptidquelle</i> .....	61
4.2.2.1	<i>Entwicklung eines MALDI-MS Autosamplers</i> .....	62
4.2.2.1.1	Mechanische Hardware.....	62
4.2.2.1.2	Elektronische Hardware .....	63
4.2.2.1.3	Software .....	63
4.2.2.1.4	Zusammenfassung.....	64
4.2.3	<i>Erzeugung von MS/MS-Fingerprints</i> .....	65
4.2.3.1	<i>Hardware</i> .....	65
4.2.3.2	<i>„Alcatrap“: Ein Softwaretool zur automatischen Erstellung von</i>	

	<i>Arbeitslisten für die Ionenfalle</i> .....	66
4.2.3.2.1	Überführung von einfach geladenen „MALDI-Ionen“ in mehrfach geladene „ESI-Ionen“ bei nativen Peptiden .....	67
4.2.3.2.2	Bestimmung der Kollisionsenergien für die automatische MS/MS-Messung von nativen Peptiden mit Hilfe einer Ionenfalle (LCQ).....	69
4.2.3.2.3	Funktionen von Alcatrap.....	72
4.2.3.2.4	Die Schnittstelle zwischen Fragmentspektrenerzeugung und Datenbanksuche: „D2M“ .....	78
4.2.4	<i>Identifizierung von Peptiden mittels Datenbanksuche</i> .....	79
4.2.4.1	<i>Die Datenbanksuchmaschine MASCOT™</i> .....	79
4.2.5	<i>Archivierung der Ergebnisse: PeTrA (Peptide Tracking Application)</i> .....	82
4.2.6	<i>Zusammenfassung der bisherigen Ergebnisse</i> .....	84
4.3	Erstellung eines Inventories für humanes Blutfiltrat .....	86
4.3.1	<i>Rechromatographie und Feinkartierung der Peptidbank</i> .....	86
4.3.2	<i>Erstellung von Fragmentspektren</i> .....	89
4.3.3	<i>Datenprozessierung und Datenbanksuche</i> .....	89
4.3.4	<i>Auswertung der Sequenzierung</i> .....	91
4.3.5	<i>Diskussion der Ergebnisse</i> .....	95
4.3.5.1	<i>Einfluss der Massengenauigkeit auf die Effizienz der Identifizierung von Peptiden mit Hilfe der Datenbanksuche</i> .....	98
4.4	Erweiterung des Ansatzes auf Quadrupol-time-of-flight Analysatoren .....	104
4.4.1	<i>Anpassung der Kollisionssenergie</i> .....	104
4.4.2	<i>Identifizierung großer Peptide und kleiner Proteine</i> .....	107
4.5	Anwendung der erweiterten Strategie auf Hämofiltrat als Proof- of-Concept.....	114
4.5.1	<i>Versuchsdesign</i> .....	114
4.5.2	<i>Ergebnisse</i> .....	115
4.5.2.1	<i>Diskussion</i> .....	117
5	ZUSAMMENFASSUNG .....	121

6	AUSBLICK .....	124
7	ABBILDUNGSVERZEICHNIS .....	125
8	LITERATUR .....	130
9	EIDESSTATTLICHE VERSICHERUNG.....	140
10	DANKSAGUNG.....	141

## 1 Einleitung

Erhebliche Fortschritte sind in den letzten 20 Jahren in der molekularbiologischen Wissenschaft erzielt worden. Erstmals wurde es möglich, nicht nur einzelne Gene, sondern ganze Genome systematisch und mit hoher Geschwindigkeit zu analysieren. Begonnen wurde dabei mit der Analyse einfacher Organismen wie Bakterien, die über einen begrenzten Satz an Genen verfügen und keine ausgesprochenen Intergen-Sequenzen oder Intron-Strukturen verfügen. Die Anstrengungen, die seit 1987 im *Human-Genome-Project* von Einrichtungen der öffentlichen Hand und im privaten Bereich von der Firma *Celera* unternommen wurden, um das menschliche Genom systematisch zu entschlüsseln, erreichten Anfang 2001 mit der Veröffentlichung des kompletten menschlichen Genoms einen Höhepunkt (Lander *et al.* 2001).

Inzwischen hat sich die weltweite Erkenntnis schnell weiter entwickelt. Die noch vor einigen Jahren herrschende Erwartung, dass mit Kenntnis des Genoms eine hinreichende Basis zur Erklärung und nachfolgend auch Behandlung von Erkrankungen gegeben sei, hat sich nicht erfüllt. In den meisten Fällen ist die Betrachtung der Gensequenz nicht ausreichend. Es ist z.B. nicht möglich, vom Spiegel der mRNA auf den Gehalt an daraus abgeleiteten Proteinen zu schließen (Anderson *et al.* 1997). Viele für den Organismus wichtige Funktionen werden durch strukturelle Modifikationen erst zu einem späteren Zeitpunkt, während und vor allem nach der Translation, durch z.B. posttranslationale Modifizierung oder durch weitergehende Prozessierungsschritte der ursprünglichen Genprodukte bereitgestellt. Somit können viele Funktionen und Krankheiten, an denen Proteine und Peptide beteiligt oder für die sie sogar verantwortlich sind, auch nach der Entschlüsselung des humanen Genoms nicht oder nur teilweise verstanden werden.

Um Funktionen auf molekularer Ebene zu entschlüsseln und Krankheiten zu erforschen, ist es notwendig, die Genprodukte, also Proteine und Peptide, zu beleuchten. Aus diesem Grund müssen (ebenfalls systematische) Methoden

---

entwickelt werden, die folgenden Ansprüchen genügen:

1. Extraktion und hochauflösende Darstellung aller Proteine eines Organismus oder eines Teils eines Organismus.
2. Identifizierung und (relative) Quantifizierung jeder Komponente.
3. Vergleich, Analyse und Darstellung von Veränderungen in diesem komplexen System.

Es entwickelte sich, folgend der Terminologie von Genomics und Transscriptomics, eine Richtung in der biologischen Forschung, die als *Proteomics* bezeichnet wird. Der Begriff *Proteom* setzt sich zusammen aus den Begriffen **Protein** und **Genom** und wurde von M. Wilkens eingeführt. Er ist definiert als die „komplette quantitative Inhaltsangabe des Proteinexpressionsmusters eines Genoms“, welches aus mehreren Zehntausend bis Hunderttausend Proteinen und deren Isoformen besteht (Wilkins *et al.* 1996; Wilkins *et al.* 1996; James 1997). Ein Proteom ist um ein vielfaches komplexer und dynamischer als ein Genom, da es von vielen Faktoren (äußere Bedingungen, Entwicklungsstand des Organismus etc.) abhängig ist. Ein gutes Beispiel hierzu liefert Abbildung 1-1, die einen *Orgyia antiqua* einmal als Larve und einmal als Falter zeigt. Beide Organismen haben das gleiche Genom, aber unterschiedliche Proteome, was sich in ihrem Phänotyp ausdrückt. Aus diesen Vorgaben erwachsen zwei Grundvoraussetzungen, die in der Proteomforschung unabdingbar sind, um verwertbare Ergebnisse zu erzielen. Einerseits die Notwendigkeit, die Umgebungsvariablen für ein Proteom genau zu definieren und andererseits über hochautomatisierte Systeme zur Visualisierung, Identifizierung und Datenanalyse zu verfügen.



Abbildung 1-1 Larve und Falter des *Orgyia Antiqua*. Ein Genom und zwei Proteome.

Die Basistechnologien für die Proteomanalyse bilden zur Zeit die zweidimensionale Gelelektrophorese (2D-Gel) als bevorzugte Trenntechnik in Verbindung mit der Massenspektrometrie zur Identifizierung der Analyten.

Neben einigen großen Vorteilen, wie z.B. der Auflösung der 2-D Gelelektrophorese, hat diese allerdings auch einige klare Nachteile zu verzeichnen. So ist der dynamische Bereich gering, so dass sich gering konzentrierte Proteine nicht neben hoch konzentrierten darstellen lassen. Ferner ist auch die Reproduzierbarkeit der Gele nicht besonders hoch, was deren Einsatz bei differentiellen Vergleichen erschwert. Desweiteren schließt die 2D-Gelelektrophorese mit den Peptiden eine wichtige Substanzgruppe bei der Proteomanalyse aus. Unterhalb eines Molekulargewichts von ca. 10.000 Da können Peptide bzw. kleine Proteine mit dieser Methode nicht effektiv fokussiert werden, d.h. Informationen über sie gehen im klassischen *Proteomics*-Ansatz verloren. Dies ist besonders aufgrund der wichtigen Funktion dieser

Komponenten im Organismus (vgl. Kap. 2.4, S.39) eine starke Limitierung dieser Technik.

Die Massenspektrometrie als Analysemethode kann die Basisanforderungen der Proteomanalyse - großer dynamischer Bereich, hohe Automatisierbarkeit, Möglichkeit der Analyse komplexer Proben, zuverlässige Identifizierung der Analyten sowie Analyse von posttranslationalen Modifikationen – in hohem Maße befriedigen und ist deshalb die Methode der Wahl in der Peptid- und Proteinanalyse.

Für die Substanzklasse der Peptide und kleinen Proteine wurden von unserer Arbeitsgruppe Methoden entwickelt, die auf die 2D-Gelelektrophorese als Basistechnologie verzichten und stattdessen flüssigkeitschromatographische Techniken in Verbindung mit massenspektrometrischen Methoden anwenden, um ein *Peptidom* darzustellen.

Der *Peptidomics*-Ansatz ergänzt den *Proteomics*-Ansatz um die Stoffklasse der Proteine und Peptide mit einem Molekulargewicht bis 15 kDa. Der Begriff Peptidom ist zusammengesetzt aus **Peptid** und **Genom** und beschreibt die qualitative und quantitative Analyse einer Peptidquelle zu einem definierten Zeitpunkt. Erreicht wird dies durch Einsatz von zwei sich ergänzenden Technologien. Dies ist zum einen die Kartierung von Peptidquellen und deren systematische Identifizierung (**Peptide Trapping**, PT) (Schulz-Knappe *et al.* 1996), und zum anderen die differentielle Analyse zweier miteinander in Beziehung stehender Proben mit anschließender Identifizierung der Unterschiede (**Differential Peptide Display**, DPD<sup>TM</sup>) (Schulz-Knappe *et al.* 2001). Im letzteren Fall werden zwei Proben unterschiedlichen Status (z.B. pathologische Probe und gesunde Kontrollprobe) qualitativ und quantitativ hinsichtlich ihres Peptidgehalts miteinander verglichen.

Sowohl der Proteomics- als auch der Peptidomics–Ansatz eröffnen der Medizin ein großes Potential in der Erforschung von Krankheiten. Mit der Durchführung von differentiellen Analysen eröffnet sich die Möglichkeit, diagnostische Marker oder auch Therapeutika zu entdecken (Jungblut *et al.* 1999;

---

Banks *et al.* 2000; Chambers *et al.* 2000; Carrette *et al.* 2003).

Von hoher wissenschaftlicher Bedeutung ist die umfassende Analyse von Peptidomen, insbesondere von Proben, die bei Patienten gewonnen werden (Clinical Proteomics, Clinical Peptidomics). Damit ist es erforderlich, mit teilweise sehr geringen Probenmengen und entsprechend hoher Empfindlichkeit vorzugehen. Die Fülle an natürlich vorkommenden Peptiden legt nahe, von Beginn an mit automatisierten Methoden einen hohen Durchsatz zu verwirklichen und für die universelle Nutzung der umfangreichen Daten diese in Datenbanksystemen abzulegen, was für ein sinnvolles Retrieval der Daten eine Voraussetzung ist.

Ein in der Proteomanalyse übliches technisches Hilfsmittel ist für die Peptidomanalyse nicht sinnvoll: Der Verdau der nativen Peptide durch entsprechende Proteasen, i.d.R. Trypsin. Es ist davon auszugehen, dass sich in einer Probe mehrere Varianten des gleichen Peptides befinden, z.B. N- oder C-terminal teilweise prozessierte Peptide. Diese haben unter Umständen unterschiedliche biologische Relevanz, und können daher als „informative Moleküle“ angesehen werden. Ein Verdau würde diese Information stark beeinträchtigen und so dem Ziel, einen biologischen Status aus der Probe zu erheben, zuwiderlaufen.

Ziel dieser Arbeit ist es, die Identifizierung von nativen, d.h. nicht artifiziell veränderten Peptiden aus Körperflüssigkeiten mit massenspektrometrischen Methoden so weiterzuentwickeln, dass diese Methode universell einsetzbar wird und den oben aufgeführten Voraussetzungen entspricht. Um die umfassende Charakterisierung einer Peptidquelle besser beschreiben zu können, wird der Begriff *Peptide-Inventory* (*Peptid-Inhaltsverzeichnis*) eingeführt. Die Erstellung eines *Peptide-Inventorys*, das als Weiterentwicklung des *Peptide Trapping* gesehen werden kann, umfasst die Auftrennung hochkomplexer Proben in handhabbare Subfraktionen, die Visualisierung/Kartierung der Quellen, die Identifizierung per Sequenzidentität und die Archivierung aller zur Verfügung stehenden Daten einer Peptidquelle. Hier kam es im wesentlichen darauf an, ein hochautomatisiertes Identifizierungsgerüst zu entwickeln,

welches die Kartierung, die Messungen zur Identifizierung, die Datenauswertung und die Weiterverarbeitung der gewonnenen Daten in einer Datenbank in sich vereint.

## 2 Theoretische Grundlagen

### 2.1 Grundlagen der massenspektrometrischen Analyse von Peptiden

Die Massenspektrometrie hat sich in den letzten Jahren zur Standardtechnologie für die Protein- und Peptid-Analytik entwickelt. Auch im Peptidomics-Ansatz bilden massenspektrometrische Methoden die Basis der Analytik. Aus diesem Grund werden im Folgenden die Grundlagen der Massenspektrometrie im Allgemeinen und in Bezug auf die Peptidanalytik beschrieben.

Die Massenspektrometrie ist eine analytische Methode zur Bestimmung der Molekülmasse freier Ionen im Hochvakuum. Grundsätzlich müssen die Probenmoleküle dazu in einer *Ionenquelle* ionisiert werden. Die entstandenen Ionen werden dann im *Massenanalysator* nach ihrem Masse/Ladungsverhältnis getrennt und von einem *Detektor* in Signale übersetzt, die das Massenspektrum liefern. Dieses macht Aussagen darüber, welches Ion in welcher Quantität detektiert wurde.

Ein schematischer Aufbau eines Massenspektrometers ist in Abbildung 2-1 gezeigt. Die Ionisierung der Probe erfolgt dabei über Hinzufügen oder Entfernen von Elektronen aus dem Probenmolekül oder durch Protonierung/Deprotonierung des Analyten. Für die Massentrennung der geladenen Teilchen stehen mehrere geeignete Systeme mit passenden Detektoren zur Verfügung. Bei den Analysatoren kommen häufig Kombinationen aus zwei oder drei Analysatoren in sogenannten *Hybrid*-Geräten zum Einsatz, da sie die Vorteile der einzelnen Systeme in sich vereinen. Als Beispiel für in der Protein/Peptid-Analytik wichtige Hybrid-Systeme seien hier die Quadrupole-Time-of-Flight-Geräte (QTOF) genannt.

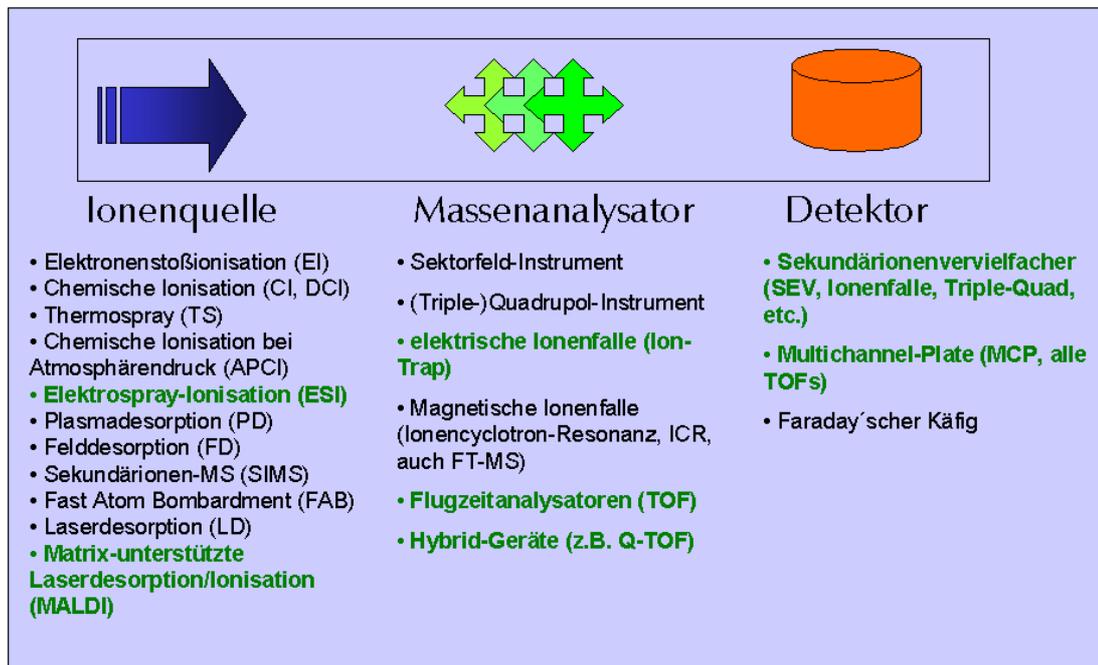


Abbildung 2-1 Komponenten eines Massenspektrometers. Grün hervorgehoben sind die in der Arbeit verwendeten Komponenten.

Bei den Ionenquellen ist zwischen *offline-* oder *online-Probenzuführung* zu unterscheiden. Bei den Offline-Methoden ist eine spezielle Probenvorbereitung erforderlich, so dass eine direkte Kopplung mit einer chromatographischen Technik aus apparatetechnischen Gründen nicht möglich ist. Beispiele hierfür sind MALDI und FAB. Die auf flüssiger oder gasförmiger Ionenzuführung beruhenden Techniken (ESI, APCI, LISIMS, EI, CI) können dagegen mit chromatographischen Techniken, wie Gas- oder Flüssigkeitschromatographie (GC, HPLC) oder Kapillarelektrophorese (CE) *online* gekoppelt werden. Dies ist bei hoch komplexen Proben der Bioanalytik oft notwendig, da die Proben erst durch vorausgehende Trennschritte der Massenspektrometrie zugänglich gemacht werden können.

Eine weitere Einteilung der Ionisierungsmethoden kann nach der Unterscheidung der Ionisierung in *hart* und *weich* erfolgen. Bei *harten* Ionisierungsmethoden (EI, CI) wird der Analyt entweder durch thermische Belastung oder den Eintrag von Energie beim Ionisationsprozess so „gestresst“, dass nur kleinere Moleküle erfasst werden können. Biopolymere wie Peptide und Pro-

teine sind mit diesen Techniken nicht zu erfassen. Hierzu bedurfte es der Entwicklung *weicher* Ionisierungsmethoden, die auch polare oder thermisch labile Peptide in die Gasphase überführen können, ohne sie zu zerstören.

Mit der Entwicklung der FAB-Ionenquelle ((Barber *et al.* 1981)) wurde ein erster Schritt in diese Richtung gemacht. Mit dieser Technik konnte ein Molekularmassenbereich bis zu ca. 8.000 Da erfasst werden. Diese Technik verlor am Ende der 80er Jahre zunehmend an Bedeutung, als zwei neue Ionisierungstechniken etabliert wurden, die auch heute noch die Methoden der Wahl sind, wenn es um die sanfte Ionisierung von Biopolymeren geht. Bei der **Elektrospray Ionization (ESI)** (Fenn *et al.* 1989) und deren Varianten mit niedrigen Flussraten (mikro-ESI und nano-ESI (Wilm *et al.* 1996)), werden die Analytionen in der flüssigen Phase präformiert und durch Versprühen in einem elektrischen Feld in die Gasphase überführt. **Matrix Assisted Laser Desorption Ionization (MALDI)** wurde zwischen 1985 und 1988 durch M. Karas und F. Hillenkamp eingeführt (Karas *et al.* 1985; Karas *et al.* 1987; Karas *et al.* 1988). Hierbei werden die in eine feste Matrix eingebetteten Analytmoleküle durch einen gepulsten Laser beschossen und durch Matrixionen ionisiert. Abbildung 2-2 gibt einen Überblick über verschiedene Substanzklassen der massenspektrometrischen Bioanalytik und zeigt, dass mittlerweile alle relevanten Biopolymere mit MALDI oder ESI der Massenspektrometrie zugänglich gemacht werden können.

Substanzklassen	Massenspektrometrie-Techniken	Typischer Molekulargewichts-Bereich der analysierten Verbindungen
Desoxyribonukleinsäuren		bis 20.000 u
Ribonukleinsäuren		bis 20.000 u
Proteine		bis 250.000 u
Peptide	GC-(E/C)MS, FD-MS, FAB-MS, MALDI-MS, LC/ESI-MS	bis 5.000 u
Kohlenhydrate	GC-(E/C)MS, FD-MS, FAB-MS, MALDI-MS, LC/ESI-MS	bis 5.000 u
Lipide	GC-(E/C)MS, FD-MS, FAB-MS, MALDI-MS, LC/ESI-MS	bis 1.000 u

Abbildung 2-2 Massenspektrometrische Ionisierungsmethoden, dargestellt anhand ihrer Anwendungsmöglichkeiten auf biochemisch wichtige Substanzklassen; nach (Lehmann 1996).

Die Daten, die mit diesen beiden Ionisierungstechniken und der geeigneten Analytortechnik gewonnen werden können, sind sehr vielfältig. Von besonderem Interesse sind drei Informationen: Die *molekulare Masse* (MW), die *Aminosäuresequenz* (inklusive der Beschreibung posttranslationaler Modifikationen wie Glykosylierung, Phosphorylierung u.v.m.) und eine *Quantifizierung* des Analyten.

Die molekulare Masse kann als monoisotopische (monoisotopic mass) oder Durchschnittsmasse (average mass) bestimmt werden. Die monoisotopische Masse wird dabei aus den exakten Massen der jeweils häufigsten Isotopen der Elemente einer Verbindung berechnet. Die Durchschnittsmasse berücksichtigt alle Isotopen einer Verbindung und bezieht den Wert aus den durchschnittlichen Atommassen der Elemente einer Verbindung (Abbildung 2-3).

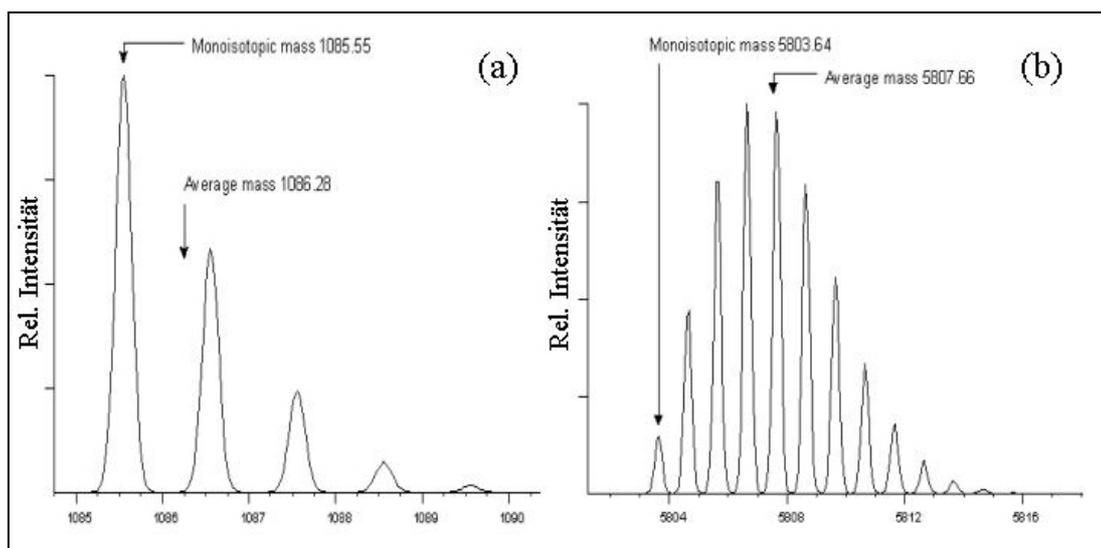


Abbildung 2-3 Massenspektren der Peptide (a) HLKTEAEMK und (b) Insulin. Gezeigt ist jeweils ein Ausschnitt der Molekülmassenregion. Eingezeichnet sind sowohl die Durchschnittsmasse (average mass) als auch die monoisotopische Masse.

Ob die monoisotopische Masse in einem Spektrum dargestellt werden kann, hängt vom *Auflösungsvermögen* der Methode und von der Masse des Analyten ab. Abbildung 2-4 zeigt das Peptid Glucagon bei den Auflösungen 1.000, 3.000, 10.000 und 30.000 nach der Definition der Halbwertsbreite (**F**ull **W**idth **H**alf **M**aximum; FWHM), die in der Flugzeit-Massenspektrometrie üblich ist. Es ist zu erkennen, dass eine Bestimmung der monoisotopischen Masse nur

bei den Auflösungen  $> 10.000$  möglich ist. Die zur Zeit am Markt befindlichen Flugzeit- und Hybrid-Massenspektrometer für die Bioanalytik können maximal eine Auflösung von  $10.000$ -  $20.000$  (FWHM) erreichen.

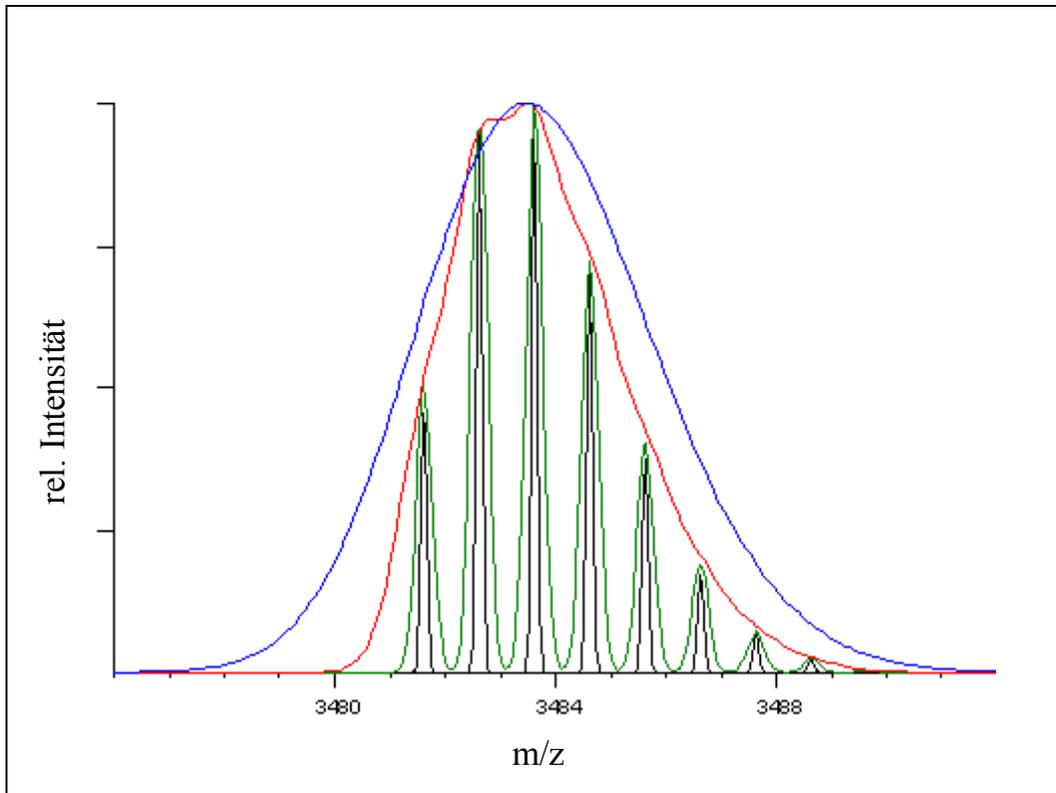


Abbildung 2-4 Massenspektren des humanen Glucagons. Gezeigt ist ein Ausschnitt der Molekülmasse-  
region mit der Auflösung  $1.000$  (Blau),  $3.000$  (rot),  $10.000$  (grün) und  $30.000$  (schwarz).

Je größer ein Molekül ist, desto höher muss die Auflösung der massenspektrometrischen Methode sein, um eine isotopenaufgelöste Darstellung zu ermöglichen.

Aber auch strukturelevante Daten lassen sich mit einem Massenspektrometer erzeugen. In Abhängigkeit vom Biopolymer (Peptid, Protein, Poly-/Oligosaccharid, Nucleinsäure) gibt es unterschiedliche Methoden zur Strukturanalyse.

Für Peptide ist es möglich, durch stoßinduzierte Fragmentierung (collision induced fragmentation; CID) aus dem erzeugten Massenspektrum auf die Aminosäuresequenz rückzuschließen (vgl. Kap. 2.1.3.2, S.33). Diese Methode wird als massenspektrometrische Sequenzierung bezeichnet. Eine weni-

ger verbreitete Methode ist das *Peptide Ladder Sequencing*, bei dem durch chemischen Abbau (Chait *et al.* 1993) oder durch Carboxypeptidasen (Patterson *et al.* 1995) Aminosäurespaltungen von Peptiden induziert werden, die über MALDI-TOF-MS dargestellt werden. Aus den gebildeten Bruchstücken wird die Aminosäuresequenz rekonstruiert. Somit verfügt man über massenspektrometrische Methoden, die zur „klassischen“ Edman-Sequenzierung komplementär sind. CID hat zusätzlich den Vorteil der Schnelligkeit und der Möglichkeit, direkt aus komplexen Gemischen zu sequenzieren, ohne einzelne Peptide vorher chromatographisch aufzureinigen zu müssen.

Für die Sequenzanalyse von Proteinen stehen interessante massenspektrometrische Methoden bereit. So werden in der Proteomanalytik die durch 2D-Gelelektrophorese aufgetrennten Proteine durch spezifische Proteasen „artifizial verdaut“ (meistens mit Trypsin). Die Massen der entstandenen Bruchstücke werden durch MALDI-MS hochgenau bestimmt und zur Identifizierung mit einer Protein- oder DNA-Datenbank abgeglichen. Zusätzlich steht hier die massenspektrometrische Sequenzierung der tryptischen Fragmente über CID zur Verfügung. Neben der Ermittlung der Primärsequenz eines Peptides/Proteins ist die Ermittlung posttranslationaler Modifikationen (Quadroni 2001; Rudd *et al.* 2001) und die Lokalisation von Disulfidbrücken von großer Bedeutung (Neitz *et al.* 2001). Zudem können mit der MALDI- und ESI-MS auch in gewissen Grenzen nichtkovalente Wechselwirkungen, beispielsweise Proteinkomplexe (Kuwata *et al.* 1998), (Lyubarskaya *et al.* 1998), untersucht werden.

Eine weitere Fragestellung, die durch massenspektrometrische Methoden zu lösen ist, sind quantitative Bestimmungen. Die absolute Quantifizierung von Peptiden in komplexen biologischen Matrices mit massenspektrometrischen Methoden stellt sich nach wie vor als schwierig dar, weil das Ionisationsverhalten auch innerhalb dieser Stoffklasse stark variiert und somit die Verwendung externer Standards i.d.R. nicht ausreichend ist. In sehr komplexen biologischen Matrices wie beispielsweise Blutplasma ergibt sich zudem das

Problem der Selektivität. Es ist durchaus nicht selten, dass falsche Komponenten quantifiziert werden, wenn ausschließlich über das Masse zu Ladungsverhältnis quantifiziert wird, da hier isobare Ionen, die koeluiieren zusammen quantifiziert werden. Dieses Problem ist am besten durch die Quantifizierung charakteristischer Fragmentionen in Zusammenhang mit dem Vorläuferion zu lösen (**selected reaction monitoring**, SRM), wobei die zu quantifizierende Substanz vorher bekannt sein muss. Am weitesten entwickelt ist die Quantifizierung von Peptiden über eine online HPLC-ESI-MS Kopplung mit Triple-Quadropol, Ion-Trap oder Quadrupol-TOF Analysatoren. Eine relative Quantifizierung, wie sie beim Peptidomics Ansatz benötigt wird, kann jedoch sowohl mit ESI-MS, als auch mit MALDI-MS realisiert werden, da hier einzig eine hohe Reproduzierbarkeit der Messungen Voraussetzung ist.

### **2.1.1 Matrix Assisted Laser Desorption Ionization Mass Spectrometry (MALDI-TOF-MS)**

#### *2.1.1.1 Ionisierung*

Die Technik der MALDI-MS ist aus der in den 70er Jahren verbreiteten Laserdesorption/Ionisation (LDI) (Hillenkamp 1983) hervorgegangen. Bei der LDI wird die auf eine Metalloberfläche aufgebrachte Probe direkt mit Photonen bestrahlt. Die erzeugten Massenspektren waren charakterisiert durch geringe Intensitäten und hohe Fragmentierung. So konnten im allgemeinen Moleküle bis zu mehreren kDA erfasst werden.

Die Verwendung einer absorbierenden festen Matrix (Nikotinsäure) ist die Grundlage der bis heute angewendeten MALDI-MS (Karas *et al.* 1985; Karas *et al.* 1987; Karas *et al.* 1988). In den folgenden Jahren wurden andere absorbierende Matrices, wie Sinapinsäure (SIN), Ferulasäure (Beavis *et al.* 1989; Beavis *et al.* 1989), 2,5-Dihydroxybenzoesäure (DHB) (Strupat *et al.* 1991) oder  $\alpha$ -Hydroxy-4-Cyanozimtsäure (HCCA) (Beavis *et al.* 1992) beschrieben. Hierbei wird der Analyt zusammen mit einem ca. 10.000 bis 100.000-fachen molaren Überschuss der bei der verwendeten Laserwellen-

länge absorbierenden gelösten Matrixsubstanz auf ein Metalltarget gebracht, so dass nach Verdunstung des Lösungsmittels Matrix und Probenmoleküle kokristallisieren.

Der Prozess der Desorption und Ionisation lässt sich mit verschiedenen Modellen erklären. Einen guten Überblick über diese Prozesse gibt Glückmann (Glückmann 2001) in seiner Dissertation. Abbildung 2-5 zeigt in schematischer Darstellung die Vorgänge bei Desorption, Ionisierung und Extraktion in der MALDI-MS.

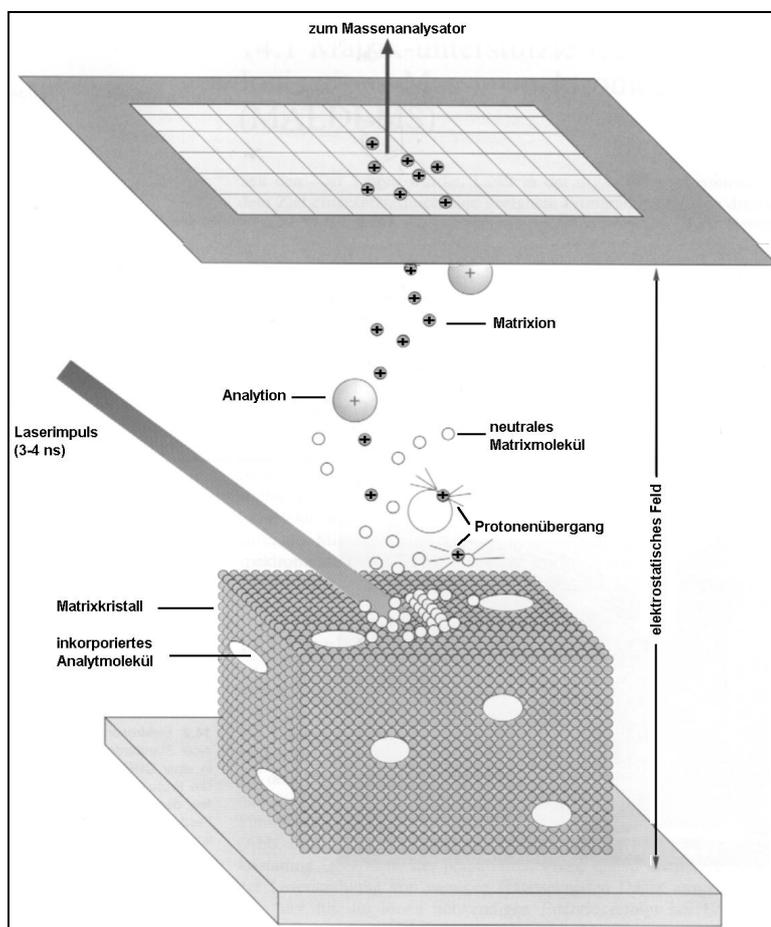


Abbildung 2-5 Prinzip der Desorption, Ionisierung und Extraktion in der MALDI-MS; nach (Eckerskorn 1998).

Als Laser im UV-Bereich werden heute sowohl Stickstoff- als auch ND-YAG-Laser (Neodym dotierte Yttrium-Aluminium-Granat Kristalle) verwendet, die bei einer Wellenlänge von 337 nm bzw. 355/266 nm arbeiten. Weniger verbreitet aufgrund höherer Kosten und geringerer Stabilität sind Infrarot-Laser

(z.B. Er-YAG). Sie besitzen aber bei einigen Problemstellungen durchaus Vorteile (Eckerskorn *et al.* 1992; Eckerskorn *et al.* 1997)

Mit der MALDI-Ionisierung ist die Voraussetzung gegeben, Peptide und sogar Proteine bis zu einem Molekulargewicht von mehreren 100 kDa schonend in die Gasphase zu überführen und zu ionisieren, so dass sie der Massenanalyse zugänglich sind.

### 2.1.1.2 Massenanalyse

In einem MALDI-TOF Massenspektrometer ist die matrixunterstützte Laser-desorption/ Ionisation als Ionisierungsmethode mit einer *Time-of-Flight* (TOF) Massentrennung gekoppelt. Dabei wird im Hochvakuum ein gepulster Ionenstrahl erzeugt und die Flugzeit der Ionen von der Quelle zum Detektor über eine sehr genaue elektronische Messung bestimmt. Der Start einer Einzelmessung wird durch den Puls des zur Desorption und Ionisierung eingesetzten Lasers vorgegeben. Zur Verbesserung des Signal-zu-Rausch-Verhältnisses können prinzipiell beliebig viele Einzelmessungen aufsummiert werden.

Grundlage der Massenanalyse in einem TOF-Detektor ist die Tatsache, dass Ionen mit unterschiedlichen Quotienten aus Masse  $m$  und Ladung  $z$  ( $m/z$ ) nach Extraktion und Beschleunigung aus der Ionenquelle zwar über die gleiche Energie verfügen, aber unterschiedliche Geschwindigkeiten besitzen. Die Geschwindigkeit zur Durchquerung der Flugröhre ist für jedes  $m/z$  verschieden: Ionen mit einem großen  $m/z$ -Verhältnis benötigen eine längere Zeit, um den Detektor zu erreichen, als solche mit einem kleineren  $m/z$ -Verhältnis.

Die klassische Bestimmungsgleichung für die kinetische Energie  $E_{kin}$  eines mit der Geschwindigkeit  $v$  bewegten Ions der Masse  $m$  lässt sich mit der an der Beschleunigungsstrecke anliegenden Potentialdifferenz  $U_{acc}$  und der Ladung des Ions  $z$  (in der Regel  $z = 1$ ) verknüpfen:

$$E_{kin} = \frac{1}{2}mv^2 = z \cdot e \cdot U_{acc} \quad \text{Gleichung 2-1}$$

Die Geschwindigkeit der Ionen  $v$  ergibt sich aus der Zeit  $t$ , die ein Ion benötigt, um die feldfreie Driftstrecke  $L$  im Flugrohr zu passieren.

$$v = \frac{L}{t} \quad \text{Gleichung 2-2}$$

Schließlich erhält man durch Einsetzen von Gl. 2-2 in Gl. 2-1 und Umformen nach  $m/z$  einen allgemeinen Zusammenhang zwischen der Flugzeit eines Ions und seinem  $m/z$ -Verhältnis:

$$\frac{m}{z} = \frac{2e \cdot U_{acc}}{L^2} t^2 \quad \text{Gleichung 2-3}$$

Das Masse-zu-Ladungsverhältnis ist also dem Quadrat der Flugzeit proportional und kann somit aus ihr bestimmt werden.

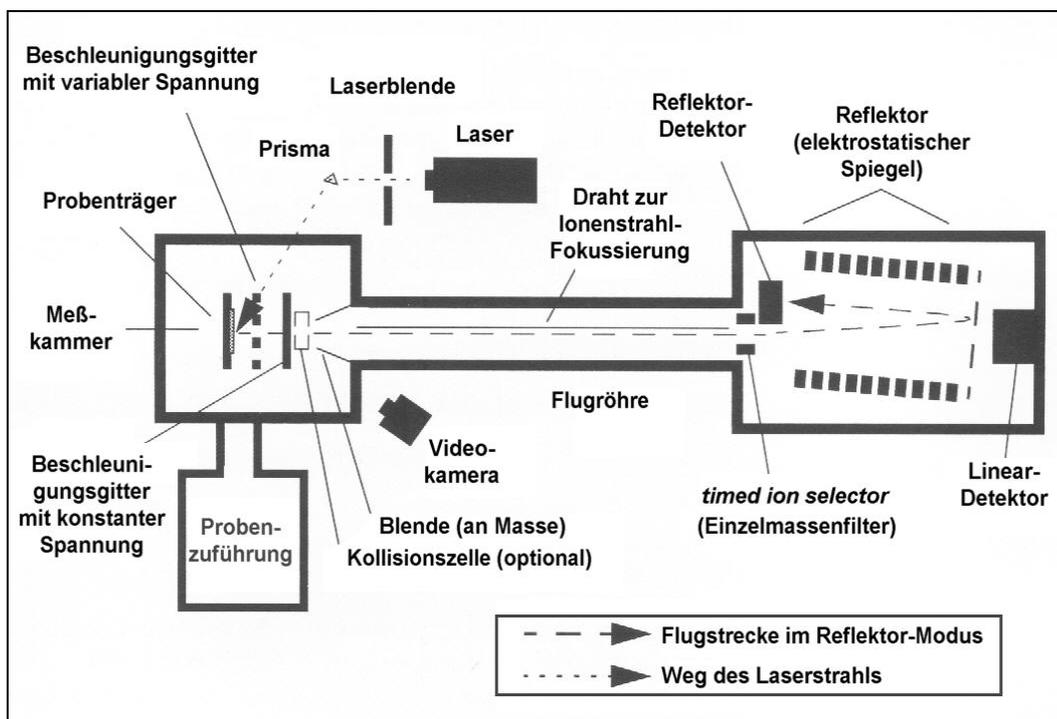


Abbildung 2-6 Schematischer Aufbau des in dieser Arbeit verwendeten Reflektor-MALDI-TOF Massenspektrometers *Voyager-DE STR* der Fa. PerSeptive Biosystems; nach (PerSeptive 1997)

Auflösungslimitierende Faktoren die durch die Verteilung in Zeit, Raum und kinetischer Energie des ursprünglich gebildeten Ionen-Paketes bedingt sind,

konnten mit Entwicklungen wie *der verzögerten Ionenextraktion* (delayed extraction<sup>1</sup>, DE) (Vestal *et al.* 1998) und der Weiterentwicklung der ursprünglich linearen TOF-Massenspektrometern zu Reflektor-TOF-Instrumenten (Mamyrin *et al.* 1973) teilweise eliminiert werden. Bei der DE wird zwischen dem Auslösen des Laserpulses und der Ionenextraktion, also dem Anlegen der Beschleunigungsspannung, eine kurze (variable) Verzögerung eingefügt (Abbildung 2-7). In diesem Zeitraum kann eine Inhomogenität der Probe, die z.B. durch die Laser-induzierte Geschwindigkeitsverteilung der Molekülonen in der Gasphase verursacht wird, bis zu einem gewissen Grad ausgeglichen werden. Anschließend können die Molekülonen mit gleichem  $m/z$ -Verhältnis unter annähernd identischen Bedingungen extrahiert werden.

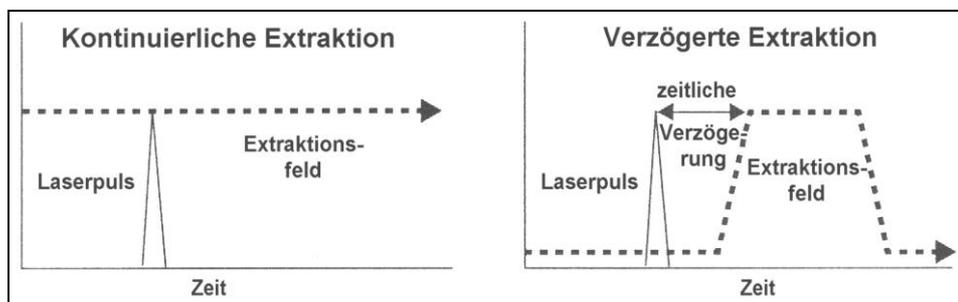


Abbildung 2-7 Prinzip von kontinuierlicher (*continuous*) und verzögerter (*delayed*) Extraktion.

Die Spiegelung der Ionen in einem elektrostatischen Ionenspiegel (Reflektor) hat eine zusätzliche Auflösungsverbesserung aufgrund einer Refokussierung von Ionen gleicher Masse aber unterschiedlicher kinetischer Energie zur Folge. Geräte auf dem heutigen Stand der Technik bieten eine Auflösung von 10.000-15.000 (FWHM) in einem Massenbereich von 800-7.000 u.

### 2.1.2 Elektrospray Ionization Mass Spectrometry (ESI-MS)

Unter dem Begriff *Elektrospray* versteht man die Dispersion einer Flüssigkeit

<sup>1</sup> DE ist ein Warenzeichen von Applied Biosystems. Andere Hersteller bezeichnen diese Technologie als „time-lag-focussing“ (TLF) oder „Pulsed Ion Extraction“ (PIE). In dieser Arbeit wird ausschließlich der Term DE verwendet.

in sehr viele kleine geladene Tröpfchen mit Hilfe eines elektrostatischen Feldes. Dieses bereits im letzten Jahrhundert beobachtete Phänomen wurde Ende der 60er Jahre zum ersten Mal zur Bestimmung der Molekülmasse genutzt (Dole *et al.* 1968). Mitte der 80er Jahre konnten darüberhinaus unabhängig voneinander die Arbeitsgruppen von J. Fenn an der Universität in Yale (Yamashita 1984) und von M. Alexandrov (Aleksandrov *et al.* 1984) an der Universität in Leningrad zeigen, daß der Elektrospray-Prozeß eine definierte Ionisierung und komplette Desolvatisierung von in Lösung versprühten Analytmolekülen bewirkt.

### 2.1.2.1 Ionisierung

Der Prozess der Elektrospray-Ionisierung kann makroskopisch wie auch mikroskopisch betrachtet werden.

Bei der makroskopischen Betrachtungsweise entsteht am Ende der Kapillare ein Flüssigkeitskegel (Taylor-Konus), der sich zu einem Flüssigkeitsfilament aus vielen Tröpfchen verzweigt. Aus diesem Konus bildet sich ein Tröpfchennebel, aus dem durch Desolvatisierung freie Analyt-Ionen entstehen (Abbildung 2-8, oben).

Der Weg der Ionen von der Lösung in die Gasphase läßt sich formal in vier Einzelschritte unterteilen (mikroskopische Betrachtungsweise):

1. die Bildung von kleinen geladenen Tröpfchen aus Elektrolyten,
2. kontinuierlicher Lösungsmittelverlust dieser Tröpfchen durch Verdampfen, wobei die Ladungsdichte an der Tröpfchenoberfläche zunimmt,
3. wiederholter spontaner Zerfall der Tröpfchen in Mikrotröpfchen (sog. *Coulomb-Explosionen*) und schließlich
4. Desolvatisierung der Analytmoleküle beim Transfer in das Massenspektrometer.

Der ESI-Prozeß (Abbildung 2-8) beginnt mit der kontinuierlichen Zuführung des gelösten Analyten über eine HPLC- oder Spritzenpumpe mit typischerweise 1-200  $\mu\text{L}/\text{min}$  an die Spitze einer Edelstahlkapillare. Die Zerstäubung des Flüssigkeitsstroms findet in einem starken elektrostatischen Feld von annähernd  $10^6$  V/m statt, das durch das Anlegen einer Hochspannung von 3-6 kV zwischen der Spitze der Edelstahlkapillare und einer 1-3 cm entfernten Gegenelektrode (*interface plate*) entsteht.

Ähnlich wie in der Elektrophorese werden Ionen unterschiedlicher Polarität in dem elektrischen Feld getrennt: Eine Ionensorte driftet dabei an die Flüssigkeitsoberfläche, die entgegengesetzt geladene Ionenspezies entfernt sich von ihr. Dieser Vorgang findet statt, bis das elektrische Feld innerhalb der Flüssigkeit durch die Umverteilung aufgehoben ist. Durch die an der Flüssigkeitsoberfläche akkumulierten Ionen einer Polarität, die weiter in Richtung der Gegenelektrode gezogen werden, entsteht eine als *Taylor-Konus* bezeichnete charakteristische Flüssigkeitsform: die Oberflächenspannung der Flüssigkeit wirkt dem elektrischen Feld entgegen. Bei ausreichend hohem elektrischen Feld ist der *Taylor-Konus* stabil und emittiert von seiner Spitze einen kontinuierlichen, filamentartigen Flüssigkeitsstrom von wenigen Mikrometern Durchmesser. In einiger Entfernung von der Gegenelektrode wird dieser instabil und zerfällt in winzige, aneinandergereihte Tröpfchen. Die Tröpfchenoberfläche ist mit Ionen einer Ladung angereichert, für die es keine entsprechende Anzahl Gegenionen im Tröpfchen gibt, so daß nach außen eine Nettoladung entsteht. Die Art der Ladung und ihre Anreicherung läßt sich über die Polarität und durch die Höhe der an der Edelstahlkapillare angelegten Spannung steuern.

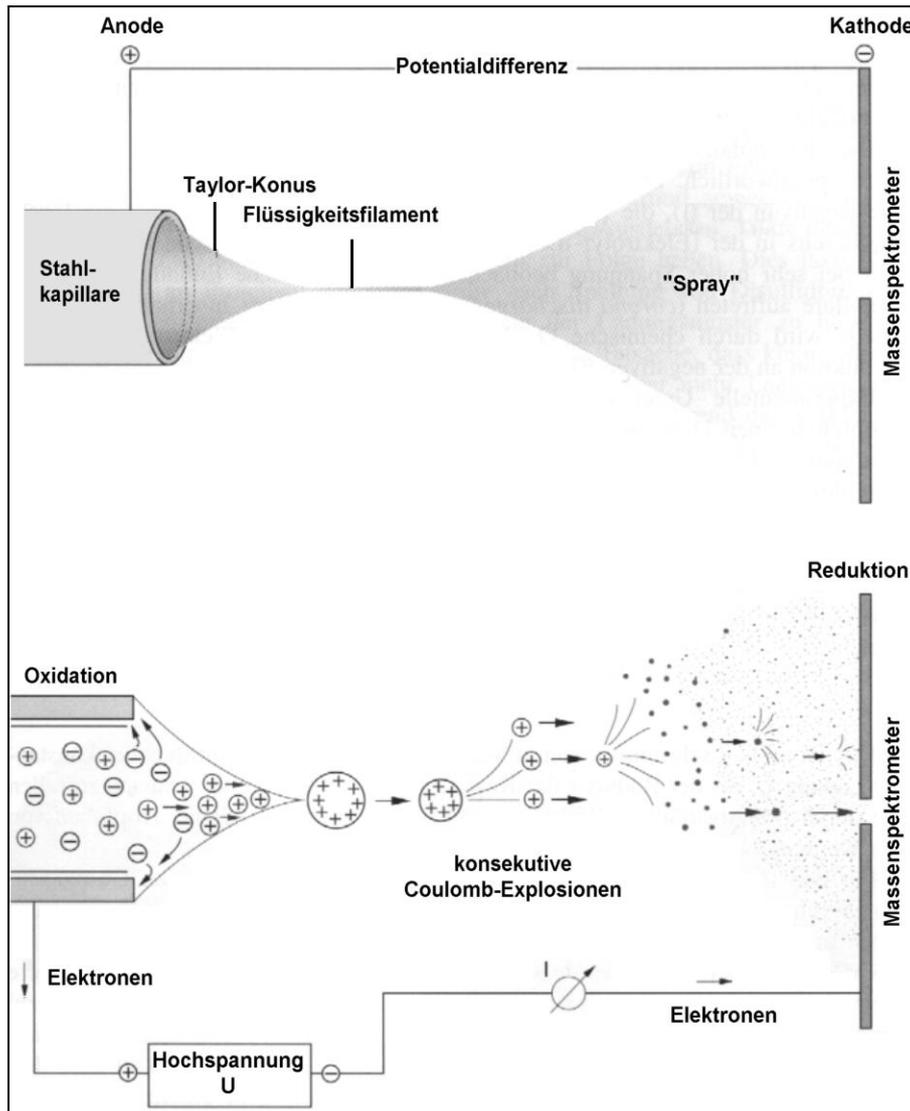


Abbildung 2-8 Prinzip des makroskopischen (oben) und mikroskopischen ESI-Prozesses (unten) (Eckerskorn 1998).

Abhängig von den Spraybedingungen und den Lösungsmiteleigenschaften haben die im ESI erzeugten geladenen Tröpfchen einen Durchmesser von 2-3  $\mu\text{m}$ . Aufgrund der Lösungsmittelverdampfung findet eine weitere Schrumpfung der Tröpfchen auf ihrem Weg in den Analysatorteil des Massenspektrometers statt. Ein beheizter, dem Tröpfchenstrom entgegengesetzter, hochreiner  $\text{N}_2$ -Gasstrom (*curtain gas*) kann diesen Vorgang unterstützen. Durch die kontinuierliche Verkleinerung der Tröpfchen erhöht sich die Ladungsdichte auf der Oberfläche bis zum sog. *Rayleigh-Limit*, bei dem die Ladungsdichte so groß wird, dass die Coulomb-Abstoßung die Oberflächenspannung überwindet. Die Rayleigh-Gleichung gibt an, wann die Ladung  $Q$

die Oberflächenspannung  $\gamma$  ausgleicht:

$$Q^2 = 64 \cdot \pi^2 \cdot \epsilon_0 \cdot \gamma \cdot r^3 \quad \text{Gleichung 2-4}$$

mit

$\epsilon_0$ : Dielektrizitätskonstante im Vakuum

r: Radius des Tröpfchens

Nach Überschreitung des Limits zerfallen die Tröpfchen durch die Abstoßung gleichnamiger Ladungen in viele kleine Tröpfchen von nur noch wenigen Nanometern Durchmesser (*Coulomb-Explosionen*, Abbildung 2-8, unten).

Für den Mechanismus, der das Entstehen von freien Ionen in der Gasphase aus geladenen Flüssigkeits-Tröpfchen beschreibt, werden nach wie vor verschiedene Theorien diskutiert. Das ältere Modell von Dole und Röllgen wird als *Single Ion in Droplet Theory (SIDT)*<sup>2</sup> bezeichnet (Dole *et al.* 1968). Es beruht auf der Überlegung, dass die durch eine Serie von Coulomb-Explosionen erzeugten Tröpfchen so klein werden, dass sie schließlich nur noch ein Analyt-Molekül enthalten. Aus diesen kleinsten Tröpfchen entstehen letztendlich durch Desolvatisierung in Folge von Kollisionen mit den Stickstoffmolekülen im Interface zum Massenspektrometer lösungsmittelfreie, gasförmige Ionen entstehen.

Im Mittelpunkt des *Ion Evaporation Model (IEM)* von Iribarne und Thomson steht die direkte Ionenemission aus hochgeladenen Tröpfchen, die noch viele Analytmoleküle enthalten. Diese Tröpfchen messen ca. 8nm im Durchmesser und enthalten ca. 70 Analytmoleküle (Iribarne *et al.* 1976). Dabei erzeugen die sehr kleinen, hochgeladenen Tröpfchen auf ihrer Oberfläche ein elektrisches Feld, das stark genug ist, um Ionen in die Gasphase zu emittieren (Abbildung 2-9). Die kontinuierliche Emission wird durch die gleichzeitige Verdampfung von Lösungsmittelmolekülen gewährleistet. Hierbei ist für den

---

<sup>2</sup> Auch „charged residue model“ (CRM).

Ladungszustand des Analyten dessen räumliche Ausdehnung entscheidend (Fenn 1993), d.h. je größer der Analyt ist, desto höher ist der mittlere Ladungszustand. Für beide Modelle gibt es experimentelle Befunde. Einen guten Überblick hierzu gibt Kebarle (Kebarle *et al.* 1993).

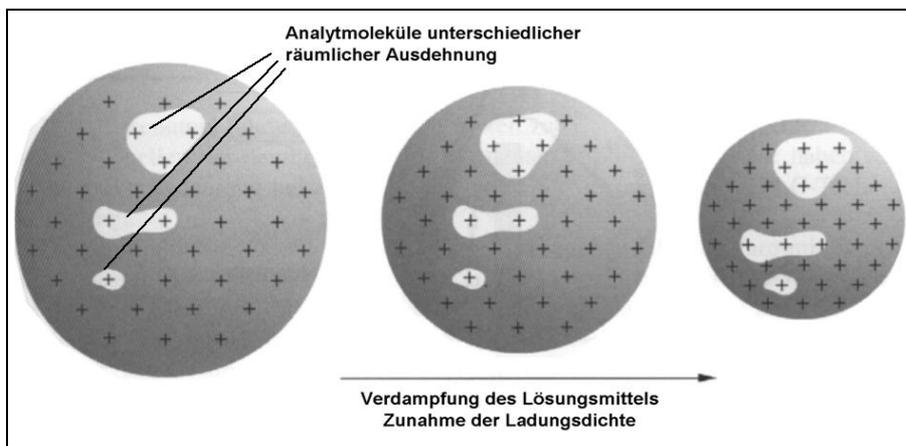


Abbildung 2-9 Modell zur Erklärung der ESI nach dem Ionenemissionsmodell, nach (Fenn 1993). Die Tröpfchen des ESI-Prozesses besitzen Überschussladungen, die auf der Oberfläche durch Coulomb-Abstoßung eine äquidistante Lage einnehmen. Die Analytmoleküle übernehmen vor der Ionenemission die Anzahl an Ladungen, die sie aufgrund ihrer räumlichen Ausdehnung aufnehmen können.

#### 2.1.2.1.1 Varianten der ESI-MS

Prinzipiell muß eine geeignete Elektrosprayquelle aus einer Kapillare bestehen, über die kontinuierlich die Analytlösung in das elektrische Feld injiziert werden kann. Für höhere Flußraten, wie sie zum Beispiel bei direkter Kopplung mit der analytischen HPLC auftreten, wurden ursprünglich pneumatisch unterstützte ESI-Quellen entwickelt (Abbildung 2-10, a).

Die Elektrosprayionisierung ist, wie Studien gezeigt haben, eine konzentrationsabhängige Methode. Bei größeren Flussraten, also größerer absoluter Menge an Analyt im Spray, ergibt sich bei gleicher Konzentration des Analyten keine Sensitivitätssteigerung. Diese Erkenntnis führte in den vergangenen Jahren zur Entwicklung von mikro-Elektrosprayquellen mit deutlich kleineren Flussraten (0,5 – 0,1  $\mu\text{L}/\text{min}$ ) (Abbildung 2-10b). Diese haben vor allem Bedeutung bei der Micro-HPLC/ESI-MS-Kopplung erreicht.

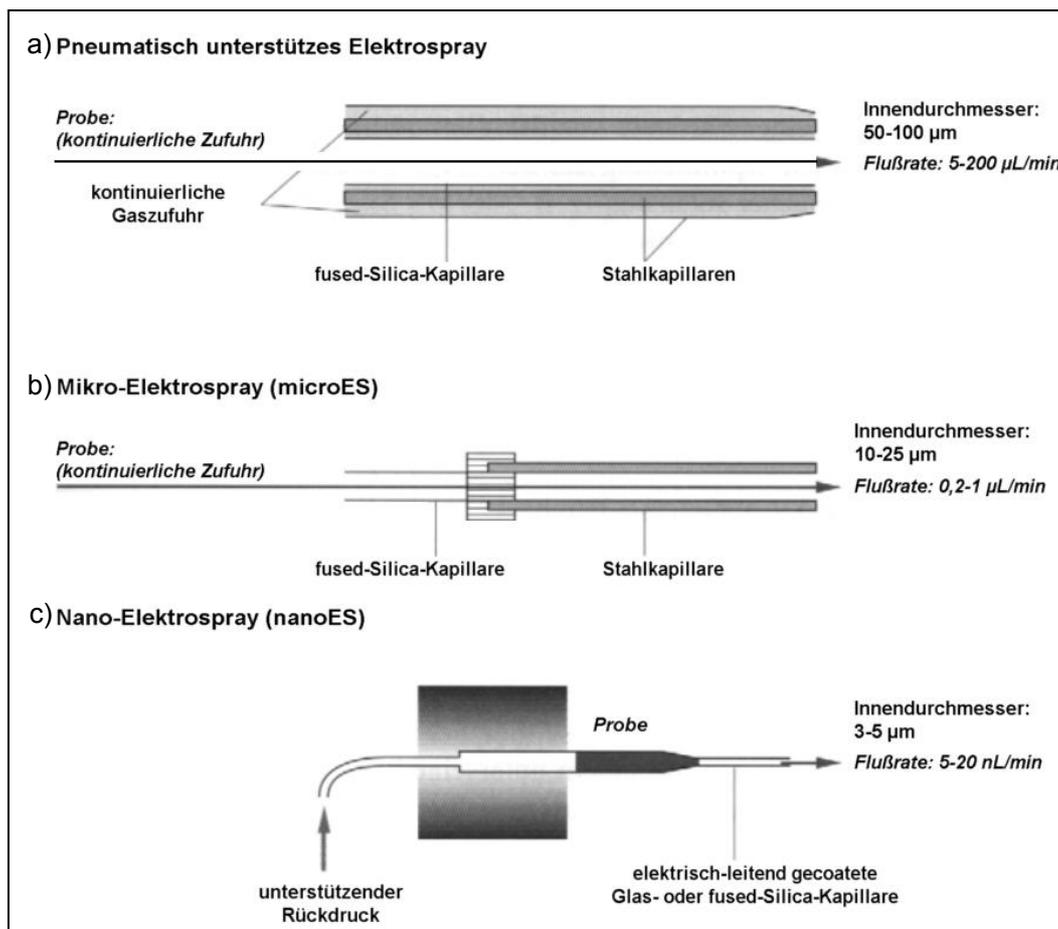


Abbildung 2-10 Schematischer Aufbau und Varianten von ESI-Quellen (nach Eckerskorn 1998). Nicht dargestellt ist die Variante mit Flüssen  $>1\text{ml}$ , die als Turbo-Ionspray bezeichnet wird und zusätzlich zur pneumatischen Unterstützung Hitze zur Desolvatisierung verwendet.

Die Entwicklung der Nano-Elektrosprayionisierung wurde begründet von Wilm und Mann (Wilm *et al.* 1996; Wilm *et al.* 1996). Die Ionenquelle besteht aus einer gezogenen Glas- oder Fused-Silica-Kapillare mit einem Durchmesser von wenigen Mikrometern an der Austrittsöffnung. Die Glasspitze ist mit einer Metallschicht (z.B. Gold oder vergleichbare Legierungen) überzogen, um das Potential an die Spitze zu führen. Unterstützt durch einen statischen Rückdruck (erzeugt etwa mit einer gasdichten Kolbenspritze) wird die Probe kontinuierlich durch das elektrische Feld mit 5-20 nL/min in den Vakuumbereich des Massenspektrometers überführt. Durch diese Anordnung entstehen sehr kleine Primärtropfen, die die Ursache für die Vorteile der nano-ESI-MS sind: Da von Anfang an nur sehr wenige Analytmoleküle je Tröpfchen vorhanden sind, treten kaum substanzspezifische Unterdrückungseffekte auf,

die etwa durch eine unterschiedliche Oberflächenaktivität hervorgerufen werden. So führen etwa die relativ unpolaren Zucker bei vergleichbaren Konzentrationen zu Signalintensitäten, die mit denen der relativ polaren Peptiden vergleichbar sind (Bahr *et al.* 1997). Weiterhin ist diese Methode im Vergleich zu konventionellem ESI-MS toleranter gegenüber Salzen und anderen Verunreinigungen in der Probe (Wilm *et al.* 1996; Juraschek *et al.* 1999).

Ein weiterer Vorteil der nanoES liegt in dem durch die extrem niedrigen Flußraten bedingten geringen Materialverbrauch begründet. So können häufig mit Probenvolumina von 1 µL (Analytmengen: >10 fmol bis wenige pmol pro µL) Meßzeiten von 30 - 60 min bei ausreichendem Signal zu Rausch Verhältnis erreicht werden. Neueste Untersuchungen von Schmidt (Schmidt *et al.* 2003) zeigen, dass erst ab einer Flussrate von < 50 nL/min die oben beschriebenen Vorteile voll zum Tragen kommen und „echtes“ Nanospray zu erwarten ist.

Nachteile dieses Aufbaus ist die Gefahr der Verstopfung durch Mikropartikel oder das Auskristallisieren des Analyten an der sehr feinen Kapillarspitze, sowie eine nicht genau justierbare und bestimmbare Flussrate.

### 2.1.2.2 Massenanalyse

#### 2.1.2.2.1 Der Quadrupol Analysator

Ein Massenspektrometer, das unter Verwendung des Quadrupols als Massenfilter arbeitet, wurde bereits 1953 von W. Paul und H. Steinwedel beschrieben (Paul *et al.* 1953). Unter vorgegebenen physikalischen Bedingungen läßt eine Quadrupolanordnung nur Ionen mit einem bestimmten  $m/z$ -Verhältnis zum Detektor durch. Er setzt sich aus einem System von vier (theoretisch hyperbolischen) Stäben zusammen, die auf einem Kreis mit dem Radius  $r$  um die  $z$ -Achse angeordnet sind (Abbildung 2-11). An den in der Praxis aus Kostengründen meist kreiszylindrischen Stäben liegt eine Gleichspannung  $U$  und eine Wechselspannung ( $V \cdot \cos 2\pi f t$ ) mit der Frequenz  $f$  an. Gegenüberliegende Stäbe verfügen dabei über die gleiche Polarität der Gleichspannung und die gleiche Phase der Wechselspannung. Dementspre-

chend besitzen nebeneinanderliegende Stäbe die entgegengesetzte Polarität und eine um  $180^\circ$  versetzte Phase. In der Nähe der  $z$ -Achse entsteht so ein elektrisches Potential  $\phi$ :

$$\phi(x, y, t) = (U + V \cdot \cos 2\pi ft) \cdot \frac{x^2 - y^2}{t^2} \quad \text{Gleichung 2-5}$$

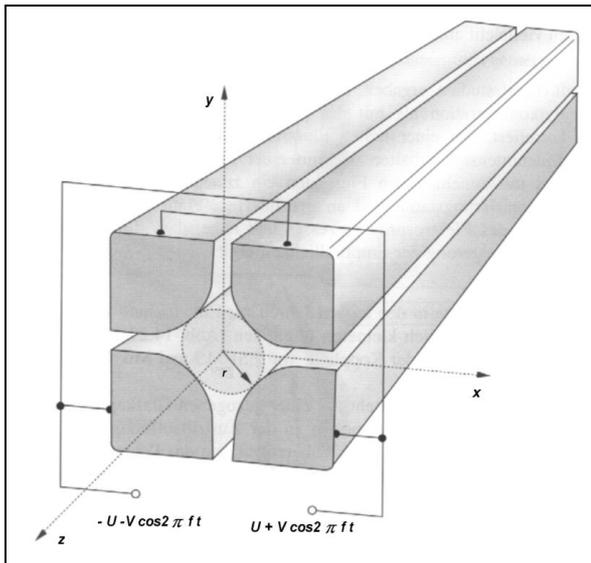


Abbildung 2-11 Geometrie eines Quadrupol-Massenfilters.

Durch eine geringe Beschleunigungsspannung von 10-20 V erhalten die Ionen eine ausreichende Translationsenergie, um in die Richtung der  $z$ -Achse in das elektrische Feld des Quadrupols zu gelangen. Ihre Bewegung in der  $xy$ -Ebene kann dabei durch einen Satz von Differentialgleichungen beschrieben werden, die im Prinzip den von Mathieu 1868 (Mathieu 1868) für die Untersuchung an schwingenden Membranen aufgestellten linearen Differentialgleichungen entsprechen.

Die *Mathieu'schen* Gleichungen haben zwei Arten von Lösungen: Eine Lösung führt zu endlichen Amplituden der Oszillationen entsprechend einer stabilen Bewegung entlang der  $z$ -Achse durch den Quadrupol, die andere führt zu in  $x$ - und/oder  $y$ -Achse exponentiell anwachsenden Amplituden und damit instabilen Trajektorien. Die numerische Auswertung der *Mathieu'schen* Gleichungen ergeben ein Stabilitätsdiagramm, in dem jedem  $m/z$ -Verhältnis ein gegebener Wertesatz  $r$ ,  $U$ ,  $V$  und  $f$  zugeordnet werden kann.

Beim Scannen des Massenbereichs werden Gleichspannung  $U$  und Amplitude  $V$  des Wechselfeldes gleichzeitig erhöht, wobei das Verhältnis  $U/V$  und die Frequenz  $f$  (im Radiofrequenzbereich) konstant gehalten werden. Dadurch werden Ionen mit verschiedenem  $m/z$ -Verhältnis nacheinander in den stabilen Bereich des Quadrupolfeldes gebracht. Da die Masse  $m$  in den Gleichungssystemen direkt proportional zu  $U$  und  $V$  ist, werden bei Veränderung von  $U$  und  $V$  lineare Massenspektren erhalten.

Kommerzielle Quadrupolmassenspektrometer haben einen maximalen Massenbereich bis etwa  $m/z = 4000$ . Die Auflösung kann variiert werden und wird in der Regel so eingestellt, dass in etwa eine Nominalmassenauflösung über den gesamten zugänglichen Massenbereich erreicht wird. Da Quadrupolmassenspektrometer über eine hohe Ionentransmission von der Quelle zum Detektor verfügen, leicht zu fokussieren sind und eine im Dauerbetrieb sehr stabile Kalibration haben, sind sie heute in der biochemischen Analytik sehr verbreitet.

Die aus dem Quadrupol austretenden Ionen werden über eine Konversionsdynode zum *Sekundärionenvervielfacher (SEV)* abgelenkt.

#### 2.1.2.2.2 Die Ionenfalle

Die Ionenfalle (Ion Trap, IT), die erstmals von Paul vorgestellt wurde (Paul *et al.* 1955; Paul *et al.* 1958), kann als dreidimensionaler, „kreisförmiger“ Quadrupol Analysator gesehen werden (Paul 1990). Zur kommerziellen Reife wurden sie aber erst in den 80er Jahren durch Stafford gebracht (Stafford *et al.* 1984). Die Ionenfalle fängt Ionen in einem dreidimensionalen elektrischen Feld und kann sie für Mikrosekunden bis Sekunden auf stabilen Bahnen halten und dann entsprechend ihrem Masse zu Ladungsverhältnis trennen.

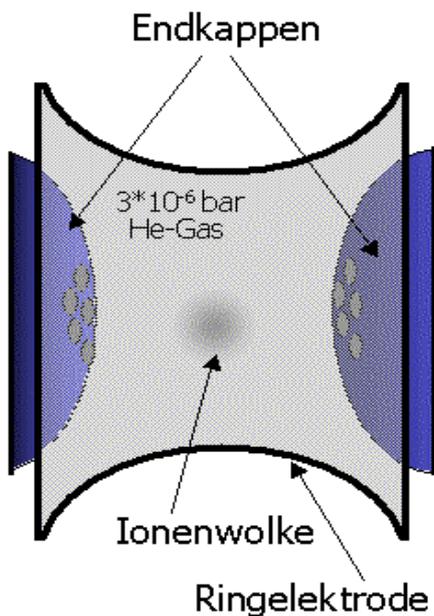


Abbildung 2-12 Schematische Abbildung einer Quadrupol-Ionenfalle (QIT) nach Paul.

Eine Ionenfalle besteht aus zwei Endkappen, an die Wechselspannungen angelegt werden, und einer Ringelektrode. Die Endkappen besitzen kleine Öffnungen zum Ionenein- und -auslass (Abbildung 2-12)



Abbildung 2-13 Kreisbahnen nach Lissajous, mit deren Hilfe man sich die Bahnen von Ionen in einer elektrischen Ionenfalle vorstellen kann.

Durch ein an der Ringelektrode angelegtes Hochfrequenzfeld werden die Ionen zunächst auf lissajousförmigen Kreisbahnen stabilisiert (Abbildung 2-13). Die Falle wird mit einem Druck von  $3 \cdot 10^{-6}$  bar betrieben. Dieser Druck wird im Allgemeinen durch Helium erzeugt, welches in die im Hochvakuum liegende Ionenfalle strömt. Das Gas hat einerseits die Aufgabe, die Ionen auf einen kleinen Bereich in der Fallenmitte zu konzentrieren (Dämpfungsgas), andererseits übernimmt es die Aufgabe des Kollisionsgases bei der massenspektrometrischen Fragmentierung (Kollisionsgas).

Wie beim linearen Quadrupol bestimmt auch hier die Lösung der Mathieu'schen Differentialgleichungen die Wertebereiche von angelegten Gleich- und Wechselspannungen, in denen die Ionen stabile Bahnen beschreiben. Das Stabilitätsdiagramm bildet sich aus den überlappenden stabilen Bereichen der Endkappen und der Ringelektrode (Abbildung 2-14).

Nach dem Einfangen der Ionen folgt die Massenanalyse. Dazu wird die Wechselspannungsamplitude  $q$  kontinuierlich erhöht, um die Ionen sukzessive aus dem Stabilitätsbereich zu drängen. Zusätzlich werden Multipolfelder durch Überlagerung von dipolaren Feldern der

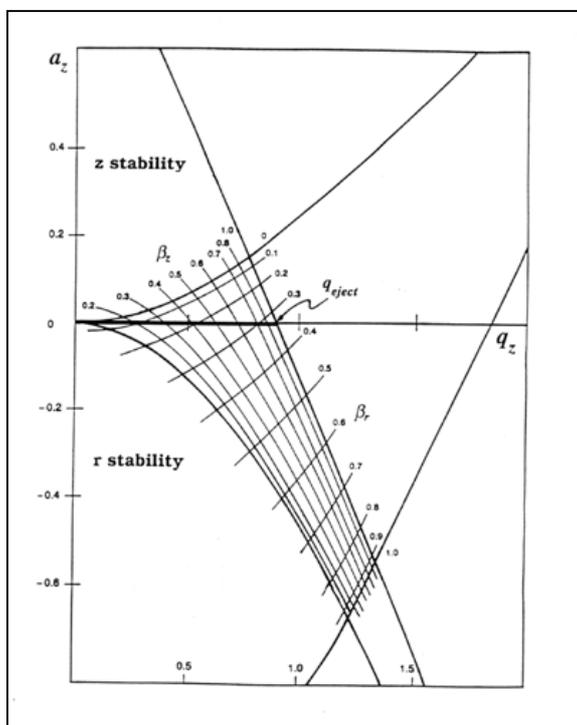


Abbildung 2-14 Stabilitätsdiagramm der Mathieu'schen Gleichungen für die elektrische Ionenfalle

Endkappen und dem quadrupolaren Feld der Ringelektrode erzeugt. Diese Felder erzeugen innerhalb des eigentlichen Stabilitätsbereiches scharfe Instabilitätsbereiche, in denen das Ion resonante Energie aufnimmt, so dass es aus der Ionenfalle katapultiert wird. So können Scangeschwindigkeiten von bis zu 60.000 amu/s erreicht werden. Der Detektor besteht aus einer

Konversionsdynode und einem Sekundärelektronenvervielfacher (SEV). Die Massenauflösung des in dieser Arbeit verwendeten Gerätes betrug ca. 1000 (FWHM) bei  $m/z=1000$ .

Die Vorteile der Ionenfalle für die Peptid/Proteinanalytik liegen im Bereich der massenspektrometrischen Sequenzierung. Durch die Robustheit und hohe Automatisierbarkeit wird sie heute als Standardinstrument - meistens als online HPLC-Kopplung - in der Proteomanalyse verwendet. In Peptidom-Projekten ist sie durch die geringe Massengenauigkeit (100- 300 ppm) und

die geringe Auflösung jedoch nur begrenzt einsetzbar (vgl. Kap. 4.3.5).

### 2.1.2.2.3 Quadrupol-TOF-Hybrid-Geräte (QToF)

In den letzten Jahren gewinnt eine weitere Technologie in der massenspektrometrischen Peptidanalytik immer mehr an Bedeutung. Die Q-TOF Technologie wurde 1996 (Morris *et al.* 1996) eingeführt und vereinigt die Vorteile der Quadrupol-MS mit denen des time-of-flight Analysators. Das Instrument (Abbildung 2-15) ist wie ein Triple-Quadrupol aufgebaut, bei dem der dritte Quadrupol durch eine orthogonal angeordnete TOF-Einheit ersetzt wurde. Während der Quadrupol dabei den Mathieuschen Gleichungen gehorcht, ist der TOF-Analysator mit den schon beschriebenen Gleichungen solcher Systeme determiniert.

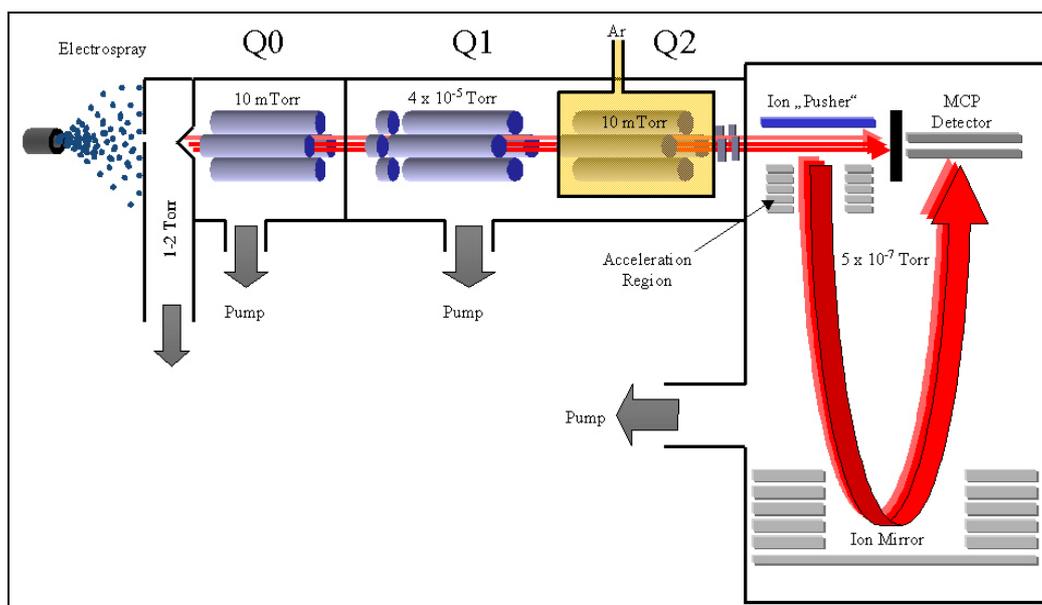


Abbildung 2-15 Schematischer Aufbau eines Quadrupol-Time-of-Flight Massenspektrometers mit Elektrospray-Ionisierung (ESI-QTOF-MS). Es besteht aus mehreren Quadrupoleinheiten zur Ionenfokussierung (Q0, Q1) und Fragmentierung (Q2) und einem orthogonal angeordneten reflektoren Flugzeitanalysator.

Die Vorteile dieser Hybrid Anordnung ist die hohe Auflösung im MS und auch im MS/MS Modus ( $\sim 10000$  FWHM bei  $m/z$  800), die relativ hohe Massengenauigkeit von  $< 20$ ppm bei externer Kalibration und eine gute Empfindlichkeit über einen großen Massenbereich.

Ein Nachteil dieser Technik ist der begrenzte dynamische Bereich. Dies ist

auf die Detektionsmethode zurückzuführen. Da bei der orthogonalen Anordnung des Flugrohrs, im Gegensatz zur linearen Anordnung von üblichen MALDI-MS Instrumenten, die Anzahl der Ionen, die den Detektor erreichen, um ein Vielfaches kleiner ist, müssen einzelne Ionen registriert werden. Dies ist aufgrund des inhärenten Rauschens analoger Systeme (Transienten Recorder, TR) nicht möglich. Aus diesem Grund werden für orthogonale Systeme sog. Time-to-Digital-Converter (TDC) benutzt, die einzelne Ioneneignisse detektieren können. Beim Auftreffen des Ions auf die *Multichannelplate* (MCP) entsteht eine Totzeit, während der keine weiteren Ionen registriert werden können. Diese Zeit ist einige Nanosekunden lang – ungefähr so lang wie die Peakbreite in einem hochauflösendem System, aber nicht so lang wie die Zeit zwischen zwei natürlichen Isotopen, die aufeinanderfolgen. In der Praxis heißt das, dass intensive Signale an der Peakspitze und an der rechten Flanke schrumpfen und somit Intensität verloren geht und die centroide Darstellung eines Peaks zu kleineren  $m/z$ -Werten verschoben wird. Diese Sättigungseffekte haben also eine Auswirkung auf die Amplitude und auf die Massengenauigkeit. Man kann diese Messfehler durch einen Korrekturfaktor ausgleichen, da die Ankunftszeiten von Ionen in einem TOF der Poisson Statistik unterliegen. Ferner ist es möglich, den dynamischen Bereich zu erhöhen, indem nicht nur eine Kollektoranode zur Registration der Ionen verwendet wird, sondern mehrere voneinander unabhängige Kanäle benutzt werden (Rockwood *et al.* 1997; Barbacci *et al.* 1998). Einen guten Überblick über die Quadrupol-TOF Technologie gibt Chernushevich (Chernushevich *et al.* 2001).

### **2.1.3 Massenspektrometrische Strukturanalyse: $MS^n$ -Methoden**

Neben der Bestimmung des Molekulargewichts kann die Massenspektrometrie durch spezifische und reproduzierbare Fragmentierung Informationen über die Struktur des Analyten liefern. Der Erfolg der Methoden zur massenspektrometrischen Sequenzierung von Peptiden hängt daher in erster Linie vom Verständnis der Fragmentierungsmechanismen der Peptid-Ionen ab. Die Analyse wird durch die Tatsache vereinfacht, dass unter den energeti-

schen Bedingungen der Kollisionsaktivierung in Quadrupolgeräten<sup>3</sup> bevorzugt Bindungsbrüche zwischen dem Carbonylkohlenstoff und dem Amidstickstoff der Peptidkette, also entlang des sog. *Peptide-Backbone*, auftreten (vgl. Abbildung 2-17).

In der MALDI-MS können diese Fragmente trotz der relativ schonenden Ionisierung bereits vor der Ionen-Extraktion in einem Prozeß, der als *in-source decay* bezeichnet wird, entstehen. Allerdings ist die Rate der Fragmentierung so gering, dass die resultierenden Signale kaum zur Informationsverwertung genutzt werden können. Da es zudem keine Möglichkeit der Filterung der Vorläuferionen gibt, kann das Verfahren nur für sehr reine Proben angewendet werden (Katta *et al.* 1991). Relativ weit verbreitet ist dagegen die massenspektrometrische Sequenzierung ausgehend von den Ionen, die in der MALDI-MS entlang der Flugstrecke im *post-source decay* entstehen (Spengler *et al.* 1992; Kaufmann *et al.* 1993). Da hier die Fragmentierungsrate deutlich höher liegt, und der Prozeß zusätzlich durch Kollisions-Aktivierung (*collisionally-induced dissociation*, CID) mit einem Stoßgas unterstützt werden kann, lassen sich von geeigneten Peptiden gut auswertbare Fragmentspektren erhalten. Hinzu kommt die Möglichkeit, auch Sequenzierungen aus Mischungen durchführen zu können, da eine Filterung von engen *m/z*-Bereichen möglich ist.

Am erfolgreichsten und daher auch am weitesten verbreitet ist die massenspektrometrische Sequenzierung mittels CID in der ESI-MS, in Triple-Quadrupol-, elektrischen Ionenfallen, magnetischen Ionenfallen (FT-ICR-MS) und Hybridgeräten, wie Quadrupol-TOF und Sektorfeld-Trap. Die Vorläuferionen können mit Nominalmassen-Breite gefiltert werden, womit die Möglichkeit zur Analyse sehr komplexer Gemische besteht.

In dieser Arbeit wurde sowohl mit einer elektrischen Ionenfalle als auch mit

---

<sup>3</sup> Hierzu zählt auch die elektrische Ionenfalle, die als dreidimensionaler Quadrupol verstanden werden kann.

einem QToF-Gerät gearbeitet.

### 2.1.3.1 Die Prinzipien der Tandem-Massenspektrometrie: $MS^2$ und $MS^n$

Die Einleitung der Fragmentierung durch CID in der ESI-MS erfolgt durch Überführung der kinetischen Energie der Ionen in Schwingungsenergie durch Kollision mit den Teilchen eines Inertgases. Der CID kann als Zwei-Stufen-Mechanismus aufgefaßt werden (Mc Luckey 1992; Shukla *et al.* 1993) in

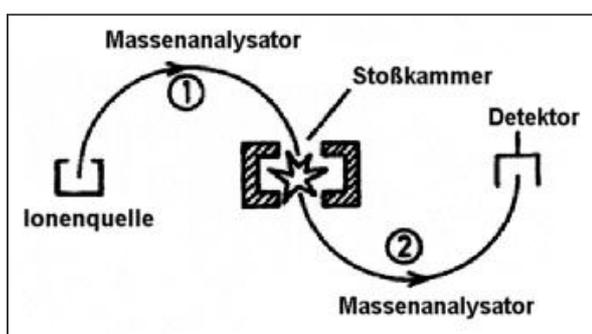


Abbildung 2-16 Vereinfachtes Prinzip der Tandem-MS (Lehmann 1996)

dem das zu fragmentierende Ion zunächst durch Stoß mit einem Molekül des Inertgases in einen schwingungsangeregten Zustand versetzt wird. Wird

beim Stoß genügend Energie übertragen, zerfällt das angeregte Ion in einer unimolekularen Reaktion in Fragmente. Je nach Ladungszustand des Vorläufers und der Lokalisation der Ladung entstehen dabei geladene und neutrale Fragmente. Diese Fragmente können ihrerseits in Folgestößen weiter zerfallen.

Das Prinzip der Tandem-MS beruht auf der Verwendung von zwei Massenanalysatoren, was den Begriff der Tandem-Massenspektrometrie ( $MS/MS$  oder  $MS^2$ ) geprägt hat und in Abbildung 2-16 veranschaulicht ist. Der erste Massenanalysator läßt nur das ausgewählte Vorläuferion passieren. In der Stoßkammer trifft das Vorläuferion auf das Inertgas und fragmentiert sequenzspezifisch. Die so erzeugten Fragmente können dann im zweiten Massenanalysator entsprechend ihrem  $m/z$ -Verhältnis getrennt werden.

Einen Sonderfall bildet hierbei die Ionenfalle, bei der die beiden Massenanalysatoren und die Stoßzelle nicht räumlich getrennt sind. Die drei beschrie-

benen Stufen zur Fragmentierung laufen allerdings zeitlich getrennt ab. Zunächst werden aus der gefüllten Falle sämtliche Ionen, außer den zu fragmentierenden, entfernt, dann wird das entsprechende Ion mit einer gewissen Amplitude aus seiner Bahn gelenkt und stößt dabei mit Gasmolekülen. Die aufgenommene Energie führt zur Fragmentierung und die entstehenden Fragmente werden ebenfalls im Quadrupolfeld gefangen. Der letzte Schritt ist das sukzessive Leeren der Falle nach  $m/z$ -Werten. Diese Vorgehensweise offenbart einen weiteren Vorteil der Ionenfalle, denn durch die „Zwischenspeicherung“ der Ionen kann nach dem ersten MS/MS-Schritt ein zweiter folgen usw.. Somit können  $MS^n$ -Experimente durchgeführt werden. Diese  $MS^n$ -Experimente können zur weiteren Strukturaufklärung des Analyten, wie auch zur *de novo*-Sequenzierung von Peptiden (Zhang *et al.* 2000) verwendet werden.

### 2.1.3.2 Fragmentierung von Peptiden

Peptide sind lineare Biomoleküle mit sich wiederholenden Struktureinheiten (Aminosäuren), die über eine Peptidbindung miteinander verbunden sind und sich nur in der Art der Seitenkette voneinander unterscheiden. Durch den bevorzugten Bruch der Bindungen unter CID Bedingungen am Rückgrat (backbone) des Peptids entstehen eine Reihe von Fragmentationen, die über ihre Massendifferenzen einen direkten Rückschluss auf die Aminosäuresequenz zulassen.

Die Bezeichnung der charakteristischen Fragmentationen erfolgt nach einer Nomenklatur von Roepstorff und Fohlman (Roepstorff *et al.* 1984), die später von Biemann modifiziert wurde (Biemann 1988). Die Fragmente werden als a, b, c und x, y, z bezeichnet werden. In Abbildung 2-17 ist die Nomenklatur schematisch dargestellt.

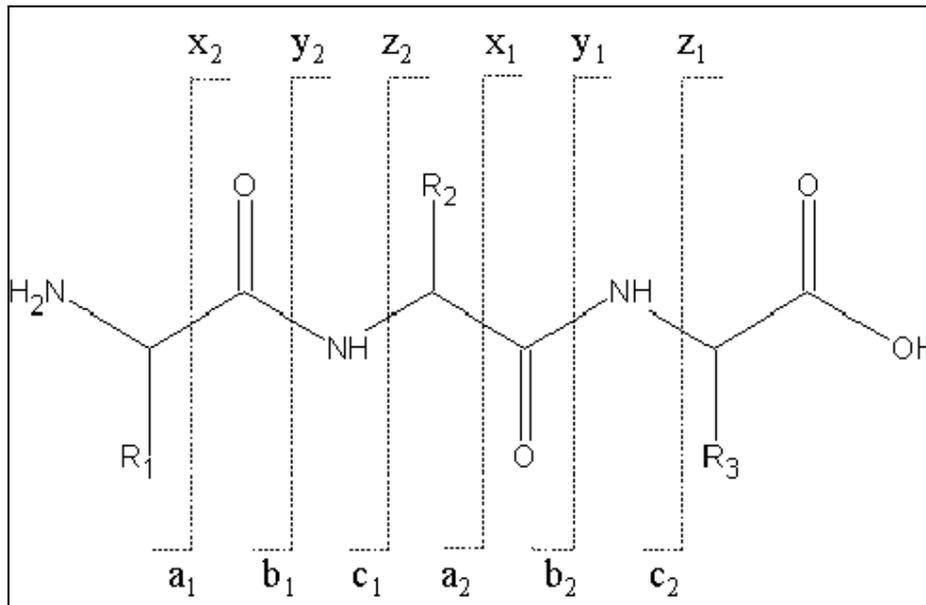


Abbildung 2-17 Nomenklatur für die Peptidfragmentierung nach Roepstorff und Fohlman (Roepstorff *et al.* 1984).

Dabei enthalten a-, b-, c-Fragmente immer den N-Terminus und die x-, y-, z-Fragmente grundsätzlich den C-terminus des Peptids. Fragmente, die aus dem Bruch der Peptidbindung resultieren, sind b- und y-Fragmente. Formale Abspaltung einer Carbonylgruppe aus einem b-Ion führt zu einem a-Fragment. Weiterhin werden häufig auch Neutralabspaltungen von Ammoniak ( $\Delta=17$  Da) und Wasser ( $\Delta=18$  Da) aus den Fragmentionen beobachtet. Die Bezeichnung für diese Ionen lautet z.B.  $[b_n-H_2O]^+$ . Der Index  $n$  gibt die Position des Fragmentions innerhalb des Peptids vom jeweiligen Terminus an. Der Ladungszustand des Fragments wird als Exponent hinzugefügt

Abbildung 2-18 zeigt ein Fragmentspektrum von Glu-Fibrinopeptid B. Aus Gründen der Übersichtlichkeit ist dieses Spektrum bereits dekonvolviert und „deisotopiert“, d.h. es besteht ausschließlich aus einfach geladenen monoisotopischen Fragmentensignalen. Deutlich zu erkennen ist eine y-Serie, die sich über das gesamte Peptid erstreckt ( $y_2$  bis  $y_{13}$ ). Fragmentionen anderen Typs sind auch im Spektrum vorhanden, wurden aber nicht annotiert. Die Bezeichnung der Abstände zwischen den y-Ionen mit Aminosäuren entspricht - von rechts nach links gelesen - der Aminosäuresequenz von Glu<sup>1</sup>-Fibrinopeptid B.

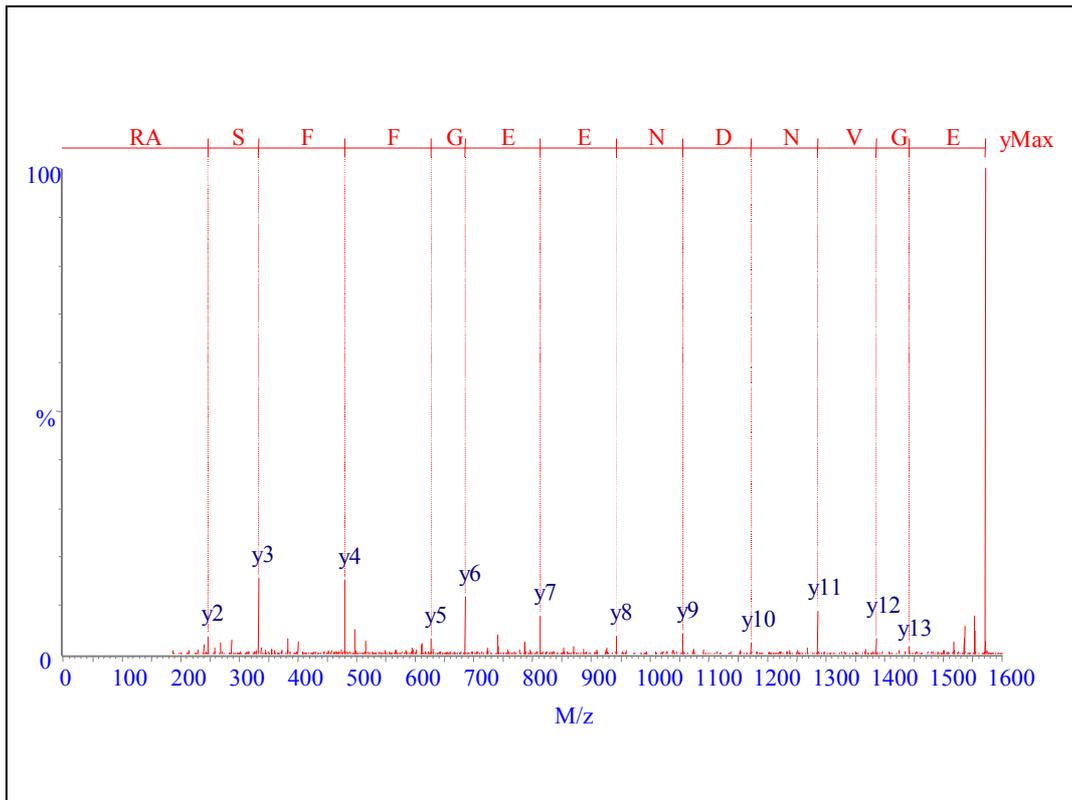


Abbildung 2-18 Dekonvolutiertes Fragmentspektrum von Glu<sup>1</sup>-Fibrinopeptid B. Annotiert ist sowohl die y-Ionen-Serie, als auch die Aminosäuresequenz des Peptids.

## 2.2 Der gängige Identifizierungsweg von Proteinen in der Proteomanalytik

In Abbildung 2-18 ist die typische Vorgehensweise in der gelbasierten Proteomanalytik gezeigt.

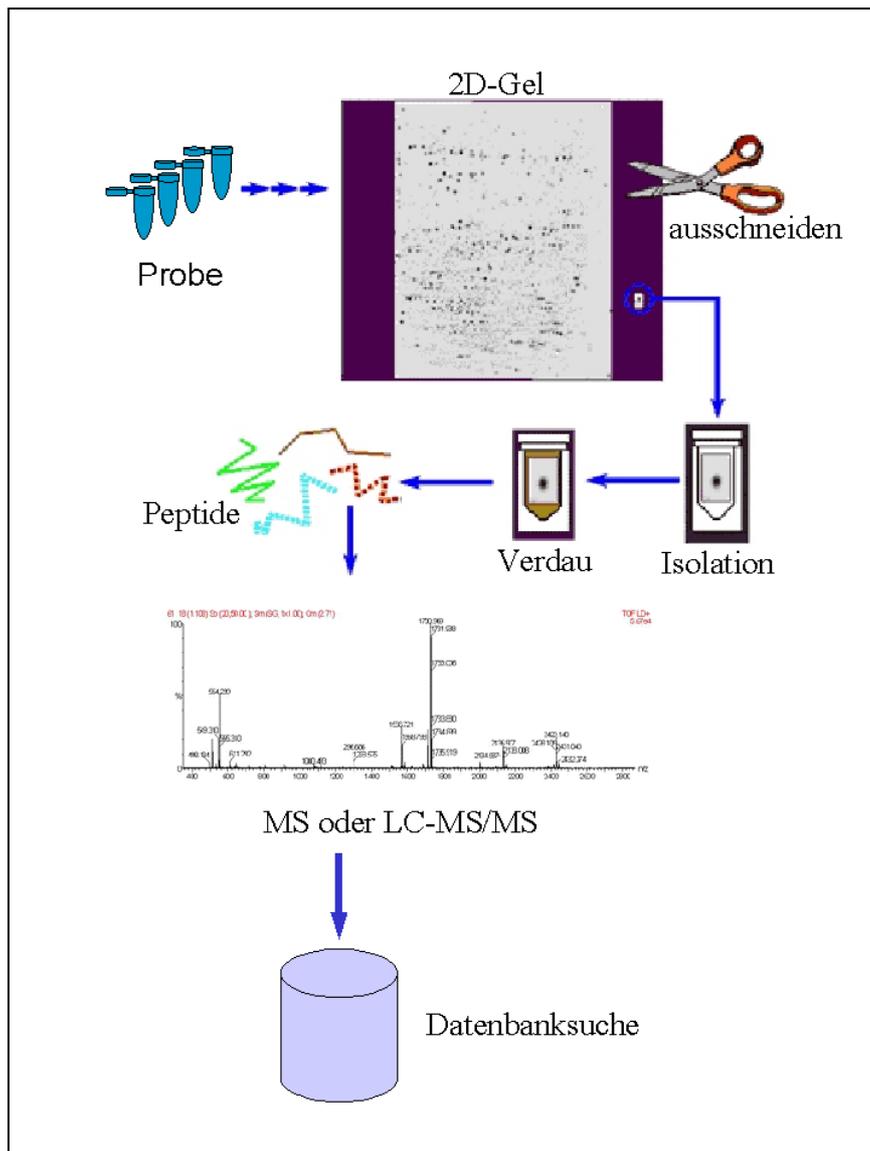


Abbildung 2-19 Gel basierte Proteomanalytik im Überblick.

Zunächst werden die Proteine der Probe durch 2D-Gelelektrophorese getrennt und so für die weitere Visualisierung und Analytik dekomplexifiziert.

Danach werden die so aufgetrennten Proteine durch Anfärbetechniken<sup>4</sup> visualisiert. Die einzelnen *Spots* werden ausgeschnitten und mit spezifischen Proteasen<sup>5</sup> verdaut. Die resultierenden tryptischen Fragmente werden dann mit MALDI-TOF-MS oder HPLC-ESI-MS(MS) vermessen und einer Datenbanksuche gegen Protein- oder DNA-Datenbanken unterzogen. Hier kann das Protein anhand seiner spezifischen Fragmentmassen identifiziert werden. Dabei ist das Ergebnis signifikanter, je mehr Fragmente einem Protein zugeordnet werden können.

### 2.3 Unterschiede der Peptidom- und Proteom-Analytik

Im Unterschied zur gängigen gelbasierten Proteinanalytik können Peptide und kleine Proteine über 2D-Gelelektrophorese nicht befriedigend charakterisiert werden, da sie zum Einen nur schwach angefärbt werden können und sich zum Anderen nicht gut im Maschengitter des Gels fixieren lassen, was zu beträchtlichen Verlusten führt. Weiterhin ist die Trennleistung des Gels für diesen Massenbereich ungenügend. Daher müssen für die Analyse dieser Verbindungen Technologien verwendet werden, die unter dem Peptidom Ansatz zusammengefasst sind. Auch hier ist eine zwei- bis mehrdimensionale Auftrennung erforderlich, da die zu untersuchenden Proben im Regelfall hoch komplex sind. Trennung von Peptiden und kleinen Proteinen geschieht vorzugsweise über flüssigchromatographische Methoden, wobei keine Reindarstellung einzelner Peptide erreicht wird, sondern weiterhin ein weniger komplexes Peptidgemisch vorliegt. Als letzte Dimension bietet sich der Einsatz der Massenspektrometrie an.

Die Probe, die im Massenspektrometer zu analysieren ist, besteht nicht, wie im oben gezeigten Proteom-Ansatz aus Peptidgemischen eines bzw. weniger

---

<sup>4</sup> Als Anfärbetechniken werden in der Proteomanalytik vorzugsweise Coomassie- oder Silberfärbungen verwendet, da diese mit der anschließenden MS-Analytik kompatibel sind.

<sup>5</sup> In der Proteomforschung wird vorzugsweise Trypsin als Protease verwendet, da diese spezifisch nach den basischen Aminosäuren Arginin (R) und Lysin (K) schneidet.

Proteine, sondern setzt sich aus Peptiden verschiedener Vorläuferproteine zusammen. Daher ist jedes Peptid eine zu identifizierende Entität und muss unabhängig von den anderen identifiziert werden. Eine Strategie, die auf redundanter Analyse mehrerer Peptide eines Vorläufers aufbaut, bleibt in der Peptidom-Analytik wirkungslos.

Zusätzlich zu bedenken ist, dass es sich bei den zu untersuchenden Peptiden nicht um spezifisch prozessierte Peptide handelt, die, wie im Beispiel des Trypsinverdau, einen genau beschriebenen Terminus (R oder K) haben. Dies hat sowohl Konsequenzen in der Qualität der erzeugten Fragmentenspektren<sup>6</sup>, als auch auf die nachgeschaltete Datenbanksuche. Hier werden durch das fehlende Kriterium der definierten Termini die Sensitivität und die Selektivität sowie die Suchgeschwindigkeit stark verringert. Das mittlere Molekulargewicht bei nativen Peptidgemischen liegt wesentlich höher als bei tryptischen Proteinverdau (4.000 Da statt ca. 1.900 Da)<sup>7</sup>. Das hat zur Konsequenz, dass die zu analysierenden Peptide aus nativen Quellen bei der Identifizierung durch ESI-MS/MS häufig höhere Ladungszustände annehmen als dies bei tryptischen Fragmenten zu erwarten ist. Diese Situation stellt große Anforderungen an die Datenauswertung und ist ein limitierender Faktor bei der Geräteauswahl.

Ausgehend von dieser Situation wurde in dieser Arbeit ein Ansatz entwickelt, der speziell die Problematik der Peptidomanalytik adressiert. Er ist zu unterteilen in den apparativen und den virtuellen Teil der Analyse. Das Letztere bezeichnet hierbei die softwarebasierte Auswertung der erzeugten Daten.

---

<sup>6</sup> In der Regel werden wesentlich fragmentärmere Spektren erzeugt als bei tryptischen Peptiden, da die positiven Ladungen nicht vorzugsweise an einem basischen Terminus lokalisiert sind.

<sup>7</sup> Die Durchschnittswerte für das Peptidom wurden aus den MALDI-Kartierungen bestimmt, den Werten für den Proteinverdau liegt ein *in silico* Verdau der SwissProt Datenbank (Version 39) zugrunde. Hierbei wurde jeweils nur der Massenbereich ab 1.000 Da berücksichtigt. Dies entspricht dem bei der Peptid-Kartierung mit MALDI-TOF-MS relevanten Bereich.

## 2.4 Peptide als funktionelle Einheiten im Körper

In einem komplexen Organismus von Mehr- oder Vielzellern ist es eine unabdingbare Voraussetzung, dass die lokal voneinander getrennten Systeme im Organismus miteinander kommunizieren können, um z.B. Leistung und Wachstum aufeinander abzustimmen. Zu diesem Zweck haben sich im Laufe der Evolution zwei verschiedene Systeme entwickelt, das Nervensystem und das hormonale System. Während das neuronale System der schnellen Übertragung von Nachrichten dient, beeinflussen Hormone, die häufig über Körperflüssigkeiten transportiert werden, Funktionen, die erst nach Sekunden, Minuten bzw. Stunden initialisiert werden (Crapo 1988).

Der menschliche Organismus besitzt neben niedermolekularen Signalträgern, wie Amininen, Aminosäuren und deren Derivaten hauptsächlich zwei Klassen von Hormonen: Peptidhormone und Steroidhormone. Bei beiden Gruppen handelt es sich um völlig verschiedene Substanzklassen, die sich sowohl in ihrer Darstellung, als auch ihrer chemischen Zusammensetzung und ihrer Wirkungsweise unterscheiden.

Steroidhormone leiten sich vom Cholesterin ab und haben eine polyzyklische Ringstruktur. Sie können die Zellmembran passieren und binden innerhalb der Zelle an Rezeptoren, die sie zum Zellkern transportieren. Dort wirken sie über eine Veränderung der DNA- Transkription.

Peptidhormone sind aus einzelnen Aminosäuren aufgebaut, die über eine Peptidbindung kettenartige Strukturen bilden. Als Beispiel zeigt Abbildung 2-19 Oxytocin, ein Peptid, das durch die Neurohypophyse abgegeben wird und u.a. für die Uteruskontraktion in den Wehen und die Laktation der stillenden Mutter verantwortlich ist.

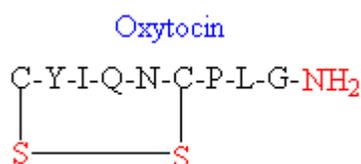


Abbildung 2-20 Aminosäuresequenz von Oxytocin

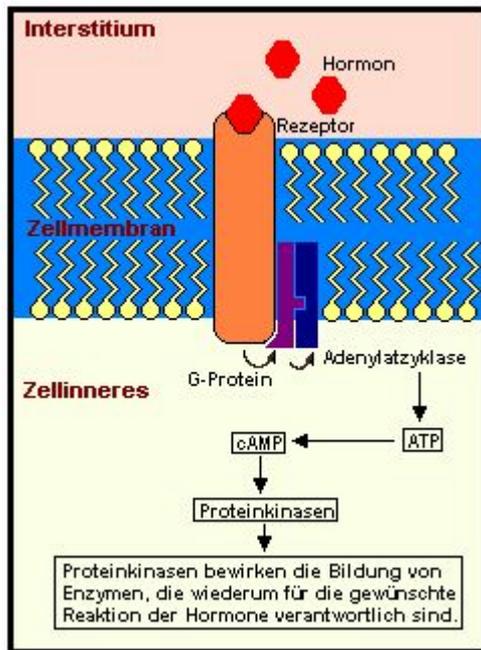


Abbildung 2-21 Schematische Darstellung der Wirkungsweise von Peptiden. Durch das Andocken an den Rezeptor wird im Zellinneren eine Reaktionskette ausgelöst: zunächst wird das G-Protein, das sich direkt neben dem Rezeptor in der Zellwand befindet, aktiviert. Das G-Protein aktiviert seinerseits die Adenylatzyklase, die sich im Inneren der Zelle befindet. Die Adenylatzyklase spaltet ATP (Adenosintriphosphat) in cAMP (zyklisches Adenosinmonophosphat) auf. cAMP aktiviert eine oder mehrere Proteinkinasen in der Zelle, die dann für die gewünschte Hormonreaktion der Zielzelle sorgen.

Peptidhormone durchdringen im Regelfall die Zellmembran nicht von extern, sondern binden an einen spezifischen,

membranständigen Rezeptor, der dann im Inneren der Zelle über Botenstoffe die biologische Information transportiert (Abbildung 2-21).

Die von Peptidhormonen übernommenen regulatorischen Funktionen im Körper sind sehr vielfältig. Sie wirken als Hormone oder Transmitter auf ihre Zielorgane oder -zellen, wobei Entstehungsort und Wirkort unterschiedlich sein können. Die Wirkmechanismen werden unterschieden in 1. *endokrin*, 2. *parakrin* und 3. *autokrin* (Koolman *et al.* 2003) (Abbildung 2-22).

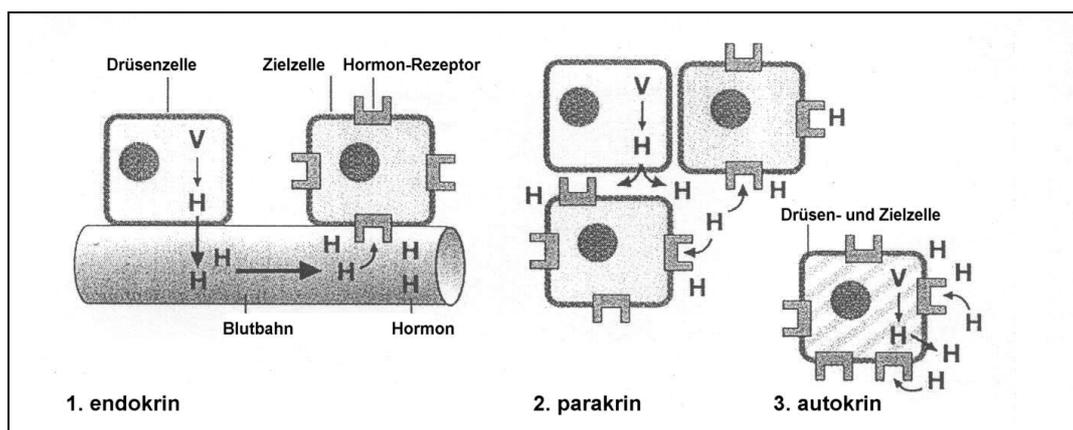


Abbildung 2-22 Spezialisierte Drüsenzellen synthetisieren das Hormon (H) aus Vorstufen (V) und schütten es bei Bedarf in den Extrazellularraum aus. Nach Diffusion oder Transport erfolgt die Wirkung auf eine Zielzelle. Gezeigt sind hier endokrine, parakrine und autokrine Hormonwirkung (Koolman *et al.* 2003).

Bei der *endokrinen* Funktionsweise werden die Hormone von einer Drüsen-

zelle ins Blut sezerniert und wirken auf andere, weiter in Organismus entfernte Zellen. Ist der Wirkort der Peptide in unmittelbarer Nähe zum Entstehungsort und daher ein Transport über die Blutbahn nicht erforderlich, spricht man von *parakrin*. Die Wirkung auf die hormonproduzierende Zelle selbst bezeichnet man als *autokrin*. Weiterhin ist die neurokrine Wirkung beschrieben, bei der Peptide in den synaptischen Spalt abgegeben werden und dort als Neurotransmitter oder Modulatoren auf andere Neuronen wirken.

Peptidhormone werden vom Organismus in einem mehrstufigen Prozess dargestellt. In einem ersten Schritt wird die genetische Information durch Transkription von der DNA abgelesen und die entsprechende RNA wird durch Spleißen von Intron-Sequenzen befreit. Die mRNA kodiert das Präpropeptid, welches wesentlich größer ist als das eigentliche Hormon. Linear abgebildet sind hier die Informationen für ein Signalpeptid, ein Propeptid und für ein oder mehrere Peptidhormone. Das Signalpeptid wird im endoplasmatischen Retikulum durch eine spezifische Signalpeptidase abgespalten und es bleibt eine Vorstufe des Hormons, das Prohormon übrig. Durch weitere gezielte Proteolyse und posttranslationale Modifikationen (z.B. Amidierung der C-Termini, Glykosylierungen, Pyroglutamatbildung, Phosphorylierung etc.), die häufig einen großen Einfluss auf die Aktivität haben, reift das Peptid dann zum aktiven Hormon heran (Krishna *et al.* 1993).

Um eine Homöostase im Organismus zu gewährleisten, ist die oben beschriebene, auf den wechselnden Bedarf ausgelegte Bildung und Abgabe, sowie nachfolgend auch die Inaktivierung der Hormone erforderlich. Diese erfolgt meistens durch proteolytische Spaltungen der Aminosäurekette. Dabei sind eine Vielzahl von Proteasen beteiligt, die sich an bestimmten Sequenzmotiven orientieren. Hierdurch entsteht eine sehr große Anzahl von Protein- und Peptidfragmenten mit unterschiedlicher biologischer Aktivität und Relevanz, die zum gleichen Zeitpunkt in einer Probe vorkommen können. Die Vorläufer können dabei eine deutlich andere Aktivität besitzen als die Bruchstücke. Ein Studium dieser Bruchstücke kann zum Verständnis von enzymatischer Degradation im Körper führen.

Somit ist klar ersichtlich, dass man nicht unmittelbar von der Gensequenz auf eine biologische Aktivität schließen kann. Um Funktionen im Körper und Krankheiten besser verstehen zu können, ist es deshalb unerlässlich, die Genprodukte, also Proteine und Peptide, systematisch und umfassend zu untersuchen.

Mit der Entwicklung leistungsfähiger, automatisierter analytischer Techniken wie zum Beispiel der High Performance Liquid Chromatography (HPLC), des Edman Abbaus (Edman 1956; Hewick *et al.* 1981) und der Entwicklung weicher Ionisationsverfahren in der Massenspektrometrie (Barber *et al.* 1981; Karas *et al.* 1988; Fenn *et al.* 1989) ging ein exponentieller Anstieg der Entdeckung und Charakterisierung unbekannter Peptide einher. Waren 1988 erst ca. 60 Sequenzen von biologisch aktiven Peptiden beim Menschen veröffentlicht (Sundler *et al.* 1988) waren im Jahre 2001 bereits über 500 Einträge in SwissProt (Bairoch *et al.* 1996; Bairoch *et al.* 2000; Boeckmann *et al.* 2003) (Release 39) mit bis zu 120 Aminosäuren, die dem humanen Organismus entstammen, verzeichnet. Trotz der raschen Weiterentwicklung geeigneter Methoden ist bis jetzt nur ein Bruchteil dieser Substanzklasse beschrieben.

## 3 Material und Methoden

### 3.1 Material

#### 3.1.1 *Peptidbank aus humanem Blutfiltrat*

Die Firma BioVisioN AG hat sich zum Ziel gesetzt, erstens neue krankheitsrelevante Peptide zu entdecken und zu charakterisieren und zweitens bereits bekannte Peptide in einen krankheitsassoziierten Kontext zu stellen. Dabei spielt sowohl die Identifizierung eines einzelnen biologisch aktiven Peptids eine Rolle, als auch das Zusammenspiel mehrerer krankheitsassoziierter Peptide. In beiden Fällen bietet sich die Möglichkeit neue, auf Peptiden basierende Substanzen für die Therapie oder Diagnose humaner Erkrankungen zu entdecken.

Zu diesem Zweck wird, neben der Isolierung von Peptiden aus Zellen und Organen, in großem Maßstab die Extraktion und Sequenzbestimmung von Peptiden aus Blutflüssigkeit durchgeführt. Das Blut bietet sich als zentrale Körperflüssigkeit insbesondere deshalb an, weil darin neben einer Vielzahl von großen Proteinen vollständig prozessierte kleine Moleküle mit systemischer und möglicherweise lokaler Wirkung zirkulieren. Wissenschaftlich und wirtschaftlich attraktiv ist der mögliche pharmakologische Wert, da diese Substanzen sowohl im Blut als Biomarker und Diagnostika nachgewiesen, als auch als Therapeutika in das Blut appliziert werden können.

Da zu Beginn der Arbeiten Blut für die Isolierung und Charakterisierung von Peptiden nicht in ausreichenden Mengen zur Verfügung gestellt werden konnte, wurde eine alternative Quelle in Anspruch genommen, um zuerst über die Herstellung einer Peptidbank eine stofflich ausreichende Basis für eine umfangreiche Peptid-Charakterisierung zu sichern. Die Präparation dieser Peptidbank erfolgte aus Hämofiltrat, welches bei der Blutentgiftung von chronisch niereninsuffizienten Patienten entsteht (Quellhorst *et al.* 1977; Henderson *et al.* 1986) und als ausreichend gutes Äquivalent zum „norma-

len“ Blut in Bezug auf den Peptidinhalt gesehen werden kann. Verschiedene Untersuchungen haben gezeigt, dass im Hämofiltrat viele Peptide in ähnlichen Konzentrationen wie in Blutplasma enthalten sind (Schepky *et al.* 1994). Die für diese Arbeit eingesetzte Probenbank ist im Rahmen einer Diplomarbeit am Niedersächsischen Institut für Peptid-Forschung (IPF) in Hannover im Jahr 1998 erstellt worden und dient seither der Firma BioVisioN AG als eine der Hauptquellen zur Identifizierung und Kartierung von Blutpeptiden (Neitz 1998).

Gewonnen wird das Hämofiltrat im Nephrologischen Zentrum Niedersachsen (NZN) in Hannoversch-Münden, wo es in großen Mengen von etwa 2000 L/Woche anfällt. Dort wird es gesammelt, sofort auf 4 °C heruntergekühlt und mit verdünnter Salzsäure auf einen pH-Wert von 2.6 eingestellt, um enzymatische Abbauprozesse durch Proteasen und eine Kontamination durch Mikroorganismen zu verhindern.

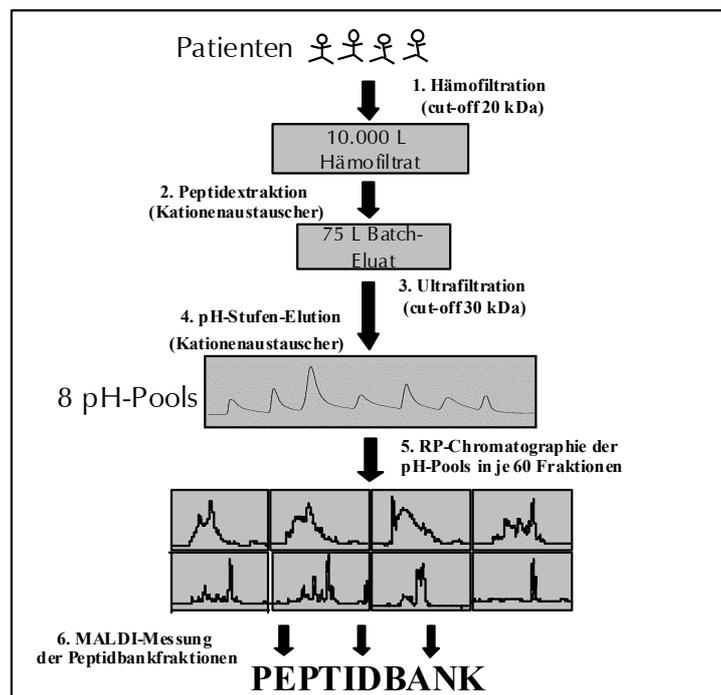


Abbildung 3-1 Schematischer zur Vorgehensweise bei der Herstellung der Peptidbank 970303 aus 10.000 L humanem Blutfiltrat.

Das in der Abbildung dargestellte Schema beschreibt eine Peptidbank-Präparation aus 10.000 L Hämofiltrat. Nach Ultrafiltration zur Entfernung hö-

hermolekularer Proteine, die zum Teil die am Patient verwendeten Hämofilter passieren können, werden verschiedene chromatographische Trennschritte durchgeführt, die über 500 gefriergetrocknete Fraktionen mit Peptiden ergeben, die bei einer Temperatur von  $-28\text{ °C}$  gelagert werden. Aus den einzelnen Fraktionen werden nach Bedarf die geeigneten Mengen entnommen und zur weiteren Verarbeitung im entsprechendem Lösungsmittel aufgenommen. Im weiteren Verlauf der Arbeit wird diese Peptidbank als „970303“ bezeichnet. Diese Probenquelle wurde im Lauf der Arbeit umfangreich massenspektrometrisch charakterisiert und liefert die Basis für weitere, hier vorgestellte Forschungstätigkeiten.

### **3.1.2 Lösungsmittel und Chemikalien**

Alle Chemikalien besitzen die Reinheitsqualität *pro analysi*. Abweichungen davon werden in den jeweiligen Listen angegeben. Alle Prozentangaben beziehen sich, wenn nicht anders angegeben, auf Volumenprozent.

- Acetonitril, *gradient grade* (Merck, Darmstadt)
- Trifluoressigsäure, *sequencing grade* (Applied Biosystems, Weiterstadt)
- *alpha*-Cyano-4-hydroxymizsäure ( $\alpha$ CHC), Reinheitsgrad 97% (Sigma-Aldrich Chemie, Deisenhofen)
- L-Fucose, Reinheitsgrad 99% (Sigma-Aldrich Chemie)
- Ameisensäure, 98-100% (Merck)
- Methanol, *gradient grade* (Merck)

### **3.1.3 Peptidstandards**

Im folgenden werden kurz die Kalibrationssubstanzen aufgeführt und Me-

thoden vorgestellt, wie sie für die verwendeten Massenspektrometer benutzt wurden.

### 3.1.3.1 MALDI-MS

Peptidstandards werden zur externen Kalibrierung des MALDI-TOF-MS eingesetzt. Dafür werden die Peptide in den in Tabelle 3-1 angegebenen Konzentrationen mit 0,1% TFA in 50% Acetonitril gelöst.

Tabelle 3-1 Standardpeptide für die Kalibrierung der massenspektrometrischen Messungen. Alle Peptide werden von Sigma-Aldrich bezogen.

Peptid	Konzentration [pmol/ $\mu$ L]	Mittleres MW als [M+H <sup>+</sup> ] <sup>+</sup>	Monoisotopisches MW als [M+H <sup>+</sup> ] <sup>+</sup>
Insulin (bovin)	10	5734,59	5730,61
ACTH 7-38 (human)	2	3660,16	3657,93
Neurotensin (human)	1	1673,97	1672,92
Angiotensin I (human)	1	1297,44	1296,69
Bradykinin (human)	1	1061,22	1060,58

### 3.1.3.2 ESI MS

Zur Kalibration des LCQ Classic wurde die von Thermo Finnigan mitgelieferte Kalibrationslösung verwendet. Diese besteht aus Koffein, MFRA (Peptid) und Ultramark (Polyfluorierter Kohlenwasserstoff). Kalibriert wurde vierteljährlich, wie es vom Hersteller empfohlen wird.

Das QTOF-Ultima wurde täglich mit Glu<sup>1</sup>-Fibrinopeptid (Sigma-Aldrich, Darmstadt) im MS/MS-Modus kalibriert. Hierzu wurde das Peptid in 50% Methanol, 0,1% Ameisensäure gelöst (100 fmol/ $\mu$ L) und per Spritzenpumpe oder Nanospray-Nadel der Messung zugeführt.

### 3.1.3.3 Massenspektrometer

In dieser Arbeit wurden folgende Massenspektrometer verwendet:

1. MALDI-MS : Voyager DE STR (Applied BioSystems, Framingham, USA)
2. ESI-MS/(MS): LCQ Classic (ThermoFinnigan, San Jose, USA)  
QTOF-ULTIMA (Micromass, Manchester, UK)  
QSTAR Pulsar (Applied Biosystems/Sciex, Toronto, Canada)

## 3.2 Methoden

### 3.2.1 Reversed Phase Chromatographie

Die *Reversed Phase* Chromatographie wird zur Auftrennung von komplexen Peptidgemischen eingesetzt. Ziel für einen großen Teil der Arbeit ist, die Peptidbankfraktionen trotz der bereits erfolgten Trennung auf 500 Fraktionen weiter in Subfraktionen aufzutrennen, damit eine möglichst empfindliche Messung mit MS/MS durchgeführt werden kann.

Alle chromatographischen Trennungen werden auf einer Chemstation HP1100 (Agilent Technologies, Waldbronn) durchgeführt, die wahlweise mit einem als Mehrkanal-UV-Detektor verwendeten Dioden-Array-Detektor (DAD) oder einem variablen Wellenlängen-Detektor (VWD) ausgestattet ist. Die UV-Absorption wird beim DAD bei den Wellenlängen 214 und 280 nm, beim VWD bei 214 nm aufgenommen. Die Anlage ist mit einem Säulenofen ausgestattet, der die Säulen während der Trennungen auf einer Temperatur von 33 °C hält.

Aus jeder verwendeten Fraktion der Peptidbank 970303 werden 5 mg Peptidgemisch für die Rechromatographie abgewogen und in 1 ml Laufmittel A gelöst und manuell über ein Rheodyne-Ventil (Abimed, Langenfeld) injiziert. Als wässriges Laufmittel A dient, wenn nicht anders erwähnt, bei allen Trennungen 0.06 % TFA. Als organische Phase B für die Gradientenelution wird

0.05 % TFA in 80 % Acetonitril eingesetzt. Die für die Trennung hauptsächlich eingesetzte Trennsäule war eine Phenomenex/Nucleosil (RP C18; 20.0 x 250 mm; 5 µm, 100 Å).

Die chromatographischen Trennungen werden mit einem Gradienten durchgeführt, bei welchem der Anstieg des Laufmittels B 0.5% pro Minute beträgt. Nach einer kurzen Äquilibrationsphase von 5 - 10 Minuten, bei der die Säule mit 100 % Laufmittel A gespült wird, wird der Gradient gestartet.

Das Sammeln der Proben erfolgt mit einem Fraktionssammler der Firma Gilson (Abimed Analysen-Technik, Langenfeld, Germany) in Mikrotiterplatten im 96-deep well-Format.

### **3.2.2 Resuspendierung der Proben**

Nach Fraktionierung wurden die Proben zuerst in Portionen zu 100µl (5x) und 500 µl (2x) aliquotiert, dann bei -80 °C tiefgefroren und anschließend lyophilisiert und bei -28° C gelagert. Die 100 µl Aliquots wurden für die massenspektrometrischen Messungen eingesetzt. Unmittelbar vor der massenspektrometrischen Messung wurden die Proben mit 50% Methanol, 1% Ameisensäure resuspendiert. Das Volumen des Lösungsmittels wurde anhand der UV-Absorption der vorangegangenen Rechromatographie so abgeschätzt, dass ungefähr gleichartige Konzentrationen an gesamt Peptidgehalt pro Fraktion erhalten werden. Fraktionen, die eine Absorption unter 500 mAu zeigten, wurden in 50 µl aufgenommen, von 500 bis 1000 mAu 100 µl, darüber wurden 200 µl verwendet. Zur Resuspendierung wurde die jeweilige Lösungsmittelmenge in die Mikrotiter-Platten pipettiert und zur Unterstützung der Solubilisierung für 5 Minuten ins Ultraschallbad gestellt.

### **3.2.3 Massenspektrometrische Methoden**

#### **3.2.3.1 MALDI-MS**

Die Messungen werden auf einem *Voyager-DE STR* (Applied Biosystems) im

Linear- sowie Reflektron-Modus durchgeführt. Die Länge der Flugröhre beträgt im linearen Modus 2 m und im Reflektron-Modus 3 m. Das Gerät ist mit einem 337 nm Stickstofflaser ausgestattet, der auf eine Pulslänge von 3 ns mit einer Wiederholungsrate von 3 Hz eingestellt ist.

Positive Ionen werden je nach ausgewählter Methode mit 20 bzw. 25 kV beschleunigt und mit einer Kombination aus Microchannelplate und Sekundärelektronenvervielfacher detektiert. Das verwendete Massenspektrometer ist mit der *delayed extraction*-Technologie (DE) ausgestattet, mit der eine hohe Massenauflösung erreicht wird. Die Energie des Lasers ist über eine mechanische Blende variierbar und wird so eingestellt, dass sie eine nur leicht über der zur Ionisierung der Probenmoleküle benötigten Intensität besitzt. Das Vakuum in der Probenkammer beträgt etwa  $2,7 \times 10^{-5}$  Pa und in der Flugröhre  $6,7 \times 10^{-7}$  Pa.

Die Proben werden auf eine mit 100 Positionen ausgestattete, spiegelpolierte Edelstahlplatte aufgetragen, auf der nur die inneren 64 Positionen bestückt werden. Der Proben- und Matrixauftrag wird mit einem Pipettierroboter *Symbiot I* (Applied Biosystems) im modifizierten *dried droplet*-Verfahren durchgeführt (Beavis *et al.* 1996): nach Aufbringen von 0,5  $\mu$ L Matrixlösung (je 5 oder 10 g/L  $\alpha$ CHC plus L-Fucose in 50% Acetonitril/0,1% TFA) wird das gleiche Volumen der Probe auf die noch feuchte Matrix pipettiert und durch zweimaliges Aufziehen in die Pipettenspitze vermischt. Das Trocknen der Proben erfolgt in einem Luftstrom bei Raumtemperatur. Die Kalibrierungsstandards werden jeweils in die Mitte zwischen vier Probenpositionen auf die gleiche Weise pipettiert. Wird Sinapinsäure als Matrixsubstanz verwendet, wird der Auftrag nach dem gleichen Protokoll durchgeführt.

Bei den Messungen fährt der Laserstrahl die einzelnen „Probenspots“ ab und nimmt jeweils pro Spot eine bestimmte Zahl von Messungen nach einem bestimmten Abtastmuster auf. Pro Messung wird eine vorgegebene Anzahl von Scans (zwischen 30 und 50) aufsummiert und anschliessend für die Datenspeicherung akkumuliert. Die Parameter der verwendeten Messmethoden wie etwa die verwendete Laserintensität sind an komplexe Peptidgemische

angepasst (Jürgens 1999).

Die primäre Signalverarbeitung und Digitalisierung erfolgt mit einem Digital-speicher-Oszilloskop TDS 540A (Tektronix, Beaverton, Oregon, USA), welches wie die gesamte Datenaufnahme über die Software *Voyager RP BioSpectrometry Workstation* (Applied Biosystems) gesteuert wird.

### 3.2.3.2 ESI-Quadrupol-Ion-Trap-MS (ESI-QIT-MS)

Um eine schnelle und effektive Erzeugung von Fragmentspektren aus den im 96-well-Format vorliegenden Proben zu gewährleisten wurde eine Anordnung gewählt, die auf einer Fließinjektionsanalyse (FIA) beruht. Dazu wurde das zur Verfügung stehende LCQ Classic mit einer HPLC-Pumpe der Firma ABI (Modell 140c) und dem Endurance-Autosampler (Spark, Niederlande) gekoppelt. Die benutzte Probenschleife hatte ein Volumen von 10µl. Der Fluss der Transportflüssigkeit (49% Methanol-, 50 Wasser-, 1% Ameisensäure- Gemisch) wurde auf 2µl/min gesetzt. Damit ergibt sich eine Messzeit von ca. 5 Minuten pro Fraktion, in der verschiedene Peptide in dieser Probe bearbeitet werden können. Die Proben wurden ebenfalls in der Transportflüssigkeit gelöst. Abbildung 3-2 zeigt den schematischen Aufbau der Anlage.



jektionsmethode (FIA) verwendet. Allerdings ist hier der Fluss mit Hilfe eines Splitters auf 250 nl/min reduziert worden. Die Sprayspannung war auf 1,5 – 2 kV eingestellt. Weitere Details befinden sich in den jeweiligen Kapiteln.

#### 3.2.3.4 *Darstellung der MALDI-MS Spektren mit der Software Spectromania*

Die Darstellung von kompletten Spektrensätzen, etwa aus einem Set gesammelter Fraktionen einer Chromatographie, wird mit der von BioVision entwickelten Software *Spectromania* durchgeführt. Dieses Programm erlaubt die schnelle Visualisierung und den Vergleich von vielen Spektren auf einem Blick, wobei Molekülmassen, Signalintensitäten und entsprechende Fraktionsnummern berücksichtigt werden. Die einzelnen Spektren werden in diesem Programm derart verarbeitet, dass jedes Spektrum um 90° gedreht wird und man so eine eindimensionale Ansicht eines Einzelspektrums „von oben“ bekommt.

Durch die Anordnung der Fraktionen in der zweiten Dimension ergibt sich von einer biologischen Probe ein individueller, zweidimensionaler Fingerprint in Form einer Peptidkarte, die zum einen das spezifische Elutionsverhalten in der Chromatographie und zum anderen die detektierten Massen in den einzelnen Fraktionen berücksichtigt. Eine weitere Möglichkeit bietet dieses Programm bei der vergleichenden Analyse von mehreren Spektren. Verschiedene Datensätze lassen sich so miteinander kombinieren und vergleichen, indem sie übereinandergelegt werden und Unterschiede farblich sichtbar gemacht werden können.

#### 3.2.3.5 *Aminosäure-Sequenzanalyse nach Edman*

Die Bestimmung der Aminosäuresequenzen der aufgereinigten und carboxyamidomethylierten Peptide nach Edman (Edman 1956) wird auf einem 494- (*Procise*)-Sequenzierer (Applied Biosystems, Weiterstadt) im *fast cycle*-Protokoll nach Angaben des Herstellers durchgeführt.

Zwischen 10 und 100 pmol der Probe werden auf eine mit *PolyBrene Plus* (Applied Biosystems) vorbehandelte Glasfasermembran aufgetragen. Die Analyse der abgespaltenen Phenylthiohydantoin-Aminosäuren erfolgt bei einer Wellenlänge von 269 nm über RP-Chromatographie, die an den Sequenzer standardmäßig gekoppelt ist. Die Datenaufnahme und Auswertung wird mit der Software *Model 610A Data Analysis* (Applied Biosystems) durchgeführt.

### 3.2.3.6 Modifizierungen am Cystein

Da in der Arbeit ausschließlich native Peptide bearbeitet werden sollten, kam eine Carboxyamidomethylierung der kompletten Peptidbank 970303 nicht in Frage, da hier Hetero- und Homo- Di und Polymere verloren gingen (z.B. Insulin). Bei der Entwicklung der Methode zur massenspektrometrischen Sequenzierung großer Peptide zeigte sich allerdings, dass in diesen Fällen die Vorteile die Nachteile überwiegen. Deshalb wurde bei der Entwicklung dieser Methode eine Reduktion und Carboxyamidomethylierung (CAM) der Proben durchgeführt.

Die Probe wird in 10 mM Ammoniumhydrogencarbonat, pH 8.2 oder alternativ in 200 mM 4-Methylmorpholin des gleichen pH-Wertes in einer Konzentration zu 0,1 – 1,0 µg/mL gelöst. Nach Begasung der Probe mit Stickstoff wird die frisch angesetzte, wässrige 400 mM DTT-Lösung zugegeben. Die Inkubation erfolgt bei einer Temperatur von 45 °C für eine Stunde auf dem Thermomixer. Die Zugabe des Reduktionsmittels erfolgt in ca. 50 fachem, molarem Überschuss zur Probe.

Zur Carboxyamidomethylierung werden die Proben nach Reduktion der Disulfidbrücken und anschließender Abkühlung auf Raumtemperatur im etwa 100fachen molarem Überschuss mit 500 mM wässrigem Iodacetamid versetzt. Die Inkubation erfolgt für 30 Minuten bei Raumtemperatur im Dunkeln.

### 3.2.3.7 Entsalzung der Proben

In Filtertips werden jeweils 100 µL methanolische Poros-Suspension pipetiert. Nach Abzentrifugation des Lösungsmittels wird das Material konditioniert, indem es 2 - 3 mal mit 100 µL 0,1% TFA gespült wird. Nach Probenauftrag wird 2-3 mal mit 100 µL 0,1% TFA gewaschen. Die Elution der Probe erfolgt mit 100 µL Elutionspuffer (70% Acetonitril/0,1% TFA).

Alternativ wurden bei größeren Probenmengen Festphasen-Kartuschen verwendet. Die Kartuschen werden mit 1-2 Säulenvolumina (CV) Methanol konditioniert, danach mit 2 CV 0,1% TFA äquilibriert und nach dem Probenauftrag mit 3 CV 0,1% TFA gewaschen. Die Elution der Probe erfolgt mit 70% Acetonitril/0,1% TFA.

### 3.2.3.8 Datenbankrecherchen

Je nach Art der durchgeführten Analyse werden für die Identifizierung der cysteinhaltigen Peptide verschiedene Suchmaschinen und Protein- bzw. Nukleotid- und EST-Datenbanken genutzt.

Nach erhaltener Sequenz aus der Aminosäure-Sequenzierung nach Edman wird die Aminosäureabfolge in das Software-Tool *MS-Edman* aus dem Datenbank-Suchprogramm *Protein-Prospector* (Version 3.2.1) eingegeben und gesucht (Clauser *et al.* 1996). Verschiedene Datenbanken lassen sich dort kaskadiert abfragen. Die Erstsuche wird standardgemäß in der Datenbank *Swiss-Prot* (ab Release 37) durchgeführt, in der ausführlich annotierte, weitgehend nicht-redundante Proteinsequenzen mit Funktionen, Domänenstrukturen und posttranslationalen Modifizierungen abgelegt sind. Lassen sich dort keine Ergebnisse finden, werden weitere Suchen in den Datenbanken *dbEST* (Boguski *et al.* 1993) und Genbank (Benson *et al.* 1999) generiert.

Massenspektrometrisch erzeugte Fragmentspektren aus dem ESI-Quadrupol-Iontrap Massenspektrometer werden in dem Suchprogramm *MASCOT* (Matrix Science, London, UK) automatisch nach Import der Spekt-

---

rendaten in Form von Peaktabellen im Batch-Verfahren gesucht (Perkins *et al.* 1999). Die Details der Suche sind im Kapitel 4.2.4 detailliert beschrieben.

## 4 Ergebnis und Diskussion

### 4.1 Problemstellung

In dieser Arbeit sollte ein Konzept erarbeitet und exemplarisch umgesetzt werden, die es erlaubt, eine Peptidquelle umfassend zu charakterisieren. Dieses Ansatz baut auf der Methodik des „Peptide-Trappings“ auf (Schulz-Knappe *et al.* 1996). Eine konsequente Weiterführung und Erweiterung dieser Methodik führt zu einem Inhaltsverzeichnis einer biologischen Probe auf Peptidebene (*Peptide-Inventory*). Ein solches Inhaltsverzeichnis wird in fünf Schritten erstellt:

1. Peptidextraktion aus einer biologischen Quelle
2. Dekomplexifizierung der Probe durch chromatographische Methoden
3. Kartierung der peptidischen Inhaltsstoffe (Visualisierung)
4. Identifizierung der einzelnen Peptide
5. Archivierung der Ergebnisse

Aufbauend auf diesen fünf Grundsritten galt es, die Identifizierung und Archivierung unter Berücksichtigung der folgenden Punkte weiter zu entwickeln:

- Möglichst umfassende Charakterisierung<sup>8</sup> von kartierten Peptiden.
- Plattformunabhängige<sup>9</sup> Auswertung und Archivierung der Messdaten.
- Übertragbarkeit der Methode auf verschiedene Peptidquellen.

Damit eine Methode diese Aufgabe erfüllen kann, ist es zunächst nötig, die

---

<sup>8</sup> Die Charakterisierung eines Peptides in diesem Zusammenhang wird detailliert in einem späteren Kapitel beschrieben.

<sup>9</sup> Plattformunabhängig bedeutet in diesem Fall die Unabhängigkeit der erzeugten Daten von den angewendeten Techniken und von speziellen Geräteherstellern. Dies ist besonders wichtig in Bezug auf die Datenformate der erzeugten Daten.

Quelle, aus der der Analyt (Peptid) stammt und dann den Analyten selbst zu beschreiben.

Wie schon in der Einleitung bemerkt, wurde in dieser Arbeit humanes Hämofiltrat zur Etablierung der Methodik eingesetzt. Diese Quelle repräsentiert den Großteil der im Körper befindlichen hormonell wirksamen Peptide. Zu diesem Teil addiert sich der Anteil der Peptide, der unspezifisch aus größeren Proteinen abgebaut wurde, die aber durchaus auch als diagnostische Marker relevant sein können (Anderson *et al.* 2002).

Wie die Kartierungen dieser Peptidquelle (Schritt 3 des Inventories) zeigen, handelt es sich um eine hoch komplexe Probe (ca. 50.000 bis 100.000 Peptidsignale<sup>10</sup>). Bei einer systematischen Identifikation der kartierten Peptide fällt der Automatisierung also ein großer Anteil an der Lösung der Aufgabe zu. Diese Automatisierung ist sowohl bei der Analyse selbst, als auch bei der Auswertung der erzeugten Daten wichtig, da durch die vor der Kartierung durchgeführte Dekomplexifizierung (Schritt 2) durch chromatographische Methoden eine Vielzahl von Proben entstehen, die bei der nachgeschalteten Identifizierung große Datenmengen erzeugen. Weiterhin muss dafür Sorge getragen werden, dass bei mehrstufigen Prozessen keine Informationen verloren gehen oder verfälscht werden (*sample tracking*).

Die massenspektrometrische Analyse der Proben wurde aus verschiedenen Gründen als Basistechnologie für dieses Vorhaben gewählt. Zum einen sind massenspektrometrische Verfahren sehr sensitiv, und zum anderen können sie mit relativ wenig Aufwand einen hohen Grad der Automatisierung erlangen. Wie an anderer Stelle beschrieben, wird die MALDI-TOF-MS bereits sehr erfolgreich zur Kartierung von Peptidquellen eingesetzt (Schrader *et al.* 2001; Schulz-Knappe *et al.* 2001). Anfänge bei der Identifizierung solcher

---

<sup>10</sup> Die mit MALDI-TOF-MS aufgenommenen Peptidkarten zeigen nach Feinkartierung ca. 50.000-100.000 peptidische Signale, die jedoch zum Teil redundant sind, so dass mit dieser Methode eine Anzahl von ca. 15.000-25.000 verschiedener Peptide sichtbar gemacht werden können.

kartierten Peptide wurden bereits gemacht, wobei Urin als biologische Quelle benutzt wurde (Raida *et al.* 1999). Weiterhin nutzt ein weiteres großes Applikationsfeld, welches sich im Laufe der Arbeit als wohl wichtigste Disziplin der postgenomischen Forschung etablierte, genau diese Technologien: Die *Proteomforschung*. Ziel dieser Disziplin ist es, den gesamten Proteingehalt von Zellen, Körperflüssigkeiten oder Organen zu charakterisieren (Kapitel 1.1). Bei einer Gegenüberstellung der Versuchsschemata von Proteom- und Peptidomforschung wird man auf den ersten Blick viele Ähnlichkeiten feststellen. So beinhalten beide die grundlegenden Prozessschritte: Extraktion, Dekomplexifizierung, Visualisierung, Identifizierung und Archivierung. Bei einer detaillierteren Betrachtung der angewendeten Methoden zeigen sich allerdings deutliche Unterschiede, die zumeist ihre Ursache in den zu untersuchenden Molekülen (Proteine bzw. Peptide) haben.

#### **4.2 Entwicklung eines allgemeinen Ablaufschemas zur Erstellung eines *Inventories***

In Abbildung 4-1 sind die Informationen zusammengestellt, die zu einer umfassenden Beschreibung einer Peptidquelle in das Inventory einfließen. Hieraus ergeben sich die Techniken und Methoden, die zur Erstellung herangezogen werden.

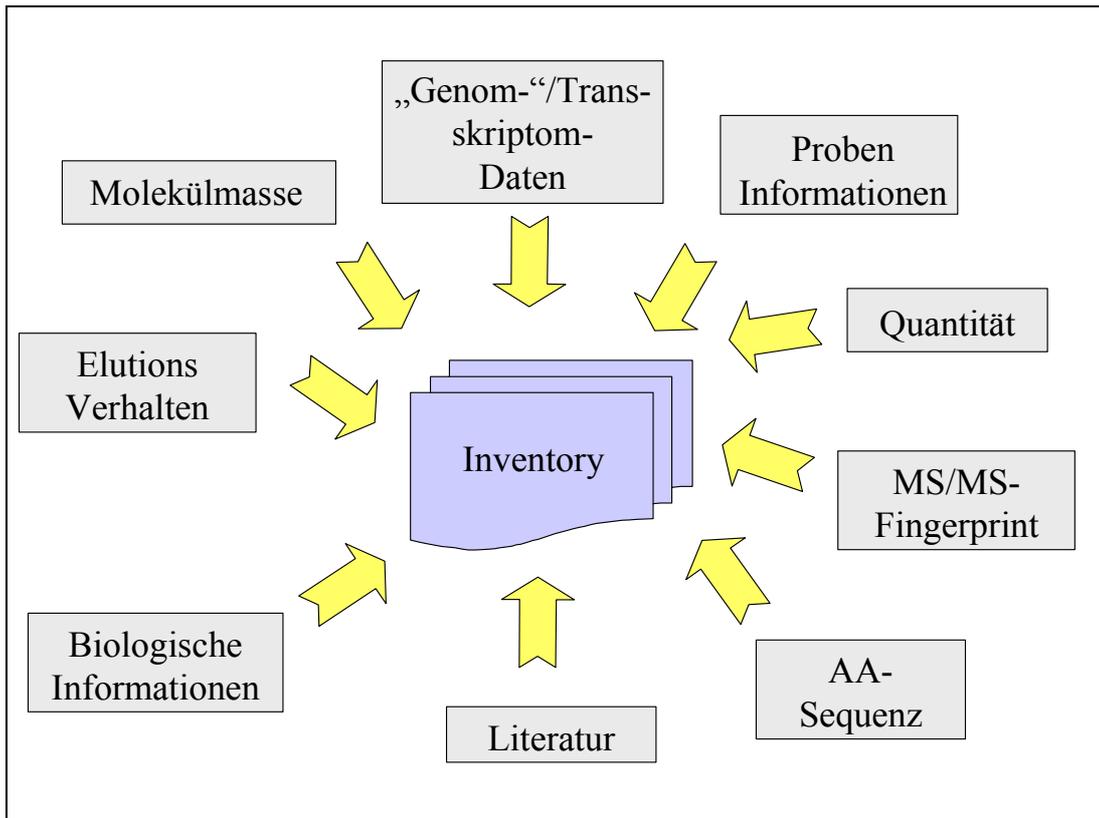


Abbildung 4-1 Überblick über die Informationen, die bei einer umfassenden Charakterisierung einer Peptidquelle in einer Datenbank gesammelt werden.

Die in das *Inventory* einfließenden Daten können in primär und sekundär ordnende Daten unterteilt werden. Diese Einteilung ist sinnvoll, um später ein geeignetes Datenbankmodell für die Archivierung der Ergebnisse zu entwerfen. Für die Messung selbst ist sie nicht von Bedeutung. Primär ordnende Daten werden als solche bezeichnet, wenn sie direkt durch eine Messung der Probe erzeugt werden. Hierzu gehören:

- Elutionsverhalten
- Molekülmasse
- MS/MS-Fingerprint
- Aminosäuresequenz

Sekundär ordnende Daten werden dementsprechend aus externen Datenbeständen akquiriert. Im Falle eines Inventories sind dies:

- Probeninformationen (z.B. Probenquelle, pathologische Informationen etc.)
- Genom- und Transkriptom- Daten (Gen, mRNA, mglw. auch EST oder SNP-Daten)
- Biologische Informationen (z.B. Spezies, Proteinvorläufer, Zellstamm etc.)
- Literaturzitate

Diese Informationen sind zumeist vor der peptidomischen Untersuchung bekannt oder werden nach der Identifizierung einzelner Peptide recherchiert.

Im ersten Schritt soll beschrieben werden, wie die primär ordnenden Daten erzeugt und in einer Datenbank abgelegt werden. Ein allgemeines Schema zur Erzeugung dieser Daten ist in Abbildung 4-2 gegeben.

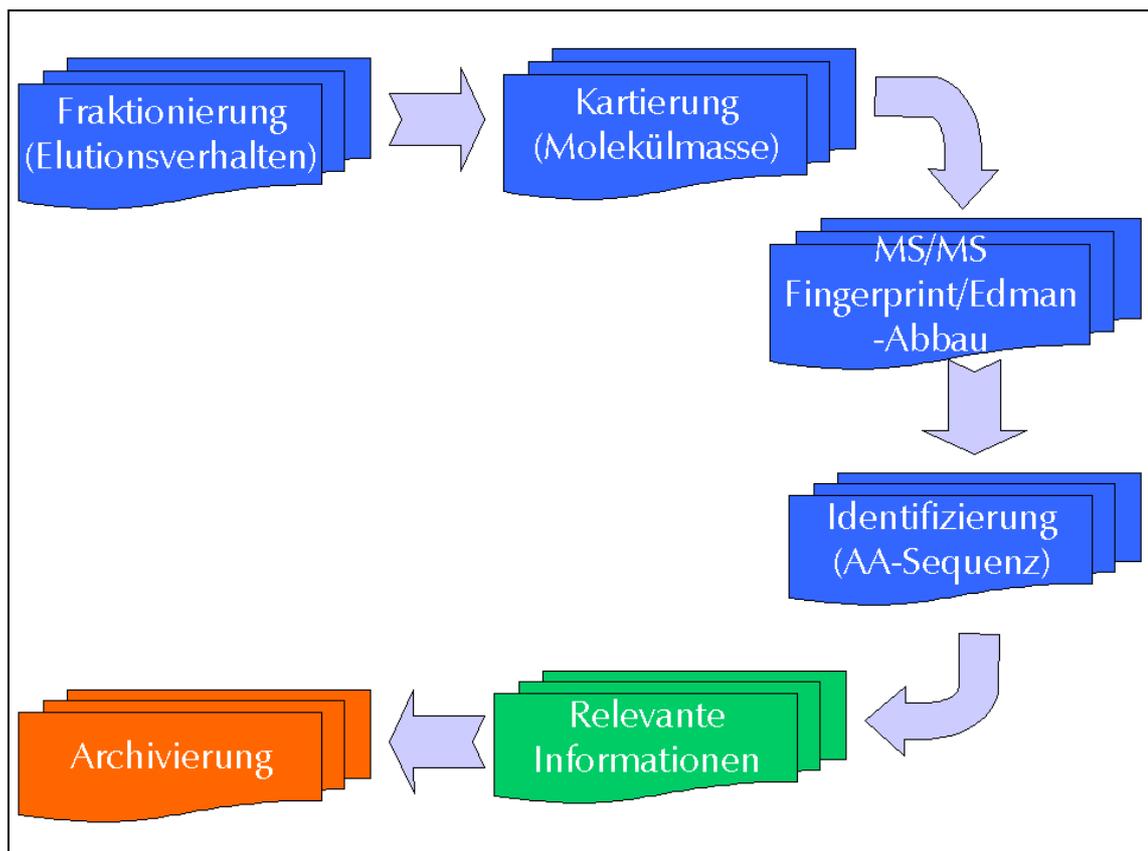


Abbildung 4-2 Ablaufschema zur Erzeugung von primär und sekundär ordnenden Daten innerhalb eines Inventories. Blau gekennzeichnet sind die zur Erzeugung von primären Daten notwendigen Schritte, grün markiert, die Sekundärdatenschritte. Rot die Zusammenführung und Archivierung der Daten in einer Datenbank.

Die ersten beiden Schritte dieses Ansatzes sind bereits an anderer Stelle ausführlich beschrieben (Jürgens 1999). Hier wird der Fokus auf die Schritte zur Erzeugung der Schritte zur Erzeugung der MS/MS-Fingerprints, der Identifizierung der Peptide und auf die Automatisierung des Gesamtkomplexes<sup>11</sup> gelegt.

#### **4.2.1 Fraktionierung der Probe**

Die Dekomplexifizierung der Proben wurde nach dem im Material und Methodenteil beschriebenen Prinzip durchgeführt, so dass eine vollständige Automatisierung, sowohl der Probenaufgabe, als auch der Fraktionierung möglich geworden ist. Das hier gezeigte Beispiel beruht jedoch noch auf dem Ansatz, wie er in Abbildung 3-1 gezeigt ist. Hier wurde nach orientierenden Untersuchungen der Peptidbankfraktionen selbst aufgrund der hohen Komplexität der Einzelfraktionen ein weiterer Fraktionierungsschritt etabliert. Jede dritte Fraktion<sup>12</sup> aus der ursprünglichen Peptidbank wurde rechromatographiert, so dass die Peptidbank insgesamt in ca. 10.000 Fraktionen unterteilt wurde. Die einzelnen Fraktionen wurden dann so bearbeitet, wie es in Abbildung 4-2 dargestellt ist.

#### **4.2.2 Kartierung der Molekülmassen einer Peptidquelle**

Die prinzipielle Vorgehensweise ist bereits ausführlich in einer früheren Arbeit (Jürgens 1999) beschrieben worden. An dieser Stelle war es allerdings aufgrund des hohen Probenaufkommens nötig, im Sinne einer weiteren Fortentwicklung auf Basis der vorliegenden Konzepte die massenspektrometri-

---

<sup>11</sup> Hierbei wird nur kurz auf die Automatisierung des ersten und letzten Schrittes eingegangen, da diese Teile von anderer Seite bei BioVisioN umgesetzt wurden. Eine Beschreibung ist dennoch notwendig, um den Gesamtprozess verstehen zu können.

<sup>12</sup> Aufgrund der verhältnismäßig geringen Trennleistung der präparativen Chromatographie (RP-Säulen mit 10 cm Durchmesser) und damit verbundenen Redundanzen im Peptidgehalt einzelner Fraktionen wurde darauf verzichtet, jede Ursprungsfraktion zu rechromatographieren.

sche Hardware weiter zu automatisieren.

#### 4.2.2.1 Entwicklung eines MALDI-MS Autosamplers

Durch das sehr große Probenaufkommen (vgl. Kap.4.2.1) war es im Zuge der Arbeit notwendig geworden, nicht nur eine Maldiprobeplatte à 400 Probenpositionen ohne die Interaktion eines Benutzers zu vermessen, sondern gleich mehrere dieser Platten nacheinander automatisch dem Gerät zuzuführen. Hierdurch ergibt sich die Möglichkeit, sowohl die Nacht als auch die Wochenenden in den Messplan einzubinden und es ist somit eine wesentlich höhere Auslastung des Gerätes zu erzielen.

Der zu entwickelnde Autosampler (*DiskJockey*) (Walter *et al.* 2002) sollte folgende Kriterien erfüllen:

- Unbeaufsichtigte Messung im 24/7 Betrieb.
- Möglichst wenig invasive Aufrüstung der Voyager STR-Systeme.
- Vollständige Software-Kontrolle.
- Benachrichtigungssystem bei Fehlern per e-mail/SMS an den Operator.

##### 4.2.2.1.1 Mechanische Hardware

Der Prototyp der Plattenwechslers besteht als Probenreservoir aus einem Magazin mit einer Kapazität von 20 MALDI Probenplatten des Voyager Systems mit jeweils 400 Probenpositionen, so dass maximal 8000 (20 x 400) Proben ohne Unterbrechung das System durchlaufen können. Dies entspricht bei einer mittleren Messdauer von 360 min pro Probenplatte einer Messzeit von 60 h pro Magazin. Das Magazin wird, wie in Abbildung 4-3 gezeigt, horizontal zur Be-/Entlademimik von einem Linearantrieb bewegt. Das individuelle Plattenhandling wird von einem Magnetgreifer, der an einem pneumatischen Zylinder montiert ist, realisiert. Die dafür benötigte Druckluft liefert ein Kompressor des Voyager Systems. Um die korrekte Platzierung der Probenplatten während des Lade- und Entladevorgangs zu überwachen, wird ein optischer Sensor mit einer Auflösung < 1 mm benutzt.

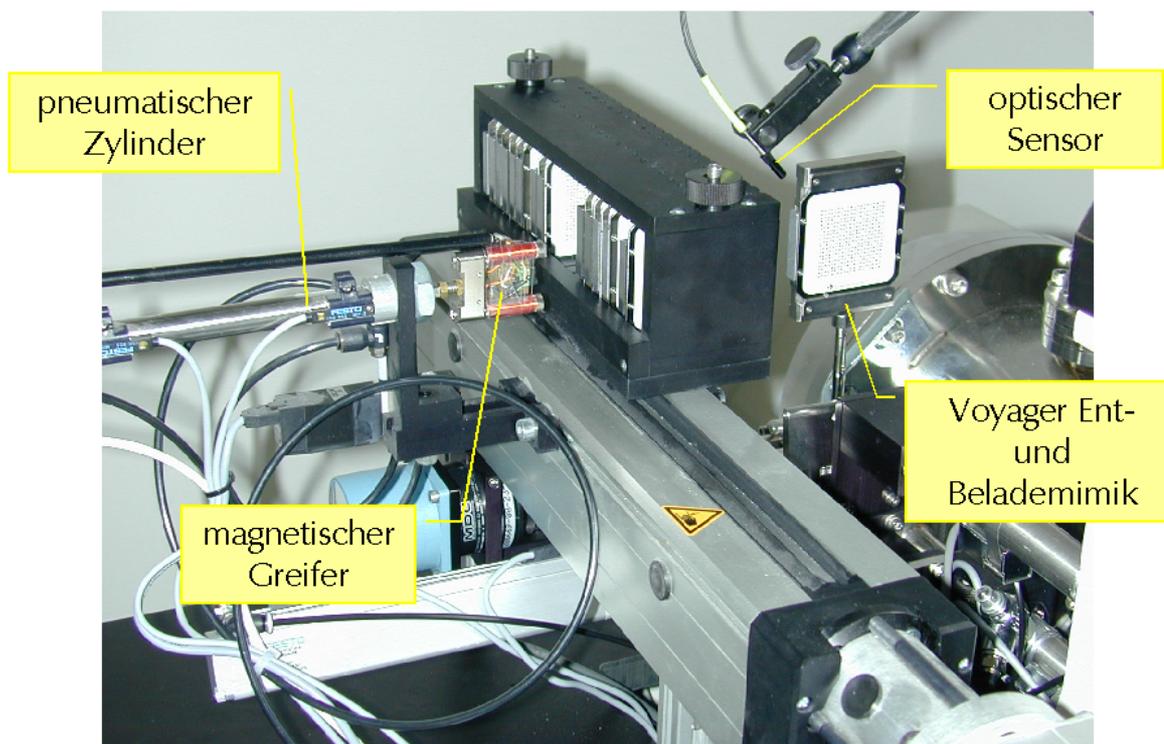


Abbildung 4-3 Fotografische Darstellung des Autosamplers-Prototyps „DiskJockey I“ mit Bezeichnung der einzelnen Komponenten.

#### 4.2.2.1.2 Elektronische Hardware

Der pneumatische Zylinder, der Greifer, der Sensor und der Linearantrieb werden von einem 8051 basierten Amtel  $\mu$ -Controller gesteuert. Die Kommunikation mit dem Steuerrechner erfolgt über eine serielle RS 232 Schnittstelle. Die komplette elektronische Hardware, inklusive der Spannungsversorgungen, ist in einem 19'' Rack untergebracht.

#### 4.2.2.1.3 Software

Die größte Herausforderung in diesem Projekt bestand darin, den Plattenwechsler in die bestehende Softwareumgebung zu integrieren. Die Softwareanpassung sollte dabei so wenig invasiv wie möglich stattfinden. Zu diesem Zweck wurde eine neue Applikation entwickelt, das DiskJockey Control Panel (DJCP), welche als „Treiber“ für den DiskJockey fungiert, d.h. alle Funktionen des Wechslers steuert. Über eine graphische Bedienoberfläche können zusätzlich alle Funktionen zum Einrichten und für Wartungsarbeiten ausgewählt

werden.

Das eigentliche Herz der Anordnung bildet weiterhin das Sequence Control Panel (SCP), welches Bestandteil der kommerziellen Voyager Software ist. Es fungiert als „Master“ des Systems und wurde um Funktionen erweitert, die die stapelweise Verarbeitung des MALDI-Targets ermöglichen. Der „batch-Betrieb“ wird über eine ASCII-Datei gesteuert, in der Informationen über Messparameter, der im Magazin befindlichen Platten abgelegt sind. Diese Parameter sind im speziellen die Position des Probenträgers im Magazin, die zugehörige Parameterdatei jeder Probenplatte, der Status der Messung und die Identifikationsnummer der Platte. Eine Übersicht über die Steuerung des DiskJockeys und des MALDI-MS gibt Abbildung 4-4.

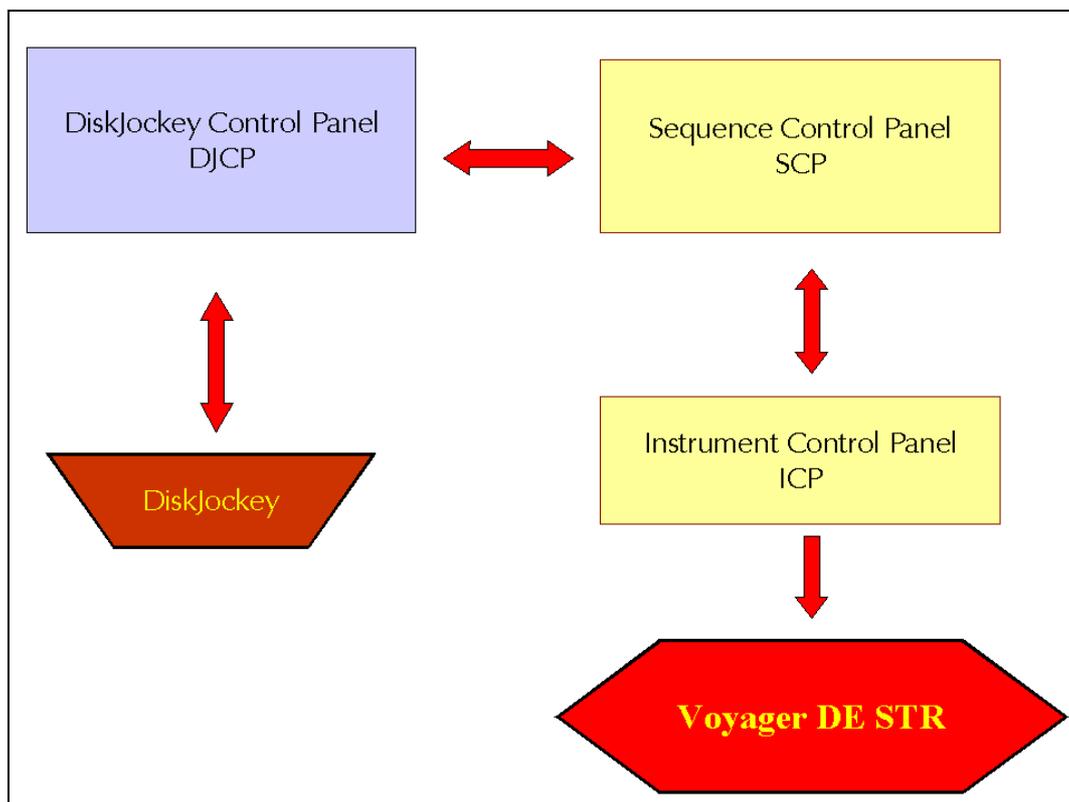


Abbildung 4-4 Schematische Darstellung der DiskJockey Steuerung. Das DiskJockey-Control-Panel (DJCP) interagiert mit der Hardware des Autosamplers und mit der Hardware des MALDI-MS über das Sequence Control Panel (SCP) und das Instrument Control Panel (ICP). Das SCP übernimmt hierbei die Funktion des „Masters“ und steuert die „Slaves“ DiskJockey und DJCP.

#### 4.2.2.1.4 Zusammenfassung

Durch die Entwicklung des MALDI-Autosampler Systems kann man den

Durchsatz der Voyager STR MALDI-MS-Geräte um Faktor drei steigern. Das System läuft sehr zuverlässig als eine 24/7 Einheit und benachrichtigt den Bediener im Falle eines Fehlers per email oder SMS via Handy.

### **4.2.3 Erzeugung von MS/MS-Fingerprints**

Da jedes sich in einer biologischen Quelle befindliche Peptid eine eigene, individuelle Entität darstellt, muss es auch unabhängig von anderen in der Quelle befindlichen Peptiden identifiziert werden. Eine isolierte Analyse kann auf zwei verschiedenen Wegen erreicht werden. Einerseits kann man das Peptid über chromatographische Methoden gezielt aufreinigen und es der Analyse (z.B. Edman-Abbau) zuführen,. Andererseits kann es aus einer komplexen Probe über die Trennmöglichkeiten eines Massenspektrometers isoliert und anschließend analysiert werden. Beide Methoden haben Vor- und Nachteile. So ist die Analyse nach einer Aufreinigung qualitativ hochwertiger und bei Reindarstellung eines Peptids frei von Interferenzen und liefert somit sehr verlässliche Ergebnisse. Allerdings ist jeder Aufreinigungsschritt zeitintensiv und mit einem Verlust an Probe verbunden. Eine direkte Analyse aus einer komplexen Probe ist um ein vielfaches schneller, zeigt aber häufig Grenzen bei sehr komplexen biologischen Proben. Aus diesen Gründen wurde eine Kombination aus beidem gewählt:

Die teilweise Dekomplexifizierung mit chromatographischen Methoden erfolgte kombiniert mit der massenspektrometrischen Sequenzierung über selektiv vorgefilterte MS/MS-Fringerprintenspektren.

Das QIT-MS wurde vor allem aufgrund seiner Robustheit und Schnelligkeit für diesen Ansatz ausgewählt. Das Ziel der Methodenentwicklung war, möglichst schnell einen umfassenden Überblick über die Komponenten einer Peptidquelle zu bekommen.

#### **4.2.3.1 Hardware**

Um eine schnelle und effektive Erzeugung von Fragmentspektren aus den im

96-well-Format vorliegenden Proben zu gewährleisten, wurde eine Anordnung gewählt, die auf einer Fließinjektionsanalyse (FIA) beruht. Die genaue Beschreibung der Methodik befindet sich in Kapitel 3.2.3.2.

#### *4.2.3.2 „Alcatrap“: Ein Softwaretool zur automatischen Erstellung von Arbeitslisten für die Ionenfalle*

Für eine Abarbeitung der Proben nach dem oben gezeigten Prinzip musste eine softwareseitige Lösung gefunden werden, die den Umfang der kommerziellen Gerätesoftware überstieg. Die zu erarbeitende Lösung sollte folgenden Anforderungen genügen:

- Importmöglichkeit von Daten aus dem Kartierungsschritt (MALDI-TOF-MS Peaklisten).
- Überführung der einfach geladenen Vorläufer-Ionen aus der MALDI-MS Messung in theoretisch mögliche mehrfachgeladene Ionen, wie sie in der Elektrospray Ionisation üblich sind.
- Erstellung einer importierbaren Arbeitsliste für das LCQ.
- Möglichst hohe Zukunftscompatibilität (z.B. Verwendung der Software mit anderen Massenspektrometern).

Das allgemeine Ablaufschema für die Softwarelösung ist in Abbildung 4-5 abgebildet. Dabei werden zunächst die  $n$  intensivsten Signale aus einer MALDI-MS Peakliste ausgewählt. Danach erfolgt die Berechnung in die typischen mehrfach geladenen Ionen der Elektrosprayionisation, um mit diesen dann Methoden-Dateien zu erstellen. Ausgehend hiervon wird eine Arbeitsliste für das LCQ erzeugt.

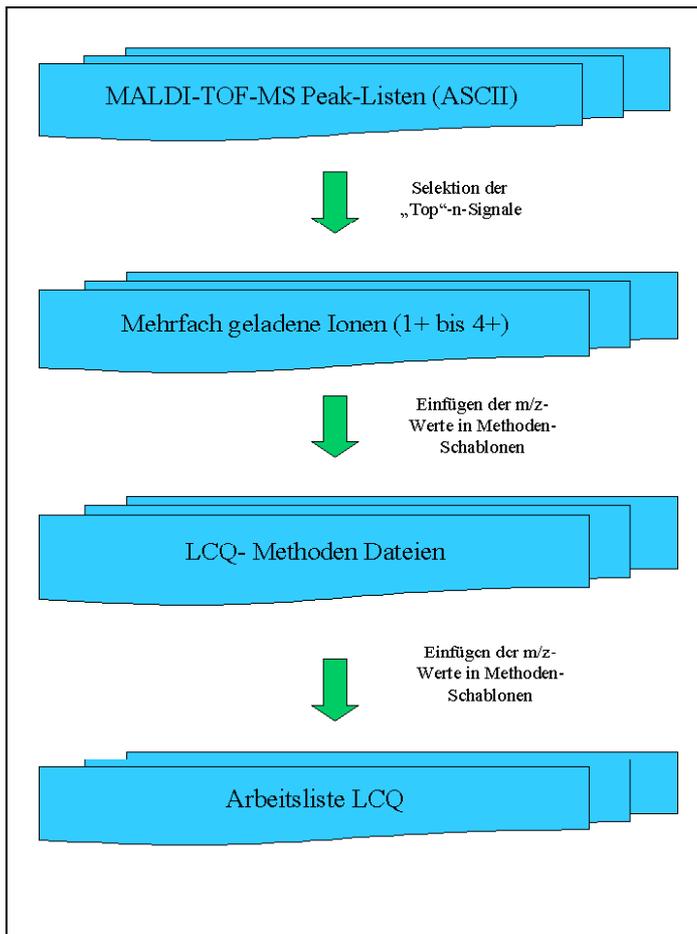


Abbildung 4-5 Schematischer Aufbau der Berechnungen innerhalb der Alcatrap-Software. Zunächst werden Peaklisten aus MALDI-TOF-MS Messungen eingelesen. Dann werden die intensivsten „n“-Signale ausgewählt und in verschiedene Ladungszustände umgerechnet. Hieraus erstellt Alcatrap Methodendateien für das LCQ und eine importierbare Arbeitsliste für die automatische Abarbeitung.

#### 4.2.3.2.1 Überführung von einfach geladenen „MALDI-Ionen“ in mehrfach geladene „ESI-Ionen“ bei nativen Peptiden

Zunächst muss festgestellt werden, in welchen Ladungszuständen native Peptide typischerweise bei der Elektrospray-Ionisation in der Ionenfalle auftreten. Dieser Parameter ist sowohl abhängig von der Größe des Peptides<sup>13</sup>, als auch von der Art der Ionisierung (Flussrate, Desolvatisierung). In der Lite-

<sup>13</sup> Genauer, von der Anzahl der basischen Gruppen innerhalb des Peptids, die statistisch mit der Größe des Peptids zunehmen.

ratur finden sich hier verschiedene Ansätze zur Beschreibung und Theorie der Ionenbildung im Elektrospray. Um die Ladungsverteilung nativer Peptide zu bestimmen, wurden stichprobenartig jeweils drei Proben aus den verschiedenen pH-Pools des Hemofiltrats (HF) untersucht. Die Spektren der Proben wurden hierbei sowohl mit MALDI-MS vermessen, als auch in der Ionenfalle im „Fullscan“-Modus aufgezeichnet. Die jeweils 3 intensivsten Signale im  $m/z$ -Intervall 1000-2000, 2000-3000 und 3000-4000 des MALDI-TOF-MS wurden ausgewählt und mit den Intensitäten der korrespondierenden mehrfach geladenen Signale korreliert. In Abbildung 4-6 sind die absoluten Häufigkeiten der Ladungszustände mit Intensitäten  $>30\%$  des Vorläuferions, unterteilt in Massenbereiche, dargestellt. Hieraus kann abgeleitet werden, mit welcher Wahrscheinlichkeit ein Peptid mit einer bestimmten Masse in den unterschiedlichen Ladungszuständen im Elektrospray-Spektrum vorkommt<sup>14</sup>.

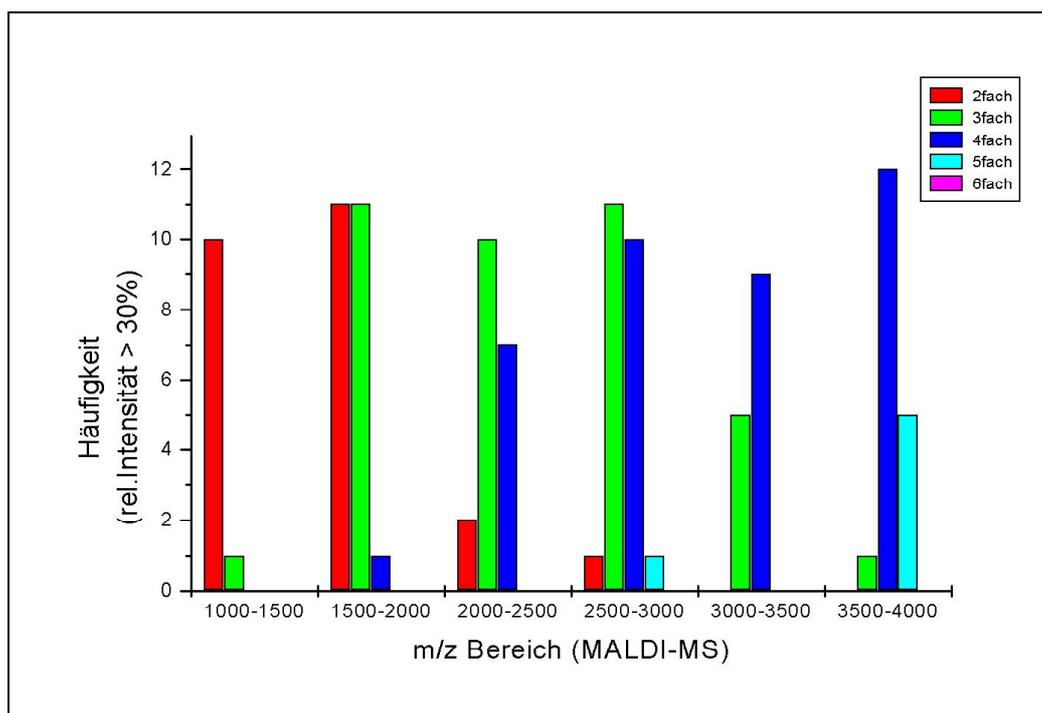


Abbildung 4-6 Verteilung der nativen Peptide aus verschiedenen Massenbereichen (MALDI-TOF-MS) auf verschiedene Ladungszustände im ESI-Spektrum..

<sup>14</sup> Angemerkt sei hier, dass es sich nicht um definierte Grenzen handelt, sondern um fließende Übergänge, die von Peptid zu Peptid variieren können.

Wie zu erwarten nimmt der Ladungszustand mit zunehmender Masse auch zu. Deutlich zu erkennen ist, dass jeweils zwei Ladungszustände im jeweiligen Massenfenster dominieren.

Mit Blick auf die fließenden Übergänge zwischen den Massenbereichen und der Möglichkeit, positive Identifizierungen anhand der unabhängigen Identifizierung eines anderen Ladungszustandes des selben Peptids intern validieren zu können, wurden jeweils die beiden am häufigsten auftretenden Ladungszustände für die massenspektrometrische Fragmentierung herangezogen. Im einzelnen bedeutete dies eine Messung folgender Ladungszustände: Bereich 1.000 Da bis 2.000 Da zweifach und dreifach, Bereich 2.000 Da bis 3.500 Da dreifach und vierfach und 3.500 bis 4.500 vierfach bis fünffach. Die sechsfach geladenen Ionen spielen in dem betrachteten Massenfenster keine Rolle.

#### 4.2.3.2.2 Bestimmung der Kollisionsenergien für die automatische MS/MS-Messung von nativen Peptiden mit Hilfe einer Ionenfalle (LCQ)

Um ein möglichst informationsreiches und gleichzeitig rauscharmes Spektrum zu erhalten, muss der Wert der Kollisionsenergie für ein Peptid angepasst werden. Bei tryptischen Proteinfragmenten ist dieser Wert innerhalb eines Ladungszustandes von der Masse des Peptids linear abhängig. Um diese Funktion für native Peptide zu bestimmen, wurden exemplarisch 56 Ionen nativer Peptide aus Hämofiltrat der Ladungszustände  $2^+$ - $4^+$  mit einem  $m/z$ -Verhältnis von 400 – 1200 untersucht. Die Ionen wurden so fragmentiert, dass jeweils 50%, 30% und 10% des Vorläuferions erhalten blieb. Die korrespondierenden Kollisionsenergien<sup>15</sup> wurden aufgezeichnet. Tabelle 4-1 zeigt die Übersicht für eine Restintensität des Vorläufers von 30%.

---

<sup>15</sup> Die Einheiten für die Kollisionsenergie sind arbiträr und nicht in Energien umrechenbar. Es gilt: Je größer der Wert, desto höher ist die Kollisionsenergie.

Tabelle 4-1 Kollisionsenergien für 30% Restintensität des Vorläufers

2-fach geladen		3-fach geladen		4-fach geladen	
m/z	CE	m/z	CE	m/z	CE
533.05	29	423.51	26	409.19	28
539.32	27	431.85	30	448.05	27
569.79	28	438.09	27	461.01	29
576.37	27	486.72	27	501.76	31
598.78	29	606.04	25	626.87	29
634.01	28	616.94	25	667.02	26
791.17	30	622.71	25	684.23	28
791.85	33	675.3	28	701.2	30
805.66	30	767.5	30	706.67	29
851.06	33	855.21	37	712.09	26
907.57	37	882.32	29	712.63	26
962.72	36	978.57	34	751.16	29
1001.63	31	1025.26	35	762.5	27
1011.54	34	1026.17	31	838.7	36
1034.48	33	1097.43	32	855.37	30
1041.95	32	1103.36	36	963.31	33
1083.66	32	1158.67	35	1021.47	34
1150.99	29	1188.2	33	1102.21	35
1159.17	35	1188.99	32		

Zunächst wurde untersucht, welche Abhängigkeit der Kollisionsenergie vom Ladungszustand besteht. Hierzu wurden Regressionsanalysen für jeden der drei Datensätze bei 30% Restintensität durchgeführt (Abbildung 4-7 und Tabelle 4-2).

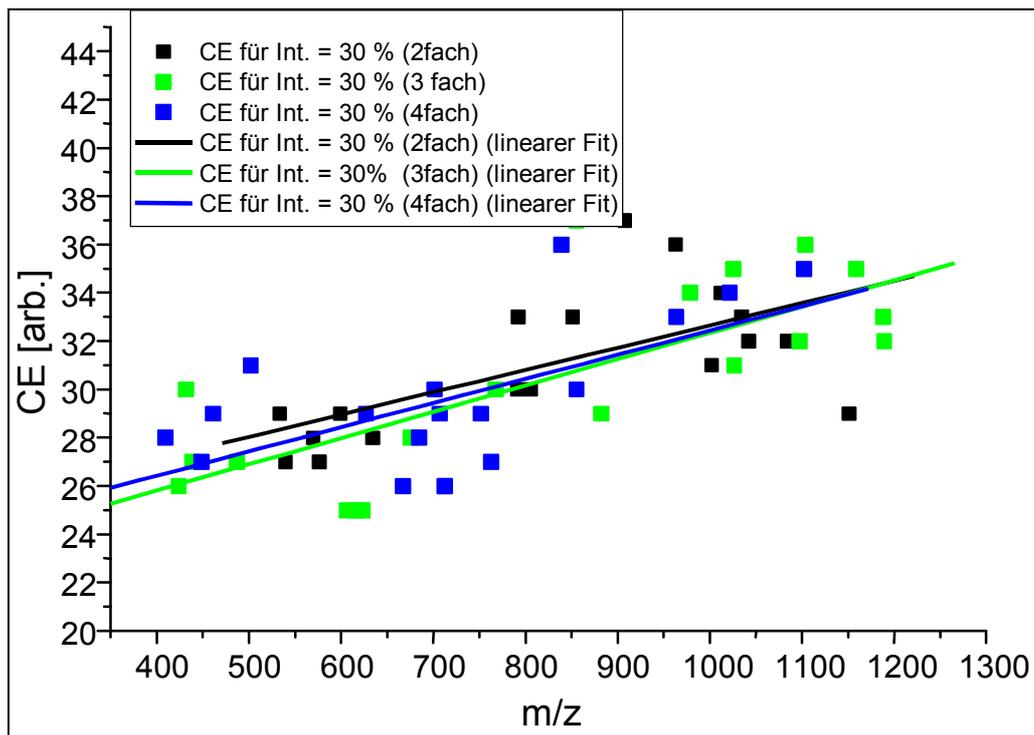


Abbildung 4-7 Regressionsgeraden der Kollisionsenergien (CE) bei 30% Restintensität des Vorläuferions. Schwarz gekennzeichnet sind die 2-fach geladenen, grün die 3-fach geladenen und blau die 4-fach geladenen Ionen. Die Ergebnisse der Regressionsanalyse sind in Tabelle 4-2 dargestellt.

Tabelle 4-2 Resultate der Regressionsanalyse aus Abbildung 1-9. Mit A=Achsenabschnitt, B= Steigung, R=Regressionskoeffizienten nach Pearson, SD= Standardabweichung und N=Anzahl der Datenpunkte.

$y = B \cdot x + A$	2-fach	3-fach	4-fach
<b>A</b>	23.42	21.45	22.41
<b>B</b>	0.00922	0.01088	0.01003
<b>R</b>	0.66393	0.76124	0.6247
<b>SD</b>	2.32	2.63	2.49
<b>N</b>	19	19	18

Die beobachteten Kollisionsenergien liegen alle zwischen 24 und 38. (2fach: 27 bis 37, dreifach: 25 bis 37 und vierfach: 26 bis 36). Die für die einzelnen Ladungszustände durchgeführten linearen Regressionsanalysen zeigen eine

Steigung zu größeren  $m/z$ - Werten. Die Regressionskoeffizienten sind mit 0,62 bis 0,76 nicht besonders gut und auch die Standardabweichungen liegen mit ca. 10% der Kollisionsenergie recht hoch. Klare Abhängigkeiten zwischen der Kollisionsenergie und dem Ladungszustand sind nicht zu erkennen. Vielmehr sind die optimalen Kollisionsenergien bei den zu untersuchenden nativen Peptiden weit gestreut. Ein Trend zu höheren Kollisionsenergien bei zunehmendem  $m/z$ -Wert ist allerdings deutlich zu erkennen. Ausgehend von diesen Ergebnissen wurde der  $m/z$ -Bereich in zwei Bereiche unterteilt ( $m/z$  400-800 und  $m/z$  800-1200), um diese jeweils mit einer möglichst optimalen Aktivierungsenergie auszustatten. Dazu wurden die Mittelwerte der Kollisionsenergien im jeweiligen  $m/z$ -Intervall ohne Berücksichtigung des Ladungszustandes gebildet. Für 400-800 ist dies  $CE = 28$ , für 800-1200 ist dies  $CE = 33$ . Diese Werte wurden für die automatische Erstellung der Methoden verwendet. Zusätzlich wurde für jedes der untersuchten Peptide die Intensität für das Vorläuferion mit den oben ermittelten Kollisionsenergien unter zu Hilfe nahme der  $CE$ -Werte bei 50% und 10% Restintensität interpoliert. Unter der Annahme, dass die Kollisionsenergien in dem betrachteten Bereich linear verlaufen, zeigte sich, dass 95% der Ionen im Bereich von 10% bis 50% Restintensität liegen, was für eine ausreichende Fragmentierung ausreichend sein sollte.

#### 4.2.3.2.3 Funktionen von Alcatrap

Da die Werte für die Kollisionsenergien nicht direkt über die Gerätesoftware dynamisch gesetzt werden können, wurden Schablonen (Templates) von Methodendateien mit den oben beschriebenen Kollisionsenergien erzeugt. In diesen Templates sind alle konstanten Werte, die für die Messung notwendig sind, bereits vorgegeben. Die Werte für das Vorläuferion ( $m/z$ ), und den Messbereich werden dann dynamisch – je nach zu untersuchendem Peptid – auf Binärebene von dem selbst erstellten Softwaretool *Alcatrap* in das Template eingefügt. Das Resultat dieses Patchvorgangs ist eine Methodendatei für jeden zuvor in der Software spezifizierten Ladungszustand der  $n$  intensivsten Signale eines MALDI-MS-Spektrums. Um die Abarbeitung von

ganzen Mikrotiterplatten zu ermöglichen, können mehrere MALDI-MS-Spektren stapelweise verarbeitet werden. Eine gleichzeitig erstellte Arbeitsliste kann in die Gerätesoftware importiert werden. Im Laufe dieser Arbeit ist die beschriebene Grundfunktionalität von Alcatrap noch um einige zusätzliche Funktionen erweitert worden, die im folgenden anhand der Bedieneroberfläche erläutert werden (Abbildung 4-8 bis Abbildung 4-11).

Nach dem Starten von Alcatrap muss der Benutzer verschiedene Eingaben zur zu messenden Probe machen. Die Seite „*General*“ (Abbildung 4-8) dient zur Eingabe allgemeiner Daten. Mit dem Feld „*Inputfile*“ können einer oder mehrerer Peaklisten aus MALDI-MS-Spektren ausgewählt werden, die von Alcatrap prozessiert werden sollen. Weiterhin muss festgelegt werden, welche Methoden-Schablonen benutzt werden sollen und wo diese hinterlegt sind, und es wird festgelegt, wie die vom LCQ zu importierende Arbeitsliste bezeichnet werden soll und wo die prozessierten Methodendateien abgelegt werden sollen. Mit Hilfe der Funktion „*Calculate Charges from... ..to*“ wird festgelegt welche Ladungszustände berechnet werden sollen.

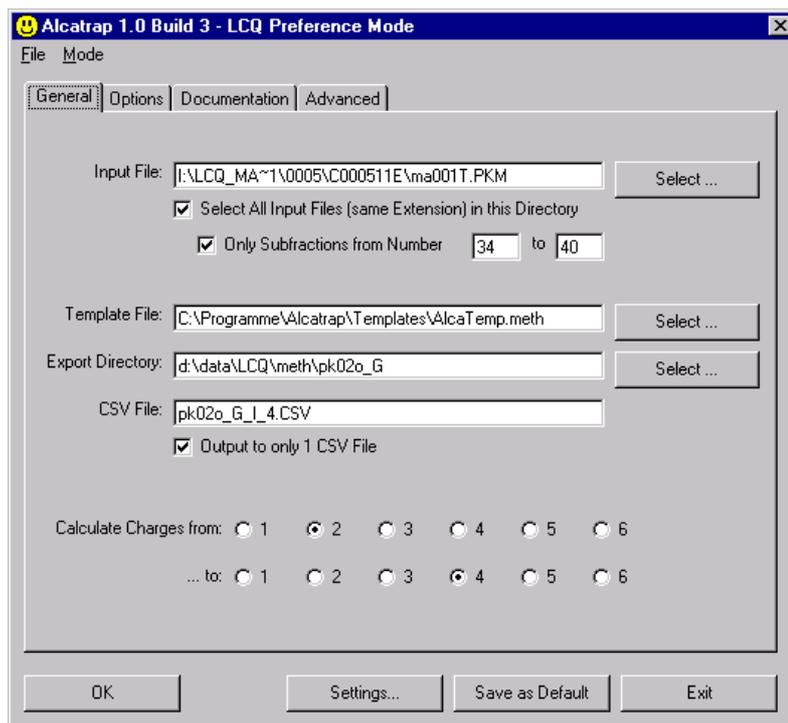


Abbildung 4-8 „General“-Seite von Alcatrap. Hier werden Einstellungen vorgenommen, welche MALDI-Peaklisten verarbeitet werden sollen und mit welchen Parameterdateien dies stattfinden soll. Außerdem werden Speicherorte und zu berechnende Ladungszustände festgelegt.

Die Seite „Options“ (Abbildung 4-9) legt weitere Optionen fest. So wird sowohl der  $m/z$ -Bereich der MALDI-MS Spektren festgelegt („Set original mass range“), als auch der  $m/z$ -Bereich der ESI-Vorläuferionen festgelegt („none“, „fixed Range“, „charge dependent Range“), um unsinnige<sup>16</sup> Vorläuferionen auszuschließen. Es wird festgelegt, wie viele der MALDI-MS Signale verarbeitet werden sollen („Create exactly  $n$  EMD<sup>17</sup> file(s) per input file“). Diese Entscheidung wird aufgrund der Intensität der MALDI-MS Signale im „Inputfile“ getroffen<sup>18</sup>. Es kann entschieden werden, ob ein einziges Methoden-Template verwendet wird (statische Kollisionsenergie) oder ob die Kollisions-

<sup>16</sup> Als unsinnig werden hier diejenigen Vorläuferionen bezeichnet, die nicht in das in Kapitel 4.2.3.2.1 erarbeitete Schema passen.

<sup>17</sup> EMD-Dateien sind die Methoden-Dateien der LCQ-Software. Mit dieser Option wird die Anzahl der zu berücksichtigen  $n$  intensivsten Signale des MALDI-TOF-MS Spektrums festgelegt.

<sup>18</sup> Es werden dabei die Signale von höchster zu niedrigster Signalintensität nacheinander bearbeitet.

energie dynamisch nach den bereits oben beschriebenen Kriterien gesetzt werden („*Multiple Template Files*“).

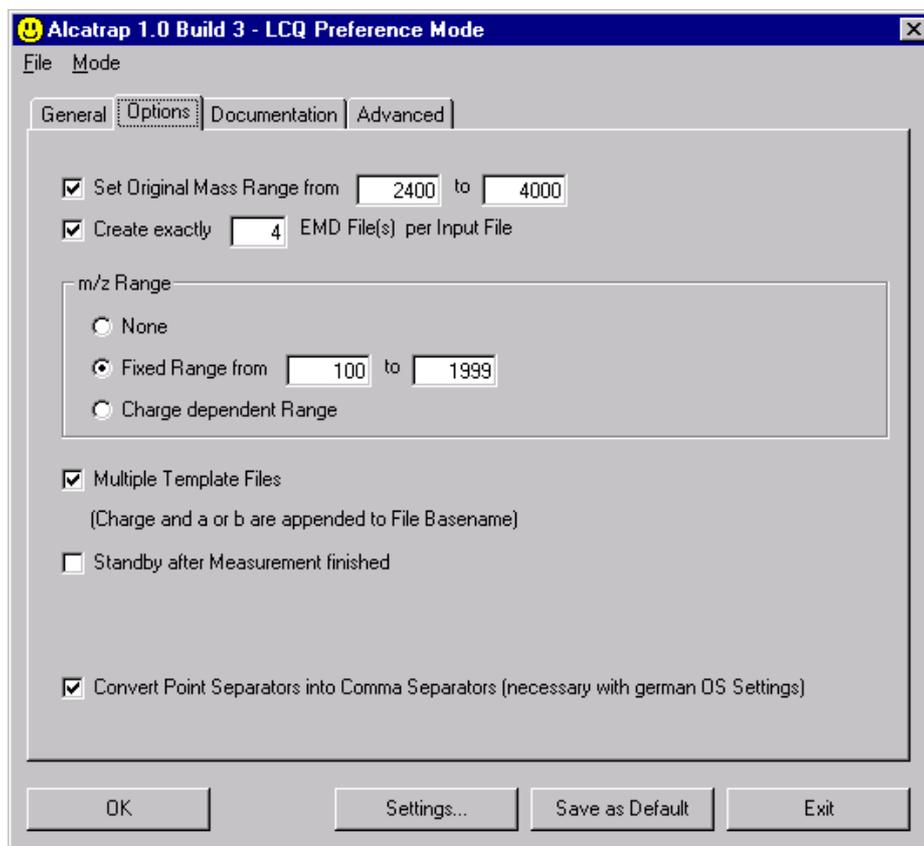
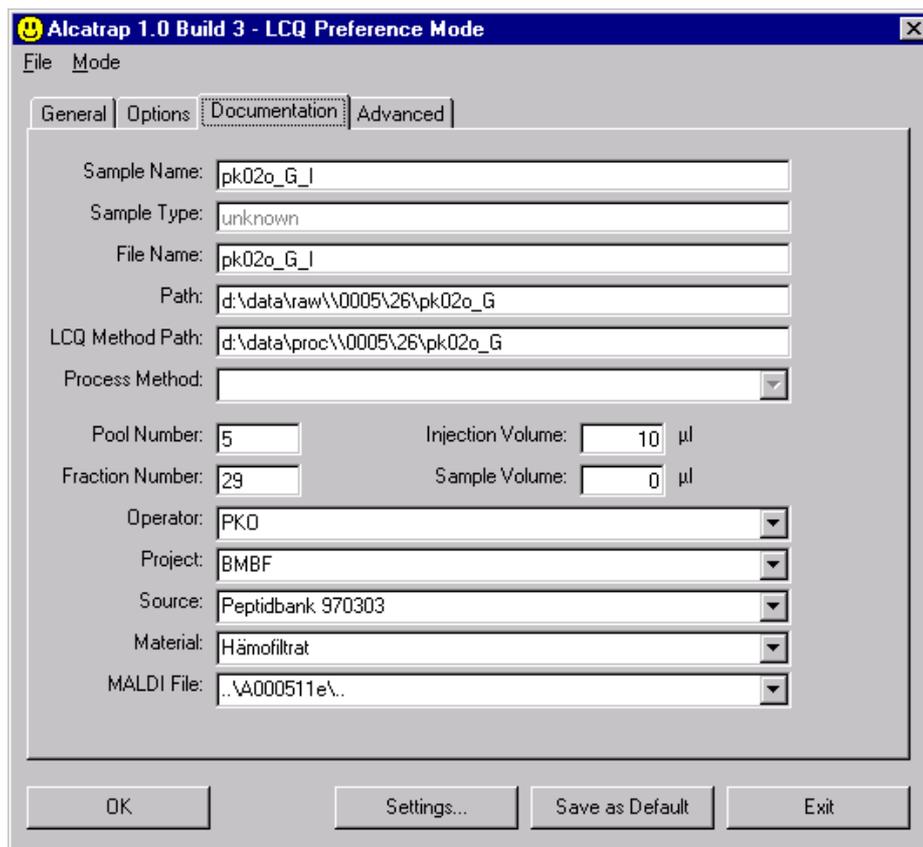


Abbildung 4-9 „Options“-Seite von Alcatrap. Hier können die Massenbereiche für die Messung festgelegt werden. Es kann die Anzahl der zu untersuchenden Peptid-Signale pro MALDI-MS Spektrum festgelegt werden und entschieden werden, ob die Kollisionsenergie dynamisch oder statisch gesetzt werden soll.

In die „*Documentation*“-Seite (Abbildung 4-10) werden ausschließlich Daten aufgenommen, die bei der Prozessierung in die Arbeitsliste des LCQs aufgenommen werden. Diese Informationen beinhalten sowohl den Dateinamen der aufzunehmenden Rohdaten, wie auch die jeweils zur Messung korrespondierende Methodendatei. Ferner können hier probenrelevante Daten, wie Probenname, Injektionsvolumen, Projektname, Operator etc. eingetragen werden, die dann im weiteren Verlauf des Prozesses zur Verfügung stehen<sup>19</sup>.

---

<sup>19</sup> Diese Eingaben sind besonders im Hinblick auf die lückenlose und automatisierte Dokumentation der Messungen von Bedeutung.



The screenshot shows the 'Alcatrap 1.0 Build 3 - LCQ Preference Mode' dialog box with the 'Documentation' tab selected. The dialog has a menu bar with 'File' and 'Mode'. Below the menu bar are four tabs: 'General', 'Options', 'Documentation', and 'Advanced'. The 'Documentation' tab contains the following fields:

- Sample Name: pk02o\_G\_I
- Sample Type: unknown
- File Name: pk02o\_G\_I
- Path: d:\data\raw\0005\26\pk02o\_G
- LCQ Method Path: d:\data\proc\0005\26\pk02o\_G
- Process Method: (dropdown menu)
- Pool Number: 5
- Injection Volume: 10 µl
- Fraction Number: 29
- Sample Volume: 0 µl
- Operator: PKO
- Project: BMBF
- Source: Peptidbank 970303
- Material: Hämofiltrat
- MALDI File: ..\A000511e\..

At the bottom of the dialog are four buttons: 'OK', 'Settings...', 'Save as Default', and 'Exit'.

Abbildung 4-10 „Documentation“-Seite von Alcatrap. Hier werden Daten für die „samplelist“ des LCQ erfasst.

Die „Advanced“-Seite (Abbildung 4-11) gibt dem Benutzer Möglichkeiten, bestimmte Vorläuferionen von der Prozessierung und damit der Messung auszuschließen. Der Ausschluss von Vorläuferionen kann auf zwei verschiedene Arten erfolgen. Zum einen ist es möglich, Signale von der Verarbeitung auszunehmen, die in zwei aufeinanderfolgenden Fraktionen einer Chromatographie vorhanden sind. Hier wird der redundanten Analyse von über mehrere Fraktionen eluierenden Peptiden vorgebeugt. Zum anderen können bestimmte Signale in einer Datei definiert werden, die ausgeschlossen werden sollen. Hierbei handelt es sich zumeist um Matrixionen oder bereits vorher identifizierte Peptide. Zusätzlich können Ausschlusslisten für Folgemessungen automatisch erstellt werden („PKO Special: Create/Append to Mass List ASCII File“). Des Weiteren ist die Möglichkeit gegeben „average“ in monoisotopische Massen umrechnen zu lassen. Diese Möglichkeit kommt dann zum Tragen, wenn die zugrundeliegende MALDI-TOF-MS-Messung im linear Mo-

aus durchgeführt wurde und somit mittlere Molekulargewichte zur Verfügung stehen.

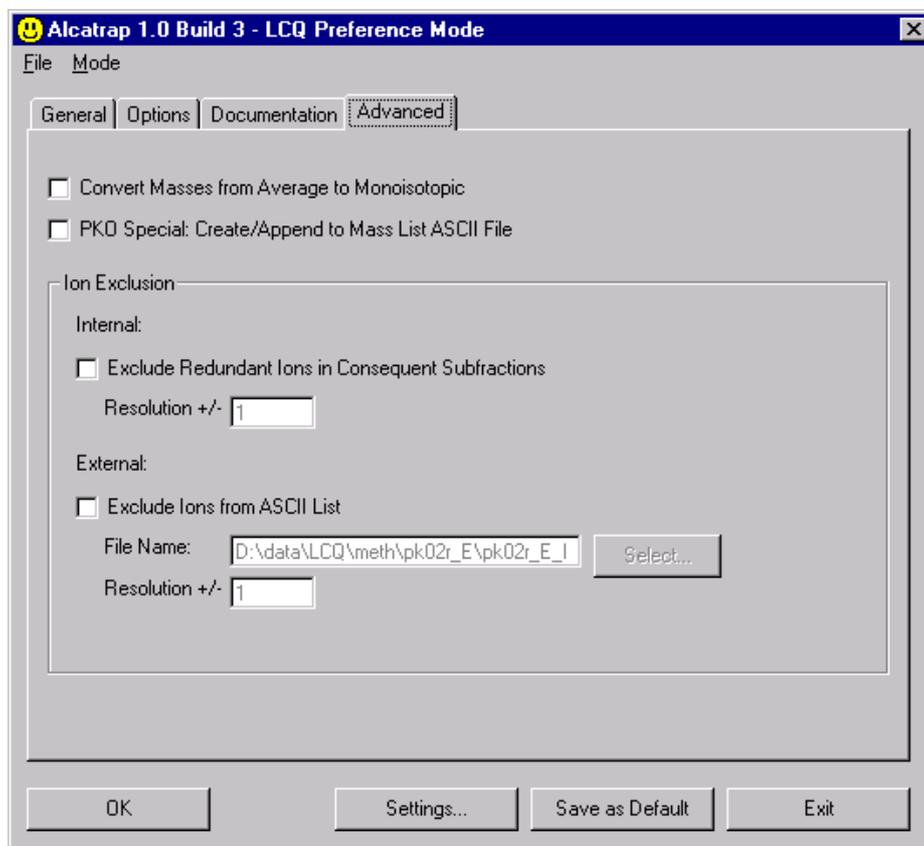


Abbildung 4-11 „Advanced“-Seite von Alcatrap. Hier können bestimmte Signale von der Prozessierung ausgenommen werden, mittlere in monoisotopische Vorläuferionen überführt werden und Ausschlusslisten für Folgemessungen erstellt werden.

Für die Berechnung wurde für 1500 zufällig ausgewählte Peptide die monoisotopische gegen die mittleren Massen aufgetragen (Abbildung 4-12). Eine lineare Regression liefert den Zusammenhang für die Umrechnung von mittleren in monoisotopische Massen für Peptide

$$MW_{\text{mono}} = 0.9994 * MW_{\text{av.}} - 0.034$$

Die Anwendung der Gleichung auf native Peptide, die aus der PeTrA-Datenbank ausgewählt wurden, zeigt eine Standardabweichung von 27,9 ppm. Diese Ungenauigkeit ist mit ca. 10% der Messungengenauigkeit eher als klein einzustufen, so dass diese Näherung beim Fehlen genauer Reflektrom-MS Daten genutzt werden kann, um möglichst genaue mehrfach geladene

ESI-Ionen zu berechnen.

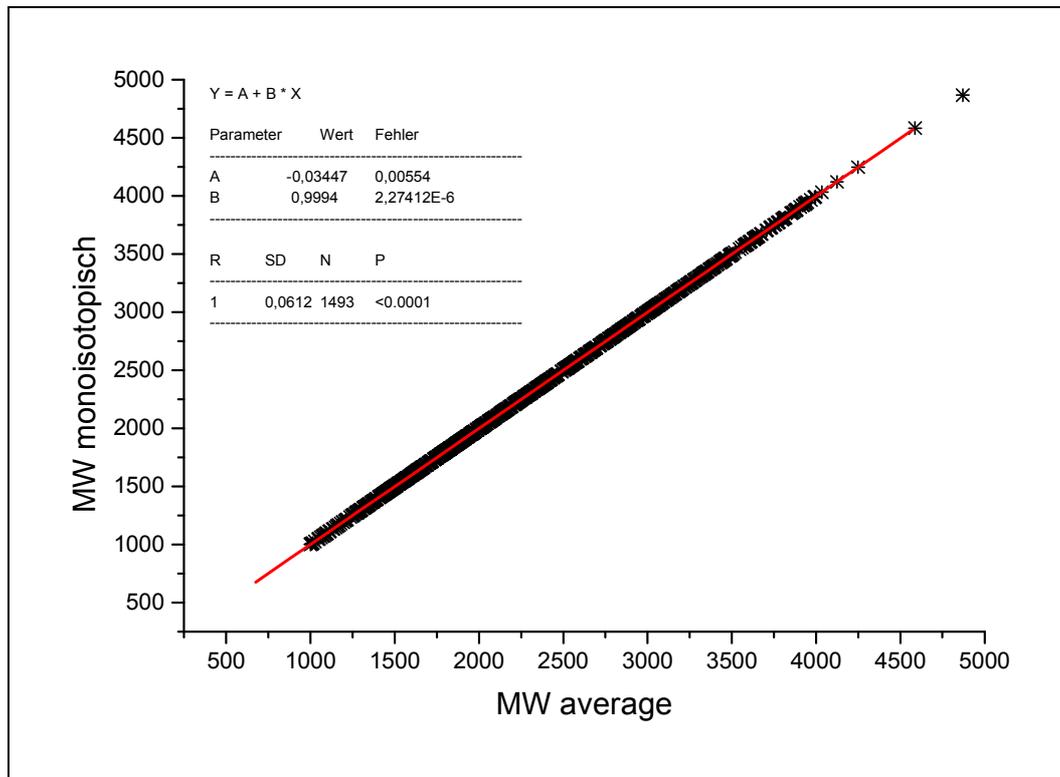


Abbildung 4-12 Auftragung der mittleren gegen die monoisotopischen Massen von 1.500 nativen Peptiden. Aus der linearen Regression ergibt sich die Funktion zur Umrechnung von average zu monoisotopischen Massen der Vorläuferionen.

#### 4.2.3.2.4 Die Schnittstelle zwischen Fragmentspektrenerzeugung und Datenbanksuche: „D2M“

Nachdem die in der Alcatrap-Software berechneten Vorläuferionen in der Ionenfalle fragmentiert wurden und für jeden einzelnen der berechneten Ladungszustände eine Datei mit dem Fragmentspektrum im proprietären Dateiformat vorliegt, müssen diese Dateien in ein Format überführt werden, welches von einer Datenbanksuchmaschine gelesen werden kann. Das zu diesem Zweck entwickelte Werkzeug mit dem Namen „D2M“ (Data to MASCOT) liest sowohl die Rohdaten, als auch Informationen aus der Arbeitsliste des LCQ ein und erstellt hieraus Dateien im MASCOT™ Format (Abbildung 4-13).

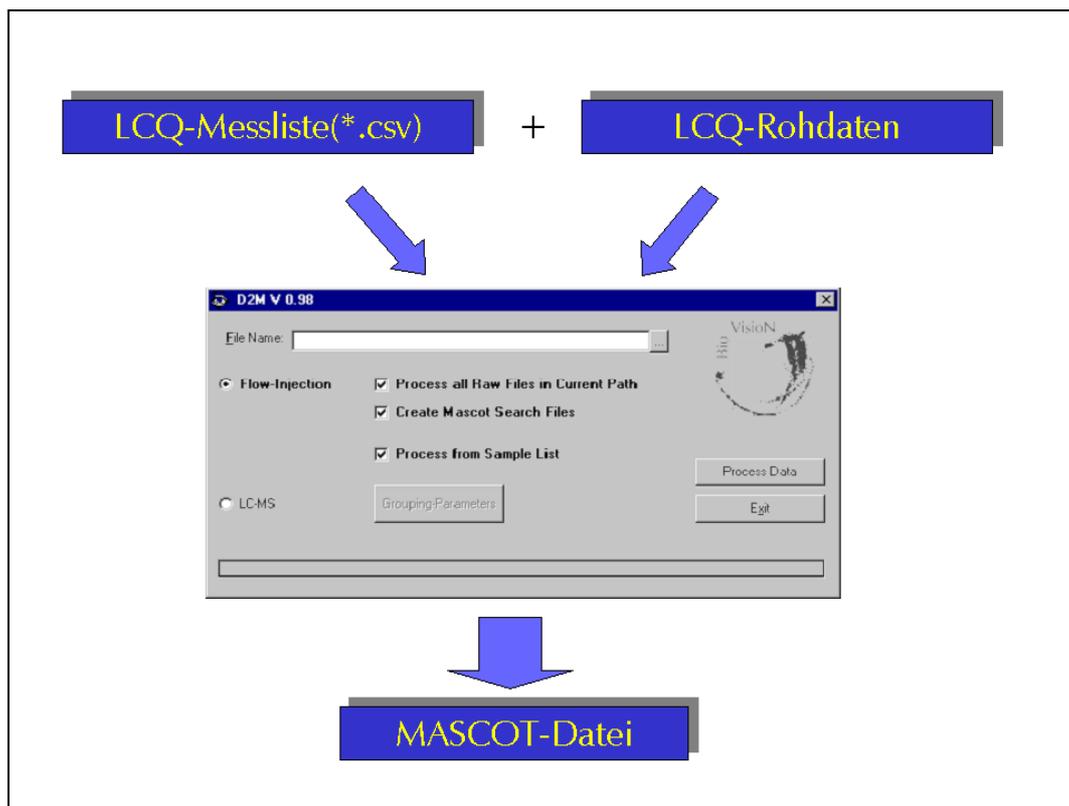


Abbildung 4-13 Schematische Darstellung des D2M-Werkzeugs zur Erstellung von MASCOT suchbaren

Die *D2M* zugrunde liegende Visual Basic Routine ist auf einer ActiveX-Bibliothek aufgebaut (Xcalibur.ocx), die es dem Benutzer erlaubt, auf alle relevanten Daten innerhalb von LCQ-Rohdaten zuzugreifen und diese zu verwerten. Auf eine weitere Datenreduktion der bereits zentroidisiert gespeicherten MS/MS-Dateien wurde verzichtet, da bei Ion-Trap-Daten aufgrund der Einheitsauflösung, weder der Ladungszustand der Fragmente, noch das monoisotopische Signal eines Isotopenclusters bei Ladungszuständen  $> 2^+$  sicher bestimmt werden können. Der  $m/z$ -Wert des Vorläuferions und dessen Ladungszustand wird dabei aus der Messliste extrahiert.

#### **4.2.4 Identifizierung von Peptiden mittels Datenbanksuche**

##### **4.2.4.1 Die Datenbanksuchmaschine MASCOT<sup>TM</sup>**

Der nächste Schritt bei der Erstellung eines Peptid-Inventories ist die Identifi-

zierung der einzelnen Peptide mit Hilfe der zuvor gewonnenen MS/MS-Spektren. Für Fragmentspektren gibt es hier mehrere Ansätze. Zum Einen können uninterpretierte Daten als Basis für eine Suche in Protein-oder Oligonukleotiddatenbanken dienen, zum Anderen gibt es den Ansatz des Sequenz-Tags, wobei eine kurze Aminosäuresequenz in Verbindung mit der Molekülmasse des Vorläuferions als Eingabe für die Suche benutzt wird. In dieser Arbeit wurde der Ansatz gewählt, der uninterpretierte Spektren verarbeiten kann, da die zweite Methode immer manuelle Interpretationsschritte benötigt, was bei einer hohen Probenanzahl nur bedingt möglich ist. Ferner sollte die Lösung weitere Kriterien erfüllen:

- Plattformunabhängiger Datenfluss
- Sensitivität (Vermögen auch schwach intensive oder verrauschte Daten mit hohen Scores zu belegen, falls die Identifikation „richtig-positiv“ ist.)
- Selektivität (Vermögen bei falsch positiven Hits niedrige Scores zu erzeugen.)
- Vergleichbarkeit der Scores untereinander
- Mögliche kaskadierende Suchen in mehreren Datenbanken
- Vollständige Automatisierbarkeit
- Schnelle Suche (skalierbar)
- Einbeziehung von möglichen (posttranslationalen) Modifikationen

Die Verwendung von uninterpretierten Daten lässt sich mit verschiedenen zur Verfügung stehenden Algorithmen durchführen. So gibt es reine Filter, die im Rahmen von vorgegebenen Fehlertoleranzen die zu durchsuchende Datenbank filtern (z.B. MS-TAG (Clauser *et al.* 1996)). Diese Filter liefern keinen Score, an dem der Benutzer erkennen kann, ob ein Ergebnis signifikant ist, oder ob es sich um einen zufälligen Hit handelt. Weiterhin gibt es Suchmaschinen, die theoretisch ermittelte Spektren mit den tatsächlich gemessenen vergleichen und das am besten passende Spektrum ermitteln. Hierbei werden die Spektren entweder miteinander korreliert (SEQUEST<sup>TM</sup> (Yates *et al.* 1995)) oder es wird kalkuliert, wie hoch die Wahrscheinlichkeit ist, einen

Zufallstreffer zu haben. Diese letztere Möglichkeit hat den Vorteil, dass die Scores normierbar und damit vergleichbar sind. Dies ist die beste Möglichkeit, falsch positive Ergebnisse weitgehend auszuschließen. Mit MASCOT™ (Perkins *et al.* 1999) gibt es im kommerziellen Bereich solch eine Lösung. MASCOT™ beruht weitestgehend auf dem MOWSE-Computer Programm (Pappin *et al.* 1993), das um einige Funktionen, wie ein wahrscheinlichkeitsbasiertes Scoring-System und die Möglichkeit, MS/MS-Daten zu verarbeiten, erweitert wurde. Weiterhin bietet MASCOT die Möglichkeit, fast alle MS/MS-Daten sowohl im proprietären Format als auch als Peaklisten im ASCII-Format einzulesen. Die Lösung ist in Bezug auf den Durchsatz beliebig skalierbar, wobei die Suchgeschwindigkeit linear ( $R = 0,99978$ ) mit der Rechnerleistung zunimmt. Dieser Benchmark wurde mit einem Testdatensatz von 25 MS/MS-Spektren in der dBEST-Datenbank (13.595.532 Sequenzeinträge) durchgeführt. Dafür wurde für verschiedene Suchen die Rechnerleistung nach und nach erhöht und die Suchdauer aufgezeichnet (Abbildung 4-14). Sowohl posttranslationale als auch chemische Modifikationen sind frei konfigurierbar und bis zu neun Modifikationen sind parallel suchbar. Über einen *Daemon* ist die Suche sowohl kaskadierend, als auch vollautomatisch von Client-Rechnern möglich.

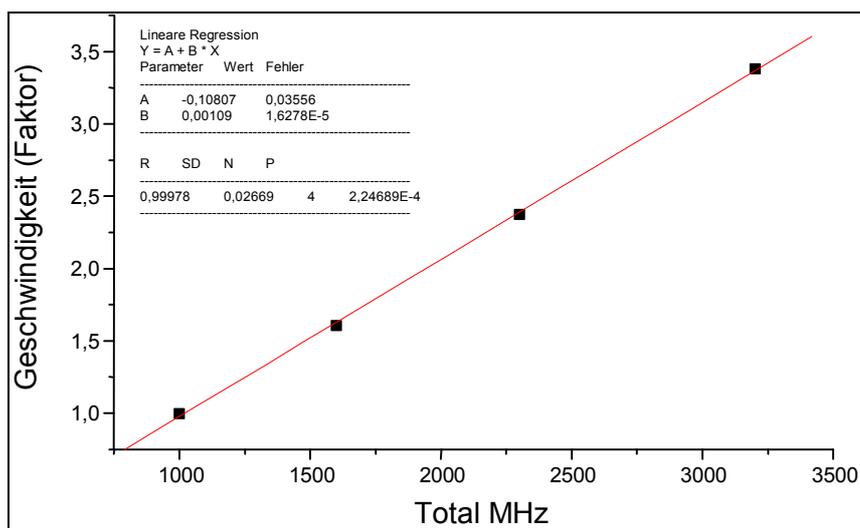


Abbildung 4-14 Benchmark der MASCOT™ MS/MS-Datensuche. Es wurde jeweils ein Datensatz von 25 Spektren in die Suchmaschine gegeben (Datenbank: dBEST (13.595.532 Sequenzeinträge)). Die Rechnerleistung wurde sukzessiv erhöht. Es zeigt sich eine lineare Abhängigkeit zwischen Suchgeschwindigkeit und Rechengeschwindigkeit in MHz.

Da die MASCOT™-Suchmaschine von Matrix Science die gestellten Anforderungen erfüllt und überdies auch noch Möglichkeiten bietet, eigene Anpassungen vorzunehmen, wurde sie als Suchmaschine ausgewählt. Eine schematische Darstellung der Funktionsweise dieser Suchmaschine ist in Abbildung 4-15 gegeben.

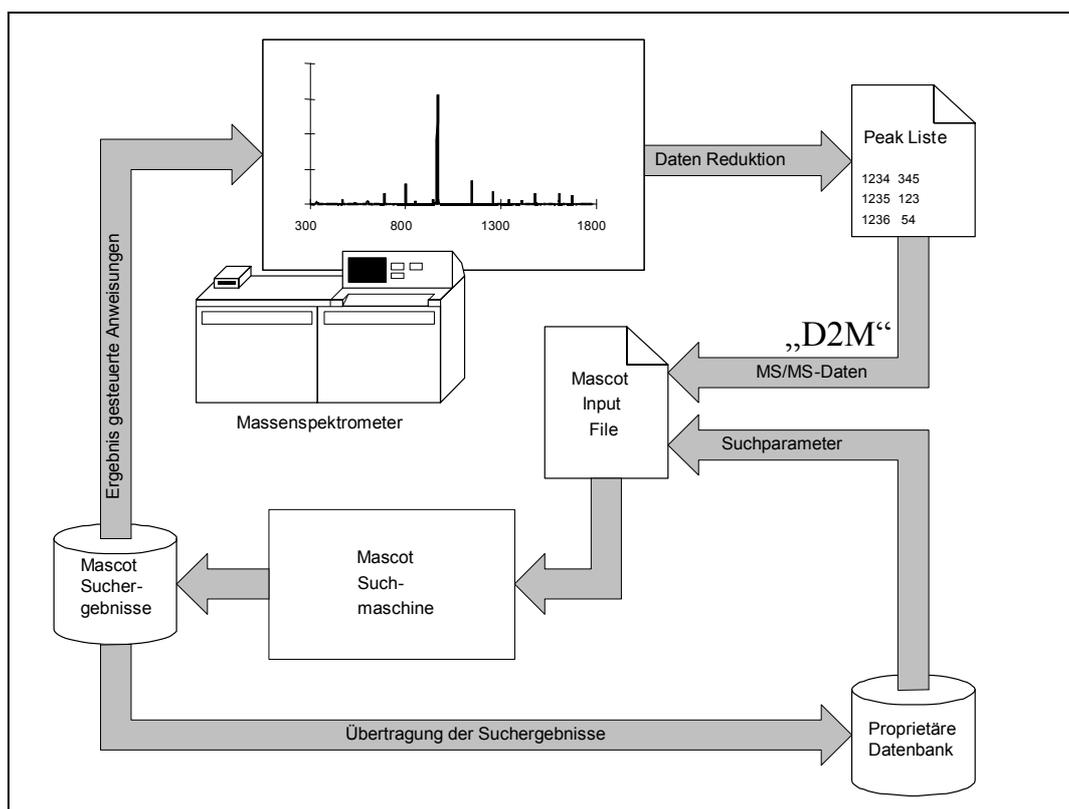


Abbildung 4-15 Schematische Darstellung der Datenbanksuchung mit der MASCOT-Suchmaschine unter Einbeziehung von D2M.

#### 4.2.5 Archivierung der Ergebnisse: PeTrA (Peptide Tracking Application)

Die Arbeiten zur Archivierung der Daten wurden nicht im Rahmen dieser Arbeit durchgeführt, sollen aber, da in dieser Lösung sämtliche Ergebnisse des „Inventories“ beschrieben sind, kurz hier beschrieben werden. Hierzu werden im folgenden einige Passagen aus dem Abschlussbericht des BMBF-Projekts „Umfassende Identifizierung humaner Blutpeptide zur Erarbeitung neuer diagnostischer und therapeutischer Leitsubstanzen“ (BMBF-Förderkennzeichen:

0311832) zitiert. Der Verfasser hat im Rahmen der hier vorliegenden Arbeit hauptsächlich für dieses Projekt gearbeitet.

Im Rahmen des Projektes sollte eine Datenbank geschaffen werden, die Daten über die analysierten Peptide aufnimmt. Dabei sind Suchfunktionen zu implementieren, so dass eine Identifikation eines eingetragenen Peptides anhand von seinen Eigenschaften möglich ist.

Es wurde dazu das SQL-Datenbanksystem Interbase der Firma Borland eingesetzt. Dabei handelt es sich um ein Client-Server System, welches über den Datenbank-Abfragesprache SQL (Structured Query Language) gesteuert wird. Alle Daten befinden sich dabei auf einem zentralen Server. Von den Clients können mittels SQL Abfragen gemacht werden, wobei nur die Ergebnismenge dieser Abfragen über das Netz verschickt wird.

Die Datenbank „PeTrA“ besteht aus insgesamt 41 Tabellen, von denen jede einem von vier verschiedenen Blöcken zuzuordnen ist:

- Peptid-spezifische Informationen (Angaben zum Peptid, seinem Vorläufer, Literaturangaben etc.) (grün)
- Verwaltung von Proben und Fraktionen in Form eines hierarchischen Baumes, des sogenannten Probenbaumes (gelb)
- die Tabelle „Identifikationen“, die Angaben darüber enthält, welches Peptid in welcher Fraktion gefunden wurden (blau)
- Hilfstabellen z.B. für Schlüsselworte (orange)

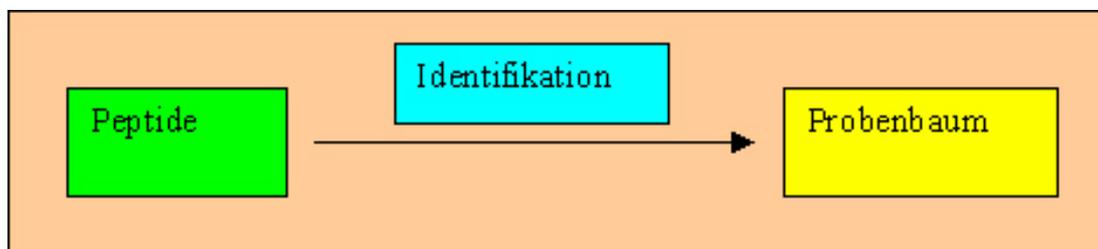


Abbildung 4-16 Schematischer Aufbau der PeTrA Datenbank zur Archivierung von Peptide-Inventory-Daten.

Da die ausschließliche Verwendung von SQL-Kommandos zur Steuerung und Abfrage der Datenbank zu umständlich ist, wurde mit der Programmier-

sprache Delphi ein Programm entwickelt, welches die Informationen der Datenbank sichtbar macht und Eingaben sowie Abfragen erlaubt, ohne dass der Benutzer SQL-Kenntnisse oder Kenntnisse der Datenbankstruktur benötigt:

- **Eingabekontrolle:** Das Programm „PeTrA“ führt Prüfungen der eingegebenen Werte durch, um Fehleingaben zu verhindern. Außerdem existieren in den Eingabemasken Pflichtfelder, die den Benutzer zur Erstellung eines vollständigen Datensatzes anhalten.
- **Visualisierung:** Logisch zusammenhängende Daten, die aus technischen Gründen in verschiedenen Tabellen abgelegt wurden, werden durch das Programm zu visuellen Einheiten zusammengefasst dargestellt.
- **Suche:** Durch verschiedene Suchmasken sowie die Möglichkeit, einen SQL-Suchtext frei einzugeben, werden die verschiedenen Kenntnisstände der Benutzer berücksichtigt. Es ist auch möglich, ohne Kenntnis des Datenbankmodells Peptide zu finden.
- **Literaturverweise:** Durch die Verknüpfung von neu erworbenem Wissen über die analysierten Peptide mit vorhandenem Wissen aus der Literatur wird die Datenbank zu einer umfassenden Wissensbasis.

#### **4.2.6 Zusammenfassung der bisherigen Ergebnisse**

Das Ziel, Hard- und Software zur Erstellung eines Peptide-Inventories in einem Ansatz so zu kombinieren, dass sämtliche Informationen über eine Peptidquelle zentral erfasst werden können, ist durch die Integration von bereits existierenden chromatographischen (2D-Chromatographie), massenspektrometrischen (ESI-QIT-MS/MS), peptidchemischen (Edman-Abbau) und bioinformatischen Methoden (Datenbanksuche) mit neu zu entwickelnden Methoden zur Automatisierung (MALDI-MS-Autosampler) und Datenverarbeitung (*Alcatrap*, *D2M*, *PeTrA*) erreicht worden. Der Fokus hierbei lag hierbei auf der automatischen Hochdurchsatz-Identifizierung von Peptiden aus biologischen Quellen. Dabei ist auf die besonderen Erfordernisse bei der Bearbeitung von nativen Peptiden gegenüber durch tryptischen Verdau entstandenen Peptide eingegangen worden. Es ist möglich, große Probenmen-

---

gen, wie sie bei der Erstellung von Peptid-Inhaltsangaben anfallen, zu bearbeiten und diese durch den Prozess zu schleusen, bis hin zur Archivierung in einer proprietären Datenbank. Im folgenden Schritt soll dieser Ansatz auf eine Peptidquelle angewendet werden, um Vor- und Nachteile zu diskutieren, damit dann die Methodik verfeinert werden kann.

### 4.3 Erstellung eines Inventories für humanes Blutfiltrat

Ausgehend von einer Hämofiltrat Peptidbank (Bezeichnung: „970303“)<sup>20</sup>, wurde begonnen, ein *Peptide-Inventory* nach dem zuvor beschriebenen Schema anzufertigen.

#### 4.3.1 Rechromatographie und Feinkartierung der Peptidbank

Die Peptidbank besteht aus 7 pH-Pools und umfasst insgesamt 292 Fraktionen. Erste Analysen zeigten jedoch, dass der Grad der Komplexität der einzelnen Fraktionen zu hoch ist, um damit effektiv nach dem schon beschriebenen Schema zu verfahren. Durch die hohe Anzahl von Peptiden und die Verteilung eines jeden Peptids auf mehrere Ladungszustände kommt es häufig zu Überlagerungen von zwei oder mehreren verschiedenen Peptiden. Da die Vorläuferionenauflösung bei elektrischen Ionenfallen im MS/MS-Modus zu gering ist, um diese getrennt voneinander zu analysieren, entstehen auf diese Weise Mischspektren aus mehreren Spezies, die einer automatischen Datenbanksuche und häufig auch einer manuellen Interpretation nicht zugänglich sind. Die Lösung dieser Problematik besteht in einer weiteren Subfraktionierung der Probe.

Hierfür wurde die gesamte Peptidbank einer erneuten Reversed-Phase-Chromatographie unterzogen und nach der bereits beschriebenen Methodik mit MALDI-TOF-MS kartiert. Alle Messungen wurden sowohl im linear Modus (maximale Sensitivität) als auch im Reflektron Modus (maximale Massengenauigkeit und Auflösung) durchgeführt. Aufgrund der häufig auftretenden Redundanzen durch über mehrere Fraktionen eluierende Peptide in der ursprünglichen Peptidbank wurde jeweils nur jede dritte Fraktion in Form einer virtuellen Peptidkarte in einer Datenbank hinterlegt. Die so entstandenen 97 Peptidkarten aus 9.312 Fraktionen enthalten ca. 100.000 Signale (30.000 im

---

<sup>20</sup> Die Herstellung dieser Peptidbank ist im Methodenteil beschrieben.

Reflektron Modus). Die geringere Anzahl an Species im Reflektron-Modus stellt eine bekannte Beobachtung dar, da insbesondere höhermassige Species im Reflektor-Modus verloren gehen und insgesamt die Empfindlichkeit schlechter ist. Auch hier ergeben sich Redundanzen durch breiter eluierende Peptide, so dass schließlich davon auszugehen ist, dass ca. 40.000-50.000 verschiedene zirkulierende Peptide (10.000-15.000 im Reflektron Modus) mit dieser Methode erfasst werden können. Abbildung 4-17 zeigt exemplarisch ein Spektrum der Fraktion 26 aus pH-Pool 6 der Blutfiltratbank 970303 (unten) und die durch Rechromatographie entstehende Peptidkarte (oben).

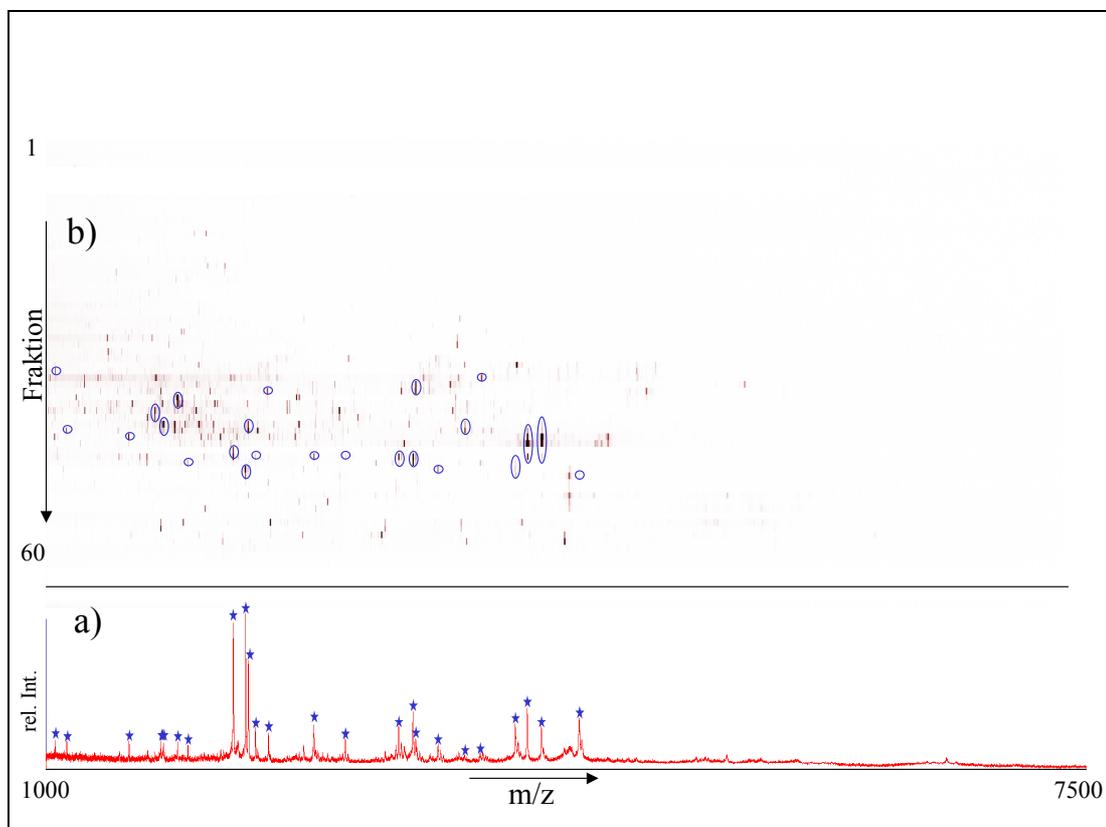


Abbildung 4-17 Visualisierung der RP-Rechromatographie einer Peptidbankfraktion aus „970303“ (Pool 6 Fraktion 26). a) MALDI-TOF-MS der ursprünglichen Fraktion und b) aus der Rechromatographie resultierende Peptidkarte. Blau gekennzeichnet sind die Signale in der Ursprungfraktion und ihre Position in der Karte der Rechromatographie. Zu erkennen ist die erneute Dekomplexifizierung und das damit verbundene Auftreten zusätzlicher Signale.

Im Spektrum sind 25-40 Signale von Peptiden abgebildet. Durch die weitere Auftrennung erhöht sich die Zahl der detektierten Peptide auf ca. 500, wie an den zusätzlichen Signalen in der über dem Spektrum gezeigten Peptidkarte der Rechromatographie zu sehen ist (Abbildung 4-17). Die in der Ursprungs-

fraktion enthaltenen Signale lassen sich ausnahmslos in der Peptidkarte der Rechromatographie wiederfinden. Exemplarisch ist dies für einige Komponenten der Peptidbank gezeigt. Die im Spektrum mit Sternen markierten Signale wurden in der Peptidkarte mit Kreisen markiert. Dieses Beispiel zeigt, dass eine massenspektrometrische Messung nur eine begrenzte „Kapazität“ hat, Inhaltsstoffe einer Probe darzustellen. Die Kapazität einer Messung ist durch den dynamischen Bereich (Fähigkeit, niedrig konzentrierte Analyten neben hoch konzentrierten abzubilden)<sup>21</sup> und die Anzahl der Messkanäle determiniert und liegt bei MALDI-TOF-MS Messungen bei ca. 50-200 Peptiden pro Spektrum. Eine weitere Auftrennung der Probe hat eine Erhöhung des dynamischen Fensters und eine Verteilung der Peptide auf mehr Messkanäle zur Folge. Es werden folglich zusätzlich Peptide detektiert, die ohne Rechromatographie nicht sichtbar waren. Durch die zusätzliche Dimension in der Aufreinigung entstehen aber auch Probleme. So ist die Zuordnung der ursprünglichen Peptidsignale in der neu erzeugten Peptidkarte bei linear TOF-MS Messungen aufgrund von isobaren Ionen nicht trivial. Hier liefert die Aufnahme von hochgenauen Reflektrom Daten wesentlich mehr Sicherheit.

Zur Erstellung des Inventories wurden nur die Peptidsignale enthaltenden Fraktionen der rechromatographierten Peptidbank 970303 eingesetzt, d.h. ca. 40–50 % der Fraktionen besaßen nach der weiteren RP-Chromatographie bei Betrachtung der MALDI-TOF-MS Spektren keine Peptidsignale und wurden nicht bei der automatischen Identifizierung berücksichtigt. Im gezeigten Beispiel (Abbildung 4-17) zeigen nur die Fraktionen von 15 bis 60 Signale und wurden für die Erstellung des Inventories verwendet. Peptide, die nach der Rechromatographie bereits isoliert in einer Fraktion vorliegen, wurden in Einzelfällen mit Hilfe der Edman Sequenzierung identifiziert. Die restlichen Fraktionen wurden mit Hilfe der oben beschriebenen Methodik bearbeitet.

---

<sup>21</sup> Bei MALDI-TOF-MS Spektren liegt der dynamische Bereich im Idealfall bei drei Größenordnungen.

### **4.3.2 Erstellung von Fragmentspektren**

Um einen möglichst umfassenden Überblick über die Hauptkomponenten des Hemofiltrats zu gewinnen, wurden aus 4.813 Fraktionen jeweils die 4 intensivsten Signale zwischen 1.000 und 3.500 amu des MALDI-MS-Spektrums (Reflektron-Modus) mit Alcatrap extrahiert und von denen jeweils zwei verschiedene Ladungszustände fragmentiert. Wie bereits erwähnt, hat eine Ionenfalle im MS/MS-Modus keine Isotopenauflösung, und es ist deshalb nicht möglich, den Fragmentionen Ladungszustände zuzuweisen. Für die automatische Datenbanksuche bedeutet das, dass für jedes Fragmention jeder mögliche Ladungszustand angenommen werden muss, was eine unspezifische Aufweitung der Datenmatrix zur Folge hat und zu einer Herabsetzung der Sensitivität der Datenbanksuche führt. Daher wurde auf eine Fragmentierung von Peptiden mit potentiell höheren Ladungszuständen ( $MW > 3.500 \text{ mu}$ ) von vornherein verzichtet<sup>22</sup>. Die Kollisionsenergie wurde dabei in Abhängigkeit des  $m/z$ -Wertes dynamisch gesetzt (Kap. 4.2.3.2.2). Auf diese Art und Weise wurden ca. 38.000 Fragmentspektren erzeugt und ausgewertet.

### **4.3.3 Datenprozessierung und Datenbanksuche**

Nachdem durch „D2M“ Peaklisten aus den Rohdaten erzeugt und in das MASCOT™ Format überführt wurden, wurde eine Datenbanksuche mit Hilfe von MASCOT™ und MASCOT Daemon™ initialisiert.

---

<sup>22</sup> Zu Anfang der Arbeit war die Grenze bei 5.000 Da (11 Rechromatographien) bevor sie aus den beschriebenen Gründen auf 3.500 Da zurückgesetzt wurde.

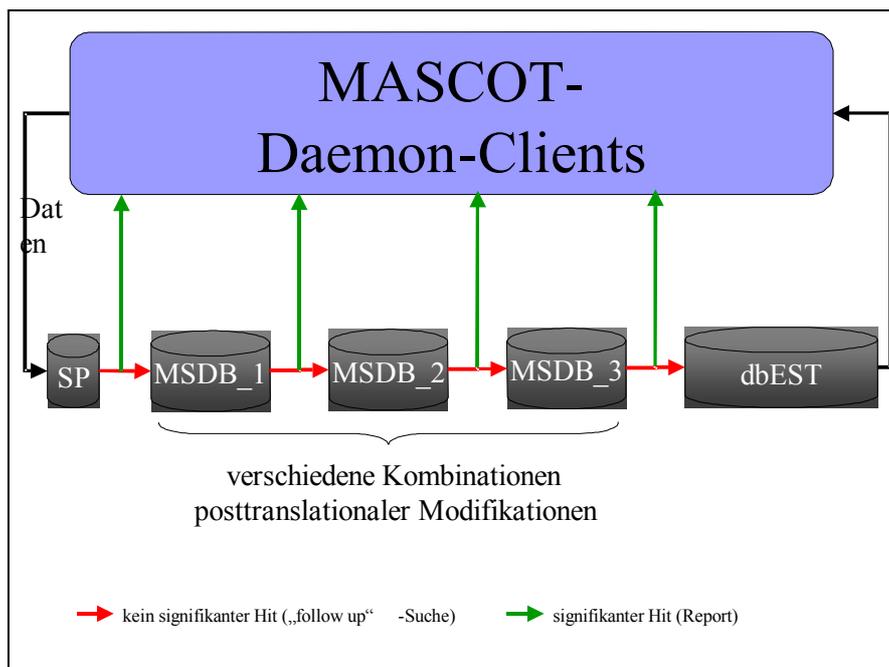


Abbildung 4-18 Suchstrategie für die automatische Datenbanksuche von MS/MS-Daten mit Hilfe der MASCOT Suchmaschine. Es wird eine kaskadierende Suche in mehreren Datenbanken mit verschiedenen möglichen posttranslationalen Modifikationen durchgeführt.

Tabelle 4-3 Übersicht über die verwendeten Datenbanken und posttranslationalen Modifikationen.

Kürzel	Datenbank	Version	Taxonomie	Anzahl der Einträge	Variable Modifikationen
SP	SwissProt	38	human	6 840	Oxidation (M)
MSDB_1	Mass Spectrometry Database	010523	human	25 456	Oxidation (M) PyroGlu (Q, E) Amidierung (C-Term) Acetylierung (N-Term)
MSDB_2	Mass Spectrometry Database	010523	human	25 456	Hydroxylierung (P) Phosphorylierung (S,T,Y)
MSDB_3	Mass Spectrometry Database	010523	alle	301 345	Oxidation (M) PyroGlu (Q, E) Amidierung (C-Term) Acetylierung (N-Term)
dBEST	EST-Database (NCBI)	010217	human	2 893 765	Oxidation (M)

Um möglichst viele der bekannten Protein- und Nukleotid-Datenbanken zu

erfassen, wurde eine kaskadierende Suchstrategie über mehrere nicht redundante Protein- und Oligonukleotid-Datenbanken gewählt, die zusätzlich noch einige mögliche posttranslationale Modifikationen mit einbezog. Um die Recherchezeit zu minimieren, wurde dabei immer von der kleinsten in Richtung zur größten Datenbank gesucht und häufig auftretende Modifikationen vor weniger häufigen behandelt. In Abbildung 4-18 ist das Suchschema wiedergegeben. Tabelle 4-3 zeigt eine Übersicht über die verwendeten Datenbanken und posttranslationalen Modifikationen.

Der „MASCOT Daemon™“ übernimmt hierbei die Automatisierung der Suche und leitet nicht erfolgreiche Suchen an die nächste Datenbanksuche weiter, bis alle Stufen der Kaskade durchlaufen sind. Erfolgreiche Suchen werden sofort beendet und in Form eines HTML-Reports abgelegt.

#### **4.3.4 Auswertung der Sequenzierung**

Es wurden insgesamt 37.248 MS/MS-Spektren erzeugt, wobei 1.755 Peptide mit Hilfe der Ionenfalle über MS/MS und Datenbanksuche identifiziert wurden. 88 Peptide wurden über Edman- Sequenzierung identifiziert. Insgesamt konnten also 1.843 kartierte Signale mit Sequenzinformationen nativer Peptide versehen werden. Nach Abzug redundanter Identifikationen ergibt sich eine Zahl von 1225 verschiedenen nativen Peptiden aus 146 Proteinvorläufern übrig.

Abbildung 4-19 zeigt eine Übersicht der 25 am häufigsten vorkommenden Peptide in Bezug auf den Vorläufer an. Dabei ist der Fibrinogen alpha und alpha-E chain precursor (FIBA\_human) mit 552 Fragmenten der am häufigsten gefundene Vorläufer. Danach folgen mit 172 und 153 Fragmenten das humane Serum Albumin und Beta-2-Mikroglobulin. Eine genaue Aufstellung aller Fragmenthäufigkeiten in Bezug auf das Vorläuferprotein gibt Tabelle 4-4.

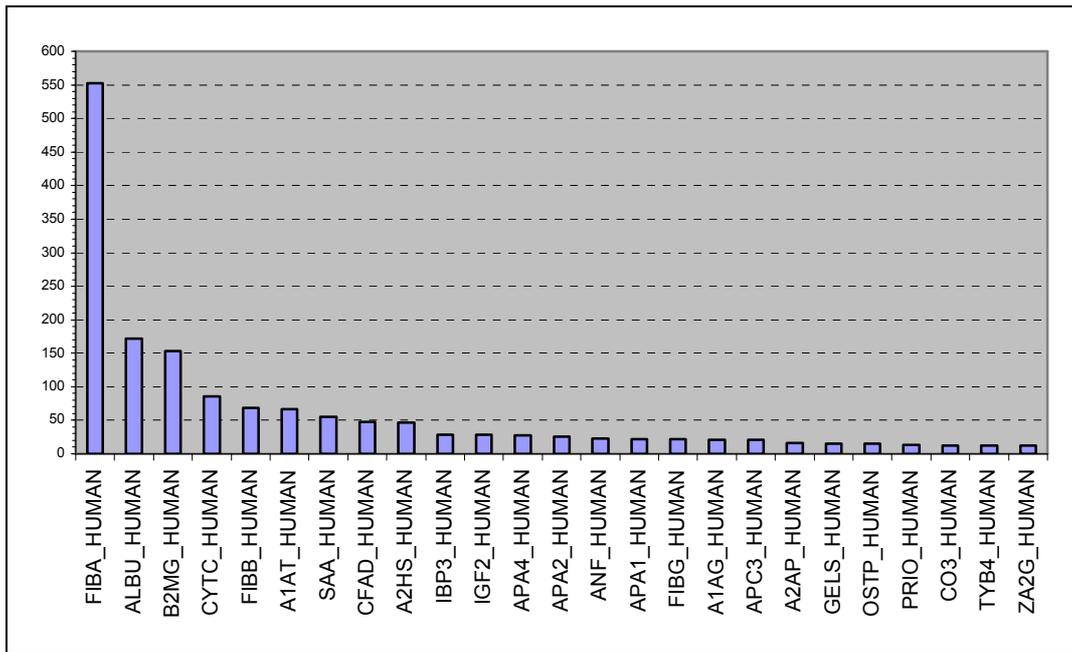


Abbildung 4-19 Fragmenthäufigkeiten der Proteinvorläufer. Erfasst sind die 25 am häufigsten identifizierten Vorläuferproteine (bezeichnet mit dem SwissProt-Kürzel).

Tabelle 4-4 Liste der identifizierten Vorläuferproteine und deren Häufigkeit. Der Proteinvorläufer wurde mit dem SwissProt-Kürzel bezeichnet.

Abkürzung Vorläufer	Häufigkeit								
FIBA_HUMAN	552	MYG_HUMAN	9	1C01_HUMAN	2	CA17_HUMAN	1	MAS2_HUMAN	1
ALBU_HUMAN	172	CLUS_HUMAN	8	A2GL_HUMAN	2	CA21_HUMAN	1	MIC2_HUMAN	1
B2MG_HUMAN	153	MGP_HUMAN	7	BD01_HUMAN	2	CATL_HUMAN	1	MLEV_HUMAN	1
CYTC_HUMAN	86	RETB_HUMAN	7	CADB_HUMAN	2	CB45_HUMAN	1	NO, O75256	1
FIBB_HUMAN	68	A1AH_HUMAN	6	DSC3_HUMAN	2	CBG_HUMAN	1	NO, Q96MG2	1
A1AT_HUMAN	67	APC1_HUMAN	6	IGFA_HUMAN	2	CBPN_HUMAN	1	NO, Q9P2N5	1
SAA_HUMAN	55	CYTB_HUMAN	6	ILBP_HUMAN	2	CFAB_HUMAN	1	OSTC_HUMAN	1
CFAD_HUMAN	48	HPT1_HUMAN	6	ISK5_HUMAN	2	CRIP_HUMAN	1	PDGA_HUMAN	1
A2HS_HUMAN	47	LEA2_HUMAN	6	MYHB_HUMAN	2	DEFN_HUMAN	1	PENK_HUMAN	1
IBP3_HUMAN	29	PGHD_HUMAN	6	PEDF_HUMAN	2	DSG2_HUMAN	1	PEPA_HUMAN	1
IGF2_HUMAN	29	AMBP_HUMAN	5	TBA1_HUMAN	2	ELAF_HUMAN	1	PEPC_HUMAN	1

Abkürzung Vorläufer	Häufigkeit	Abkürzung Vorläufer	Häufigkeit	Abkürzung Vorläufer	Häufigkeit	Abkürzung Vorläufer	Häufigkeit	Abkürzung Vorläufer	Häufigkeit.
APA4_HUMAN	28	SAA4_HUMAN	5	TIG2_HUMAN	2	FABA_HUMAN	1	PF4L_HUMAN	1
APA2_HUMAN	26	SG1_HUMAN	5	VIME_HUMAN	2	FIBH_HUMAN	1	PGBM_HUMAN	1
ANF_HUMAN	23	AHNK_HUMAN	4	1A23_HUMAN	1	FKB1_HUMAN	1	PGS1_HUMAN	1
APA1_HUMAN	22	CAG1_HUMAN	4	1B02_HUMAN	1	H12_HUMAN	1	PRPC_HUMAN	1
FIBG_HUMAN	22	CMGA_HUMAN	4	1CXX_HUMAN	1	H2AO_HUMAN	1	PTHY_HUMAN	1
A1AG_HUMAN	21	HLAE_HUMAN	4	3MG_HUMAN	1	HEPC_HUMAN	1	PTN_HUMAN	1
APC3_HUMAN	21	IBP2_HUMAN	4	A1BG_HUMAN	1	HG17_HUMAN	1	RB32_HUMAN	1
A2AP_HUMAN	16	LV2E_HUMAN	4	A4_HUMAN	1	HPTR_HUMAN	1	RL24_HUMAN	1
GELS_HUMAN	15	TYB0_HUMAN	4	AACT_HUMAN	1	IBP4_HUMAN	1	RNP_HUMAN	1
OSTP_HUMAN	15	ANGT_HUMAN	3	AMYC_HUMAN	1	IBP5_HUMAN	1	RUN2_HUMAN	1
PRIO_HUMAN	13	CD14_HUMAN	3	ANFC_HUMAN	1	IPST_HUMAN	1	S109_HUMAN	1
CO3_HUMAN	12	CO4_HUMAN	3	APE_HUMAN	1	ITH4_HUMAN	1	S5A1_HUMAN	1
TYB4_HUMAN	12	DCD_HUMAN	3	APF_HUMAN	1	K2C8_HUMAN	1	SAA3_HUMAN	1
ZA2G_HUMAN	12	FA9_HUMAN	3	BMP2_HUMAN	1	LCT2_HUMAN	1	SG2_HUMAN	1
KNG_HUMAN	10	GUAN_HUMAN	3	C04_HUMAN	1	LTBS_HUMAN	1	SODC_HUMAN	1
LAC_HUMAN	10	GUAU_HUMAN	3	C1S_HUMAN	1	LUM_HUMAN	1	SYUA_HUMAN	1
TRFE_HUMAN	10	OIF_HUMAN	3	CA11_HUMAN	1	LV2C_HUMAN	1	TP3B_HUMAN	1
TTHY_HUMAN	10	1B55_HUMAN	2	CA13_HUMAN	1	LV2D_HUMAN	1	UBIQ_HUMAN	1
								VGf_HUMAN	1

Die Ionenfalle deckt bei diesem Ansatz einen Massenbereich bis 3.500 Da ab. Größere Peptide (bis 16.000 Da), die in Hemofiltrat bekanntlich durchaus häufig vorkommen, konnten nur über die sehr zeit- und arbeitsaufwändige Edman-Sequenzierung identifiziert werden. Abbildung 4-20 zeigt die Massenverteilungen der identifizierten Peptide mit ESI-QIT-MS/MS und Edman-Abbau.

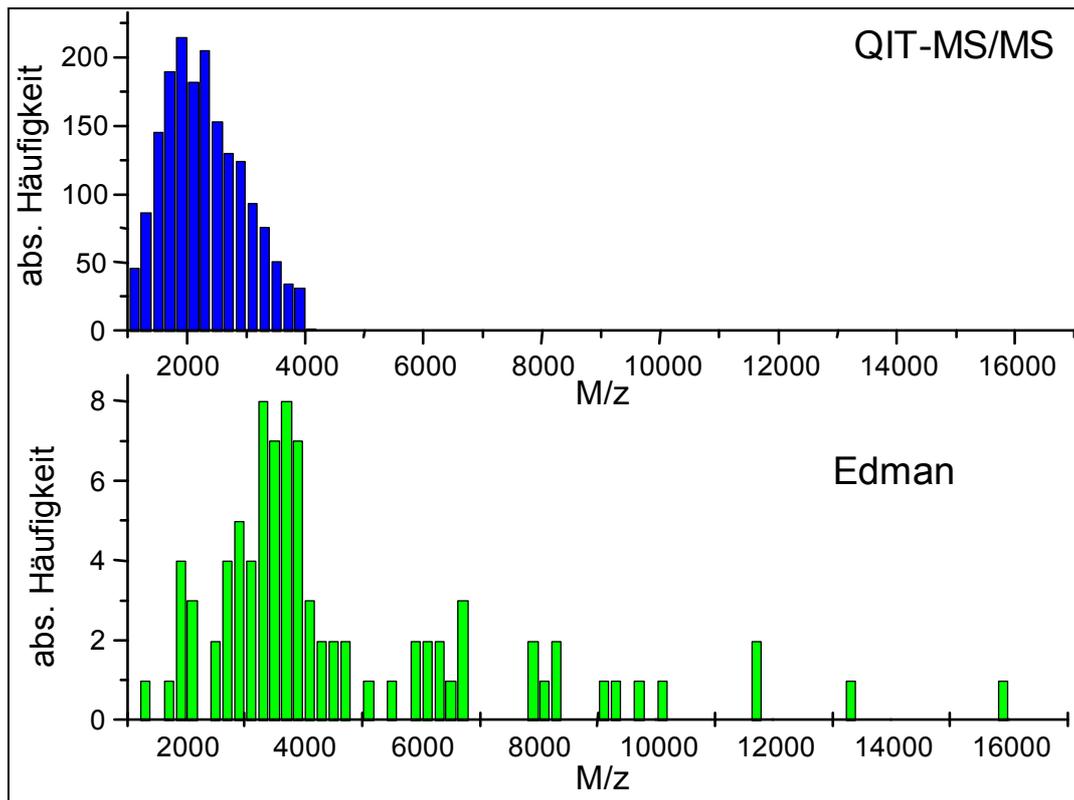


Abbildung 4-20 Massenverteilungen der identifizierten Peptide mit ESI-QIT-MS/MS und Edman-Abbau.

Der durch die Sequenzierung abgedeckte Konzentrationsbereich in Bezug auf die Vorläuferproteine ist extrem groß und reicht von  $10^{-4}$  mol/L für Albumin (s. Tabelle 4-4: ALBU\_HUMAN) bis in den picomolaren Bereich ( $10^{-12}$  mol/L) für Guanylin (s. Tabelle 4-4: GUAN\_HUMAN). Dieser Bereich von 7-8 Größenordnungen ist identisch mit den in Blut natürlich vorkommenden Konzentrationsspannen von  $10^{-4}$  bis  $10^{-12}$  mol/l (Albumin bzw. Interleukin 6, (Anderson *et al.* 2002)). Abbildung 4-21 zeigt die Plasmakonzentrationen für einige in Hemofiltrat identifizierte Proteinvorläufer in blau. In grün eingezeichnet sind beispielhaft Konzentrationen von Peptidhormonen, die für diagnostische Zwecke herangezogen werden.

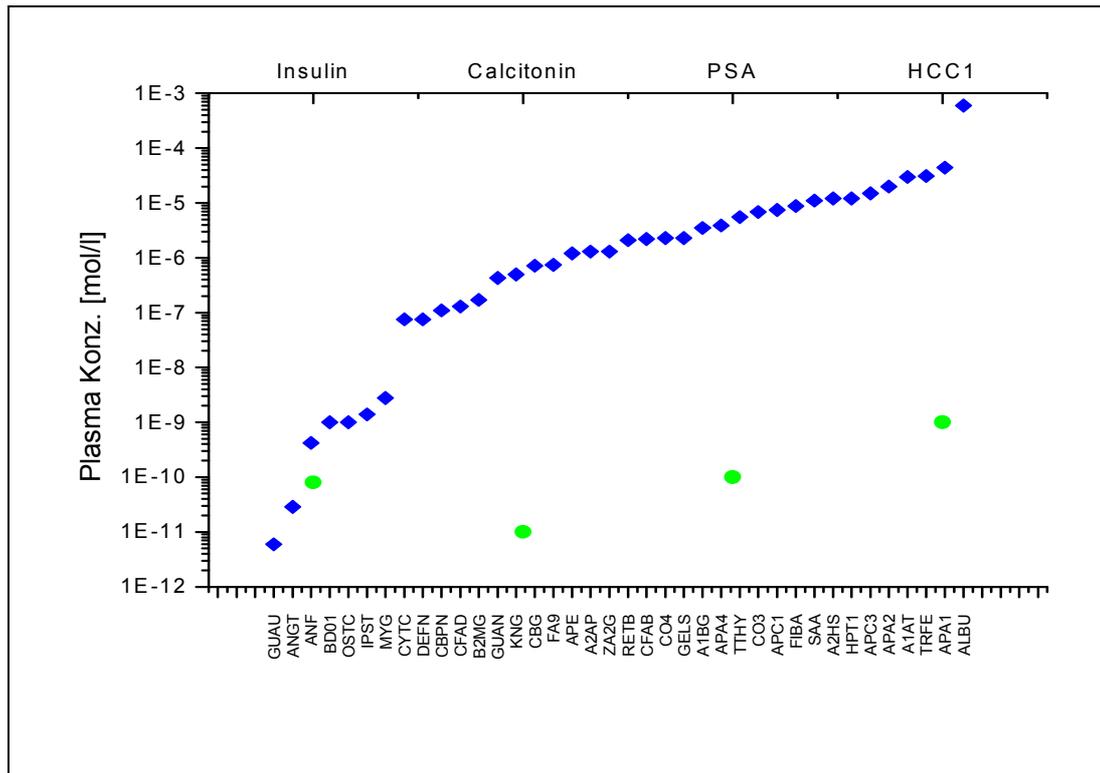


Abbildung 4-21 Konzentrationsbereich in Blutplasma in Bezug auf das Vorläuferprotein einiger ausgewählter Peptide des Inventories (blau). Konzentrationswerte einiger Peptidhormone, die zu diagnostischen Zwecken herangezogen werden (grün, nicht identifiziert). Die Konzentrationsangaben stammen aus (Anderson *et al.* 2002) und Zitaten hieraus.

#### 4.3.5 Diskussion der Ergebnisse

Bei Betrachtung der Ergebnisse kann festgestellt werden, dass das Ziel, eine Methodik zu entwickeln, die einen guten Überblick über die peptidischen Inhaltsstoffe einer biologischen Quelle gibt, erreicht wurde. Es wurde ein in dem Umfang bisher einmaliges Inhaltsverzeichnis von im Blut (zumindest von nierenkranken Patienten) vorkommenden Peptiden erzeugt. In diesem Verzeichnis ist sowohl der natürliche Konzentrationsbereich von ca. 8 Größenordnungen als auch die Diversität der Quelle „Blut“ repräsentativ, wenn auch nicht erschöpfend, abgebildet (1.225 verschiedene Peptide aus 146 Proteinvorläufern). Der Umfang dieser Studie ist vergleichbar mit den Arbeiten von Pieper (Pieper *et al.* 2003), Adkins (Adkins *et al.* 2002) und Tirumalai (Tirumalai *et al.* 2003), die sich durchweg mit dem Plasma bzw. Serum-Proteom auseinandersetzen. Bei Pieper werden 325 Proteine aus humanem

Serum identifiziert, bei Adkins sind es 490. Die beiden ersten Arbeiten betrachten ausschließlich die Proteine größer 15.000 Da und setzen zur Identifizierung nach mehrdimensionaler Auftrennung und tryptischem Verdau den Peptide Mass Fingerprint bzw. LC-MS/MS-Methoden zur Identifizierung ein. Tirumalai kommt dem in dieser Arbeit beschriebenen Ansatz am nächsten, denn er betrachtet ausschließlich das „Low molecular weight human Serum Proteome“ (Peptide/Proteine < 30.000 Da) und kann hier 340 distinkte Proteine nachweisen. Allerdings verwendet auch er einen tryptischen Verdau, bevor die Identifizierung stattfindet. Dieser Eingriff verhindert die spezifische Darstellung (Kartierung) und Identifizierung von nativ im Körper vorkommenden Peptiden und kleinen Proteinen, da jeweils nur Fragmente aus den zirkulierenden Formen nachgewiesen werden und spiegelt damit nicht die reale Situation im Körper wider. Der in der vorliegenden Arbeit beschriebene Ansatz stellt demgegenüber die natürlich vorkommenden Peptide im Blut dar und beinhaltet dabei auch die Identifizierung der Peptide, so wie sie im Organismus vorkommen. Dies bietet deutliche Vorteile bei der Findung von diagnostischen Markern, Therapeutika und bei der Aufklärung von Prozessierungswegen von Proteinen, um beispielsweise die (Neben-) Wirkungen von Medikamenten zu charakterisieren, da direkt das native Produkt betrachtet wird und nicht ein unter Umständen wirkungsloses artifizielles Fragment dieses Peptids/Proteins.

Kürzlich sind aber auch Arbeiten beschrieben worden, die sich auf native Peptide aus biologischen Quellen beziehen. So zeigen Minamino (Minamino *et al.* 2003) das Peptidom aus Schweinehirn. Hier werden ca. 10.000 Peptide detektiert, von denen über 1.000 mittlerweile identifiziert wurden. Minamino hat allerdings durch eine multidimensionale Auftrennung den Aspekt der reproduzierbaren Probenvorbereitung, der Voraussetzung für einen differentiellen Vergleich ist, in den Hintergrund gestellt. Die Gruppe um Andren beschrieb kürzlich eine Methode, die sowohl die Kartierung – in diesem Falle mit LC-MS – aufgreift und mittels einer automatisierten MS/MS-Methode Neuropeptide aus Hirngewebe identifiziert (Skold *et al.* 2002; Svensson *et al.* 2003). Auch wenn in diesen Fällen bisher wenig Peptide identifiziert wurden

(25-50 je Arbeit), kommt sie dem Peptidomicsansatz, wie er hier beschrieben worden ist, am nächsten.

Die in dieser Arbeit erzielten Ergebnisse sind sowohl in technischer Hinsicht, als auch aus biologischer Sicht bisher für den Bereich der Peptide in diesem Umfang einzigartig und bilden den Grundstein für weitere Entwicklungen in diesem Sektor. Erste Ergebnisse für menschliche Cerebrospinalflüssigkeit (CSF) liegen bereits vor (Kellmann *et al.* 2003).

Trotz der großen Zahl von 1.843 identifizierten Peptiden fällt auf, dass die Ausbeute von identifizierten Peptiden im Verhältnis zur Anzahl der durchgeführten MS/MS-Experimenten (37.248) eher gering ist (4,9%). Ein wichtiger Grund für die geringe Effizienz der Identifizierung liegt in der geringen Auflösung und der geringen Massengenauigkeit der Ionenfalle begründet. Die geringe Auflösung hat zur Folge, dass die Isotopenmuster größerer und damit auch höher geladener Peptide nicht mehr aufgelöst werden können. Eine Bestimmung des Ladungszustandes von Mutter- und Tochterionen ist in der Regel mit der Ionenfalle im MS/(MS)-Modus nicht möglich<sup>23</sup>. Eine Massendekonvolution, d.h. eine Transformation aller mehrfach geladener Ionen auf eine Massenskala mit ausschließlich einfach geladenen Ionen oder eine korrekte Annotierung des Ladungszustandes, kann hier nicht stattfinden. Die Suchmaschine muss demzufolge für jedes Signal jeden möglichen Ladungszustand annehmen. Durch die starke Erweiterung der Berechnungsmatrix verliert die Suche an Sensitivität, so dass es häufig zu nicht relevanten Suchergebnissen kommt, obwohl die Information im Spektrum bei richtiger Annotation der Ladungszustände<sup>24</sup> für eine Identifizierung ausreichen würde. Ein weiterer Grund für die niedrige Sequenzausbeute ist, dass die berechne-

---

<sup>23</sup> Mit speziellen scan Modi, wie dem sog. „zoom-scan“, können Ladungszustände bis zu 5-fach automatisch erkannt werden. Diese Scans decken im Regelfall einen Massenbereich von 10 Da ab und werden in einer Scangeschwindigkeit von 550 Da/sec. durchgeführt. Aufgrund dieser Einschränkungen können diese Modi nicht standardmäßig über den kompletten Massenbereich eingesetzt werden.

<sup>24</sup> Eine Massendekonvolution des gesamten Spektrums hat den selben Effekt.

ten Ladungszustände zwar statistisch untermauert sind, es aber keine Kontrolle über ihr Auftreten im Elektrospray-Massenspektrum gibt, d.h. es ist nicht klar, ob ein errechneter Ladungszustand im ESI-Spektrum real existiert oder nicht. Weiterhin wurde bewußt auf eine Reduktion der Cysteinbrücken vor der MS/MS-Fragmentierung verzichtet, um natürlich vorkommende, über Cysteinbrücken zusammenhängende Dimere (z.B. Insulin) als Ganzes bei der Kartierung zu erfassen. Da aber über Cysteinbrücken verknüpfte Peptide extrem schlecht fragmentieren oder überdurchschnittlich viele interne Fragmente bilden, werden diese bei der Datenbanksuche selten identifiziert (Neitz *et al.* 2001)

#### *4.3.5.1 Einfluss der Massengenauigkeit auf die Effizienz der Identifizierung von Peptiden mit Hilfe der Datenbanksuche*

Um den Einfluss der Massengenauigkeit auf das Ergebnis der Datenbanksuche zu bestimmen, werden die beiden Parameter Vorläuferionentoleranz<sup>25</sup> und Fragmentionentoleranz<sup>26</sup> untersucht. Zu diesem Zweck wurde ein „ideales“<sup>27</sup> Spektrum eines Peptids mit der Masse 3.860,66 Da und der Aminosäuresequenz *TSESSVSGSTGQWHSESGSFRPDSPGSGNARPNNPDW* mehrfach mit veränderten Toleranzen, aber sonst identischen Parametern einer Datenbanksuche mit MASCOT™ unterzogen. Es wurden der Grenzwert für die Ähnlichkeit (qplughole), der Grenzwert für die Identität und der Score des Hits (Peptide-Score) aufgezeichnet und grafisch dargestellt (Abbildung 4-22-24)

---

<sup>25</sup> In MASCOT™ wird diese Toleranz als „peptide-tolerance“ bezeichnet und in Da angegeben. Sie ist das Maß für die Massengenauigkeit des Vorläuferions.

<sup>26</sup> In MASCOT™ wird diese Toleranz als „MS/MS-tolerance“ bezeichnet und in Da angegeben. Über sie ist die Massengenauigkeit der Fragmentionen beschrieben.

<sup>27</sup> „Ideal“ meint hier, dass ein Spektrum aus theoretischen Fragmentionen künstlich erstellt wurde. Dazu wurden einige a-, b- und y-Ionen ausgewählt und mit zufälligen Intensitäten versehen. Es wurde für diesen Vergleich ein artifizielles Spektrum benutzt, da so die Vergleichbarkeit zwischen den Expe-

Abbildung 4-22 zeigt den Ähnlichkeitsgrenzwert („qplughole“) einer Suche bei unterschiedlichen Einstellungen für die Toleranz der Fragmentationen (MS/MS-Toleranz) und bei verschiedenen Toleranzen für das Vorläuferion („peptide-tolerance“). Der „qplughole“ gibt an, ob eine Ähnlichkeit zwischen dem gesuchten Spektrum und dem theoretisch generierten Spektrum aus der Datenbank besteht. Er gibt dem Auswerter eine erste Idee, um welches Aminosäuresequenz es sich handeln könnte. Diese Ergebnisse müssen im Allgemeinen aber durch manuelle Auswertung der Rohdaten verifiziert werden. Gut zu erkennen sind zwei Trends. Zum Einen ist „qplughole“ je kleiner, desto niedriger die Vorläuferionentoleranz eingestellt ist. Dies bedeutet, je größer die Massengenauigkeit der Messung für das Vorläuferion ist, desto leichter ist es, eine Homologie zwischen Spektrum und Datenbankeintrag zu entdecken. Zum Anderen hat die MS/MS-Toleranz auch einen erheblichen Einfluss auf den Schwellenwert für die Homologie. Ab einer bestimmten Grenze, die wiederum von der Vorläuferionen-Toleranz abhängig ist, fällt der Wert exponentiell auf ein Minimum ab. Dieser Abfall findet bei Werten der Vorläuferionentoleranz größer als 0,5 Da erst bei einer Fragmentationentoleranz von 0,25 Da statt. Bei Vorläuferionentoleranzen von 0,1 Da oder 0,05 Da liegen diese Barrieren mit 1,0 Da bzw. 1,25 Da wesentlich höher. Dies bedeutet wiederum, dass eine hohe Vorläuferionenmassengenauigkeit tolerabel gegenüber Abweichungen der Fragmentationen ist.

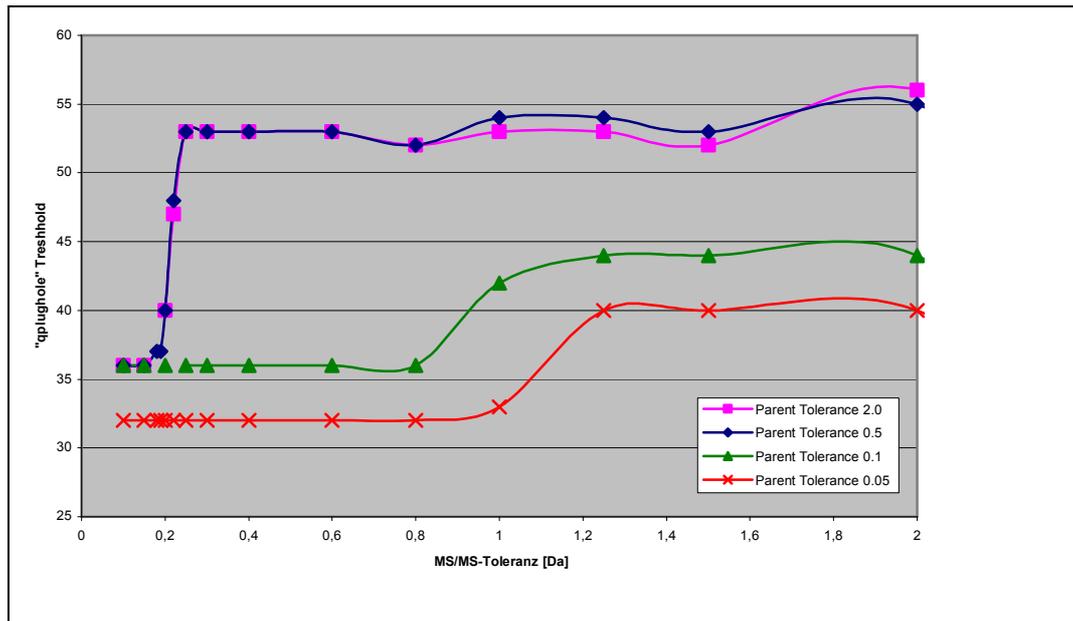


Abbildung 4-22 Darstellung des Homologie-Grenzwertes (qplughole) in Abhängigkeit der Vorläuferionen-Toleranz („Parent Tolerance“) und der Fragmentionen-Toleranz („MS/MS-Toleranz“).

Der zweite zu beachtende Wert bei der MASCOT-Datenbanksuche ist der Identitäts-Grenzwert („Identity-Threshold“). Er gibt an, ob ein gefundenes Ergebnis mit 5% Wahrscheinlichkeit ein Zufallsereignis ist, oder nicht. Dieser Wert kann über Gleichung 4.1 ermittelt werden:

$$\text{Identity Threshold} = -10 \log_{10}((C) * 1/\text{DB-Peptide}) \quad \text{Gleichung 4-1}$$

C ist hierbei die Wahrscheinlichkeit für einen Zufallshit (im Falle von 5% ist der Wert 0,05). DB-Peptide ist die Anzahl der zu durchsuchenden Datenbankeinträge nach Filterung der Taxonomie und der Vorläuferionenmasse. Er ist folglich nur abhängig von der Anzahl der zu durchsuchenden Datenbankeinträge<sup>28</sup>. Dies ist in Abbildung 4.23 dargestellt. Hier ist der Grenzwert für die Identität bei der kleineren Datenbank (SwissProt (human): 75.780 Einträge; NCBIInr: 1.112.467 Einträge) systematisch niedriger. Zusätzlich kann die Anzahl der „scorenden“ DB-Einträge aber noch verringert werden, indem die Peptid-Toleranz, also der Massenfehler des Vorläuferions, verkleinert

<sup>28</sup> Üblicherweise wird der Wert für die Wahrscheinlichkeit, einen Zufallshit zu erhalten, bei 0,05 konstant gehalten.

wird. Dieser Effekt kann besonders drastisch unterhalb von 0,5 Da Massengenauigkeit beobachtet werden, wo der Grenzwert exponentiell abfällt.

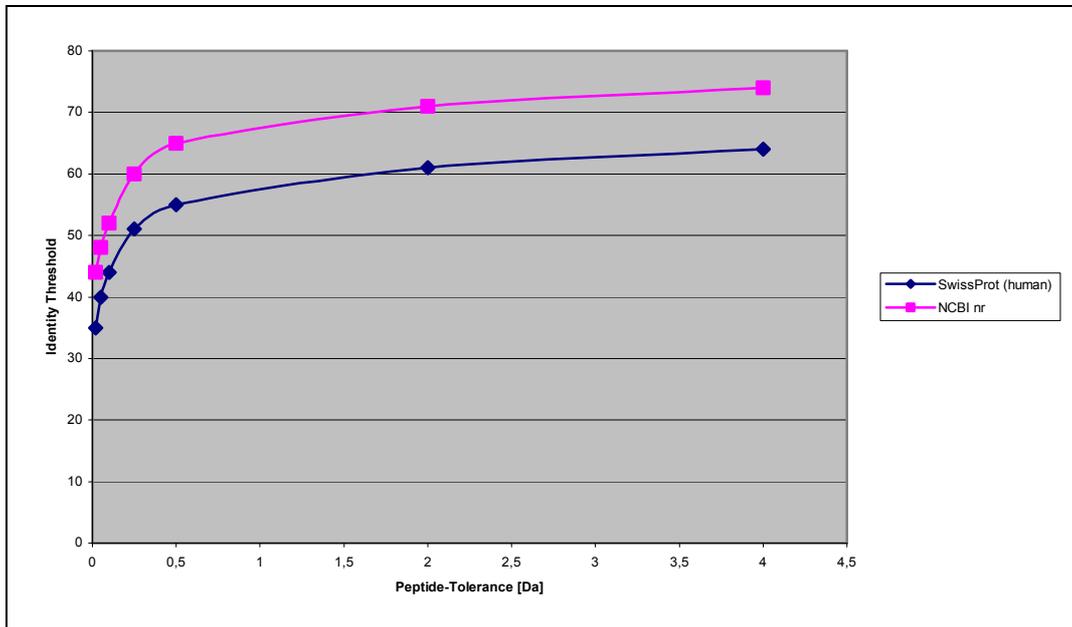


Abbildung 4-23 Abhängigkeit des „Identity-Score“ von der Datenbankgröße und von der Vorläuferionen-Toleranz („Peptide Tolerance“). Als Datenbanken haben hier SwissProt und NCBI nr gedient.

Die dritte Stellschraube bei der Datenbanksuche ist der Score selbst. Je höher dieser für einen Datenbankhit ist, desto unwahrscheinlicher ist es, ein Zufallsergebnis erhalten zu haben. Wie Abbildung 4-24 zeigt, ist der Score von der MS/MS-Toleranz abhängig, von der Peptid-Toleranz jedoch unabhängig. Je genauer die Massengenauigkeit der Fragmentationen angegeben werden kann, desto höher ist der Score einer MASCOT Suche.

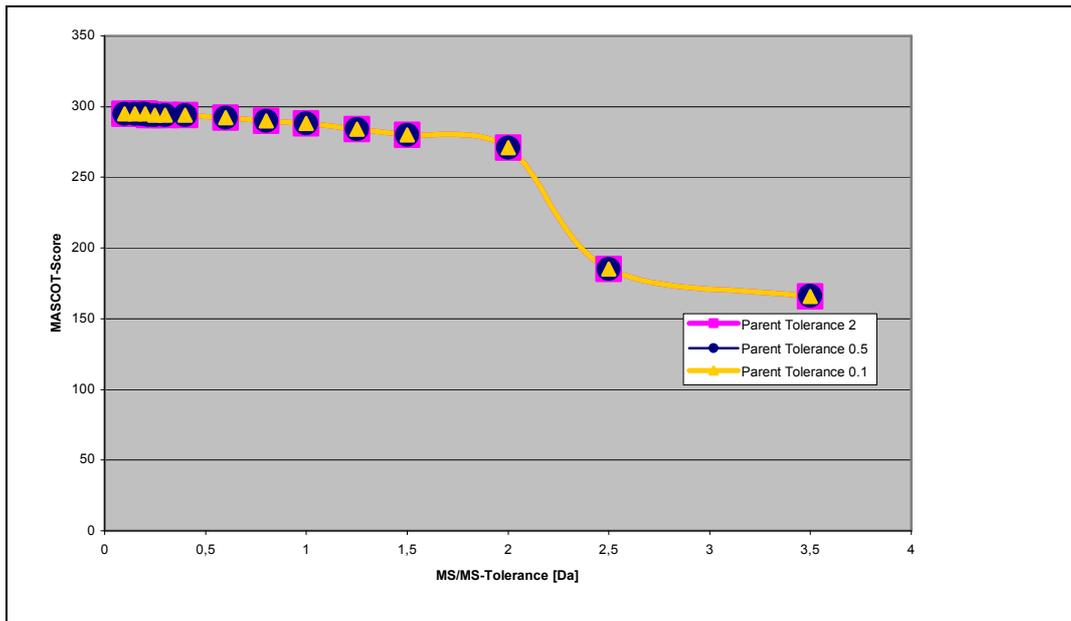


Abbildung 4-24 Abhängigkeit des Mascot Scores von der „Peptide-Tolerance“ und von der „MS/MS-Tolerance“ für ein Beispielpeptid in der SwissProt Datenbank.

Zusammenfassend lässt sich sagen, dass die Massengenauigkeit, mit der die Suche parametrisiert wird, einen erheblichen Einfluss auf das Ergebnis hat. So werden durch eine große Massengenauigkeit der Vorläufer- und Fragmentationen sowohl die Grenzwerte für Homologie und Identität gesenkt, wie auch gleichzeitig der Score erhöht.

Für die Ionenfalle kann eine Peptid-Toleranz von 0,8 Da und eine MS/MS- (Fragmentationen) Toleranz von 0,8 Da für die Suche angenommen werden. Die Suche verläuft bei Daten, die ausschließlich mit QIT-MS erzeugt wurden, folglich nicht unter optimalen Bedingungen, und es gehen Identifizierungen verloren. Durch die Übernahme der wesentlich genaueren Vorläufermassen aus Reflektrom MALDI-MS Messungen durch Alcatrap (vgl. Kap. 4.2.3.1) konnte die Peptid-Toleranz für die Suchen auf 0,1 Da reduziert werden, so dass der Ähnlichkeits-Grenzwert bei einer Tochterionen-Toleranz von 0,8 Da im optimalen Bereich liegt (vgl. grüner Graph in Abbildung 4-22) und der Identitätsgrenzwert in einen Bereich mit deutlich niedrigeren Werten fällt (vgl. Abbildung 4-23). Der Score selbst bleibt von dieser Maßnahme unbeeinflusst, hier ist noch Potential zur Verbesserung durch eine bessere Fragmentationengenauigkeit gegeben.

Ein weiteres Gebiet, welchem beim Inventarisieren von Peptidquellen eine größere Rolle zufällt, ist das *de novo*-Sequenzieren im Falle einer fehlgeschlagenen Datenbanksuche. Eine Datenbanksuche kann trotz qualitativ ausreichendem Spektrum fehlschlagen, wenn entweder kein Datenbankeintrag vorhanden ist, oder wenn z.B. durch einen Aminosäureaustausch oder posttranslationale Modifikation(en) ein Peptid mit abweichender Sequenz bzw. Masse in der Datenbank verzeichnet ist. In einigen Fällen wurde im Laufe dieser Arbeit versucht, eine teilweise oder komplette *de novo*-Sequenzierung von Peptiden anhand der mit der Ionenfalle erzeugten Fragmentspektren durchzuführen, falls die Datenbanksuche allein nicht erfolgreich war. Dies war, jeweils bedingt durch eine ungenügende Massengenauigkeit und geringe Auflösung nicht möglich. Eine *de-novo*-Sequenzierung ist insgesamt erforderlich, da unter Umständen gerade diese Peptide von großer biologischer Bedeutung sein können, da Veränderungen in der Aminosäuresequenz die Ursache für Krankheiten sein können. Ein Beispiel hierfür liefert z.B. ein Polymorphismus von Faktor XIII, der in Zusammenhang mit einem erhöhten Risiko eines Schlaganfalls steht (Tammen *et al.* 2003; Tammen *et al.* 2004).

Eine weitere Limitierung der vorgestellten Technologie besteht auch darin, dass größere Peptide ausschließlich mit der Edman-Sequenzierung identifiziert werden konnten. Dies ist ein zwar bewährtes, für die Menge an zu identifizierenden Stoffen zu aufwändiges und zudem nicht sehr sensitives Verfahren, da die zu sequenzierenden Peptide durch weitere Chromatographieschritte aufgereinigt werden müssen. Eine (schnelle und sensitive) massenspektrometrische Methode zur Identifikation von Peptiden bis ca. 10 kDa ist auch hier wünschenswert. Dies ist aus den schon genannten limitierenden Gründen (Massengenauigkeit und Auflösungsvermögen) mit einer elektrischen Ionenfalle nicht möglich, kann aber durch den Einsatz von MS-Geräten mit größerer Leistungsstärke erreicht werden.

## 4.4 Erweiterung des Ansatzes auf Quadrupol-time-of-flight Analysatoren

Eine Möglichkeit, die beschriebenen Limitierungen bezüglich der Auflösung und Massengenauigkeit zu umgehen, ist der Wechsel auf ein Quadrupol-Time-of-Flight-Gerät (Q-TOF). Bei Q-TOF-Geräten handelt es sich um Hybridgeräte, die eine Auflösung von  $R=10.000$  (FWHM) erreichen (s. Kap. 2.1.2.2.3). Zudem ist die Massengenauigkeit von  $< 20$  ppm bei externer Kalibration und  $< 5$  ppm bei interner Kalibration sehr viel besser als bei der Ionenfalle. In den letzten Jahren haben sich diese Geräte soweit entwickelt, dass auch sie in einem automatisierten Routinebetrieb einsetzbar sind. Die Integration von Q-TOF-Geräten in die Strategie zur Erstellung von Peptide-Inventories ist daher ein logischer und notwendiger Schritt.

### 4.4.1 Anpassung der Kollisionsenergie

Um die für die Ionenfalle beschriebene Methodik auf QTOF-Massenspektrometer anzuwenden, ist wiederum, die Kollisionsenergie in Abhängigkeit des  $m/z$ -Verhältnisses abzuschätzen, um eine möglichst optimale Fragmentierung eines jeden Peptids zu gewährleisten. Um die Abhängigkeit der Kollisionsenergie von der Größe und dem Ladungszustand der Peptide zu bestimmen, wurden 44 Signale aus verschiedenen Elektrospray-Spektren im  $m/z$ -Bereich von 400 – 1200 Da mit Ladungszuständen von 2+ bis 6+ zufällig ausgewählt. Bei diesen wurde im MS/MS-Modus die Kollisionsenergie kontinuierlich erhöht, bis der Basispeak auf ca. 20% der ursprünglichen Intensität zurückgegangen war. Die Werte für die Kollisionsenergie wurden nach  $m/z$  und Ladungszustand getrennt gegen die Kollisionsenergie aufgetragen (Abbildung 4-25).

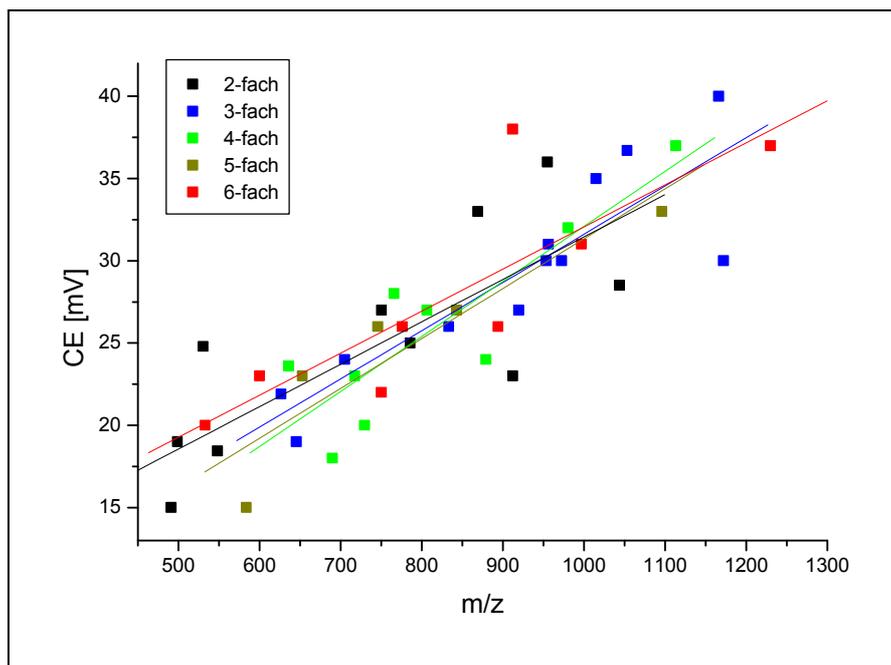


Abbildung 4-25 Darstellung der Kollisionsenergien (CE) in Abhängigkeit der  $m/z$ -Werte und des Ladungszustandes. Die Regressionsgeraden sind in der jeweiligen Farbe des Datensatzes markiert.

Anhand der Regressionsgeraden ist keine systematische Abhängigkeit der Kollisionsenergie vom Ladungszustand erkennbar, wohl aber eine Abhängigkeit der Kollisionsenergie vom  $m/z$ -Verhältnis. Wie erwartet, wird mit zunehmendem  $m/z$ -Wert - wie auch schon bei der Ionenfalle beobachtet - eine zunehmende Kollisionsenergie zur Fragmentierung nötig. Außerdem ist zu erkennen, dass die Kollisionsenergien für alle Ladungszustände über einen großen Bereich gestreut sind. Diese Streuung wird durch die Standardabweichungen beschrieben, die mit Werten zwischen 15% und 28% der Kollisionsenergie als hoch einzustufen sind (Tabelle 4-5).

Tabelle 4-5 Lineare Regressionsanalyse der Kollisionsenergien für die Ladungszustände  $2^+$  bis  $6^+$  und für die Gesamtheit der Datenpunkte.

$y = B \cdot x + A$	2fach	3fach	4fach	5fach	6fach	gesamt
A	7,12	2,29	1,37	1,01	6,50	3,13
B	0,024	0,029	0,033	0,030	0,026	0,028
R	0,77	0,88	0,86	0,92	0,84	0,84

$y = B \cdot x + A$	2fach	3fach	4fach	5fach	6fach	gesamt
SD	4,44	3,04	3,18	2,94	3,98	3,67
N	10	12	9	5	8	44

Da die Abhängigkeit vom Ladungszustand vernachlässigt werden kann, wurden alle Messungen zusammengefasst und wiederum eine lineare Regressionsanalyse durchgeführt (Tabelle 4-5, gesamt). Um der mit größer werdendem m/z-Wert steigenden Kollisionsenergie, sowie der großen Streubreite gerecht zu werden, wurde der Massenbereich von 400 m/z Da bis 1.400 m/z Da in hunderter Schritte unterteilt und für jeden dieser Massenbereiche eine Rampenmethode erstellt, bei der pro m/z-Bereich drei Kollisionsenergien (CE1 bis CE3) repetitiv durchlaufen werden. Hierzu wurde die Geradengleichung aus Tabelle 4-5 auf die jeweils mittleren m/z-Werte eines jeden Bereichs angewendet (450, 550, 650 etc.) und dieser Wert für den gesamten, 100 Da umfassenden Bereich als Kollisionsenergie (CE2) gesetzt. CE1 und CE3 kommen durch Subtraktion bzw. Addition von jeweils 4 mV des Wertes von CE2 zustande. Es entsteht das in Abbildung 4-26 gezeigte „Treppenbild“. In Tabelle 4-6 sind die Werte für die Kollisionsenergie tabellarisch erfasst.

In Abbildung 4-26 ist der Bereich dargestellt, der sich bei Anwendung der Rampenmethode abdecken lässt (zwischen der roten und der blauen Treppe). Es ist zu erkennen, dass mehr als 80% der Peptide sich im Bereich zufriedenstellender Fragmentierung befinden. Die übrigen liegen teilweise sehr nah an den Grenzlinien, so dass auch hier von einer guten Fragmentierung ausgegangen werden kann.

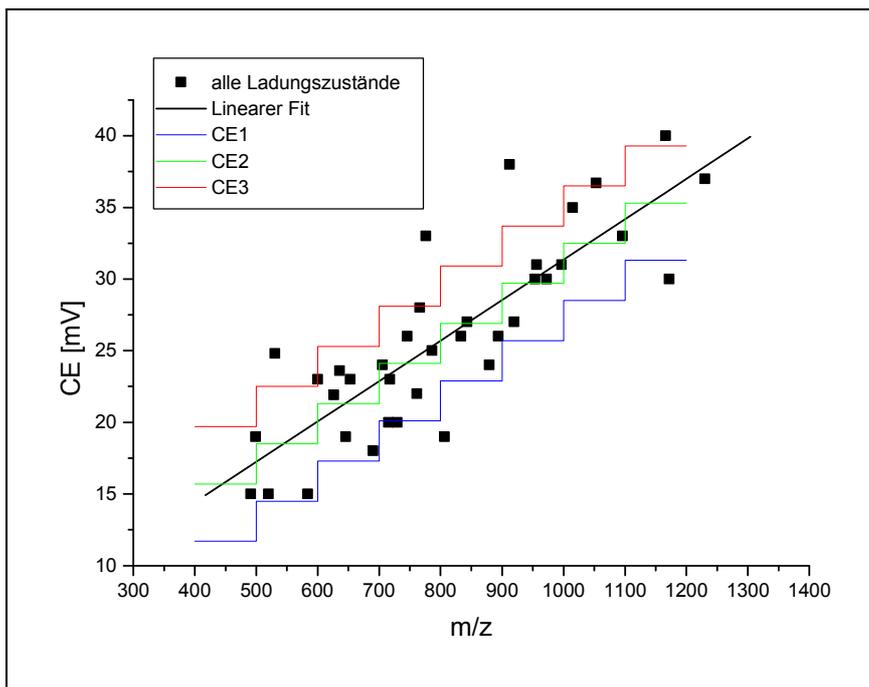


Abbildung 4-26 Darstellung der Kollisionsenergien (CE) aller Ladungszustände in Abhängigkeit der m/z-Wertes und deren lineare Regression (schwarz). Die blauen, grünen und roten Graphen beschreiben die tatsächlich in der Methode angewendeten Werte für die CE.

Tabelle 4-6 Zuordnung der Kollisionsenergien zu den m/z-Bereichen für die Rampenmethode.

m/z-Bereich	CE 1	CE 2	CE 3	M/z-Bereich	CE 1	CE 2	CE 3
400-499	11.7	15.7	19.7	800-899	22.9	26.9	30.9
500-599	14.5	18.5	22.5	900-999	25.7	29.7	33.7
600-699	17.3	21.3	25.3	1.000-1.099	28.5	32.5	36.5
700-799	20.1	24.1	28.1	1.100-1.200	31.3	35.3	39.3

#### 4.4.2 Identifizierung großer Peptide und kleiner Proteine

Durch die Nutzung eines Q-ToF-MS-Gerätes sollte der Massenbereich der zu identifizierenden Peptide zumindest signifikant erhöht werden, da hier eine Auflösung von ca. 10.000 (FWHM) bei einer Massengenauigkeit von 5-20 ppm erreicht wird. Die obere Grenze liegt hier bei ca. 8.500 bis 9.000 Da, da

hier die Isotopenauflösung gerade noch ausreichend ist, um diese Signale zu dekonvolvieren bzw. ihnen automatisch einen Ladungszustand zuzuweisen. Als theoretisches Beispiel sei hier das Fibrinogen alpha Fragment 361 bis 440 genannt, welches als neunfach geladenes Ion bei einer Auflösung von 10.000 (FWHM) gezeigt ist. Das Fragment hat ein  $m/z$ -Wert von 942,44 Da und wird durch Anwendung verschiedener Dekonvolutionsalgorithmen (MaxEnt3, MASCOT Distiller, BioAnalyst) auf die richtige Molekularmasse von 8472,87 Da projiziert.

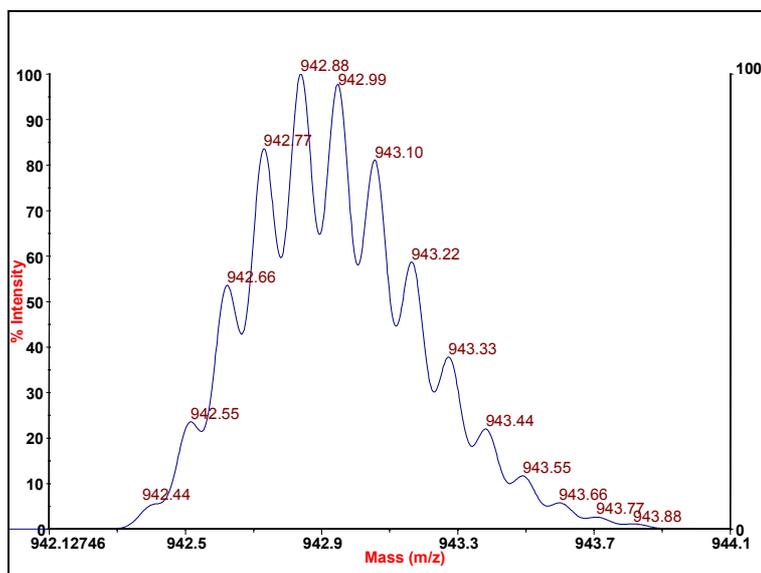


Abbildung 4-27 Darstellung der Auflösung für ein Fibrinogen A Fragment (361-440) bei einer Auflösung von 10.000 (FWHM). Dargestellt ist das  $[MH_9]^{9+}$  Ion.

Um in der Praxis eine Methode zu etablieren, die eine automatische Abarbeitung dieser großen Peptide gewährleistet, ergeben sich allerdings zwei Probleme. Einerseits liegen die monoisotopischen Signale mit ihrer Intensität, wie schon in Abbildung 4-27 zu erkennen ist, in einem Bereich mit niedrigem Signal zu Rausch/Verhältnis, was die Bestimmung des monoisotopischen Signals erschwert. Zum Anderen besitzen große Peptide häufig Cysteinbrücken, die eine Fragmentierung erschweren oder gar verhindern. Im ersten Fall liefern Algorithmen, die die theoretischen Isotopenpattern in das Peak Picking einbeziehen, die besten Ergebnisse. Hier sind die Maximum Entropie Algorithmen der MassLynx Software (Waters) und BioAnalyst-Software (Sciex),

sowie Maximum Likelihood Algorithmen (z.B. Mascot Distiller) verwendet worden, mit denen auch monoisotopische Peaks erkannt werden, die sich nur knapp über dem Rauschlevel befinden. Wenig intensive Peaks, bei denen der monoisotopische Peak vollständig im Rauschen verschwunden sind, werden aber häufig auch hiermit nicht erkannt. Hier verspricht dann nur eine manuelle Bestimmung der monoisotopischen Masse oder die Aufweitung der Massentoleranzen bei der Suche, die mit Kompromissen im Bereich der Sensitivität und Selektivität einher geht, Aussicht auf Erfolg. Das Problem der Cysteinbrücken ist nur durch eine Reduktion und Carboxyamidomethylierung (CAM) der Cysteine zu lösen. Obwohl hier ein Kompromiss bezüglich der Messung von nativen Peptiden zu schließen ist, muss dieser Prozessierungsschritt bei der Identifizierung großer Peptide getan werden, um einen Großteil der Peptide überhaupt zu erfassen. Insbesondere entziehen sich Hetero- und Homodimere, die über Cysteinbrücken miteinander verknüpft sind, der Analytik.

In einem ersten Ansatz wurden nach CAM mehrerer Peptidbankfraktionen Peptide mit einem Molekulargewicht von 5.000 – 8.600 Da automatisch in einem Q-Tof-Gerät fragmentiert, die Daten prozessiert und einer Datenbanksuche unterzogen (Kellmann *et al.* 2001) . Alle 5 untersuchten Peptide konnten eindeutig identifiziert werden, wobei die Massengenauigkeit nach Dekonvolutierung für Vorläufer- und Fragmentationen jeweils unter 50 ppm lagen. Es wurden die in Tabelle 4-7 aufgeführten Peptide identifiziert.

Tabelle 4-7 Automatisiert über ESI-QTOF-MS/MS identifizierte Peptide aus der Peptidbank 970303 mit einem Molekulargewicht > 5.000 Da.

<b>Vorläuferion</b>	<b>Ladungszustand</b>	<b>MH<sup>+</sup></b>	<b>Protein Vorläufer</b>
<b>[m/z]</b>		<b>[Da]</b>	
1.004,1	6+	6.019,41	Complement Faktor D
904,0	9+	8.128,18	Complement Faktor D
929,1	6+	5.569,72	Fibrinogen Alpha
1.071,1	8+	8.561,62	Ubiquitin
1.004,5	8+	8.028,84	Clara Cell Phospholipid Binding Protein

Am Beispiel des Ubiquitins sei hier gezeigt, dass die Kombination aus ESI-QTOF-MS/MS und hoch entwickelten Software Algorithmen es direkt ermöglicht, zuverlässig große Peptide aus einer biologischen Probe zu identifizieren.

Zunächst wurde ein Übersichtsscan (survey scan) bei hoher Auflösung aufgenommen, der den Inhalt der Probe wiedergibt. In diesem Beispiel handelt es sich um eine recht komplexe Probe, die sowohl kleinere (z.B. 4<sup>+</sup>) Ionen als auch eine größere Komponente enthält. Ionen mit Ladungszuständen größer als 4<sup>+</sup> werden als Vorläuferion für das folgende MS/MS-Experiment ausgewählt (im 952,6<sup>9+</sup>). Im Fragmentspektrum sind verschiedene Ionenserien mit Ladungszuständen von 2<sup>+</sup> bis 8<sup>+</sup> zu erkennen. Die direkte Datenbanksuche ohne weitere Präprozessierung ergibt kein signifikantes Suchergebnis. Werden die Daten vor der Datenbanksuche allerdings dekonvolutiert, ergibt sich ein Hit, der einen Score von 246<sup>29</sup> aufweist. Die Sequenzabdeckung für b- und y-Ionen ist mit 91% auch sehr gut, es können problemlos Sequence-Tags von bis zu 14 Aminosäuren Länge aus dem Spektrum abgelesen werden.

<sup>29</sup> Dies entspricht nach Gleichung 4.1 einer Wahrscheinlichkeit von  $1 : 3 \cdot 10^{19}$  kein Zufallshit zu sein.

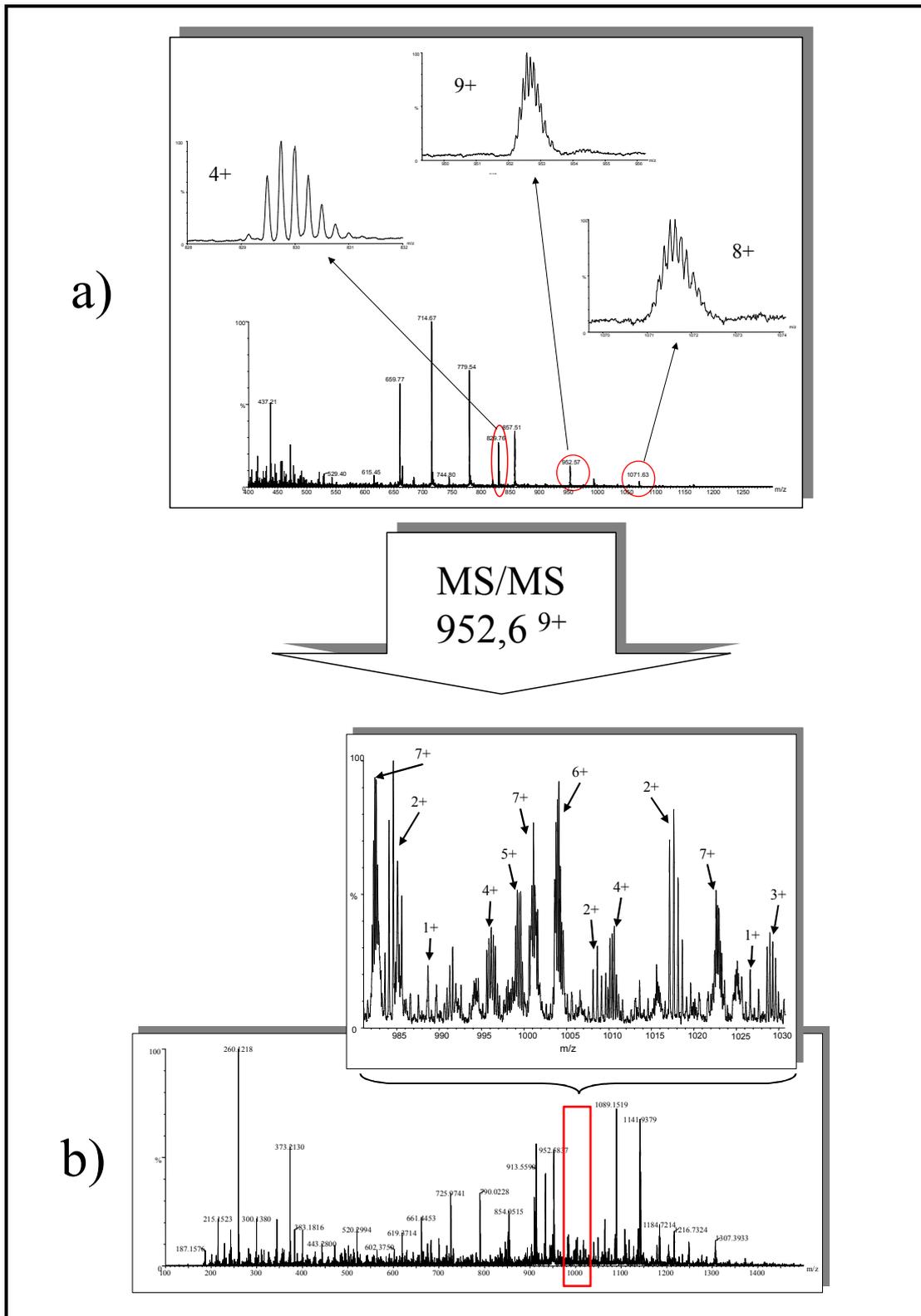


Abbildung 4-28 Direkte Identifizierung großer Peptide mit Hilfe von ESI-MS/MS. a) Spektrum einer biologischen Probe im MS-Modus. Die gezoomten Regionen zeigen das achtfach und neunfach geladene Ion des Ubiquitins, sowie ein unbekanntes vierfach geladenes Ion. b) zeigt das MS/MS-Spektrum des neunfach geladenen Ions des Ubiquitins. Der Zoom zeigt eine Region des Spektrums mit Fragmentionen der Ladungszustände 1+ bis 7+.

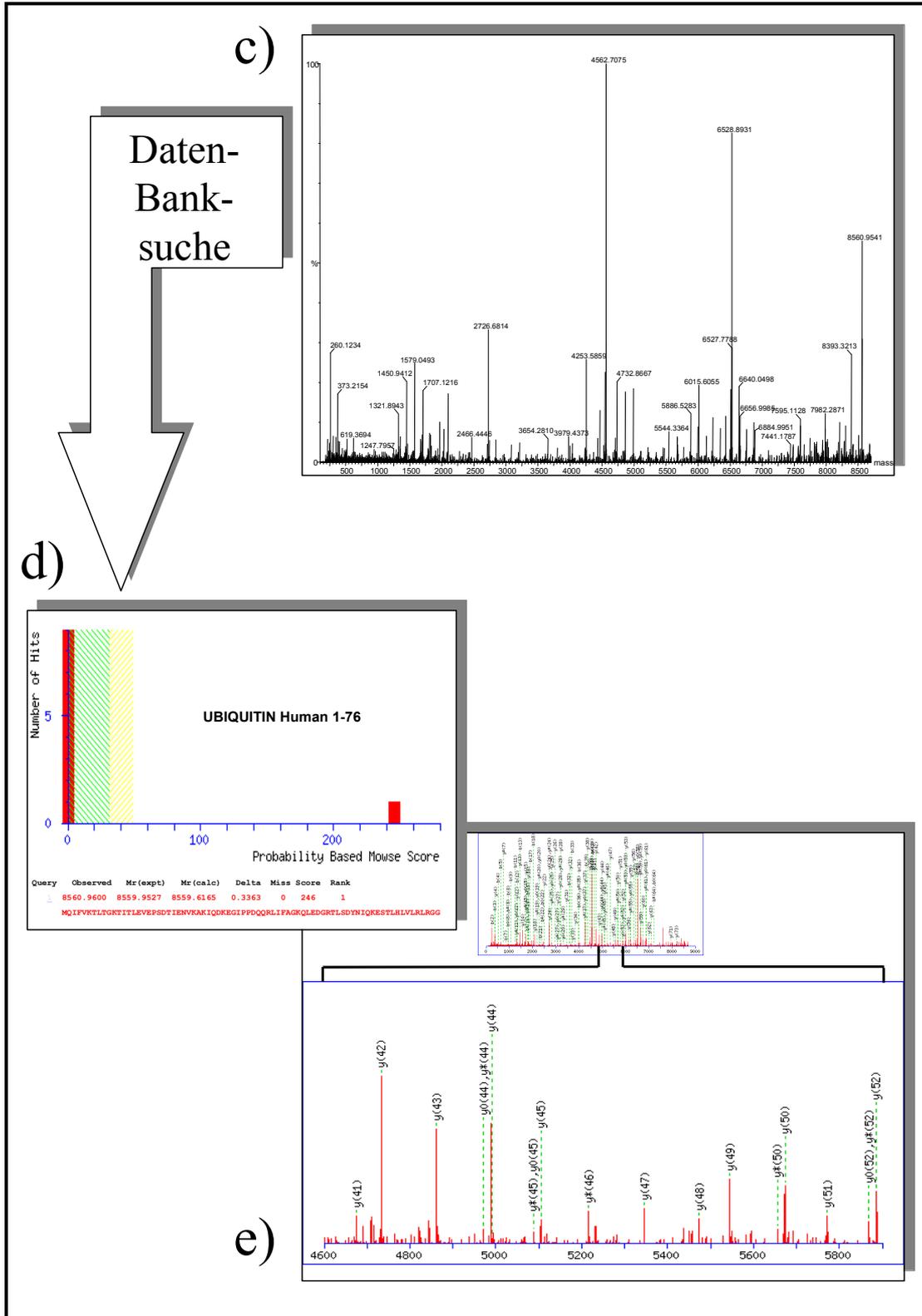


Abbildung 4-29 c) zeigt das dekonvolutierte und deisotopierte Spektrum aus b), welches zur Datenbanksuche mit Mascot verwendet wird. d) zeigt den Datenbankhit der MASCOT-Suche und e) die Sequenzabdeckung des Fragmentspektrums für Ubiquitin (human) 1-76.

Diese grundlegenden Experimente wurden bereits soweit weiterentwickelt, dass es mittlerweile bei BioVision zur Routine gehört, Peptide bis ca. 9 kDa direkt aus Fraktionen von unterschiedlichen biologischen Proben (Plasma, Urin, etc.) ohne weitere Vorreinigung mit Hilfe der Massenspektrometrie zu identifizieren.

Zeitgleich mit der Erstellung dieser Arbeit gab es Entwicklungen in der Proteomforschung, die noch einen Schritt weitergehen, nämlich die Identifizierung ganzer Proteine mit Hilfe der Fourier-Transform-Massenspektrometrie (FT-MS) (Ge *et al.* 2002). Diese Geräte liefern Auflösungen von 100.000 und mehr und können mit gasfreien Fragmentierungstechniken wie Electron Capture Dissociation (ECD) oder Infrared Multiphoton Dissociation (IRMPD) betrieben werden. Die bisherigen Veröffentlichungen zeigen, dass diese als „Top-Down“-Identifizierung bezeichnete Technik bisher nicht oder nur in sehr beschränktem Maße automatisiert werden konnte. Hier liegt die Ursache in den bisher schwer zu handhabenden FT-MS-Geräten. Neuste Entwicklungen von Hybrid-FT-MS-Instrumenten wie das QFT-7 (Ionspec, USA), APEX-Q (Bruker, Deutschland) oder LTQ-FT (ThermoElectron, Deutschland) sind soweit fortgeschritten, dass sie während eines LC-MS Laufes datenabhängig MS/MS-Spektren bei großer Auflösung ( $> 100.000$ ) aufzeichnen können. Dies ist eine wichtige Voraussetzung, um in naher Zukunft große Peptide und Proteine automatisch direkt über MS/MS in Verbindung mit Datenbanksuchen zu identifizieren. Durch Entwicklungen wie der W-Optik von Micro-mass/Waters, mit denen Massenauflösungen von ca. 17.000 (FWHM) an QTOF-Systemen erreicht werden, wird der Massenbereich für die direkte Identifizierung in Peptidomics Anwendungen auf ca. 16 kDa bis 18 kDa erweitert werden können. Dies entspricht dem Massenbereich, der mit der MALDI-MS-Kartierung erreicht wird und ist deshalb für die hier beschriebene Peptidomics Strategie ausreichend.

## **4.5 Anwendung der erweiterten Strategie auf Hämofiltrat als Proof-of-Concept**

### **4.5.1 Versuchsdesign**

Um den erweiterten Ansatz mit dem in Kapitel 4.3 beschriebenen vergleichen zu können, wurden exemplarisch 16 rechromatographierte Fraktionen (Pool 6 Fraktion 26 aus 970303) mit dem QTOF bearbeitet. Es wurde ebenfalls eine Anordnung mit einer Flussinjektionsmethode (FIA) gewählt. Der Fluss betrug 250 nl/min, so dass bei einer Injektion von 5 µl 20 Minuten für MS/MS-Experimente pro Fraktion zur Verfügung standen. Die Methode zur Fragmentierung wurde so gewählt, dass zunächst ein Übersichtsscan (survey scan) im MS-Modus aufgenommen und hieraus das intensivste Ion zur Fragmentierung ausgewählt wird. Jedes Vorläuferion wird für 2 Minuten rollierend mit den in Kapitel 4.4.1 bestimmten Kollisionsenergien fragmentiert. Nach dieser Zeit wird wiederum ein MS-Spektrum aufgenommen und das nächst intensivere Signal wird für eine Fragmentierung ausgewählt. Dieser Zyklus setzt sich 20 min fort bis die nächste Fraktion bearbeitet wird. Es werden folglich die 10 intensivsten Signale pro Fraktion fragmentiert. Diese Abhängigkeit des MS/MS-Experiments vom vorausgehenden Übersichtsscan wird als DDE (data dependent experiment) oder DDA (data dependent acquisition) bezeichnet. Damit das Massenspektrometer in den MS/MS-Modus wechseln kann, müssen bestimmte Kriterien erfüllt sein, die der Benutzer frei wählen kann. Für das hier durchgeführte Experiment muss das Vorläuferion eine Mindestintensität von 20 counts/s und einen Ladungszustand zwischen 2<sup>+</sup> und 6<sup>+</sup> aufweisen. Sollten diese Kriterien nicht erfüllt werden, wird kein Fragmentspektrum aufgezeichnet, sondern das Gerät bleibt im MS-Modus. Dies ist nicht relevant bei den hier verwendeten FI-Analysen, aber grundlegend bei LC-MS/MS Kopplungen, da hier die Zeit der Elution als weitere Dimension hinzukommt.

In dem hier beschriebenen Beispiel wurde nach optischer Begutachtung der

Peptidkarte jede zweite Fraktion der Fraktionen von 30 bis 60 vermessen, so dass 160 Fragmentspektren erzeugt wurden. Nach der Messung wurden die Rohdaten dekonvolviert und mittels MASCOT einer Datenbanksuche in SwissProt und MSDB zugeführt. Die Massentoleranzen bei der Suche betragen für das Vorläuferion und für die Fragmentationen 0,1 Da.

#### 4.5.2 Ergebnisse

Nach Auswertung der Datenbanksuchergebnisse ergab sich, dass 114 der 160 Fragmentspektren zu eindeutigen Ergebnissen führten. Dies entspricht einer Erfolgsquote von ca. 70 %. Es wurden 53 verschiedene Peptidsequenzen aus 30 verschiedenen Proteinvorläufern identifiziert. Zusätzlich wurden drei Methionin Oxidationen identifiziert und lokalisiert. Die gefundenen Peptide weisen dabei einen Massenbereich von 1.018 Da bis 5.248 Da auf. Tabelle 4-8 gibt einen Überblick über die mit diesem Ansatz identifizierten Peptide.

Tabelle 4-8 Übersicht der mit dem QTOF identifizierten Peptide aus Pool 6 Fraktion 26 (970303). In blau hervorgehoben sind die mit dem LCQ ebenfalls identifizierten Peptide. Zusätzlich angegeben sind die Scores der Datenbanksuche in SwissProt.

Nr.	AA-Sequenz	Kürzel	MW <sub>mono</sub> [Da]	Score QTOF	Score LCQ	Multi- plizität
1	SGTHYWRL	HEMO_HUMAN	1018.50	27/ +/-		2
2	YMWHLVGH	SCTM_HUMAN	1041.49	30/ +/-		1
3	WVQRALHF	CYTN_HUMAN	1055.57	32/ +/-		2
4	FISPQQRW	MGP_HUMAN	1060.55	35/ +/-		1
5	TFLKKYLY	ALBU_HUMAN	1074.61	28/ +/-		2
6	TARHWFQAGY	LYOX_HUMAN	1235.58	38/ +/+		1
7	KFGERAFKAW	ALBU_HUMAN	1238.65	36/ +/-		2
8	SVEEEGVRRALDF	CYTC_HUMAN	1505.75	31/ +/-		4
9	SWYQQHPGKAPKL	LV2E_HUMAN	1538.80	34/ +/+		2
10	LAWYQQKPGKAPKL	KV1F_HUMAN	1626.92	30/ +/+		1
11	YHLIQNWPHYRSP	PGRP_HUMAN	1709.84	43/ +/+		1
12	VVRARKQIVAGVNYF	CYTC_HUMAN	1718.99	30/ +/+		3
13	FVSWYQQHPGKAPKL	LV2K_HUMAN	1784.91	34/ +/+		5
14	YVSWYQQHPGKAPKL	LV2E_HUMAN	1800.93	48/ +/-		11
15	YRDWATYKQFGSRL	FCN2_HUMAN	1846.91	38/ +/+		1
16	SSKITHRIHWESASLL	CO3_HUMAN	1864.00	47/ +/+		1
17	ARENIQRFFGHGAEDSL	SAA_HUMAN	1945.94	67/ +/+		1
18	FSSPHGKDLLFKDSAHEG	TRFE_HUMAN	1988.97	47/ +/+		2
19	SSKITHRIHWESASLLR	CO3_HUMAN	2020.10	78/ +/+		2
20	IILYNRHGGRQGQIIY	GELS_HUMAN	2063.11	51/ +/+		2
21	FFKGEFVWVSKHWKWDRE	HEMO_HUMAN	2125.05	63/ +/+		2
22	VSGWGRNANFKFTDHLKY	HPT_HUMAN	2139.06	55/ +/+		4
23	LQFFRYNSKDRKSQPMGL	ZA2G_HUMAN	2214.15	42/ +/-		1
24	LQFFRYNSKDRKSQPM <sub>0</sub> GL	ZA2G_HUMAN	2230.13	40/ +/-		2
25	SRALQVVRARKQIVAGVNYF	CYTC_HUMAN	2274.29	31/ +/-		2
26	LAAPPGHQLHRAHYDLRHTFM	A2HS_HUMAN	2467.24	41/ +/+	27/ +/-	2
27	LKNGERIEKVEHSDLSFSKDW	B2MG_HUMAN	2516.27	52/ +/+	28/ +/-	4
28	LAAPPGHQLHRAHYDLRHTFMG	A2HS_HUMAN	2524.27	61/ +/+	45/ +/+	3
29	DAHKSEVAHRFKDLGEENFKAL	ALBU_HUMAN	2540.28	31/ +/+		1
30	LAAPPGHQLHRAHYDLRHTFMGV	A2HS_HUMAN	2623.33	59/ +/+		4

Nr.	AA-Sequenz	Kürzel	MW <sub>mono</sub> [Da]	Score QTOF	Score LCQ	Multi- plizität
31	FKIDNVKARVQVAGKKYFIDF	KNG_HUMAN	2712.55	65/ +/+	38/ +/+	1
32	YIGSDKYFHARGNYDAAKRGPGGVW	SAA_HUMAN	2784.35	48/ +/+	35/ +/+	3
33	YMWHLVGHQRNRRQVTLVSGAEP	SCTM_HUMAN	2820.39	66/ +/+		2
34	PSRGKSSSYKQFTSSTSYNRGDSTF	FIBA_HUMAN	2860.32	43/ +/+	28/ +/-	3
35	LHLQVLSTRNENKLLPKPHLVRQ	DSG2_HUMAN	2869.65	62/ +/+	44/ +/+	1
36	VQTSRLAKSKMLTPEHVFIHPGWKL	C1S_HUMAN	2902.60	41/ +/+	29/ +/-	1
37	WKEKKKGSYYDSFKLQTKFQVLK	CAG1_HUMAN	3007.62	64/ +/+	45/ +/-	2
38	YSMKTTMKIIPFNRLTIGEGQQHHLG	FIBG_HUMAN	3127.64	53/ +/+	39/ +/-	1
39	RKNPKFMETVAEKALQEYRKKHREE	CLUS_HUMAN	3144.66	54/ +/+	32/ +/-	3
40	RKNPKFM <sub>ox</sub> ETVAEKALQEYRKKHREE	CLUS_HUMAN	3160.66	98/ +/+		1
41	IILYNRYHGGRGQQIYNWQGAQSTQD	GELS_HUMAN	3178.57	78/ +/+	43/ +/+	1
42	VSETESRGSSEGITNTKESSSHHPGIAEF	FIBA_HUMAN	3206.48	117/ +/+		1
43	GQPHGGGWGQPHGGGWGQGGGTHSQWNKPSKP	PRIO_HUMAN	3217.47	84/ +/+	56/ +/-	1
44	SRKNPKFMETVAEKALQEYRKKHREE	CLUS_HUMAN	3231.69	96/ +/+		3
45	SRKNPKFM <sub>ox</sub> ETVAEKALQEYRKKHREE	CLUS_HUMAN	3247.69	66/ +/+	43/ +/+	2
46	YSMKTTMKIIPFNRLTIGEGQQHHLGGA	FIBG_HUMAN	3255.70	95/ +/+	55/ +/+	1
47	YSMKTTMKIIPFNRLTIGEGQQHHLGGAK	FIBG_HUMAN	3383.80	50/ +/+	33/ +/-	2
48	GQPHGGGWGQPHGGGWGQGGGTHSQWNKPSKPKT	PRIO_HUMAN	3446.62	103/ +/+	46/ +/+	2
49	GGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQ	PRIO_HUMAN	3456.48	134/ +/+	65/ +/+	1
50	SVSGSTGQWHSESGSFRPDSPGSGNARPNPDW	FIBA_HUMAN	3456.51	129/ +/+		1
51	SSVSGSTGQWHSESGSFRPDSPGSGNARPNPDW	FIBA_HUMAN	3543.54	103/ +/+		2
52	YSMKTTMKIIPFNRLTIGEGQQHHLGGAKQAG	FIBG_HUMAN	3639.91	86/ +/+		1
53	FAVGEYNKASNDMYHSRALQVVRARKQIVAGVNY	CYTC_HUMAN	3853.98	50/ +/+		2
54	TSESSVSGSTGQWHSESGSFRPDSPGSGNARPNPDW	FIBA_HUMAN	3860.82	56/ +/+		1
55	WKGDKVKAHYGGFTVQNEANKYQISVNKYRGTAGNALMDG	FIBB_HUMAN	4415.18	117/ +/+		2
56	TWKPSSGPGSTGWSNWSGSGTGTGNQNPSPRPGSTG TWNPGSSERGSAGHW	FIBA_HUMAN	5242.31	115/ +/+		1
						114

In blau gezeigt sind die Peptide, die sowohl mit dem LCQ (17 Peptide aus 13 Proteinvorläufern; eine Methionin Oxidation), als auch mit dem QTOF-Ansatz identifiziert wurden. Herauszustellen ist, dass sämtliche mit dem LCQ-Ansatz gefundenen Peptide auch mit dem QTOF-Ansatz eindeutig identifiziert werden konnten. Zusätzlich in der Tabelle erfasst sind die Multiplizitäten der einzelnen Sequenzen, die angeben, wie häufig das Peptid im QTOF-Ansatz identifiziert werden konnte. Dies schließt sowohl die Identifizierung über mehrere Ladungszustände, als auch die Identifizierung in mehreren Fraktionen der Rechromatographie innerhalb der Präparation ein. Weiterhin sind die MASCOT-Scores gezeigt, mit denen die Peptide in der Datenbank gefunden worden sind. Besonders interessant sind in diesem Zusammenhang die Scores der in beiden Ansätzen gefundenen Peptide (blau), da hier der direkte Vergleich zwischen den Methoden gezeigt werden kann. In Abbildung 4-30 sind die MASCOT Scores der in beiden Ansätzen gefundenen Peptide grafisch dargestellt. Es ist zu erkennen, dass die Scores der QTOF-Messungen durchweg höher sind. Da die Scores nur individuelle Werte wiedergeben, die relativ zum Homologie-/Identitätsgrenzwert zu sehen sind, müssen diese Grenzwerte mit in die Bewertung einzubezogen werden. Hierzu sind alle Scores kategorisiert worden, d.h. Hits, die im Homologiebereich der

MASCOT-Suche liegen, wurden mit einem „+/-“ gekennzeichnet und mussten manuell anhand der Rohdaten verifiziert werden. Hits, die laut MASCOT eine Wahrscheinlichkeit <5% haben, kein Zufallsergebnis zu sein (Identitätsbereich), wurden mit „+/+“ beurteilt und werden ohne Verifizierung in die PeT-rA-Datenbank aufgenommen. 45 der 56 QTOF-Hits liegen im Identitätsbereich, darunter alle der mit beiden Ansätzen identifizierten. Dagegen waren bei der Suche mit LCQ-Daten nur 9 von 18 Hits im Identitätsbereich.

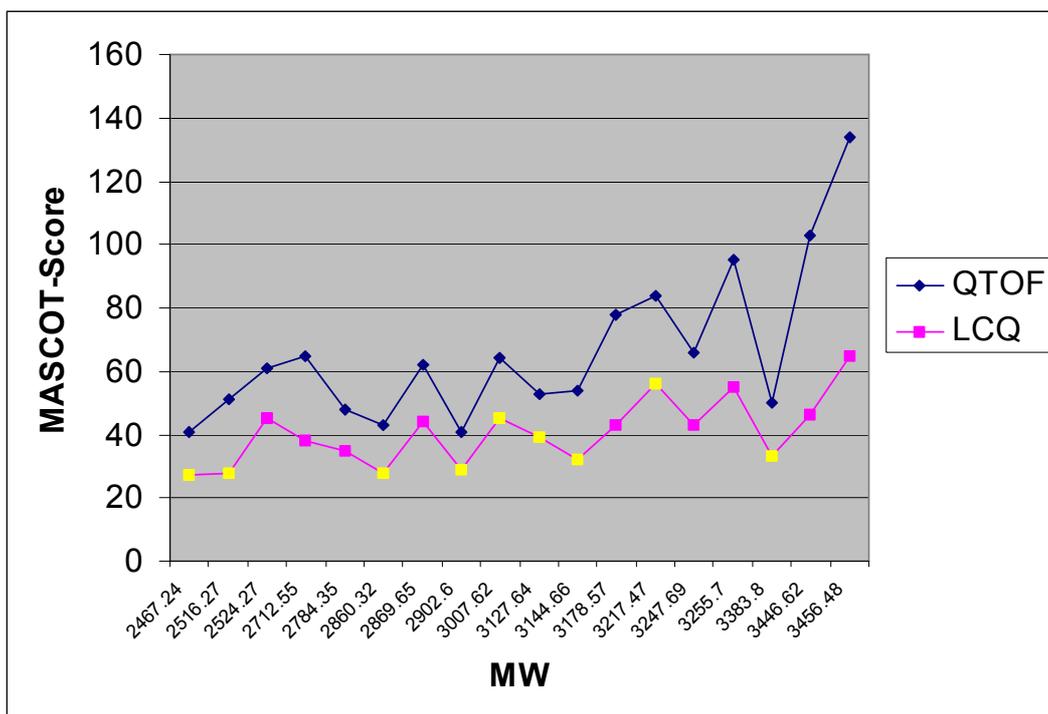


Abbildung 4-30 MASCOT-Scores der 18 mit beiden Ansätzen identifizierten Peptide. In blau, die QTOF-Daten, in pink die LCQ-Daten. Gelb markiert sind Hits, die im Homologiebereich der MASCOT Suche liegen.

#### 4.5.2.1 Diskussion

Wie erwartet, kann in diesem Ansatz durch die QTOF-Messung und die verfeinerte Datenprozessierung eine Verbesserung der Identifizierung nativer Peptide eines Peptidoms erreicht werden.

Absolut wird eine Verdreifachung der identifizierten Peptide erzielt. Zudem ergibt sich über die Erhöhung der Scores und das Absenken der Grenzwerte

eine erheblich höhere Sicherheit der Ergebnisse im Vergleich zu den LCQ Daten. Für dieses Beispiel bedeutete das, dass die 9 Hits, die im LCQ-Ansatz im Homologiebereich (+/-) der MASCOT-Suche lagen, nun im Identitätsbereich zu finden sind. Dies ist die Bestätigung der schon vorher diskutierten Einflüsse von Auflösung und Massengenauigkeit. Gerade im Bereich der kleinen und großen Peptide ist ein möglichst niedriger Threshold ausschlaggebend für eine erfolgreiche Identifizierung des Peptids. Im Falle der kleinen Peptide ist der maximale Score klein, da er von der Größe der Sequenzabdeckung und damit auch von der absoluten Länge des Peptids abhängt, so dass hier niedrige Signifikanzgrenzwerte von Nöten sind, um das Peptid überhaupt signifikant zu scoren. Im Bereich der größeren Peptide (ab 5.000 Da) können vergleichbare Effekte eine Rolle spielen. Hier ist häufig eine Fragmentierung zu beobachten, die nur Teile der Aminosäuresequenz abdeckt, was sich später in einem niedrigen Score niederschlägt. Auch hier sind niedrige Grenzen für die Homologie bzw. Identität wünschenswert. Dass überhaupt Peptide über 3.500 Da identifiziert werden können, hängt mit der hohen Auflösung des QTOF-Systems und der damit verbundenen Möglichkeit zur Massen-Dekonvolution zusammen. Einen zusätzlichen Beitrag zur Verbesserung des Ergebnisses leistet auch die um eine Größenordnung bessere absolute Sensitivität des QTOFs gegenüber dem LCQ Classic, da hiermit auch qualitativ ausreichende Spektren von niedrig konzentriert vorliegenden Peptiden erzeugt werden können.

Abschließend soll noch einmal auf die Erstellung eines Peptid-Inventories eingegangen werden, bei der der Identifizierung sicherlich die Schlüsselrolle zufällt. Abbildung 4-31 zeigt das MALDI-MS-Spektrum der Fraktion 26 aus Pool 6 der Peptidbank 970303. Die zu identifizierten Peptiden gehörenden Signale sind in grün und mit der Indexnummer aus Tabelle 4-8 markiert. Die mit dem LCQ und dem QTOF-Ansatz identifizierten Peptide sind zusätzlich blau hinterlegt. Fehlende Indices bedeuten das Fehlen eines Signals in der MALDI-MS Messung für ein via ESI-MS/MS identifiziertes Peptid. Das Fehlen ist einerseits durch unterschiedliche Ionisierungsmethoden bei der Kartierung und Identifizierung zu erklären, kann aber auch mit der zusätzlich gewonne-

nen Sensitivität durch die Rechromatographie zusammenhängen. Nicht markierte Peaks sind entweder noch nicht fragmentiert worden oder sind, trotz Fragmentierung, nicht über eine Datenbanksuche identifiziert worden. Dieses kann entweder durch ein qualitativ schlechtes Fragmentspektrum oder auch durch das nicht Vorhandensein der Peptidsequenz in der Datenbank, z.B. durch Austausch von Aminosäuren oder spezieller Modifikationen, verursacht werden. Abhilfe kann hier die Suche in anderen bzw. größeren Datenbanken oder auch chemische Modifikation der Probe (z.B. Carbamidomethylierung von Cys-Brücken; Deglycosylierung) schaffen. Durch repetitive Messung der gleichen Fraktionen unter Verwendung von Ausschlusslisten oder durch Messung aller, und nicht, wie durchgeführt, jeder zweiten Fraktion, könnten sicherlich fast alle Signale des Spektrums erfasst werden.

Zusammenfassend kann gesagt werden, dass die Kombination aus Rechromatographie und Identifizierung mit der QTOF-Methode nahezu alle kartierten Signale der ursprünglichen Peptidbank einer Sequenzierung zugänglich macht. Durch den Einsatz dieser Methode ist ein erheblicher Informationsgewinn bei der Erstellung eines Inventories zu erwarten, denn es können sowohl kleine als auch große Peptide bis 8 kDa mit einer schnellen, automatisierten, sensitiven und robusten massenspektrometrischen Methode identifiziert werden.

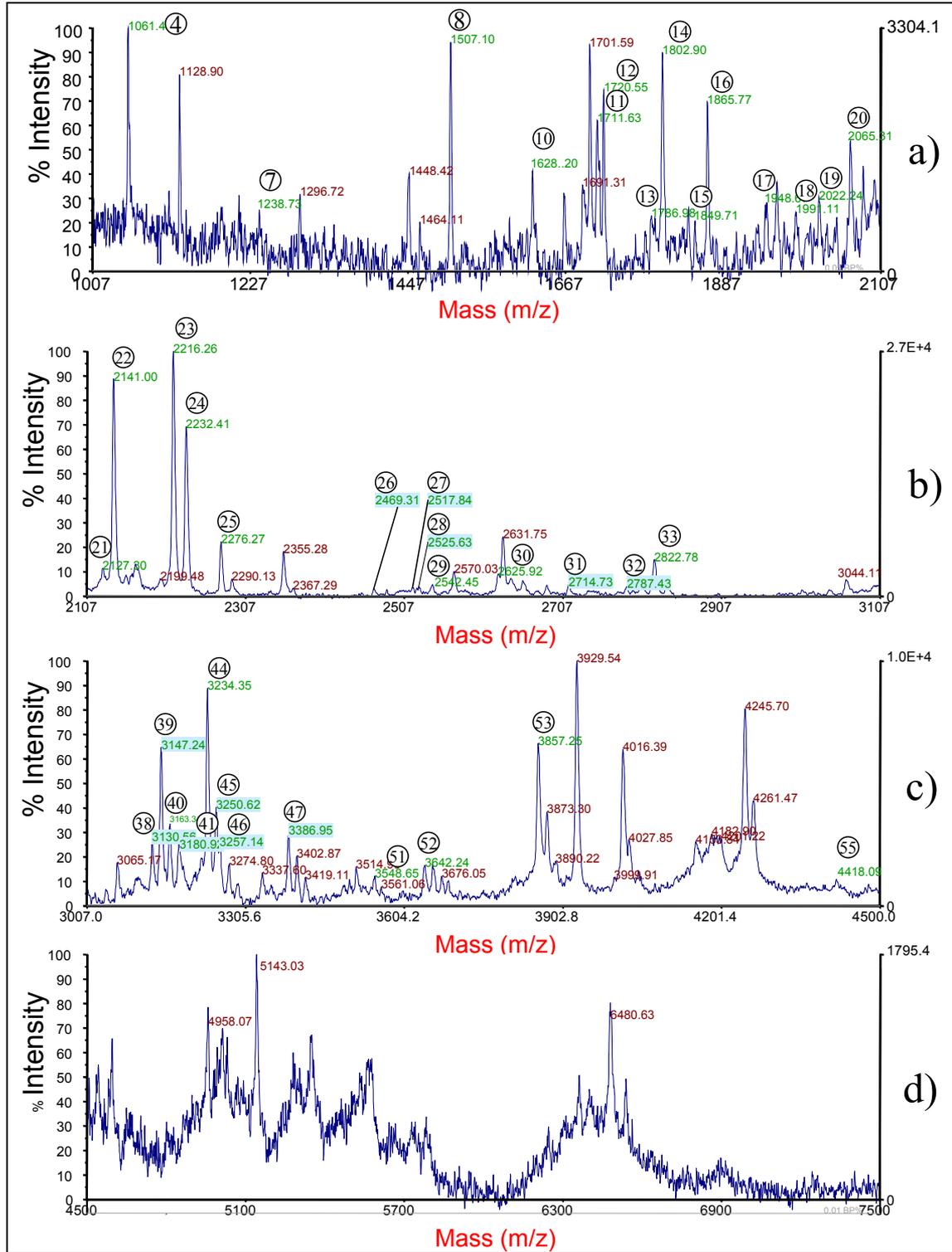


Abbildung 4-31 Annotiertes MALDI-Spektrum der Fraktion 26 aus Pool 6 der Peptidbank 970303.

## 5 Zusammenfassung

Im Rahmen der Arbeit wurde eine Methodik zur umfassenden und automatischen Identifizierung von nativen Peptiden aus Extrakten komplexer biologischer Quellen entwickelt und am Beispiel des humanen Hämofiltrats angewendet.

Die Firma BioVisioN führt seit der Gründung Forschung auf dem Gebiet der Peptide und kleinen Proteine mit dem Ziel, diagnostisch und therapeutisch relevante Substanzen zu finden, durch. Die Analyse sogenannter Peptidome, der qualitativen und quantitativen Beschreibung aller nativen Peptide und kleinen Proteinen bis ca. 15 kDa, ist als Analogon zur Proteom-Analytik zu verstehen, welche sich umfassend mit Proteinen beschäftigt. Quantitativ wird ein Peptidom über die Kombination chromatographischer Trennung und Fraktionierung mit der massenspektrometrischen „Inventarisierung“ der einzelnen Peptidspezies sowie den relativen Konzentrationen in der Probe erfasst, welches an anderer Stelle ausführlich beschrieben ist. Das Ergebnis einer solchen Peptidomanalyse besteht aus einer Peptidkarte, die die Peptide der Probe repräsentativ abbildet. Die Qualität eines Peptidoms wird über die Massen der Peptide, deren Aminosäuresequenzen und Elutionsverhalten und weitere biologische Informationen beschrieben. Zu diesem Zweck wird der Begriff des *Peptide-Inventory* eingeführt. Ein Inventory besteht aus allen verfügbaren Informationen nativer Peptide einer Quelle.

Zu Beginn der Arbeit standen mit der reproduzierbaren Erstellung von Peptidbanken und deren systematischer Kartierung über MALDI-MS bereits zwei Werkzeuge zur Verfügung, um Daten eines Inventories zu erheben. Lediglich für die Erzeugung der Sequenzdaten wurde mit dem Edman-Abbau zwar eine etablierte, aber viel zu langsame und wenig sensitive Methode eingesetzt. Hier wurde der Ersatz durch eine schnelle und leicht zu automatisierenden Methode angestrebt. Die massenspektrometrische Sequenzierung über MS/MS mit anschließender Datenbanksuche ist ein aus der Proteom-Analytik durchaus bekanntes Verfahren, musste aber für die Analyse der na-

tiven Peptide in einigen Bereichen modifiziert werden.

Obwohl die Massenspektrometrie Peptide direkt aus komplexen Mischungen heraus identifizieren kann, stellte sich schnell heraus, dass die vorliegende biologische Beispielprobe (Peptidbank „970303“ aus humanem Blutfiltrat) zu komplex war, um sie direkt für die geplante Sequenzierung einzusetzen. Es wurde eine erneute Auftrennung und Feinkartierung der Peptidbank notwendig. Jede dritte Fraktion wurde rechromatographiert und anschließend wiederum kartiert. Die entstandenen 97 Peptidkarten enthalten im Linear Modus ca. 100.000 Signale und im Reflektron Modus ca. 30.000 Signale, so dass nach der Feinkartierung > 10.000 native, im menschlichen Körper zirkulierende Peptide durch diese Methode dargestellt werden können.

Zur Automatisierung der Kartierungsmessungen mit MALDI-MS wurde im Laufe dieser Arbeit ein Autosampler für ein Voyager STR (Applied Biosystems) System entwickelt und implementiert (*DiskJockey*). Das System hat eine Kapazität von 20 MALDI-Targets und ist vollständig über die Gerätesoftware steuerbar, so dass MALDI-MS-Messungen vollständig automatisiert über mehrere Tage ablaufen können.

Zur Erzeugung der Sequenzdaten wurde eine Kombination aus elektrischer Ionenfalle und Datenbanksuche angewendet. Um zu gewährleisten, dass nur die zuvor kartierten Peptide einer MS/MS-Messung unterzogen werden, wurde das Softwaretool *Alcatrap* entwickelt. Diese Software übernimmt aus MALDI-MS-Messungen die Werte der monoisotopischen Peaks, überführt sie in mehrfach geladene Spezies und erstellt sowohl eine Methode für das MS-Gerät, als auch eine in die Gerätesoftware importierbare Arbeitsliste. Hierbei wird jeweils die Kollisionsenergie in Abhängigkeit der  $m/z$ -Wertes dynamisch gesetzt. Um den Datentransfer vom Fragmentspektrum zur Datenbanksuchmaschine zu gewährleisten, wurde mit „D2M“ eine weitere Softwareschnittstelle entwickelt.

Die Erstellung des Inventories aus humanem Hämofiltrat liefert 1.225 verschiedene native, im menschlichen Blut zirkulierende Peptide aus 146 Prote-

invorläufern. Darunter befinden sich mehrere Peptide, die neben ihrer physiologischen Bedeutung auch eine potenzielle medizinische Relevanz besitzen. Weiterhin ist auch der dynamische Bereich von acht Größenordnungen, in dem Sequenzen erzeugt wurden, vergleichbar mit dem des menschlichen Blutes. Eine Charakterisierung von Peptidomen ist in diesem Umfang bisher noch nicht durchgeführt worden und stellt eine Neuerung in der Proteomforschung dar. Die Studie ist durchaus vergleichbar mit kürzlich veröffentlichten umfangreichen Proteomstudien aus menschlichem Plasma und Serum und stellt eine Ergänzung dieser für den Bereich der Peptide und kleinen Proteine dar.

Limitierungen bei der Identifizierung von Peptiden mit der Ionenfalle wurden analysiert und mit der Adaptierung des Peptide-Inventory-Konzepts auf hochauflösende Quadrupol-Time-of-Flight Massenspektrometer zufriedenstellend gelöst. In diesem Rahmen wurde eine Methode entwickelt, die es ermöglicht, native Peptide bis zu einem Molekulargewicht von bis zu 8.500 Da direkt aus der komplexen Mischung zu fragmentieren und über eine Datenbanksuche zu identifizieren. Außerdem konnte exemplarisch gezeigt werden, dass durch die größere Massengenauigkeit und das größere Auflösungsvermögen des QTOF die Datenbanksuche wesentlich effektiver ist und durch die Kombination von Rechromatographie und QTOF-MS/MS nahezu jedes Signal einer Peptidbank der Identifizierung zugänglich gemacht werden kann.

Sowohl die integrierte Methodik der Herstellung eines Peptide-Inventories, als auch die in dieser Arbeit erzielten Ergebnisse von nativen Peptiden in humanem Blut stellen wissenschaftliche Neuerungen dar, auf denen eine weitergehende Peptidom-Analytik aufsetzen kann.

## **6    Ausblick**

Aufbauend auf den in dieser Arbeit erzielten Ergebnissen werden mit Methoden der Peptidomanalyse in Zukunft auch von weiteren biologischen Quellen Inventories angefertigt. Hier wird zuerst angestrebt, Quellen von hoher diagnostischer Relevanz, wie humanes Blut, Urin und Cerebrospinal-Flüssigkeit möglichst vollständig zu charakterisieren. Daneben ist abzusehen, dass auch Tiermodelle, Zellüberstände und Extrakte, sowie auch verschiedene Gewebe (Leber, Pankreas, Gehirn, Nieren und Tumoren) zumindest teilweise inventarisiert werden.

Als konsequente Entwicklung wird versucht werden, die eingesetzten Mengen zu minimieren, so dass Analysen bei Geweben im mg- und bei Flüssigkeiten im ml-Maßstab durchgeführt werden können.

## 7 Abbildungsverzeichnis

Abbildung 1-1 Larve und Falter des <i>Orgyia Antiqua</i> . Ein Genom und zwei Proteome.....	3
Abbildung 2-1 Komponenten eines Massenspektrometers. Grün hervorgehoben sind die in der Arbeit verwendeten Komponenten.....	8
Abbildung 2-2 Massenspektrometrische Ionisierungsmethoden, dargestellt anhand ihrer Anwendungsmöglichkeiten auf biochemisch wichtige Substanzklassen; nach (Lehmann 1996). ....	9
Abbildung 2-3 Massenspektren der Peptide (a) HLKTEAEMK und (b) Insulin. Gezeigt ist jeweils ein Ausschnitt der Molekülmassenregion. Eingezeichnet sind sowohl die Durchschnittsmasse (average mass) als auch die monoisotopische Masse. ....	10
Abbildung 2-4 Massenspektren des humanen Glucagons. Gezeigt ist ein Ausschnitt der Molekülmassenregion mit der Auflösung 1.000 (Blau), 3.000 (rot), 10.000 (grün) und 30.000 (schwarz). ....	11
Abbildung 2-5 Prinzip der Desorption, Ionisierung und Extraktion in der MALDI-MS; nach (Eckerskorn 1998). ....	14
Abbildung 2-6 Schematischer Aufbau des in dieser Arbeit verwendeten Reflektor-MALDI-TOF Massenspektrometers <i>Voyager-DE STR</i> der Fa. PerSeptive Biosystems; nach (PerSeptive 1997).....	16
Abbildung 2-7 Prinzip von kontinuierlicher ( <i>continuous</i> ) und verzögerter ( <i>delayed</i> ) Extraktion. ....	17
Abbildung 2-8 Prinzip des makroskopischen (oben) und mikroskopischen ESI-Prozesses (unten) (Eckerskorn 1998).....	20
Abbildung 2-9 Modell zur Erklärung der ESI nach dem Ionenemissionsmodell, nach (Fenn 1993). Die Tröpfchen des ESI-Prozesses besitzen Überschussladungen, die auf der Oberfläche durch Coulomb-Abstoßung eine äquidistante Lage einnehmen. Die Analytmoleküle übernehmen vor der Ionenemission die Anzahl an Ladungen, die sie aufgrund ihrer räumlichen Ausdehnung aufnehmen können.....	22
Abbildung 2-10 Schematischer Aufbau und Varianten von ESI-Quellen (nach Eckerskorn 1998). Nicht dargestellt ist die Variante mit Flüssen >1ml, die als Turbo-Ionspray bezeichnet wird und zusätzlich zur pneumatischen Unterstützung Hitze zur Desolvatisierung verwendet.....	23
Abbildung 2-11 Geometrie eines Quadrupol-Massenfilters. ....	25
Abbildung 2-12 Schematische Abbildung einer Quadrupol-Ionenfalle (QIT) nach Paul.....	27
Abbildung 2-13 Kreisbahnen nach Lissajous, mit deren Hilfe man sich die Bahnen von Ionen in einer elektrischen Ionenfalle vorstellen kann.....	27
Abbildung 2-14 Stabilitätsdiagramm der Mathieu'schen Gleichungen für die elektrische	

Ionenfalle .....	28
Abbildung 2-15 Schematischer Aufbau eines Quadrupol-Time-of-Flight Massenspektrometers mit Elektrospray-Ionisierung (ESI-QTOF-MS). Es besteht aus mehreren Quadrupoleinheiten zur Ionenfokussierung (Q0, Q1) und Fragmentierung (Q2) und einem orthogonal angeordneten reflektorn Flugzeitanalysator. ....	29
Abbildung 2-16 Vereinfachtes Prinzip der Tandem-MS (Lehmann 1996) .....	32
Abbildung 2-17 Nomenklatur für die Peptidfragmentierung nach Roepstorff und Fohlman (Roepstorff <i>et al.</i> 1984).....	34
Abbildung 2-18 Dekonvolutiertes Fragmentspektrum von Glu <sup>1</sup> -Fibrinopeptid B. Annotiert ist sowohl die y-Ionen-Serie, als auch die Aminosäuresequenz des Peptids....	35
Abbildung 2-19 Gel basierte Proteomanalytik im Überblick.....	36
Abbildung 2-20 Aminosäuresequenz von Oxytocin .....	39
Abbildung 2-21 Schematische Darstellung der Wirkungsweise von Peptiden. Durch das Andocken an den Rezeptor wird im Zellinneren eine Reaktionskette ausgelöst: zunächst wird das G-Protein, das sich direkt neben dem Rezeptor in der Zellwand befindet, aktiviert. Das G-Protein aktiviert seinerseits die Adenylatzyklase, die sich im Inneren der Zelle befindet. Die Adenylatzyklase spaltet ATP (Adenosintri-phosphat) in cAMP (zyklisches Adenosinmonophosphat) auf. cAMP aktiviert eine oder mehrere Proteinkinasen in der Zelle, die dann für die gewünschte Hormonreaktion der Zielzelle sorgen. ....	40
Abbildung 2-22 Spezialisierte Drüsenzellen synthetisieren das Hormon (H) aus Vorstufen (V) und schütten es bei Bedarf in den Extrazellularraum aus. Nach Diffusion oder Transport erfolgt die Wirkung auf eine Zielzelle. Gezeigt sind hier endokrine, parakrine und autokrine Hormonwirkung (Koolman <i>et al.</i> 2003).	40
Abbildung 3-1 Schematischer zur Vorgehensweise bei der Herstellung der Peptidbank 970303 aus 10.000 L humanem Blutfiltrat.....	44
Abbildung 3-2 Schematischer Aufbau der Hardware zur Erstellung eines Peptide-Inventories mit Hilfe eine Ionenfallen-MS. Blau gekennzeichnet sind Leitungen für den Lösungsmittel-/Probenfluss. Schwarz gekennzeichnet sind elektrische Leitungen zu Gerätesteuerung. Der Autosampler fungiert als Master, der sowohl HPLC-Pumpe als auch das Massenspektrometer steuert. ....	51
Abbildung 4-1 Überblick über die Informationen, die bei einer umfassenden Charakterisierung einer Peptidquelle in einer Datenbank gesammelt werden. .....	59
Abbildung 4-2 Ablaufschema zur Erzeugung von primär und sekundär ordnenden Daten innerhalb eines Inventories. Blau gekennzeichnet sind die zur Erzeugung von primären Daten notwendigen Schritte, grün markiert, die	

	Sekundärdatenschritte. Rot die Zusammenführung und Archivierung der Daten in einer Datenbank.....	60
Abbildung 4-3	Fotographische Darstellung des Autosampler-Prototyps „DiskJockey I“ mit Bezeichnung der einzelnen Komponenten.....	63
Abbildung 4-4	Schematische Darstellung der DiskJockey Steuerung. Das DiskJockey-Control-Panel (DJCP) interagiert mit der Hardware des Autosamplers und mit der Hardware des MALDI-MS über das Sequence Control Panel (SCP) und das Instrument Control Panel (ICP). Das SCP übernimmt hierbei die Funktion des „Masters“ und steuert die „Slaves“ DiskJockey und DJCP.....	64
Abbildung 4-5	Schematischer Aufbau der Berechnungen innerhalb der Alcatrap-Software. Zunächst werden Peaklisten aus MALDI-TOF-MS Messungen eingelesen. Dann werden die intensivsten „n“-Signale ausgewählt und in verschiedene Ladungszustände umgerechnet. Hieraus erstellt Alcatrap Methodendateien für das LCQ und eine importierbare Arbeitsliste für die automatische Abarbeitung. ....	67
Abbildung 4-6	Verteilung der nativen Peptide aus verschiedenen Massenbereichen (MALDI-TOF-MS) auf verschiedene Ladungszustände im ESI-Spektrum.. ....	68
Abbildung 4-7	Regressionsgeraden der Kollisionsenergien (CE) bei 30% Restintensität des Vorläuferions. Schwarz gekennzeichnet sind die 2-fach geladenen, grün die 3-fach geladenen und blau die 4-fach geladenen Ionen. Die Ergebnisse der Regressionsanalyse sind in Tabelle 4-2 dargestellt. ....	71
Abbildung 4-8	„General“-Seite von Alcatrap. Hier werden Einstellungen vorgenommen, welche MALDI-Peaklisten verarbeitet werden sollen und mit welchen Parameterdateien dies stattfinden soll. Außerdem werden Speicherorte und zu berechnende Ladungszustände festgelegt.....	74
Abbildung 4-9	„Options“-Seite von Alcatrap. Hier können die Massenbereiche für die Messung festgelegt werden. Es kann die Anzahl der zu untersuchenden Peptid-Signale pro MALDI-MS Spektrum festgelegt werden und entschieden werden, ob die Kollisionsenergie dynamisch oder statisch gesetzt werden soll. ....	75
Abbildung 4-10	„Documentation“-Seite von Alcatrap. Hier werden Daten für die „samplelist“ des LCQ erfasst.....	76
Abbildung 4-11	„Advanced“-Seite von Alcatrap. Hier können bestimmte Signale von der Prozessierung ausgenommen werden, mittlere in monoisotopische Vorläuferionen überführt werden und Ausschlusslisten für Folgemessungen erstellt werden. ....	77
Abbildung 4-12	Auftragung der mittleren gegen die monoisotopischen Massen von 1.500 nativen Peptiden. Aus der linearen Regression ergibt sich die Funktion zur Umrechnung von average zu monoisotopischen Massen der Vorläuferionen.	

---

.....	78
Abbildung 4-13 Schematische Darstellung des D2M-Werkzeuges zur Erstellung von MASCOT suchbaren .....	79
Abbildung 4-14 Benchmark der MASCOT™ MS/MS-Datensuche. Es wurde jeweils ein Datensatz von 25 Spektren in die Suchmaschine gegeben (Datenbank: dBEST (13.595.532 Sequenzeinträge). Die Rechnerleistung wurde sukzessiv erhöht. Es zeigt sich eine lineare Abhängigkeit zwischen Suchgeschwindigkeit und Rechnergeschwindigkeit in MHz. ....	81
Abbildung 4-15 Schematische Darstellung der Datenbanksuch mit der MASCOT-Suchmaschine unter Einbeziehung von D2M. ....	82
Abbildung 4-16 Schematischer Aufbau der PeTra Datenbank zur Archivierung von Peptide-Inventory-Daten. ....	83
Abbildung 4-17 Visualisierung der RP-Rechromatographie einer Peptidbankfraktion aus „970303“ (Pool 6 Fraktion 26). a) MALDI-TOF-MS der ursprünglichen Fraktion und b) aus der Rechromatographie resultierende Peptidkarte. Blau gekennzeichnet sind die Signale in der Ursprungfraktion und ihre Position in der Karte der Rechromatographie. Zu erkennen ist die erneute Dekomplexifizierung und das damit verbundene Auftreten zusätzlicher Signale. ....	87
Abbildung 4-18 Suchstrategie für die automatische Datenbanksuche von MS/MS-Daten mit Hilfe der MASCOT Suchmaschine. Es wird eine kaskadierende Suche in mehreren Datenbanken mit verschiedenen möglichen posttranslationalen Modifikationen durchgeführt. ....	90
Abbildung 4-19 Fragmenthäufigkeiten der Proteinvorläufer. Erfasst sind die 25 am häufigsten identifizierten Vorläuferproteine (bezeichnet mit dem SwissProt-Kürzel). ...	92
Abbildung 4-20 Massenverteilungen der identifizierten Peptide mit ESI-QIT-MS/MS und Edman-Abbau. ....	94
Abbildung 4-21 Konzentrationsbereich in Blutplasma in Bezug auf das Vorläuferprotein einiger ausgewählter Peptide des Inventories (blau). Konzentrationswerte einiger Peptidhormone, die zu diagnostischen Zwecken herangezogen werden (grün, nicht identifiziert). Die Konzentrationsangaben stammen aus (Anderson <i>et al.</i> 2002) und Zitaten hieraus. ....	95
Abbildung 4-22 Darstellung des Homologie-Grenzwertes (qplughole) in Abhängigkeit der Vorläuferionen-Toleranz („Parent Tolerance“) und der Fragmentionen-Toleranz („MS/MS-Toleranz“). ....	100
Abbildung 4-23 Abhängigkeit des „Identity-Score“ von der Datenbankgröße und von der Vorläuferionen-Toleranz („Peptide Tolerance“). Als Datenbanken haben hier SwissProt und NCBIInr gedient. ....	101
Abbildung 4-24 Abhängigkeit des Mascot Scores von der „Peptide-Tolerance“ und von der	

---

„MS/MS-Tolerance“ für ein Beispielpeptid in der SwissProt Datenbank....	102
Abbildung 4-25 Darstellung der Kollisionsenergien (CE) in Abhängigkeit der m/z-Werte und des Ladungszustandes. Die Regressionsgeraden sind in der jeweiligen Farbe des Datensatzes markiert. ....	105
Abbildung 4-27 Darstellung der Kollisionsenergien (CE) aller Ladungszustände in Abhängigkeit der m/z-Wertes und deren lineare Regression (schwarz). Die blauen, grünen und roten Graphen beschreiben die tatsächlich in der Methode angewendeten Werte für die CE. ....	107
Abbildung 4-28 Darstellung der Auflösung für ein Fibrinogen A Fragment (361-440) bei einer Auflösung von 10.000 (FWHM). Dargestellt ist das $[MH_9]^{9+}$ Ion.....	108
Abbildung 4-29 Direkte Identifizierung großer Peptide mit Hilfe von ESI-MS/MS. a) Spektrum einer biologischen Probe im MS-Modus. Die gezoomten Regionen zeigen das achtfach und neunfach geladenen Ion des Ubiquitins, sowie ein unbekanntes vierfach geladenes Ion. b) zeigt das MS/MS-Spektrum des neunfach geladenen Ions des Ubiquitins. Der Zoom zeigt eine Region des Spektrums mit Fragmentionen der Ladungszustände 1+ bis 7+.....	111
Abbildung 4-30 c) zeigt das dekonvolutierte und deisotopierte Spektrum aus b), welches zur Datenbanksuch mit Mascot verwendet wird. d zeigt den Datenbankhit der MASCOT-Suche und e) die Sequenzabdeckung des Fragmentenspektrums für Ubiquitin (human) 1-76. ....	112
Abbildung 4-31 MASCOT-Scores der 18 mit beiden Ansätzen identifizierten Peptide. In blau, die QTOF-Daten, in pink die LCQ-Daten. Gelb markiert sind Hits, die im Homologiebereich der MASCOT Suche liegen. ....	117
Abbildung 4-32 Annotiertes MALDI-Spektrum der Fraktion 26 aus Pool 6 der Peptidbank 970303.....	120

## 8 Literatur

Adkins, J. N., *et al.* (2002). "Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry." Mol Cell Proteomics **1**(12): 947-55.

Aleksandrov, M. L., *et al.* (1984). Dok. Akad. Nauk SSSR **277**: 379-383.

Anderson, L., *et al.* (1997). "A comparison of selected mRNA and protein abundances in human liver." Electrophoresis **18**(3-4): 533-7.

Anderson, N. L., *et al.* (2002). "The human plasma proteome: history, character, and diagnostic prospects." Mol Cell Proteomics **1**(11): 845-67.

Bahr, U., *et al.* (1997). "High-sensitivity analysis of neutral underivatized oligosaccharides by nanoelectrospray mass spectrometry." Anal Chem **69**(22): 4530-5.

Bairoch, A., *et al.* (1996). "The SWISS-PROT protein sequence data bank and its new supplement TREMBL." Nucleic Acids Res. **24**(1): 21-25.

Bairoch, A., *et al.* (2000). "The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000." Nucleic Acids Res **28**(1): 45-48.

Banks, R. E., *et al.* (2000). "Proteomics: new perspectives, new biomedical opportunities." Lancet **356**(9243): 1749-56.

Barbacci, D. C., *et al.* (1998). "Multi-anode detection in electrospray ionization time-of-flight mass spectrometry." J Am Soc Mass Spectrom **9**(12): 1328-33.

Barber, M., *et al.* (1981). "Fast Atom Bombardment of Solids (F.A.B.) : A New Ion Source for Mass Spectrometry." J. Chem. Soc., Chem. Commun. **34**: 325-327.

Beavis, R. C., *et al.* (1989). "Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins." Rapid Commun Mass

Spectrom **3**(12): 432-5.

Beavis, R. C., *et al.* (1989). "Matrix-assisted laser-desorption mass spectrometry using 355 nm radiation." Rapid Commun Mass Spectrom **3**(12): 436-9.

Beavis, R. C., *et al.* (1996). "Matrix-assisted laser desorption ionization mass-spectrometry of proteins." Methods Enzymol **270**: 519-51.

Beavis, R. C., *et al.* (1992). "a-Cyano-4-hydroxycinnamic acid as a matrix for matrix-assisted laser desorption mass spectrometry." Org. Mass. Spectrom **27**: 156-158.

Benson, D. A., *et al.* (1999). "GenBank." Nucleic Acids Res **27**(1): 12-7.

Biemann, K. (1988). "Contributions of mass spectrometry to peptide and protein structure." Biomed Environ Mass Spectrom **16**(1-12): 99-111.

Boeckmann, B., *et al.* (2003). "The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003." Nucleic Acids Res **31**(1): 365-70.

Boguski, M. S., *et al.* (1993). "dbEST--database for "expressed sequence tags"." Nat Genet **4**(4): 332-3.

Carrette, O., *et al.* (2003). "A panel of cerebrospinal fluid potential biomarkers for the diagnosis of Alzheimer's disease." Proteomics **3**(8): 1486-94.

Chait, B. T., *et al.* (1993). "Protein ladder sequencing." Science **262**(5130): 89-92.

Chambers, G., *et al.* (2000). "Proteomics: a new approach to the study of disease." J Pathol **192**(3): 280-8.

Chernushevich, I. V., *et al.* (2001). "An introduction to quadrupole-time-of-flight mass spectrometry." J Mass Spectrom **36**(8): 849-65.

Clauser, K., *et al.* (1996). (Protein Prospector). 44t Annual Conference on Mass Spectrometry and Allied Topics, Portland, USA.

Clauser, K. R., *et al.* (1996). . 44th Annual Conference on Mass Spectrometry and Allied Topics, Portland.

Crapo, L. (1988). Hormone: Die chemischen Boten des Körpers. Heidelberg, Spektrum-der-Wissenschaft-Verlagsgesellschaft.

Dole, M., *et al.* (1968). "Molecular beams of macroions." J. Chem. Phys. **49**: 2240-2247.

Eckerskorn, C. (1998). Massenspektrometrie. Bioanalytik. F. Lottspeich and H. Zorbas. Heidelberg

Berlin, Spektrum Akademischer Verlag. **1**: 323-368.

Eckerskorn, C., *et al.* (1992). "Mass spectrometric analysis of blotted proteins after gel electrophoretic separation by matrix-assisted laser desorption/ionization." Electrophoresis **13**(9-10): 664-5.

Eckerskorn, C., *et al.* (1997). "Analysis of proteins by direct-scanning infrared-MALDI mass spectrometry after 2D-PAGE separation and electroblotting." Anal Chem **69**(15): 2888-92.

Edman, P. (1956). "On the mechanism of the phenyl isothiocyanate degradation of peptides." Acta Chem Scand. **10**: 761-768.

Fenn, J. B. (1993). "Ion formation from charged droplets: roles of geometry, energy, and time." J. Am. Soc. Mass Spectrom. **4**(7): 524-535.

Fenn, J. B., *et al.* (1989). "Electrospray ionization for mass spectrometry of large biomolecules." Science **246**(4926): 64-71.

Ge, Y., *et al.* (2002). "Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry." J Am Chem Soc **124**(4): 672-8.

Glückmann, M. (2001). Untersuchungen zur Ablation und Ionenbildung bei matrixunterstützter Laserdesorption/Ionisation (MALDI). Frankfurt/M., Univer-

sity of Frankfurt/M. Dissertationsschrift.

Henderson, L. W., *et al.* (1986). Hemofiltration. Berlin, Heidelberg, New York, Tokyo, Springer.

Hewick, R. M., *et al.* (1981). "A gas-liquid solid phase peptide and protein sequenator." J Biol Chem **256**(15): 7990-7.

Hillenkamp, F. (1983). Laser Induced Ion Formation in Organic Solids. Berlin, Springer Verlag.

Iribarne, J. V., *et al.* (1976). "On the evaporation of small ions from charged droplets." J. Chem. Phys.(64): 2287-2294.

James, P. (1997). "Protein identification in the post-genome era: the rapid rise of proteomics." Q Rev Biophys **30**(4): 279-331.

Jungblut, P. R., *et al.* (1999). "Proteomics in human disease: cancer, heart and infectious diseases." Electrophoresis **20**(10): 2100-10. [pii].

Juraschek, R., *et al.* (1999). "Nanoelectrospray--more than just a minimized-flow electrospray ionization source." J Am Soc Mass Spectrom **10**(4): 300-8.

Jürgens, M. (1999). Multi-Dimensionale Kartierung von Peptiden aus humanem Blutfiltrat mittels massenspektrometrischer Methoden (Doktorarbeit). Mathematisch-Naturwissenschaftliche Fakultät. Münster, Westfälische Wilhelms-Universität.

Karas, M., *et al.* (1987). "Matrix-assisted ultraviolet laser desorption of non-volatile compounds." Int J Mass Spectrom Ion Process **78**: 53-56.

Karas, M., *et al.* (1985). "Influence of the Wavelength in High-Irradiance Ultraviolet Laser Desorption Mass Spectrometry of Organic Molecules." Anal. Chem. **57**: 2935-2939.

Karas, M., *et al.* (1988). "Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons." Anal Chem **60**(20): 2299-301.

Katta, V., *et al.* (1991). "Use of a single-quadrupole mass spectrometer for collision-induced dissociation studies of multiply charged peptide ions produced by electrospray ionization." Anal Chem **63**(2): 174-8.

Kaufmann, R., *et al.* (1993). "Mass spectrometric sequencing of linear peptides by product-ion analysis in a reflectron time-of-flight mass spectrometer using matrix-assisted laser desorption ionization." Rapid Commun Mass Spectrom **7**(10): 902-10.

Kebarle, P., *et al.* (1993). "From ions in solution to ions in the gas phase - the mechanism of electrospray mass spectrometry." Anal. Chem. **65**: 972A-986A.

Kellmann, M., *et al.* (2003). Differential Analysis and Inventory of Human Cerebrospinal Fluid: A Concept for Diagnosis and Research Applying Mass Spectrometry and Liquid Chromatography Based Peptidomics-Technology. IMSC 2003, Edinburgh.

Kellmann, M., *et al.* (2001). Routine Identification of Large Peptides (>5kDa) in Biological Sources Using an ESI-Quadrupol-TOF Mass Spectrometer. ABRF, San Diego.

Koolman, J., *et al.*, Eds. (2003). Taschenatlas der Biochemie, Thieme Verlag.

Krishna, R. G., *et al.* (1993). "Post-translational modification of proteins." Adv Enzymol Relat Areas Mol Biol **67**: 265-98.

Kuwata, H., *et al.* (1998). "Bactericidal domain of lactoferrin: detection, quantitation, and characterization of lactoferricin in serum by SELDI affinity mass spectrometry." Biochem Biophys Res Commun **245**(3): 764-73.

Lander, E. S., *et al.* (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Lehmann, W. D. (1996). Massenspektrometrie in der Biochemie. Heidelberg Berlin, Oxford, Spektrum Akademischer Verlag.

Lyubarskaya, Y. V., *et al.* (1998). "Screening for high-affinity ligands to the Src SH2 domain using capillary isoelectric focusing-electrospray ionization ion trap mass spectrometry." Anal Chem **70**(22): 4761-70.

Mamyrin, B. A., *et al.* (1973). "The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution." Sov. Phys. JETP(37): 45-48.

Mathieu, E. (1868). "Le mouvement vibratoire d'une membrane de forme elliptique." Jour. de Math. Pure et appliquées(13): 137-203.

Mc Luckey, S. A. (1992). "Principles of collision activation in analytical mass spectrometry." J Am Soc Mass Spectrom **3**: 599-614.

Minamino, N., *et al.* (2003). "Determination of endogenous peptides in the porcine brain: possible construction of peptidome, a fact database for endogenous peptides." J Chromatogr B Analyt Technol Biomed Life Sci **792**(1): 33-48.

Morris, H. R., *et al.* (1996). "High Sensitivity Collisionally-activated Decomposition Tandem Mass Spectrometry on a Novel Quadrupole/Orthogonal-acceleration Time-of-flight Mass Spectrometer." Rapid Communications in Mass Spectrometry **10**(8): 889-896.

Neitz (1998). Präparative Herstellung einer Peptidbank aus humanem Blutfiltrat (Diplomarbeit). Hannover, Universität Hannover.

Neitz, S., *et al.* (2001). "Screening for disulfide-rich peptides in biological sources by carboxyamidomethylation in combination with differential matrix-assisted laser desorption/ionization time-of-flight mass spectrometry." Rapid Commun Mass Spectrom **15**(17): 1586-92.

Pappin, D. J. C., *et al.* (1993). "Rapid Identification of Proteins by Peptide-Mass Fingerprinting." Current Biology **3**: 327-332.

Patterson, D. H., *et al.* (1995). "C-terminal ladder sequencing via matrix-assisted laser desorption mass spectrometry coupled with carboxypeptidase

Y time-dependent and concentration-dependent digestions." Anal Chem **67**(21): 3971-8.

Paul, W. (1990). "Elektromagnetische Käfige für geladene und neutrale Teilchen." Angewandte Chemie **102**: 780-789.

Paul, W., *et al.* (1955). "Das elektrische Massenfilter." Z.f.Physik( 140): 262-273.

Paul, W., *et al.* (1958). "Das elektrische Massenfilter als Massenspektrometer und Isotopenfilter." Z.f.Physik(152): 143-182.

Paul, W., *et al.* (1953). "A new mass spectrometer without a magnetic field." Z. Naturforschung(8a): 448-450.

Perkins, D. N., *et al.* (1999). "Probability-based protein identification by searching sequence databases using mass spectrometry data." Electrophoresis **20**(18): 3551-67. [pii].

PerSeptive (1997). Voyager-DE Biospectrometry Workstation Handbook - Supplement. Framingham, PerSeptive Biosystems.

Pieper, R., *et al.* (2003). "The human serum proteome: display of nearly 3700 chromatographically separated protein spots on two-dimensional electrophoresis gels and identification of 325 distinct proteins." Proteomics **3**(7): 1345-64.

Quadroni, M. (2001). Specific Detection and Analysis of Phosphorylated Peptides by Mass Spectrometry. Proteome Research: Mass Spectrometry. P. James. Berlin, Springer-Verlag. **1**: 187 - 206.

Quellhorst, E., *et al.* (1977). "Hemofiltration: Treatment of renal failure by ultrafiltration and substitution." Journal of Dialysis **1**(6): 529-543.

Raida, M., *et al.* (1999). "Liquid chromatography and electrospray mass spectrometric mapping of peptides from human plasma filtrate." J Am Soc Mass Spectrom **10**(1): 45-54.

Rockwood, A. L., *et al.* (1997). High-Speed Data Acquisition System for Time-of-Flight Mass Spectrometry. The 45<sup>th</sup> ASMS Conference on Mass Spectrometry and Allied Topics, Palm Springs.

Roepstorff, P., *et al.* (1984). "Proposal for a common nomenclature for sequence ions in mass spectra of peptides." Biomed Mass Spectrom **11**(11): 601.

Rudd, P. M., *et al.* (2001). Glycoproteomics: High Throughput Sequencing of Oligosaccharide Modifications to Proteins. Proteome Research: Mass Spectrometry. P. James. Berlin, Springer-Verlag. **1**: 207 - 228.

Schepky, A., *et al.* (1994). "Human hemofiltrate as a source of circulating bioactive peptides: determination of amino acids, peptides and proteins." Bio-med Chromatogr **8**(2): 90-4.

Schmidt, A., *et al.* (2003). "Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI?" J Am Soc Mass Spectrom **14**(5): 492-500.

Schrader, M., *et al.* (2001). "Peptidomics technologies for human body fluids." Trends Biotechnol **19**(10 Suppl): S55-60.

Schulz-Knappe, P., *et al.* (1996). "Systematic isolation of circulating human peptides: the concept of peptide trapping." Eur J Med Res **1**(5): 223-36.

Schulz-Knappe, P., *et al.* (2001). "Peptidomics: the comprehensive analysis of peptides in complex biological mixtures." Comb Chem High Throughput Screen **4**(2): 207-17.

Shukla, A. K., *et al.* (1993). "Collision activation and dissociation of polyatomic ions." Mass Spectrom Rev **12**: 211-255.

Skold, K., *et al.* (2002). "A neuroproteomic approach to targeting neuropeptides in the brain." Proteomics **2**(4): 447-54.

Spengler, B., *et al.* (1992). "Peptide sequencing by matrix-assisted laser-

desorption mass spectrometry." Rapid Commun Mass Spectrom **6**(2): 105-8.

Stafford, G. C., *et al.* (1984). "Recent improvements in and analytical applications of advanced ion-trap technology." Int. J. Mass Spectrom. Ion Process(60): 85-98.

Strupat, K., *et al.* (1991). "2,5-Dihydroxybenzoic acid: a new matrix for laser desorption—ionization mass spectrometry." Int. J. Mass Spectrom. Ion Proc. **111**: 89-102.

Sundler, F., *et al.* (1988). Peptide producing endocrine/paracrine cells in the gastroentero-pancreatic region., Elsevier Science Publishers B.V.

Svensson, M., *et al.* (2003). "Peptidomics-based discovery of novel neuropeptides." J Proteome Res **2**(2): 213-9.

Tammen, H., *et al.* (2003). Use of mass determination of factor XIIIa protein or factor XIIIa peptides for direct phenotyping of factor XIIIa polymorphisms.

Tammen, H., *et al.* (2004). "Mass Spectrometric Phenotyping of Val34Leu Polymorphism of Blood Coagulation Factor XIII by Differential Peptide Display." Clin. Chem. **50**(3):545-51.

Tirumalai, R. S., *et al.* (2003). "Characterization of the low molecular weight human serum proteome." Mol Cell Proteomics **2**(10): 1096-103.

Vestal, M., *et al.* (1998). "Resolution and mass accuracy in matrix-assisted laser desorption ionization- time-of-flight." J Am Soc Mass Spectrom. **9**(9): 892-911.

Walter, C., *et al.* (2002). DiskJockey - A MALDI Target Autoloader System for a MALDI-TOF Instrument (Voyager-DE STR). The 50<sup>th</sup> ASMS Conference on Mass Spectrometry and Allied Topics, Orlando, FL.

Wilkins, M. R., *et al.* (1996). "From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis." Biotechnology (N Y) **14**(1): 61-5.

---

Wilkins, M. R., *et al.* (1996). "Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it." Biotechnol Genet Eng Rev **13**: 19-50.

Wilm, M., *et al.* (1996). "Analytical properties of the nanoelectrospray ion source." Anal Chem **68**(1): 1-8.

Wilm, M., *et al.* (1996). "Femtomole sequencing of proteins from polyacrylamide gels by nano- electrospray mass spectrometry." Nature **379**(6564): 466-9.

Yamashita, M. F., J.B.; (1984). "Electrospray ion source. Another variation on the free-jet theme." J. Phys. Chem. **88**(20): 4451-4459.

Yates, J. R., 3rd, *et al.* (1995). "Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database." Anal Chem **67**(8): 1426-36.

Zhang, Z., *et al.* (2000). "De novo peptide sequencing by two-dimensional fragment correlation mass spectrometry." Anal Chem **72**(11): 2337-50.

## **9 Eidesstattliche Versicherung**

Ich erkläre hiermit an Eides statt, daß ich die vorgelegte Dissertation über die automatisierte Identifizierung von nativen Peptiden aus humanen Körperflüssigkeiten mittels Massenspektrometrie selbständig angefertigt und mich anderer Hilfsmittel als der in ihr angegebenen nicht bedient habe, insbesondere, dass aus Schriften Entlehnungen, soweit sie in der Dissertation nicht ausdrücklich als solche mit Angabe der betreffenden Schriften bezeichnet sind, nicht stattgefunden haben.

Harpstedt, den 31.08.2004

## 10 Danksagung

Mein Dank gilt allen, die zum Gelingen dieser Arbeit beigetragen haben.

Besonderer Dank gilt Prof. Dr. Michael Karas für die Betreuung und Übernahme des Gutachtens und PD Dr. Peter Schulz-Knappe für die Betreuung der Arbeit bei BioVisioN, für viele fruchtbare Ideen und Diskussionen, sowie die Motivation zum „Endspurt“.

Weiterhin danke ich Herrn Dr. Thomas Möhring den besten Kollegen, den man sich wünschen kann, für unzählige Diskussionen, Anregungen und natürlich für´s Korrekturlesen.

Danken möchte ich auch meinem „roten Bruder“ Peer Köster, der mich während der praktischen Arbeiten als technische Assistenz mehr als tatkräftig und kompetent unterstützt hat.

Herrn Dr. Danilo Kardel, der mir bei der Entwicklung von „Alcatrap“ mit Rat und Tat zur Seite gestanden hat.

Herrn Dipl.-Ing. Christian Walter danke ich für die tolle Zusammenarbeit am „DiskJockey“.

Mein Dank gilt außerdem Frau Dr. Susanne Neitz und Herrn Dr. Michael Jürgens für viele interessante Diskussionen.

Ferner möchte ich Herrn Dr. John Cottrell und Herrn Dr. David Creasy von Matrix Science für viele Diskussionen über Datenbanksuchen und für die Bereitschaft danken viele meiner Anregungen in Ihren Produkten zu berücksichtigen.

Natürlich danke ich auch dem übrigen BioVisioNs Team für die vielen Diskussionen und die tolle Arbeitsatmosphäre.

Ich möchte meinen Eltern für die Unterstützung während meines Studiums danken, ohne die diese Arbeit sicherlich nicht möglich gewesen wäre.

Zum Schluss gilt mein Dank meiner Frau Kerstin und meinen Kindern Annemieke und Ilvi, die viel Geduld aufbringen mussten, damit diese Arbeit fertiggestellt werden konnte und mich trotzdem zu jedem Zeitpunkt vollkommen unterstützt haben.