

1 Einleitung

In den vergangenen Jahren führte die Entwicklung innovativer Technologien zu einem enormen Fortschritt in der Analyse biologischer Systeme. Als Meilenstein der Biotechnologie ist zweifelsohne die Entwicklung der „*Polymerase chain reaction*“ (PCR) [1] anzusehen. Diese Technik ermöglicht die Untersuchung, Manipulation und Sequenzierung kleinster DNA-Mengen. Im Zuge der weiteren Entwicklung und insbesondere mit der Gründung des „*Human Genome Project*“, das sich zur Aufgabe gestellt hatte, das gesamte menschliche Genom zu entschlüsseln, wurde die Entwicklung biochemischer und bioinformatischer Techniken, die eine automatische „*Hochdurchsatz-DNA-Sequenzierung*“ ermöglichen, beschleunigt. Als weiterer Meilenstein und Ergebnis dieser Entwicklung wurde im Februar 2000 fast zeitgleich von der Firma Celera Inc. und den am „*Human Genome Project*“ beteiligten Wissenschaftlern das vollständig entschlüsselte Genom des Menschen publiziert [2,3]. Mit der Einführung der „*DNA-Chip*“-Technologie [4,5] wurde die systematische und simultane Analyse der Genexpression zweier unterschiedlicher biologischer Zustände möglich. Der Einsatz dieser Techniken ermöglicht einerseits das bessere Verständnis der molekularen Regulationsmechanismen biologischer Systeme, andererseits bleiben dabei wichtige Informationen über die Kontrolle der Genexpression, über Veränderungen in der Proteinexpression, über die Synthese und Abbauraten von Proteinen oder über „*post-translationale*“ Modifikationen verborgen. Neuste wissenschaftliche Erkenntnisse aus Experimenten mit menschlichen Leberzellen [6] oder Hefezellen [7] lassen den Schluss zu, dass die aktuell in einer Zelle vorhandene Proteinmenge mit einer mRNA-Analyse nicht erfaßt werden kann [8]. Fast kein Protein wird „*in vivo*“ so verwendet, wie es nach Translation der mRNA vorliegt, da neusynthetisierte Proteine ihre Funktionalität erst durch posttranslationale Modifikationen (d.h. es werden zum Beispiel einzelne Aminosäuren oder Peptide am N - bzw. C - Terminus

abgespalten, Kohlehydrate addiert oder Serin-, Threonin- oder Tyrosinreste phosphoryliert) erlangen. So entsteht z.B. erst durch Abspaltung eines Hexapeptides am N-Terminus von Trypsinogen das funtionsfähige Verdauungsenzym Trypsin. Aus diesem Grunde ist die alleinige Kenntnis des Genoms eines Organismus nur ein erster notwendiger Schritt. Erst wenn man in der Lage ist, das Proteom [9,10] zu jedem gewünschten Zeitpunkt zu analysieren, wird es möglich sein, komplexe Krankheiten wie Krebs, CJD oder Alzheimer, auf molekularer Ebene zu verstehen.

Das Proteom beschreibt das Proteinexpressionsmuster einer Zelle oder Gewebes eines Organismus unter genau definierten Bedingungen zu einem bestimmten Zeitpunkt. Im Gegensatz zum Genom, das als eine wohldefinierte, durch die Art, Anzahl und Anordnung seiner Nukleotide bestimmte, statische Größe ist, wird das Proteom durch viele Parameter beeinflusst (Abbildung 1) und verändert sich fortwährend.

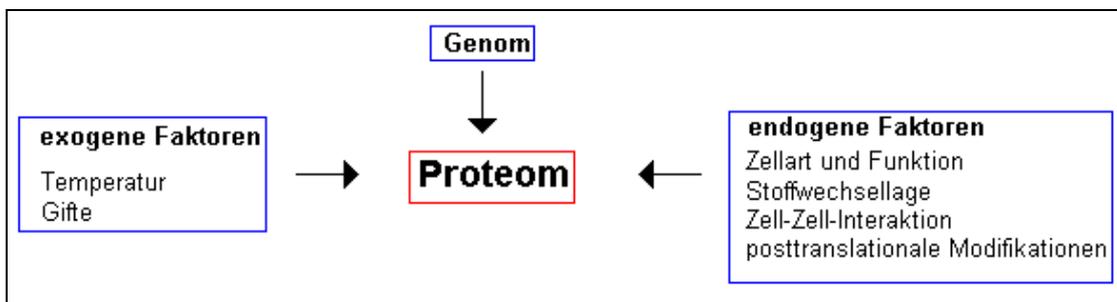


Abbildung 1 Einfluss unterschiedlicher Parameter auf das Proteom einer Zelle.

1.1 2D-Gelelektrophorese

Die Analyse eines Proteoms stellt hohe Anforderungen an die verwendeten Technologien. So müssen möglichst alle Proteine einer Zelle gleichzeitig erfaßt und voneinander getrennt werden. Die 2D-Gelelektrophorese ist momentan die einzig verfügbare Technik, die in der Lage ist mehr als 10000 Proteine aufzutrennen. Hierbei werden die Proteine zuerst nach ihrem isoelektrischen Punkt (isoelektrische Fokussierung) und anschließend nach ihrem Molekulargewicht (SDS-PAGE) getrennt [11,12] und schließlich z.B. durch Coomassie-Färbung, Silberfärbung oder mit Hilfe von Fluoreszenzfarbstoffen [13] nachgewiesen. In Tabelle 1 sind einige Vor- und Nachteile dieser Methode für die Proteomforschung aufgelistet.

Tabelle 1 Vor – und Nachteile der 2D-Gelelektrophorese für die Proteomforschung.

Vorteile	Nachteile
Hohe Auflösung (>10000 Proteinspezies)	Schwierige technische Durchführung
Mit Immobilinen in der ersten Dimension können bestimmte pH-Bereiche gespreizt werden.	Beim Transfer von erster zur zweiten Dimension können Proteine verloren gehen
Geringer Probenverbrauch	Selten vorkommende Proteine können nicht gleichzeitig neben sehr häufig vorkommenden Proteinen dargestellt werden #
Alle Proteine werden zur gleichen Zeit bearbeitet	Membranproteine sind nur begrenzt zugänglich

Seltene Proteine (z.B Transkriptionsfaktoren) liegen in einer Zelle nur in wenigen Kopien vor, andere sind um einen Faktor 10^6 höher konzentriert. Um die gleichzeitige Erfassung von Proteinen in solchen Mengenverhältnissen zu realisieren, wäre ein dynamischer Bereich von mind. 10^6 notwendig. Der real vorhandene dynamische Bereich der 2D-Gelelektrophorese liegt derzeit bei maximal 10^3 . Um diese Limitierung zu überwinden, versucht man seltene von häufigen Proteinen durch Vorfraktionierung oder durch den Einsatz von gespreizten pH-Gradienten voneinander zu separieren.

1.2 Massenspektrometrie

Massenspektrometrie ist eine Analysetechnik zur Bestimmung des Masse-Zu-Ladungsverhältnisses freier Ionen im Vakuum. Ein Massenspektrometer setzt sich aus drei Komponenten zusammen: Ionenquelle, Massenanalysator und Detektor.

Ionenquelle

In der Ionenquelle werden aus einer Substanzprobe gasförmige Ionen erzeugt. Die Ionisierung der Analytmoleküle wird entweder durch Aufnahme oder Abgabe eines Elektrons, eines Protons oder eines Kations (z.B. Na^+) herbeigeführt. Dies kann auf unterschiedliche Weise erreicht werden:

- Elektronenstoss-Ionisation (EI): Beschuss der Probe mit Elektronen
- chemische Ionisation (CI): Ladungstransfer in der Gasphase
- Fast Atom Bombardement (FAB): Beschuss mit Atomen oder Ionen bzw. Sekundärionen (SI)
- Plasmadesorption (PD): Beschuss mit hochenergetischen Teilchen
- Laserdesorption/Ionisation (LDI): Beschuss mit Photonen
- Matrix-assistierte LDI (MALDI): Beschuss mit Photonen
- Elektrospray-Ionisation (ESI): Versprühen der gelösten Probe in einem elektrischen Feld

Massenanalysator

Der Massenanalysator dient zur Trennung der Analytionen. Die Ionentrennung erfolgt nach unterschiedlichen Prinzipien:

- Sektorfeldinstrumente: Durch Kombination eines Magnetfeldes mit einem elektrischen Feld

- Quadrupolinstrumente: Im Hochfrequenzfeld eines Quadrupol-Stabsystems
- „*Time of Flight*“ (TOF) Instrumente: Nach Flugzeit in einem Messrohr in Kombination mit einer gepulsten Ionenerzeugung
- Ionenfalle (Ion trap): In elektr. Ionenfalle mit anschließender Massenanalyse über charakteristische Resonanzfrequenzen
- Ionencyclotron-Resonanz-Zelle (ICR): Magnet. Ionenfalle mit anschließender Massenanalyse über charakteristische Resonanzfrequenzen

Detektor

Zum Ionennachweis dienen folgende Detektoren:

- Faraday-Käfig
- Konversionsdynode mit Sekundärelektronenvervielfacher(SEV)
- Szintillationszähler
- Vielkanalplatten

Die Überführung von Ionen aus der kondensierten Phase in die Gasphase ist mit hohem Energieaufwand verbunden. Beispielsweise benötigt die Ionisierung von 1mol in Wasser gelöster Natriumionen:



Seit den 70er Jahren kommen für die Ionisierung auch Laser zum Einsatz. Bei der so genannten Laserdesorptions/Ionisationstechnik-Massenspektrometrie (LDI-MS) wird die Probe direkt mit einem gepulsten Laser bestrahlt, um intakte Molekülionen aus der kondensierten Phase in die Gasphase zu überführen (direkte Desorption).

Die so erzeugten Massenspektren wiesen aber nur geringe Ionenhäufigkeiten auf und durch den direkten Energieeintrag wurden große, thermisch instabile Moleküle stark fragmentiert, so dass mit dieser Technik in der Regel nur intakte Biomoleküle unter 1000 Da nachweisbar sind. Die LDI-MS war für die Analytik von großen Biomolekülen somit nur von geringer Bedeutung.

Mit der Einführung und Etablierung der schonenden Ionisierungstechniken ESI [14,15,16] und MALDI [17,18,19] gegen Ende der 80er Jahre wurde es erstmals möglich, große Moleküle intakt vom kondensierten in den gasförmigen Zustand zu überführen und zu ionisieren. Dieser Meilenstein machte die Massenspektrometrie zu einer der wichtigsten Analysemethoden für die moderne Peptid- und Proteinanalytik.

1.2.1 Elektrospray- Ionisations- Massenspektrometrie (ESI-MS)

Unter Elektrospray versteht man die Dispersion einer Flüssigkeit in kleine geladene Tröpfchen mit Hilfe eines elektrischen Feldes. Dabei wird eine durchlässige, spitz ausgezogene, mit Analytlösung gefüllte Metallkapillare oder metallbeschichtete Glaskapillare an eine Hochspannungsquelle (0,3-6 kV) angeschlossen. Im elektrischen Feld verformt sich die Flüssigkeitsoberfläche an der Austrittsstelle der Kapillare zum sogenannten „Taylor-Konus“ (Abbildung 2). Die Ionisierung erfolgt in vier Schritten:

1. Bildung kleiner geladener Tröpfchen aus Elektrolyten
2. Verdampfen des Lösungsmittels und dadurch Zunahme der Ladungsdichte an der Tröpfchenoberfläche ($\approx 10^5$ Ladungen pro Tröpfchen)
3. Wiederholter Zerfall der Tröpfchen in immer kleinere Tröpfchen (Coulomb Explosion)
4. Transfer der Ionen in die Gasphase (Desolvatisierung)

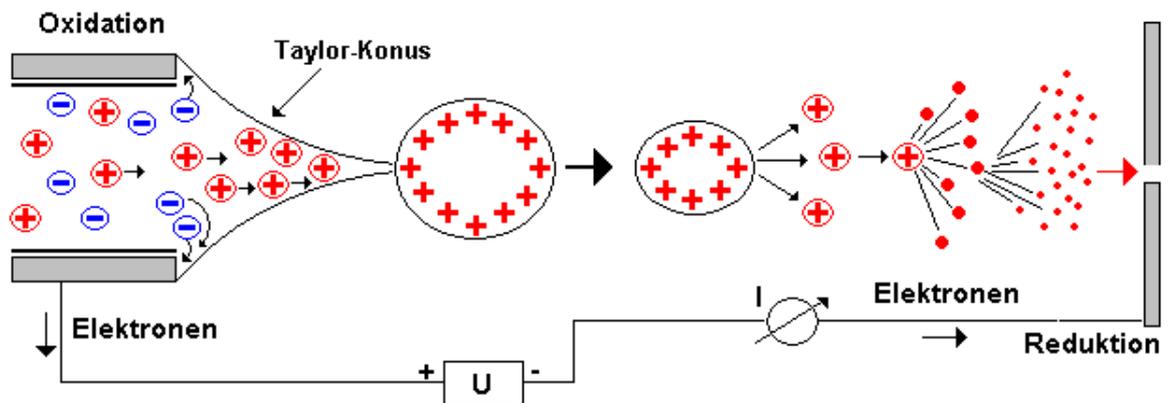


Abbildung 2 Schematische Darstellung des Ionisierungsprozesses bei der ESI-MS

Aus kleineren Analytmolekülen entstehen einfach geladene Ionen. Makromoleküle liefern hingegen bevorzugt mehrfach geladene Ionen (typischerweise 1 Ladung/1000 Da), wobei der Ladungszustand auch von der räumlichen Struktur der Analytmoleküle abhängt [20-22]. Die so entstandenen Ionen werden in ein gestaffeltes Vakuumsystem überführt und in Sektorfeld-, Ionenfallen-, Quadrupol- oder neuerdings auch in orthugonalen Flugzeit-Massenspektrometern [23] getrennt und detektiert.

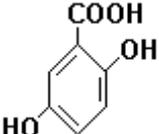
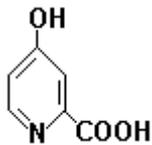
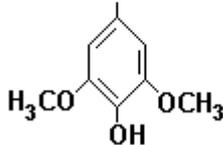
Die Erkenntnis, dass die Empfindlichkeit der ESI-MS in erster Linie von der Konzentration des Analyten abhängt und über die Abmessungen der Sprühkapillare drastisch gesteigert werden kann, führte Mitte der 90er zur Einführung der „Nano-ESI-MS“ [24], die aufgrund ihres geringen Probenbedarfs für Peptide ($\sim 10^{-15}$ mol) von großer Bedeutung für die Proteomforschung ist [25].

1.2.2 Matrix-assistierte Laserdesorptions/Ionisations-Massenspektrometrie (MALDI-MS)

Bei der MALDI-MS werden die Analytmoleküle mit einem 10^4 - 10^7 -fachen Überschuss an Matrix auf einem Probenträger präpariert. Die Matrix ist eine niedermolekulare organische Substanz (Tabelle 2), die bei der verwendeten

Laserwellenlänge eine hohe Absorption aufweist.

Tabelle 2 Typische Matrixsubstanzen für die MALDI-MS

Matrix	chem. Formel	[nm]	verwendet für
α -Cyano-4-hydroxycimtsäure (CHCA)	$N\equiv C-C=CH-COOH$ 	377, 355	Peptide
2,5-Dihydroxybenzoesäure (DHB)		266,337,355	Proteine, Peptide
3-Hydroxypicolinsäure (HPA)		337,355	Oligonukleotide
Sinapinsäure	$HC=CH-COOH$ 	266,337,355	Proteine

Beim Verdunsten des Lösungsmittels werden die Analytmoleküle in die auskristallisierende Matrix eingebaut oder an deren Oberfläche gebunden. Der genaue Vorgang dieses wichtigen Präparationsschrittes ist bis heute nicht verstanden. Alle zum Einsatz kommenden Matrices sind empirisch gefundenen worden. Das MALDI-Experiment wird gestartet, indem eine Fläche von typischerweise 50-250 μm der Probe im Hochvakuum mit einem intensiven, kurzen Laserimpuls (10^6 - 10^7 W/cm^2 , wenige ns) bestrahlt wird. Durch Absorption der Energie durch die Matrixmoleküle wird ein Teil dieser und der eingebetteten Analytmoleküle desorbiert und ionisiert (Abbildung 3). Diese beiden Vorgänge sind weder räumlich noch zeitlich klar voneinander getrennt.

Die Aufgabe der Matrix besteht darin die Analytmoleküle von der kondensierten Phase in die Gasphase zu überführen und deren Ionisation zu ermöglichen oder zu fördern.

Auf welche Weise die Desorption und Ionisierung der Analytmoleküle erfolgt, ist ebenfalls noch nicht vollständig geklärt [26]. Bei diesem Prozeß sind sowohl die Einwirkzeit des Lasers als auch die übertragene Energiemenge pro Zeiteinheit als kritische Parameter anzusehen. So führt eine zu lange Bestrahlung zur thermischen Zersetzung der Probe und ein zu hoher Energieeintrag zum spontanen Zerfall der Analytmoleküle und damit verbunden zu einer Verschlechterung des Signal-Rausch-Verhältnisses.

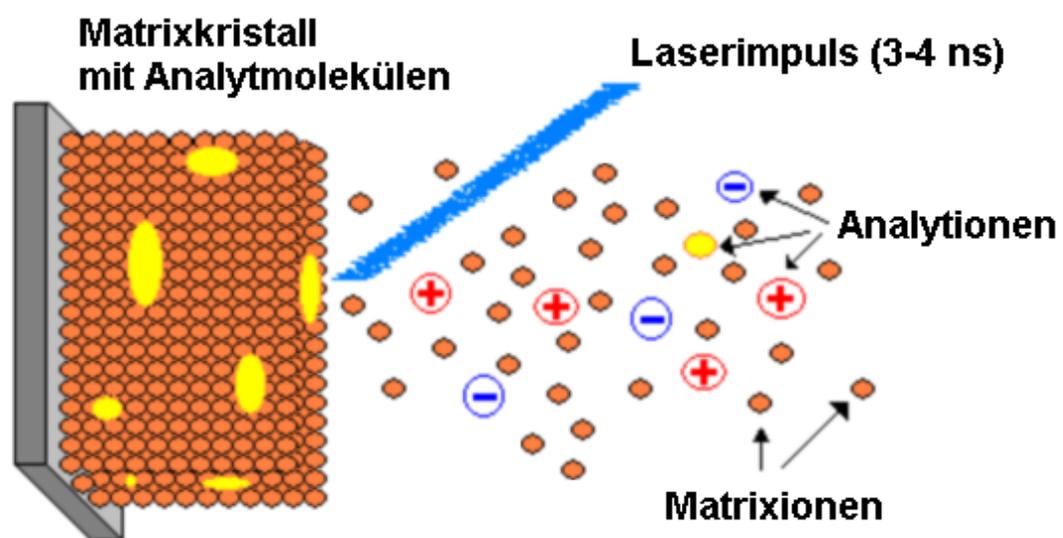


Abbildung 3 Prinzip des MALDI Prozesses:

Die Analytmoleküle (gelb) sind in einem Kristallgitter von Matrixmolekülen (rotbraun), die bei der verwendeten Wellenlänge einen hohen Absorptionskoeffizienten aufweisen, eingebettet. Der Prozeß wird durch einen kurzen Laserimpuls initiiert. Die Energieaufnahme durch die Matrixmoleküle führt zur Freisetzung dieser und der eingebetteten Analytmoleküle in die Gasphase und zur Ionisierung. Die genauen physikalisch-chemischen Ursachen für diesen Prozeß sind noch nicht geklärt.

Flugzeitmassenspektrometer (Time-of-flight - MS, TOF-MS)

Für die MALDI-MS werden bevorzugt Flugzeitmassenspektrometer eingesetzt [27-29]. Sie sind vergleichsweise günstig in der Anschaffung, empfindlich, ermöglichen die simultane Analyse aller pro Laserimpuls erzeugten positiv oder negativ geladenen Ionen, sind bezüglich des zugänglichen m/z -Bereichs nicht limitiert und leicht zu bedienen. Ein weiterer Vorteil gegenüber anderen Massenanalysatoren (wie z.B. Sektorinstrumenten oder Quadrupolinstrumenten) ist ihre hohe Transmission, d.h. ein hoher Anteil der in der Ionenquelle erzeugten Ionen erreicht intakt den Detektor und wird nachgewiesen.

Abbildung 4 veranschaulicht das zugrundeliegende Prinzip eines MALDI-TOF Instruments.

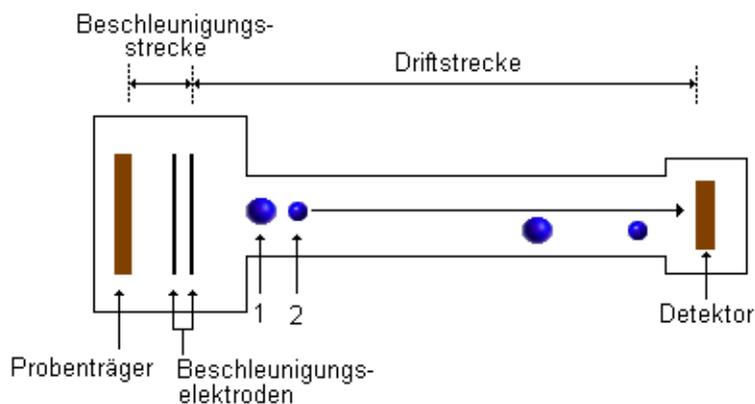


Abbildung 4 Prinzip eines linearen Flugzeitmassenspektrometers.

Zwei positiv geladene Ionen der Ladung $q=z \cdot e$ und mit den Massen m_1 und m_2 werden nach ihrer Emission durch ein zwischen dem Probenträger und den Beschleunigungselektroden angelegtes elektrisches Feld beschleunigt. Nach Durchqueren des Feldes erlangen sie die gleiche kinetische Energie und durchlaufen anschließend eine feldfreie Driftstrecke, wobei das Quadrat ihrer Geschwindigkeit umgekehrt proportional zu ihrer jeweiligen Masse ist.

Die durch Einwirkung des Laserstrahls entstandenen Ionen werden in einem elektrischen Feld beschleunigt. Ionen mit gleicher Ladung erlangen während

sie dieses Feld passieren die gleiche kinetische Energie. Ihre Geschwindigkeit ist somit eine Funktion ihrer Masse (Gleichung 1).

$$v = \sqrt{\frac{2 * z * e * U}{m}} \quad (1)$$

v = Geschwindigkeit des Ions nach der Beschleunigungsstrecke
z = Ladungszahl
e = Elementarladung
U = Beschleunigungsspannung
m = Masse des Ions

Danach durchlaufen die Ionen eine feldfreie Driftstrecke (1-3 m Länge), in der sie nach ihrem Masse/Ladungs (m/z)-Verhältnis aufgetrennt werden. Große Ionen mit hohem m/z-Wert erreichen den Detektor später als kleine Ionen (Abbildung 4). Die Geschwindigkeit der Ionen v ergibt sich aus der Gesamtflugzeit t, in der sie die Driftstrecke der Länge L passiert haben (Gleichung 2).

$$v = \frac{L}{t} \quad (2)$$

daraus folgt:

$$0.5 * m * \left(\frac{L}{t}\right)^2 = z * e * U \quad (3)$$

und

$$\frac{m}{z} = \frac{2 * e * U}{L^2} * t^2 \quad (4)$$

Das bedeutet, dass in einem Flugzeitmassenspektrometer das Verhältnis m/z dem Quadrat der Flugzeit proportional ist. Somit lässt sich die Masse der Ionen aus ihrer Flugzeit ermitteln.

Die oben erwähnten und in Abbildung 4 dargestellten Zusammenhänge setzen voraus, dass alle Ionen vom gleichen Ort und zur gleichen Zeit desorbiert und ionisiert werden. Dies ist aber aufgrund der mit dem MALDI-Prozeß verbundenen Energie-, Orts- und Zeitunschärfen nicht der Fall.

So verhindern Abschirmeffekte, dass alle Ionen zur Zeit ihrer Entstehung das elektrische Feld in seiner vollen Stärke wahrnehmen. Ebenso werden einige Ionen erst nach einer zeitlichen Verzögerung gebildet, so dass diese einen Teil der Beschleunigungsstrecke bereits als neutrale Teilchen durchlaufen haben und somit eine geringere kinetische Energie aufnehmen, als die Ionen, die unmittelbar gebildet wurden. Weiterhin entstehen Zeit- und Energiefehler durch Kollisionen mit anderen Ionen. All diese Effekte bewirken, dass Ionen gleicher Masse und gleicher Ladungszahl nach Durchlaufen des Beschleunigungsfeldes nicht zwangsläufig exakt die gleiche kinetische Energie besitzen und somit auch zu unterschiedlichen Zeiten den Detektor erreichen. Dies führt zur einer Verbreiterung des Signals und setzt somit das Auflösungsvermögen herab. Deshalb sind moderne TOF-Instrumente mit einem elektrostatischen Reflektor ausgerüstet (Abbildung 5) [30].

Der Reflektor befindet sich am Ende der Driftstrecke und bewirkt eine Richtungsumkehr der Ionen in einem elektrischen Gegenfeld. Ionen mit höherer kinetischer Energie dringen tiefer in das Gegenfeld ein und werden später hinauskatapultiert als Ionen geringerer kinetischer Energie. Die erhöhte kinetische Energie wird also durch eine verlängerte Flugpassage wieder kompensiert.

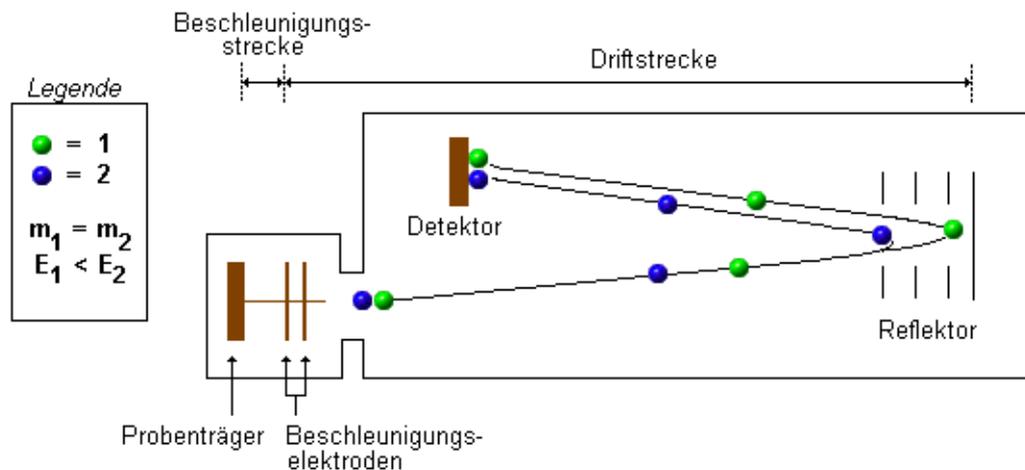


Abbildung 5 Prinzip eines Reflektor-Flugzeitmassenspektrometers.

Zwei einfach positiv geladene Ionen der gleichen Masse durchqueren die feldfreie Driftstrecke. Ein Ion (grün) besitzt bei seiner Entstehung eine höhere kinetische Energie und bewegt sich deshalb mit höherer Geschwindigkeit als das andere (blau) mit niedrigerer Anfangsenergie. Beim Eindringen in das elektrische Gegenfeld des Reflektors wird das blaue Ion aufgrund seiner geringeren kinetischen Energie früher reflektiert als das grüne Ion, welches aufgrund seiner höheren kinetischen Energie tiefer in das Reflektorfeld eindringt. Den verbleibenden Rest der Strecke holt das grüne Ion aufgrund seiner höheren Geschwindigkeit wieder auf. Am Detektor treffen dann, bei geeigneter Anordnung, beide Ionen gleichzeitig auf.

Die tiefer in den Reflektor eingedrungenen Ionen holen aufgrund ihrer höheren Geschwindigkeit bedingt durch ihre höhere kinetischen Energie die langsameren Ionen an einem bestimmten Punkt der Driftstrecke wieder ein. An diesem Punkt befindet sich der Detektor, der diese zur gleichen Zeit registriert. Durch den Einsatz des Reflektors können Energiedifferenzen von bis zu 10% kompensiert werden. Ionen gleicher Masse, die sich in ihrer kinetischen Anfangsenergie um mehr als 10% unterscheiden, führen auch beim Einsatz eines Reflektors zu verminderter Auflösung und Signalverlust. Eine weitere Verbesserung der Auflösung wird durch eine verzögerte Ionenextraktion („*delayed extraction*“) erreicht. Hierbei wird das elektrische Beschleunigungsfeld nicht permanent, sondern zeitversetzt (100-500ns) zum Laserimpuls eingeschaltet [31,32]. Ionen mit höherer Startgeschwindigkeit entfernen sich während der Verzögerungszeit weiter vom Start und erfahren nach Einschalten des Feldes eine geringere kinetische Energie als Ionen mit

niedrigerer Anfangsgeschwindigkeit. Durch die geeignete Wahl von Verzögerungszeit und Feldstärke und in Kombination mit einem Reflektor erreichen moderne TOF-Instrumente eine Auflösung >20.000 FWHM („*full width at half maximum*“, Breite des Peaks bei halber Peakhöhe) im Massenbereich bis etwa 8000 Da.

1.2.3 MALDI-MS versus ESI-MS

Die ESI-MS ist sehr empfindlich gegenüber niedermolekularen Verunreinigungen. Aussagekräftige Spektren erhält man nur dann, wenn die Probenlösungen weitgehend frei von Puffern, Salzen und Detergenzien sind. Im Gegensatz dazu ist die MALDI-MS relativ unempfindlich gegenüber kleinen Mengen an Verunreinigungen. Durch hohe Ionenausbeute, kürzere Analysezeiten und die daraus resultierende, vergleichsweise einfache Automatisierbarkeit ist die MALDI-MS die Methode der Wahl für die schnelle Analyse einer Vielzahl von Proben, wie sie z.B. in der Proteomforschung generiert werden [33-36].

1.3 „*Peptide mass fingerprinting*“ und Proteinidentifikation

Beim Einsatz traditioneller Techniken, z.B. SDS-PAGE [37], zur Bestimmung der molekularen Masse von Proteinen, wird eine Massenrichtigkeit von $\pm 5-10\%$ erreicht. Bei Verwendung von MALDI- oder ESI-MS (siehe oben) ist die Massenrichtigkeit oft besser als 0,01%.

Die massenspektrometrische Identifizierung von unbekannt Proteinen aus einer biologischen Probe in Datenbanken wird durch Verwendung von „*peptide mass fingerprinting*“ [38-40], oder mit Hilfe von sogenannten „*sequence tags*“ [41,40] oder Fragmentationen „*fingerprints*“ [43-45] durchgeführt.

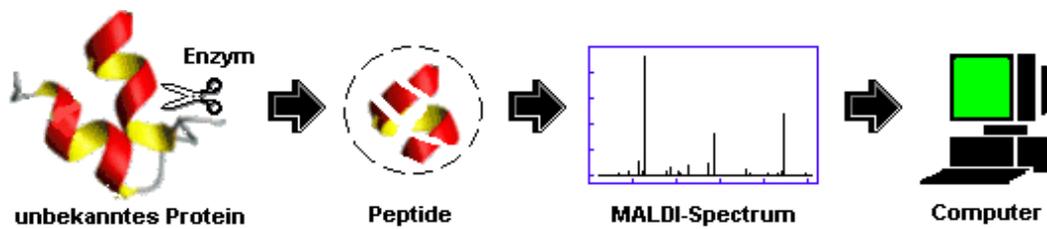


Abbildung 6 Identifizierung eines Proteins mittels MALDI-MS „*peptide mass fingerprinting*“ mit anschließender Datenbanksuche

Für die Analyse von Proteinen mittels „*peptide mass fingerprinting*“ wird das Protein zunächst nach oder vor spezifischen Aminosäureresten enzymatisch gespalten. Die Enzyme, die dafür eingesetzt werden sind vor allem Trypsin und LysC. Nach der Spaltung werden die Molekülmassen der entstandenen Peptide mit MALDI-MS bestimmt. Die Identifikation des Proteins erfolgt durch Abgleich der gemessenen Peptidmassen mit denen für das jeweilige Enzym theoretisch berechneten Peptidmassen für alle in einer Datenbank enthaltenen Proteinsequenzen (Abbildung 6). Voraussetzung dafür ist, dass das Protein sequenziert ist und in der Suchdatenbank eingetragen ist.

Dem Wissenschaftler stehen dazu eine Vielzahl frei zugänglicher Proteinsequenzdatenbanken zur Verfügung (Tabelle 3).

Tabelle 3 Liste einiger im Internet frei verfügbaren Proteinsequenzdatenbanken.

Proteindatenbank	Beschreibung	Einträge (07/2001)	Provider
SWISS-PROT	Die zuverlässigste und übersichtlichste verfügbare Datenbank.	98739	European Bioinformatics Institute (EBI)
NCBI non-redundant	Setzt sich aus mehreren Teildatenbanken, SWISS-PROT, GENPEPT u.a. zusammen.	686213	NCBI (USA)
TrEMBL	Umfasst codierte Sequenzen aus der EMBL Nucleotide Datenbank.	471750	European Molecular Biology Laboratories (EMBL)/EBI

Proteindatenbank	Beschreibung	Einträge (07/2001)	Provider
GENPEPT	Translation aller codierenden Sequenzen aus der GenBank Nukleotid Datenbank	718237	NCBI (USA)

Zur eindeutigen Identifizierung des gesuchten Proteins wird eine Mindestanzahl von gefundenen Peptidmassen vorausgesetzt. Bei der Verwendung von Trypsin oder LysC als Spaltenzym wird meist eine maximale Abweichung von 50 ppm für die Massenrichtigkeit gefordert [46].

1.4 Proteomanalyse und Automatisierung

Um ein Proteom in überschaubaren Zeiträumen zu erforschen, ist eine Automatisierung der Proteinidentifizierungsmethoden unumgänglich [47]. Eine schnelle Hochdurchsatzanalyse von mehreren tausend Proteinen, die sich nach Auftrennung und Färbung auf einem einzigen Gel als „Spots“ wiederfinden, stellt hohe zeitliche Anforderungen an die Proteinspotentnahme, Probenpräparation, massenspektrometrische Analyse und die Interpretation der generierten Daten.

1.4.1 Proteinspotentnahme

Das Ausstechen der Proteinspots aus dem Gel ist ein zeitaufwendiger Schritt. In unserem Labor wurde zur Automatisierung dieses Schrittes eine Ausstechrobotoranlage entwickelt, mit deren Hilfe im Durchschnitt 8 Proteinspots (Abbildung 7b) pro Minute aus dem Gel ausgestochen und zur weiteren Analyse in eine Mikrotiterplatte überführt werden können (Abbildung 7a) [47].

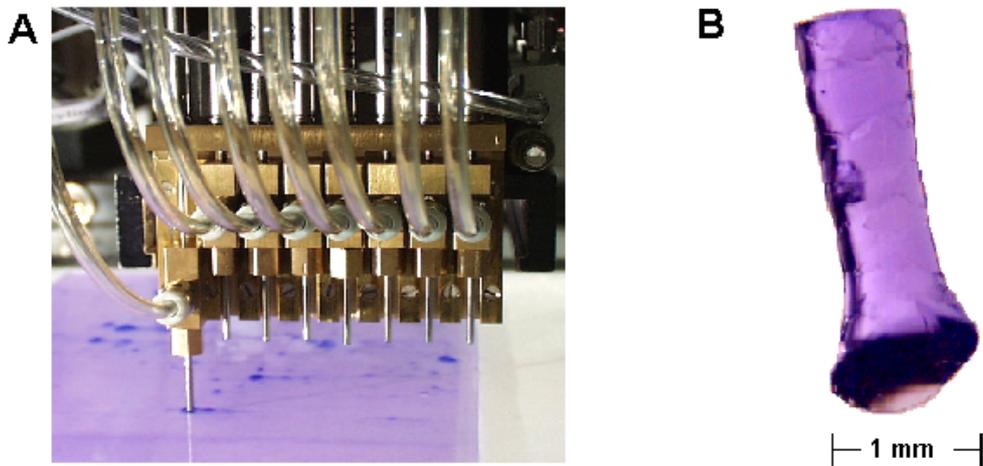


Abbildung 7 (A) Ausstechkopf des im Hause entwickelten Ausstechroboters. Nachdem der Roboter 8 Spots ausgestochen hat, werden die Gelstückchen in eine Mikrotiterplatte abgegeben, und der Ausstechvorgang beginnt von vorne. In einer Minute werden durchschnittlich 8 Proteinspots ausgestochen und in die Mikrotiterplatte überführt.
(B) Ausgestochenes Gelstückchen

Die Zugabe von Reagentien und Puffern für den „*in-situ*“ Verdau der ausgestochenen Gelproben erfolgt über eine im Hause entwickelte Dosierstation (Abbildung 8). Diese besitzt 8 Dosierkapillaren (250 μm Durchmesser), die unabhängig voneinander über ein eigenes Magnetventil (INKA4005212H; The Lee Company, USA) geöffnet und geschlossenen werden können. Mit dieser Anlage können bei Bedarf nur wenige nl oder bis zu 30 μl Lösung pro Sekunde dispensiert werden. Die Mikrotiterplatte (MTP) wird hierzu unter den Dosierkapillaren mit Hilfe eines computergesteuerten x-y Tisches (LS30-250, Cybertron GmbH, Germany) vorbeibewegt. Neben der hohen Geschwindigkeit mit der die Flüssigkeitszufuhr erfolgt, wird im Gegensatz zu den bei konventionellen Techniken üblichen Verwendung von Pipettenspitzen eine Berührung der Dosierkapillaren mit den Proben vermieden und dadurch die Gefahr einer Kontamination verringert.

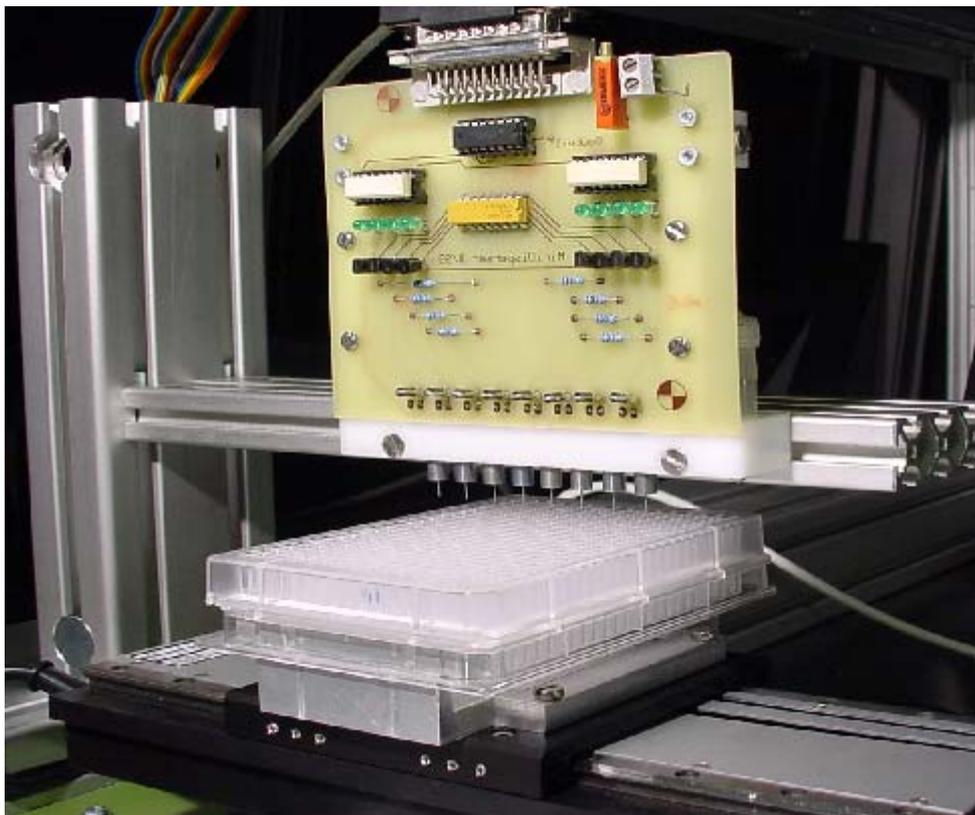


Abbildung 8 Im Hause entwickelte Dosierstation für die Zugabe von Reagentien und Puffern zu den in einer MTP befindlichen Gelproben.

1.4.2 Probenpräparation

Bei der Automatisierung der Probenpräparation stehen in erster Linie Reinheit und Ausbeute der Proteine, sowie Reproduzierbarkeit und niedrige Kosten der Methode im Vordergrund. In unserem Labor wurde dazu ein Aufreinigungsprotokoll entwickelt, dass parallel 384 Proben in nur wenigen Schritten aufarbeitet [48,49]. Die Probenpräparation wird auf einem vorstrukturierten MALDI-Probenträger (Abbildung 9-A) durchgeführt. Die verwendete kristalline Matrix, α -Cyano-4-hydroxymethylsäure (CHCA), besitzt in saurer Lösung eine hohe Bindungsaffinität zu Peptiden. Die Stahloberfläche des MALDI-Probenträgers ist mit einer ca. 5 μ m dünnen, teflonartigen Schicht (AnchorChip™ 400/384, Bruker Daltonik Bremen), die in

definierten Abständen kreisförmig (Durchmesser 400µm) unterbrochen ist, überzogen. Die an diesen Stellen freiliegende Stahloberfläche wirkt als hydrophiler Anker (Abbildung 9-B).

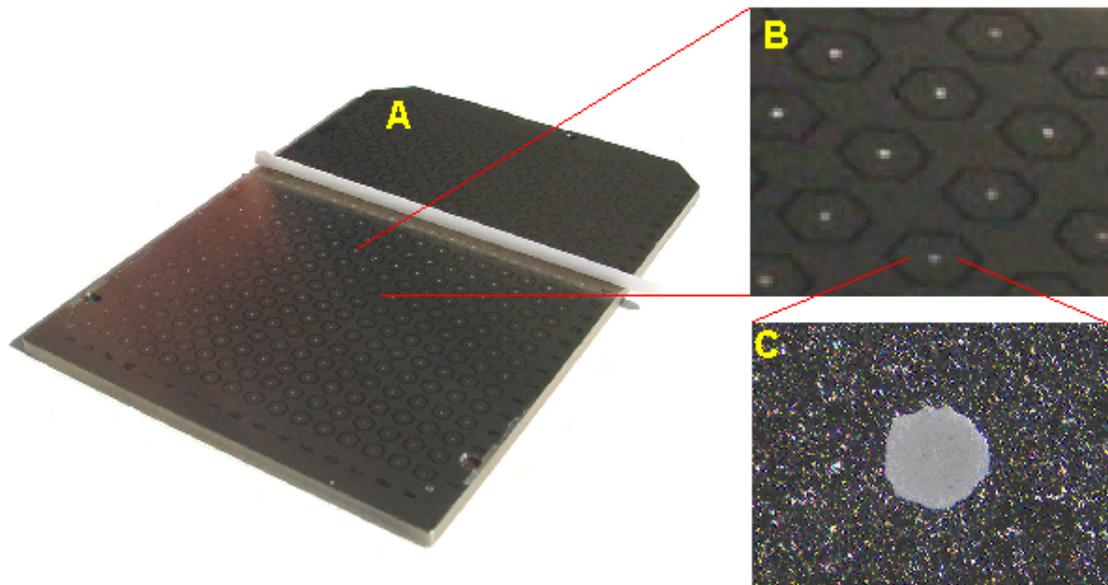


Abbildung 9 Probenpräparation auf einem vorstrukturierten MALDI-Probenträger.
Die CHCA-Matrixlösung wird mit einem Teflonstab auf dem MALDI-Probenträgers verteilt (A). Nach Verdunsten des Lösungsmittel entsteht exklusiv an den hydrophilen Ankern des Probenträgers(B) ein mikrokristalliner Film der Matrix(C).

Die Matrixlösung wird mit Hilfe eines Teflonstabes über den außerhalb der hydrophilen Ankern nicht benetzbaren MALDI-Probenträger gezogen (Abbildung 9-A). Bei diesem Vorgang bleibt an jedem hydrophilen Anker (in der Regel 400 µm) ein gleichgroßes, nur wenige nl-umfassendes Aliquot der Matrixlösung hängen. Nach Verdunsten des Lösungsmittels bildet sich exklusiv an den vorhandenen hydrophilen Ankern (Abbildung 9-B) ein mikrokristalliner Matrixfilm (Abbildung 9-C). Aliquote der Probenlösung werden anschließend mit Hilfe einer automatischen 96er Pipettierstation (CyBi well, Cy-Bio, Germany) von einem Proben-MTP auf den MALDI-Probenträger übertragen (Abbildung 10).



Abbildung 10 Übertragung der Proben von einer MTP auf den MALDI-Probenträger.

Nach drei Minuten Inkubationszeit, während der die Peptide sich an die Matrix binden, wird der Überstand abpipettiert und restliche Kontaminationen werden durch Waschen in 0.1%ige TFA entfernt. Danach sind die Proben bereit zur Messung [48].

1.5 Bioinformatik

Die Bioinformatik ist eine interdisziplinäre Wissenschaft, die eine Schnittstelle zwischen der Biochemie und der Informatik darstellt. Sie beinhaltet die Analyse, Organisation und Interpretation biologischer Daten [50,51].

1.5.1 Analyse massenspektrometrischer Daten

Die unter 1.3 zur eindeutigen Identifizierung eines Proteins in großen Datenbanken geforderte Massenrichtigkeit von mindestens 50 ppm ist nur unter Verwendung von internen Kalibranten zu realisieren. Ein großer Nachteil dieser Vorgehensweise ist jedoch, dass die Signale des Kalibranten die Signale des Analyten überdecken können und diese damit verloren gehen. Der umgekehrte Fall, dass Analytsignale die Signale des Kalibranten überdecken macht die Messung oft gänzlich unbrauchbar, da eine korrekte

Kalibrierung nicht mehr möglich ist. Bei der Verwendung einer externen Kalibrierung beträgt die Massenrichtigkeit nur im günstigsten Fall, d.h. die Probe befindet sich in unmittelbarer Nachbarschaft zu den Kalibriersubstanzen, typischerweise 100 ppm. Im ungünstigsten Fall, d.h. die Probe befindet sich an einer endständigen Position und die Kalibriersubstanzen z.B. in der Mitte des Probenträgers, können bestimmte Massen von den richtigen Werten um bis zu 500 ppm abweichen.

Die für eine hohe Massenrichtigkeit erforderliche exakte Bestimmung der Kalibrierungskonstanten ist von der Matrix/Laser-Kombination, von Geräteparametern wie Laserpulsenergie und Beschleunigungsspannung [52,53], der Art der zu untersuchenden Probe (DNA, Proteine, Peptide), der Probenpräparation (Kristallisationsbedingungen), von äußeren physikalischen Parametern (Temperatur, Luftdruck, Luftfeuchtigkeit) und von der Position an der sich die zu untersuchende Probe auf dem MALDI-Probenträger befindet [54] abhängig.

Da die Kalibrierungskonstanten, die zuvor mit Hilfe von Standardverbindungen mit genau bekannten Massen bestimmt wurden, aufgrund der oben genannten Einflüsse in der Praxis zeitlich variieren, sind alle bestimmten Massen mit mehr oder weniger großen Fehlern behaftet.

Dabei lassen sich vor allem zwei systematische Fehlerarten unterscheiden. Zum einen können alle Peaks in einem Massenspektrum um einen konstanten Wert in eine Richtung verschoben sein. Im Folgenden wird dies als Nullpunktverschiebung bezeichnet (Abbildung 11). Zum anderen kann das Spektrum und damit der Abstand zwischen zwei benachbarten Peaks unabhängig von der Nullpunktverschiebung um einen bestimmten Faktor gestreckt oder gestaucht sein (Abbildung 12). In der Praxis treten meist beide Effekte nebeneinander auf, d.h. der Nullpunkt des Spektrums ist verschoben und das Spektrum ist entlang der m/z -Achse linear gestreckt oder gestaucht.

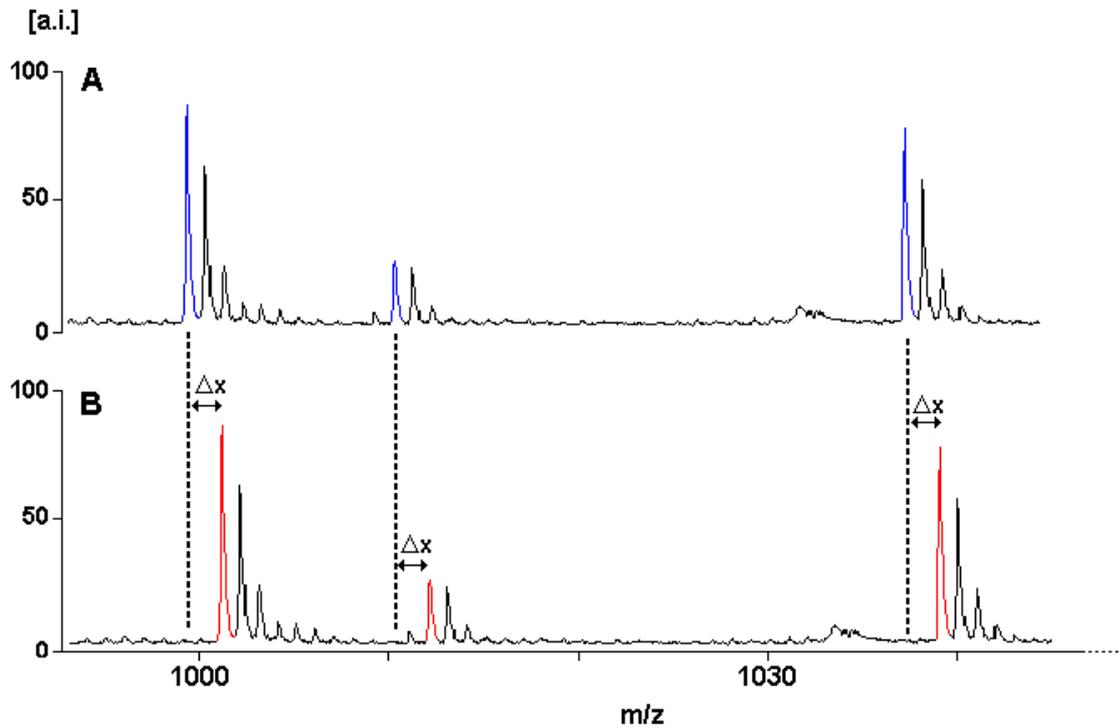


Abbildung 11 Schematische Darstellung der Nullpunktverschiebung.

Die Massenspektren A und B zeigen je drei Peptidsignale inklusive ihrer Signalfeinstrukturen (Isotopie). Jeder Peak in Spektrum B ist gegenüber dem korrespondierenden Peak in Spektrum A um den jeweils gleichen Betrag Δx zu höheren m/z -Werten verschoben. Die Verschiebung hin zu kleineren Werten (nicht abgebildet) wird ebenfalls beobachtet.

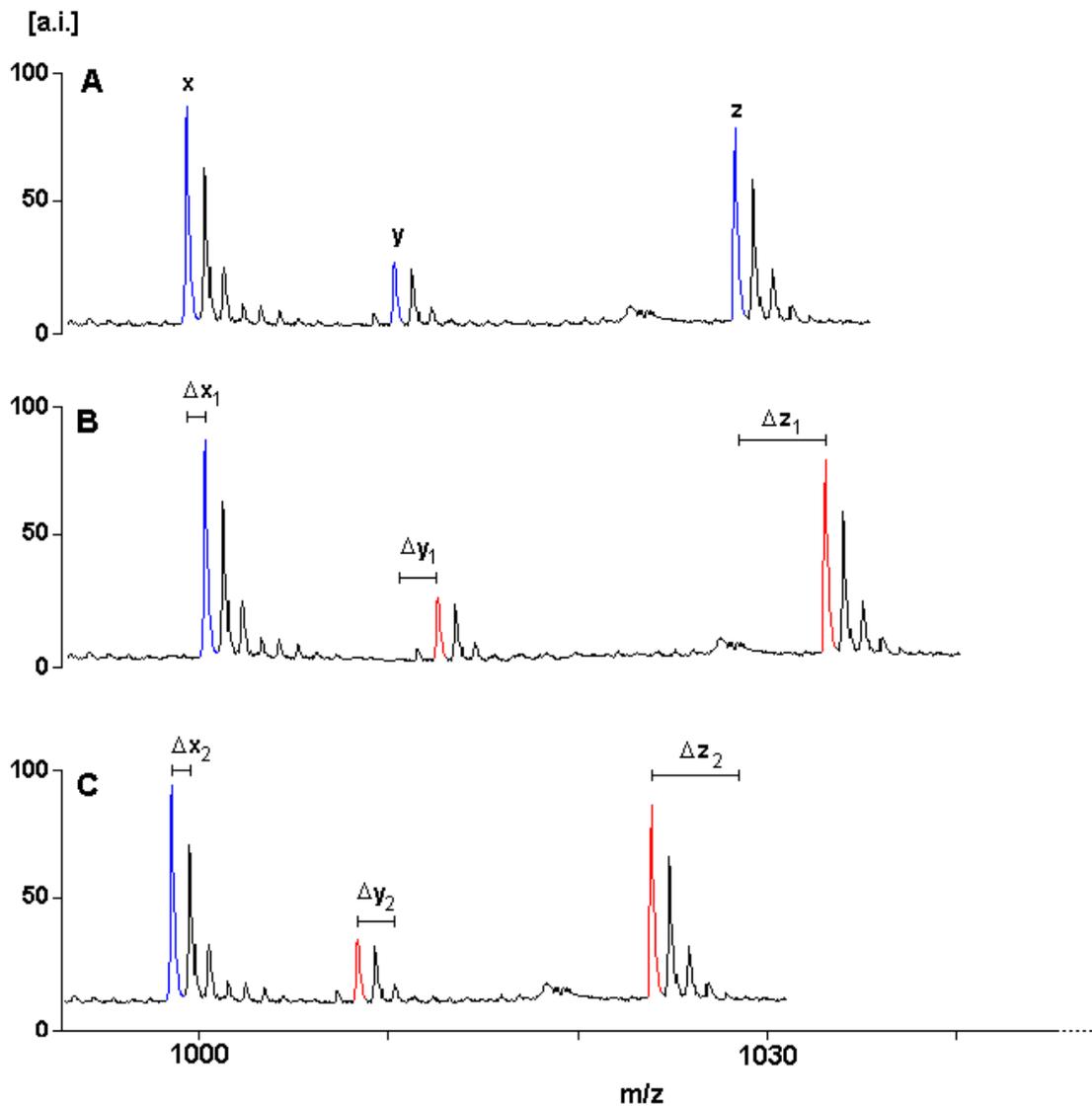


Abbildung 12 Schematische Darstellung der Stauchung bzw. Streckung eines Massenspektrums.

Die Massenspektren A, B und C zeigen je drei Peptidsignalserien. In Spektrum B sind die Peaks gegenüber Spektrum A um von links nach rechts linear zunehmende Beträge ($\Delta x_1 < \Delta y_1 < \Delta z_1$) zu höheren m/z-Werten hin verschoben (Streckung des Spektrums). Die Verschiebung hin zu kleineren m/z-Werten ist in Spektrum C gezeigt, d.h. das Spektrum ist gestaucht.

In dieser Arbeit werden von mir entwickelte Strategien und Algorithmen vorgestellt, die sich dieser Beobachtung bedienen und dadurch eine eindeutige Identifizierung von Proteinen unter Verwendung einer externen Kalibrierung ermöglichen [55].

1.5.2 Datenorganisation

Bei einer Proteomanalyse fallen sehr viele Daten an, deren Erfassung, Zuordnung und Darstellung ohne bioinformatische Methoden nicht zu realisieren ist. Hierzu gehören z.B. das automatische Einlesen, Analysieren und Auswerten von Massenspektren, die Datenbanksuche sowie eine übersichtliche Darstellung der Ergebnisse.

In dieser Arbeit wird ein von mir entwickeltes Programmpaket vorgestellt, dass diese Aufgaben erfüllt.

1.5.3 Dateninterpretation

Eine große Herausforderung bei der Identifikation von Proteinen in Datenbanken ist die Zuverlässigkeit der Ergebnisse. Da es keine eindeutige Ja/Nein-Antwort auf die Frage gibt, ob es sich bei dem in der Datenbank gefundenen Protein tatsächlich um das untersuchte Protein handelt, ist für die Bewertung der Ergebnisse die Verwendung eines geeigneten „*Scorings*“ von entscheidender Bedeutung.

Bei einem Identifizierungsexperiment treten prinzipiell drei mögliche Szenarien auf:

- 1) Das Protein wird richtig identifiziert: richtiges Ergebnis.
- 2) Es wird ein falsches Protein identifiziert: falsch positives Ergebnis.
- 3) Es wird kein Protein identifiziert: falsch negatives Ergebnis.

In der Literatur sind drei verschiedene Algorithmen beschrieben, die hinsichtlich der Frage ob ein untersuchtes Protein als eindeutig identifiziert zu betrachten ist, unterschiedliche Ansätze zugrunde legen.

Der einfachste Ansatz berücksichtigt die Anzahl der Treffer, d.h. die in der Proteinsequenzdatenbank gefundenen Peptidmassen, die innerhalb eines bestimmten, vorher festgelegten Fehlerintervalls mit den gemessenen Peptidmassen übereinstimmen, bestimmen das Kriterium wonach sortiert wird [56]. Dabei wird angenommen, dass ein Protein als eindeutig identifiziert gilt, wenn es eine Mindestanzahl von Treffern aufweist und gegenüber dem nächst folgenden Protein mindestens drei Treffer mehr erzielt.

Ein weiterer Ansatz ist der MOWSE (MOlecular Weight SEarch) Algorithmus [57,58], der größere Massen stärker gewichtet als kleinere. Eine ausführliche Beschreibung des Algorithmus und Informationen zum Aufbau einer MOWSE Datenbank sind unter „[http:// www.hgmp.mrc.ac.uk/Bioinformatics/Webapp/mowse/mowседoc.html](http://www.hgmp.mrc.ac.uk/Bioinformatics/Webapp/mowse/mowседoc.html)“ zu finden. Ein dritter Ansatz beruht auf dem Bayesschen Satz für zwei abhängige Ereignisse [59-61].

Diese Algorithmen setzen zur eindeutigen Identifizierung von Proteinen in großen Datenbanken jedoch eine Massenrichtigkeit von mindestens 50 ppm voraus.

Um bei Verwendung einer externen Kalibrierung eine möglichst hohe Massenrichtigkeit zu erzielen, muss auf dem MALDI-Probenträger in unmittelbarer Nachbarschaft jeder Probe der Peptidstandard aufgetragen, gemessen und die entsprechenden Kalibrierungskonstanten für jede Probe neu bestimmt werden. Diese Vorgehensweise ist fehleranfällig und daher für eine Automatisierung der Proteinidentifikation, d.h. für die Analyse von mehreren hundert Peptidgemischen im „batch“- Verfahren nicht geeignet. Im „batch“- Verfahren wird bevorzugt eine einmal ermittelte Kalibrierung für alle Proben verwendet (z.B. eine Kalibrierung für 384 Proben).

Diese Verfahrensweise verschlechtert jedoch die Massenrichtigkeit auf bis zu 500 ppm. Da die etablierten „*Scoring*“-Algorithmen hierfür nicht verwendet werden können war die Entwicklung und Etablierung eines neuen, auf diese bereits genannten Bedingungen angepassten „*Scoring*“-Algorithmus erforderlich.

Neben dem Problem der zu erwartenden maximalen relativen Massenabweichung, galt es hierbei zu berücksichtigen, dass bei einer Proteomanalyse die pro Protein verfügbaren Stoffmengen dramatisch schwanken. Dies bedingt, dass die Qualität der aufgenommenen Peptid-Massenspektren, insbesondere das Signal-zu-Rauschverhältnis hochsignifikant variiert. Neben Spektren hoher Qualität, wie sie in der Regel gewonnen werden, wenn das Ausgangsmaterial mehr als 1 pmol umfaßt (z.B. stark gefärbter Spot auf einem mit Coomassie G250 behandelten 2D-Gel), weisen viele Spektren eine mittlere Qualität (Ausgangsmenge: > 300 fmol und < 1 pmol, mittelstark gefärbter Spot) und andere eine schlechte Qualität (Ausgangsmenge: < 300 fmol, schwach gefärbter Spot) auf. In die letzte Kategorie fallen naturgemäß die gering exprimierten und daher oft besonders interessanten, da wenig erforschten, Proteine. In Abbildung 13 sind die drei diskutierten Beispiele gegenüber gestellt.

Die Massenrichtigkeit ist gekoppelt mit dem Signal-zu-Rauschverhältnis des zugrunde liegenden Peaks. Bei geringem Signal-zu-Rauschverhältnis (< 5) sind die Peaks verrauscht und oft unsymmetrisch. Dies hat zur Folge, dass die ermittelten Flugzeiten (Centroide des Peaks) einen größeren Fehler aufweisen. Daraus folgt, dass die Massenrichtigkeit neben den instrumentellen Parametern vor allem auch durch die präparierten Probenmengen beeinflusst wird. Da unterschiedliche Peptide bei der MALDI-MS aufgrund unterschiedlicher Einbauraten während der Probenpräparation (Einbau in die Matrixkristalle) und/oder unterschiedlicher Ionenausbeuten, mit unterschiedlicher Empfindlichkeit nachgewiesen werden, variiert das Signal-zu-Rauschverhältnis und damit die Massenrichtigkeit auch innerhalb eines MALDI-Massenspektrums erheblich. Beispiele hierfür finden sich ebenfalls in Abbildung 13. All diese Schwankungen galt es durch ein flexibles „*Scoring*“-Verfahren, d.h. ein Algorithmus der die Qualität der Rohdaten berücksichtigt, zu erfassen.

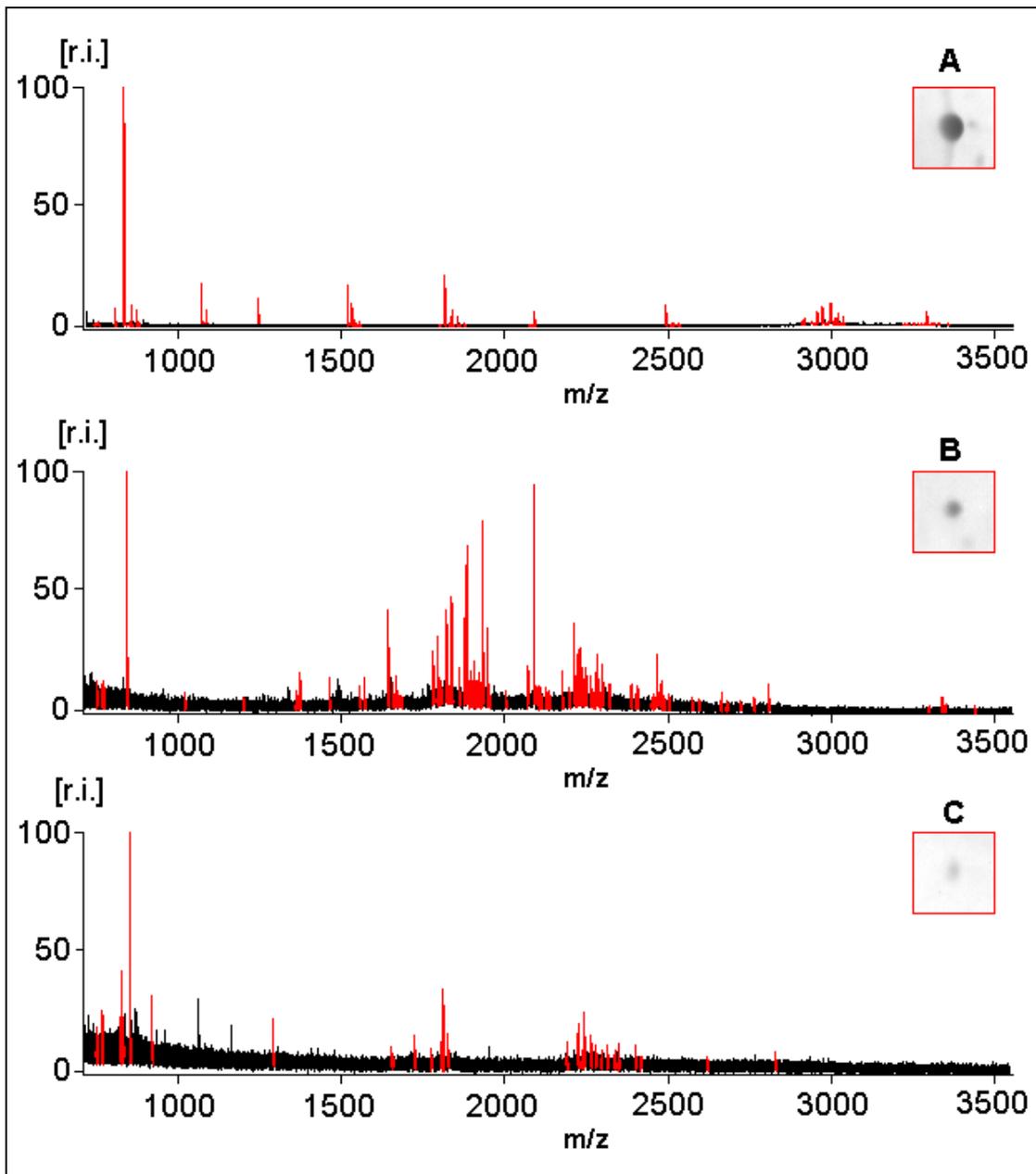


Abbildung 14 Abhängigkeit des Signal-zu-Rauschverhältnisses von der Spotgröße.

(A) Stark gefärbte Spots liefern qualitativ hochwertige Spektren, d.h. die Peptidsignale lassen sich eindeutig von den Hintergrundsignalen unterscheiden. (B) Bei der Analyse von mittelstark gefärbten Spots verschwinden bereits einige Peptidpeaks im Rauschen. (C) Bei schwach gefärbten Spots ist die Unterscheidung zwischen Peptidsignal und Hintergrundsignal schwierig und viele Signale gehen im Rauschen verloren.

In dieser Arbeit wird ein von mir entwickelter „Scoring“-Algorithmus beschrieben, der die Qualität der Rohdaten automatisch erfasst und die Zuver-

lässlichkeit der Identifizierung von Proteinen gewährleistet, auch wenn die gemessenen Massen von den erwarteten Peptidmassen um bis zu 500 ppm abweichen und der sich selbst, d.h. ohne Mitwirken des Menschen, an Schwankungen in der Qualität der Rohdaten anpaßt [62].

2 Zielsetzung der Arbeit

Das Ziel der Arbeit bestand darin, Algorithmen zu entwickeln, welche

1. die Proteinidentifikation mittels MALDI-TOF-MS „*peptide mass fingerprinting*“ verbessern und
2. die Daten, die während einer proteomweiten Analyse eines Organismus anfallen, automatisch erfassen, analysieren und geeignet darstellen.

Im Einzelnen ergeben sich folgende Aufgaben:

Wie kann durch Etablierung neuer Algorithmen die Datenerfassung und Interpretation einer MALDI-TOF-MS-Analyse von komplexen Peptidgemischen unter Verwendung einer externen Kalibrierung so weit verbessert werden, dass auf die Anwendung einer internen Kalibrierung gänzlich verzichtet werden kann. Diese Algorithmen sollen die Qualität der massenspektrometrischen Rohdaten automatisch erfassen, bewerten und die Kriterien für die Proteinidentifikation entsprechend anpassen.

Eine weitere Aufgabe besteht darin, einen „*Scoring*“-Algorithmus zu entwickeln, der die Identifikation von Proteinen in großen Sequenzdatenbanken sicherer macht.

Die Entwicklung eines Softwarepakets, das in kürzester Zeit eine große Anzahl von Peptidspektren einliest, die Proteinidentifikation durch Abgleich mit einer „*on the fly*“ aus einer ausgewählten Proteindatenbank berechneten Peptidsequenzdatenbank vornimmt und die jeweiligen Ergebnisse sinnvoll und übersichtlich darstellt. Weiterhin soll die Software, 2DE-Gelbilder einlesen und die Proteinspots den Ergebnissen aus der Datenbanksuche zuordnen.