

1. Einleitung

1.1. Die Massenspektrometrie in der Proteomforschung

Im Jahre 1909 publizierte der englische Arzt Garrold die Hypothese, daß angeborene Stoffwechselfehler auf das Fehlen spezifischer Enzyme zurückgehen, die normalerweise unter Anweisung der Wildtyp-Erbanlagen hergestellt werden (Garrold, (1909)). Obwohl diese Arbeit kaum beachtet wurde, war somit bereits die „ein-Gen*-ein-Enzym-Hypothese“ geboren. Nach Experimenten mit dem Pilz *Neurospora crassa* formulierten Beadle und Tatum ihre „Ein-Gen-ein-Protein** - Hypothese“, nach der ein definierter DNA-Abschnitt die Information für ein Protein enthält (Beadle & Tatum, (1941)). Noch im selben Jahrzehnt führte Pauling die Ursache der Sichelzellanämie auf einen genetisch bedingten Defekt im Hämoglobinmolekül zurück (Pauling, Itano *et al.*, (1949)). Später stellte man fest, daß Proteine wie z.B. Hämoglobin aus mehreren Polypeptiduntereinheiten aufgebaut sein können (Ingram, (1989)), für die funktionell zusammenhängende genetische Einheiten (Cistron, äquivalent zu Gene) kodieren. Die ursprünglichen Theorien wurden zur „ein-Cistron-ein-Enzym-Theorie“ (Benzer, (1955)) bzw. allgemeiner zur „ein-Gen-ein-Polypeptid*** - Hypothese“ präzisiert.

Tatsächlich gestaltet sich das Zusammenspiel von Genen und Proteinen wesentlich komplexer. Erst kürzlich wurde z.B. ein einzelnes eukaryontisches DNA-Transkript (= mRNA) entdeckt, mit dem ribosomal zwei verschiedene Polypeptide synthetisiert werden (Reiss, Cohen *et al.*, (1998)). Es bestehen also auch Ausnahmen von der vormals allgemeingültig angesehenen „ein-Cistron-ein-Enzym-Theorie“. Die Auswirkung und Regulation genetischer Information erfolgt auf mehreren Ebenen: der DNA, der RNA, der RNA-Ribosomen-Wechselwirkung, den Vorläuferproteinen (*protein precursors*) und der Proteine. Proteine werden oft mehrfach prozessiert und modifiziert, bis sie ihre vorbestimmte Funktionalität und Lokalisierung erhalten. Zur Aufklärung dieses komplexen Netzwerkes bildet die Kenntnis der DNA Sequenzen aller Gene eines Organismus, das „Genom“, die Grundlage. In Anlehnung an das „Genom“ wurde 1994 der mittlerweile populäre Begriff „Proteom“ von Wasinger *et al.* vom „Center for Analytical Biotechnology“ an der Macquarie Universität im australischen Sydney geprägt (Wasinger, Cordwell *et al.*, (1995)). Unter einem Proteom versteht man den quantitativen Expressionszustand sämtlicher **Proteine**, die von einem **Genom** kodiert werden. Im Gegensatz zum Genom beschreibt ein Proteom einen dynamischen Zustand. Eine wesentliche Aufgabe bei der Analyse eines Proteoms besteht in der Auftrennung, Quantifizierung und Identifizierung der Proteine eines Organismus, eines Gewebetyps, isolierter Zellen eines Typs oder bestimmten Körperflüssigkeiten unter definierten Wachstums- und Umweltbedingungen. Die Identifizierung entspricht (grob) der genetischen Zuordnung der Proteine. Die von Klose und O'Farrell unabhängig von einander entwickelte zweidimensionale Gelelektrophorese (2-DE) stellt nach wie vor das derzeit leistungsstärkste Werkzeug zur Auftrennung einer Vielzahl ($\geq 10^4$) Proteine dar (Klose, (1975), O'Farrell, (1975)). Mit dieser Methode werden Proteine in einer Polymermatrix zuerst durch isoelektrische Fokussierung nach ihrer Basizität / Acidität und anschließend nach ihrem Molekulargewicht (MW) getrennt. Die Proteine können dann durch Färbeverfahren, Fluoreszenz oder radioaktive Markierung als diskrete Spots detektiert werden. Auf der Ebene der exprimierten Proteine erlauben reproduzierbar erstellte Spotmuster beispielsweise den Vergleich von kranken und gesunden Geweben, Zellmaterial verschiedener Entwicklungsstadien oder pathogenen und harmlosen Stämmen von Mikroorganismen einer Art.

Eine wesentliche methodische Schwierigkeit besteht in der Identifizierung der getrennten Proteine, die meist nur in geringen Mengen (pmole bis fmole) erhalten werden. Anhand der Geldaten läßt

* „Gen“ - von griech.: γένος = Geschlecht, Gattung, Nachkommenschaft

** „Protein“ - von griech.: πρῶτεύω = „ich stehe an erster Stelle“

*** „poly-“ von griech.: πολυ = „viel“ und „Peptid“ von πεπτός = verdaulich

sich zwar das ungefähre Molekulargewicht einiger Proteine abschätzen, wie u.a. in dieser Arbeit gezeigt, kann dies jedoch zu falschen Annahmen führen (s. Abschnitt 3.5.1). Ebenfalls problematisch sind Interpretationen des Proteinmusters, welches durch die isoelektrische Fokussierung entsteht (s. z.B. Abschnitt 0). Die durch Edman-Sequenzierung gewonnenen Sequenzinformationen können zur Identifizierung verwendet werden (Edman & Begg, (1967)). Zahlreiche Proteine sind jedoch N-terminal modifiziert, oder nur in sehr geringen Mengen vorhanden, was die Sequenzierung nach Edman erschwert.

Ist das Genom des untersuchten Organismus bekannt, so kann die Proteinidentifizierung auf elegante Weise anhand von charakteristischen Abbaupeptiden („*peptide fingerprints*“) erzielt werden. Hierzu gewinnt man durch chemische oder enzymatische Proteolyse ein Gemisch aus Spaltpeptiden aus den isolierten Proteinen. Durch eine massenspektrometrische Analyse dieses Peptidgemisches werden die Massen der enthaltenen Abbaupeptide bestimmt. Mit den Peptidmassen lassen sich Datenbanken berechneter Massen bekannter oder theoretischer Proteine durchsuchen (Bleasby & Wootton, (1990) Cottrell, (1994)). Je nach Massengenauigkeit, Proteinreinheit und Proteinfamilie können Proteine bereits aufgrund weniger Daten (3-10 Peptidmassen) verlässlich identifiziert werden (Pappin, Højrup *et al.*, (1993), Mann, Højrup *et al.*, (1993), James, Quadroni *et al.*, (1993)). In den achtziger Jahren wurden Peptide massenspektrometrisch vorrangig mit der „*fast-atom bombardment*“-Massenspektrometrie (FAB-MS) und mit „²⁵²Cf Plasma Desorptions“- Massenspektrometrie (PD-MS) analysiert (*review* s. Lehmann, (1991)). Beide Methoden waren jedoch infolge ihrer geringen Effizienz bei der Ionenbildung, dem eingeschränkten Massenbereich ionisierbarer Peptide und auch rein praktisch, durch den erforderlichen hohen Zeitaufwand bei der Durchführung, limitiert.

Den Durchbruch für die Proteomforschung („*proteomics*“) erzielte die **Matrix-assistierte Laser Desorptions / Ionisations Massenspektrometrie (MALDI-MS;** s. Karas & Hillenkamp, (1988) Tanaka, Waki *et al.*, (1988)) und die **Elektrospray Ionisations Massenspektrometrie (ESI-MS;** s. Fenn, Mann *et al.*, (1989); *vide infra*, Abschnitt 1.1.1 und 1.1.2). Durch Fragmentierung der Peptidionen können mit beiden Methoden auch Sequenzinformationen erzeugt werden, die insbesondere in zweifelhaften Fällen (d.h. bei nicht zuzuordnenden Peptidmassen) zur Identifizierung hinzuziehbar sind (s. z.B. Abschnitt 3.4). Außerdem lassen sich posttranslationale Modifizierungen nachweisen und lokalisieren (s. z.B. Abschnitte 3.4.1 bis 3.4.3 und 3.5.5).

1.1.1. **Matrix-assistierte Laser Desorptions/Ionisations Massenspektrometrie (MALDI-MS)**

Für die MALDI-MS-Analyse wird die zu untersuchende hochmolekulare Probe (z.B. Gemische von Polymeren oder Biopolymeren wie Peptide, Proteine oder Nukleinsäuren) auf einem metallischen Probenträger (Target) mit einer niedermolekularen Matrix kokristallisiert, und in das Hochvakuum des Massenspektrometers überführt. Die Matrix, die in der Regel im bis zu 10⁴ fachen molaren Überschuss eingesetzt wird, kann die Lichtenergie eines eingestrahnten, gebündelten Laserpulses optimal absorbieren und dabei schlagartig ins Vakuum verdampfen (s. Abbildung 1-1). Wird ein Stickstofflaser eingesetzt, der bei 337 nm emittiert, so sind aromatische Matrixverbindungen besonders gut geeignet. Für Proteine und Peptide haben sich u.a. Zimtsäurederivate wie Sinapinsäure (= SA; s. Beavis & Chait, (1989)) oder 4-Hydroxy- α -cyanozimtsäure (= HCCA; s. Beavis, Chaudhary *et al.*, (1992)), aber auch Benzoesäurederivate wie 2,4-Dihydroxybenzoesäure (Huth-Fehre, Gosine *et al.*, (1992)) bewährt. Es wird angenommen, daß Analytmoleküle wie z.B. Peptide nach der Laserirradiation von der expandierenden Matrixwolke (*expanding plume*) mitgerissen und durch Matrixfragmentradikale chemisch ionisiert werden (Vertes, Irinyi *et al.*, (1993)). Hierbei kommt es hauptsächlich zur einfachen Protonierung ([M+H]⁺) oder Deprotonierung ([M-H]⁻) der Analytmoleküle (M). Im positiven Meßmodus ist eine positive, in negativen Modus eine negative Beschleunigungshochspannung (U_{acc} = 10-30 kV) an das Target

gelegt. Im elektrischen Feld werden dann (z.B. positiv) geladene Molekülonen in Richtung einer Lochplatte beschleunigt, die sich auf Massepotential (0 V) befindet. Die Ionen fliegen dabei, nach Passage eines fokussierenden Ionenlinsensystems, durch die Lochplatte in das Flugrohr. Entlang einer feldfreien Driftstrecke separieren sich nun die Ionen aufgrund ihrer unterschiedlichen Fluggeschwindigkeiten, die durch ihr jeweiliges m/z Verhältnis und die Stärke des Beschleunigungsfeldes bestimmt ist. Ein Detektor registriert das Eintreffen der Molekülonen am Ende der Driftstrecke. Mit einem Oszilloskop wird die Flugzeit (TOF = *time of flight*) der Ionen anhand von Laser („Start“) und Detektor-Ereignissen („Stops“) gemessen. In erster Näherung ist die Flugzeit proportional zur Wurzel des m/z Verhältnisses ($\text{TOF} \sim \sqrt{m/z}$).

Zur Erhöhung der massenspektrometrischen Auflösung ($m/\Delta m$) sind moderne MALDI-Massenspektrometer zusätzlich mit einem Reflektor ausgerüstet. Dieser befindet sich am Ende des Flugrohres und besteht aus mehreren Metallringen oder Gittern, an denen mehrstufige Reflektorpotentiale U_{ref} angelegt sind. Ionen gleichen m/z mit unterschiedlichen Endgeschwindigkeiten, wie sie z.B. aus verschiedenen Anfangsgeschwindigkeiten und Ionisations-Orten herrühren können, dringen unterschiedlich tief in das zunehmende elektrische Feld des Reflektors ein. Dieses Feld beschleunigt die Ionen zurück und bildet sie fokussiert auf dem zweiten Detektor (*reflektormode detector*) ab (s. Abbildung 1-1).

Ein Teil der beschleunigten Ionen zerfällt bereits während des Fluges in Fragmentationen und Neutrale nahezu gleicher Geschwindigkeit (metastabiler Zerfall). Dieses „Paket“ aus Fragmenten führt im linearen Modus ebenfalls zu einem verbreiterten Detektorsignal. Im „regulären“ Reflektormodus werden dagegen nur die Mutterionen und die größeren Fragmente, die z.B. durch Wasserabspaltung entstanden sind, zeitlich getrennt auf den Detektor abgebildet (s. z.B. Spektren in den Abschnitten 3.5.1, 3.5.2, ff.). Durch Absenkung des Reflektorpotentials unter das Beschleunigungspotential können auch leichtere Fragmentationen gezielt detektiert werden. Dieses Prinzip (*post source decomposition / decay* = PSD) läßt sich zur Gewinnung von Sequenzinformationen anhand von Fragmentationenspektren nutzen (Spengler, Kirsch *et al.*, (1991)).

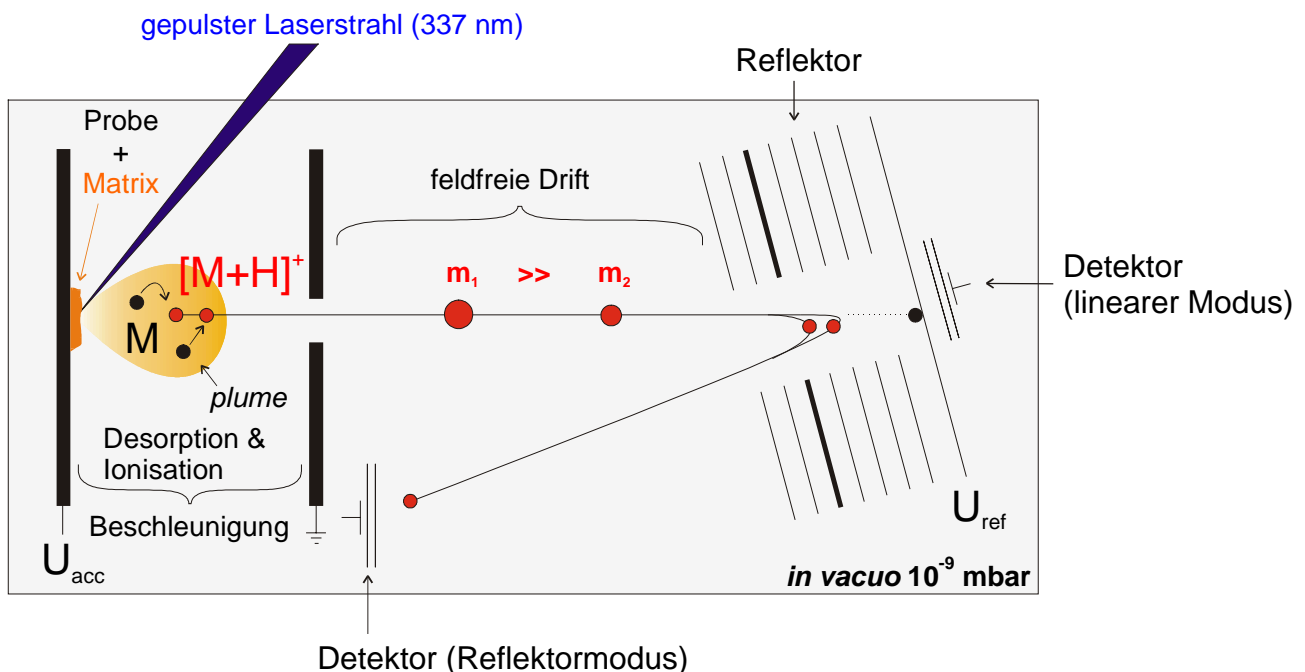


Abbildung 1-1: Funktionsweise eines MALDI-Ref-TOF-Massenspektrometers. Erläuterungen im Text.

Zur Verbesserung der massenspektrometrischen Auflösung bestehen noch weitere Möglichkeiten: Die Technik der sog. „verzögerte Ionenextraktion“ (*delayed extraction* \equiv DE oder *time lag focussing*; s. Wiley & McLaren, (1955)) wurde von mehreren Arbeitsgruppen erfolgreich auf die

MALDI-TOF-MS übertragen (Braun, (1995), King, Colby *et al.*, (1995), Whittal & Li, (1995)), so daß sie seit Mitte 1996 standardmäßig in vielen MALDI-Massenspektrometern angewendet wird. Hierbei werden die Ionen nicht direkt unter Einwirkung des vollen Beschleunigungspotentials extrahiert, sondern erst nach etwa 100 bis 500 ns Zeitverzögerung, während der ein niedrigeres Potential auf die Ionen einwirkt. Die resultierenden Spektren zeichnen sich durch geringere Fragmentierung, geringeres Rauschen und eine deutlich höhere Auflösung aus. Nach richtiger Abstimmung der Verzögerungszeit der Extraktions-, Linsen- und Reflektor-Potentiale können mit einem MALDI-Reflektor-TOF-MS z.B. Peptide von m/z 500 bis ~ 4500 isotopisch aufgelöst werden. Eine hohe Auflösung erlaubt auch eine exakte Kalibrierung der Spektren über monoisotopische Signale interner Standards (z.B. bekannte Peptide, Trypsin-Autoproteolyseprodukte, Matrixsignale, etc.). Die resultierende hohe Massengenauigkeit im Bereich von 5-35 ppm ermöglicht es, Proteine anhand weniger Signale mit hoher Zuverlässigkeit über Datenbanksuchen zu identifizieren (Jensen, Podtelejnikov *et al.*, (1996)).

Gegenüber niedermolekularen Verunreinigungen z.B. Salzen ist die MALDI-MS relativ unempfindlich. Um jedoch Suppressioneffekte und die Bildung von Alkalikationenaddukten zu vermeiden, ist besonders für Peptid- und DNA-Analysen eine geeignete Probenaufreinigung sinnvoll. Die besagte Unempfindlichkeit sowie der theoretisch unlimitierte Massenbereich der Flugzeitmassenspektrometrie haben jedoch auch neue Anwendungen ermöglicht wie z.B. die Charakterisierung von Mikroorganismen anhand kompletter Zell-Lysate (Chong, Wall *et al.*, (1997), Welham, Domin *et al.*, (1998), Dai, Li *et al.*, (1999)), was auch für diese Arbeit bedeutsam ist.

1.1.2. Elektrospray Ionisations Massenspektrometrie (ESI-MS)

Die direkte Analyse gelöster Biomakromoleküle wie Peptide und Proteine ermöglicht die ESI-MS (*reviews* s. Siuzdak, (1994), Przybylski & Glocker, (1996)). Eine Lösung dieser Analyte in wässrigen oder wasserhaltigen organischen Lösungsmitteln wird hierbei unter atmosphärischem Druck in einem elektrischen Feld fein versprüht. Hierzu wird an eine metallisierte, spitz ausgezogene, durchlässige Kapillare mit der Probenlösung gefüllt und an eine Hochspannungsquelle (0,3-6 kV) angeschlossen. Im elektrischen Feld deformiert sich die Flüssigkeitsoberfläche an der Austrittsstelle der Kapillare zum sog. „Taylor Cone“ (s. Abbildung 1-2). Dieser emittiert an seiner Spitze hochgeladene Tröpfchen. Die Aerosolbildung wird bei größerer Flüssigkeitszufuhr in der Regel mit einem zerstäubend wirkenden Gasstrom (z.B. N_2) und durch Wärmezufuhr unterstützt. Durch Stöße mit Molekülen eines entgegenströmenden Gases (*curtain gas*) erfolgt die Desolvatisierung des Aerosols. Dabei verringert sich die Tröpfchengröße so lange, bis die repulsiven Coulombkräfte in den Tröpfchen größer werden, als die Oberflächenspannung. Der Theorie nach findet entweder eine Tröpfchenspaltung oder die Emission freier Molekülionen aus den elektrisch geladenen Tröpfchen statt (s. Abbildung 1-2; zur Theorie s. Wilm & Mann, (1994)).

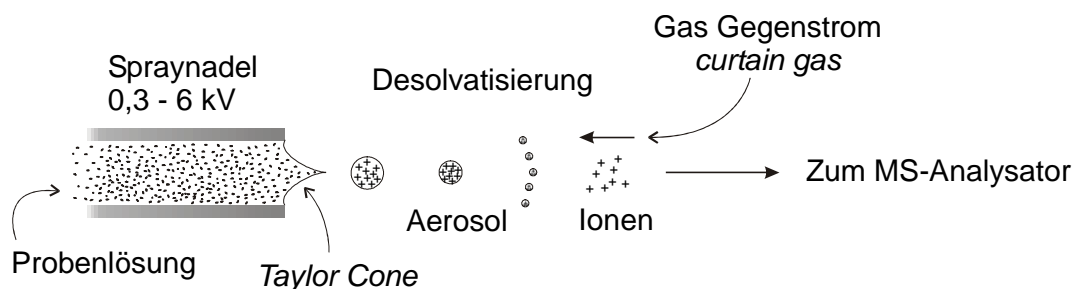


Abbildung 1-2: Desolvatisierung und Ionenemission beim Elektrosprayprozeß der ESI-MS.

Aus kleinen Analytmolekülen werden hauptsächlich einfach geladene Ionen erzeugt. Makromoleküle liefern dagegen verschiedene mehrfach geladene Ionen, wobei die Intensitätsverteilung dieser Ionenserien abhängig von der räumlichen Struktur der Analytmoleküle ist (Glocker, Borchers *et al.*, (1994), Konermann & Douglas, (1997), Gao, Yin *et al.*, (1998)). Die erzeugten Ionen werden in ein gestaffeltes Vakuumsystem überführt und in Quadrupol-, Sektorfeld-, Ionenfallen- oder gepulsten Flugzeit-Massenanalysatoren (Verentchikov, Ens *et al.*, (1994)) getrennt und nachgewiesen.

Neueste Entwicklungen von ESI-MS Ionenquellen kommen mit immer geringeren Probenmengen in sehr kleinen Probenvolumina ($\leq 1 \mu\text{l}$) aus. Wilm *et al.* zeigten, daß die Empfindlichkeit der ESI-MS hauptsächlich von der Konzentration der Analyte abhängt, was zur Einführung der „Nano-ESI-MS“ Technik führte (Wilm & Mann, (1996), Vertes, Irimyi *et al.*, (1993)) die wegen ihres geringen Probenbedarfs (Femtomole) von großer Bedeutung für die 2-DE gekoppelte Proteomforschung ist (Wilm, Shevchenko *et al.*, (1996)).

Im Gegensatz zur MALDI-MS toleriert die ESI-MS keine niedermolekulare Verunreinigung wie Salze oder Tenside. Gemische aus Abbaupeptiden müssen daher zuvor aufgereinigt werden - vorzugsweise chromatographisch z.B. über HPLC mit „reversed phase“ Materialien.

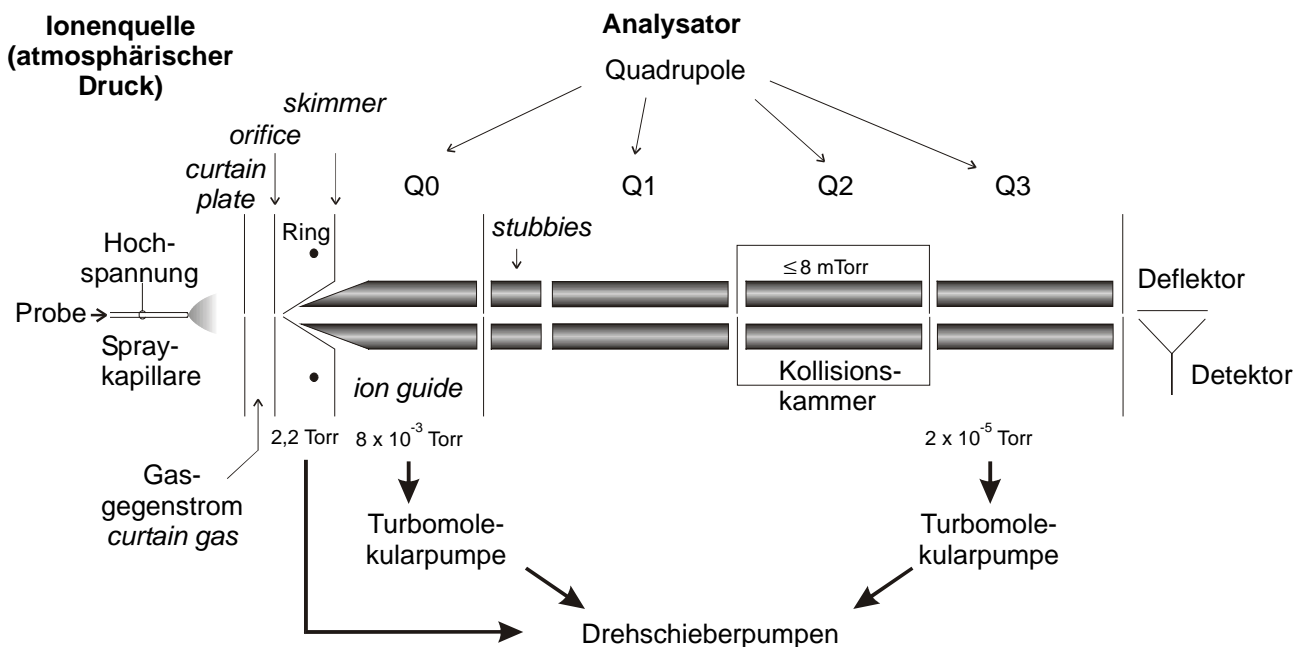


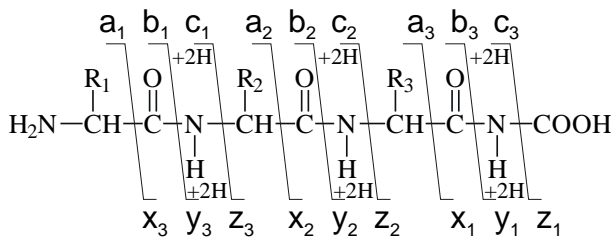
Abbildung 1-3: Funktionsweise eines Trippel-Quadrupol-Elektrospray-Massenspektrometers. Die Darstellung entspricht dem Aufbau des API-300 Gerätes von PE/Sciex (Concord, Ontario, Kanada). Erläuterungen im Text.

Zur Gewinnung sequenzspezifischer Informationen von Peptiden eignen sich Trippel-Quadrupol-ESI-MS-Geräte. Die Ionen werden über sog. „ion guide“-Quadrupole (Q0 und *stubbies* in Abbildung 1-3), die auf bestimmten Hochfrequenzen arbeiten, vom Feinvakuum- in den Hochvakuum-Bereich des Massenspektrometers überführt. Der nachfolgende Quadrupol (Q1) ist auf eine feste Frequenz abgestimmt, so daß diesen nur Ionen eines bestimmten m/z Wertes passieren können. Die so herausgefilterten, vorzugsweise doppelt oder 3 fach geladenen Ionen gelangen in die Kollisionskammer in der sie mit verdünnten Gasen wie N_2 oder Ar kollidieren und fragmentieren (*collision induced decay* = CID). Die Kollisionsenergie läßt sich über ein Gleichspannungspotential zwischen Q1 und Q3 einstellen. Der Quadrupol Q2 in der Kollisionskammer wirkt als *ion guide* und verhindert das Entweichen der Ionen. Der letzte Quadrupol (Q3 in Abbildung 1-3) dient schließlich der Auftrennung der entstandenen Fragmente. Daher wird seine

Frequenz im erwarteten m/z -Bereich durchgestimmt. Die Trippel-Quadrupol-Anordnung ermöglicht einen besonders flexiblen Betrieb. Es können sowohl Gesamtspektren (*Q1-scan*) als auch Fragmentenspektren (*Q3-scan*) aufgenommen werden. Beim sog. „parent-ion scan“ wird mit Q1 der m/z -Wert der in die Kollisionskammer eingelassenen Ionen über den zu untersuchenden m/z -Bereich variiert, während die Q3-Frequenz fest eingestellt ist, um eine erwartete Abgangsgruppe (z.B. Immoniumionen $H_2N^+=CHR$, PO_3^- , etc.) zum Detektor zu leiten und diese so dem „Mutterion“ (*parent-ion*) zuzuordnen (Metzger, Kempter *et al.*, (1994), Wilm, Neubauer *et al.*, (1996), Neubauer & Mann, (1997)).

Die Benennung der durch CID oder PSD erzeugten Fragmentionen von Peptiden erfolgt einheitlich nach der von Biemann, Johnson und Roepstorff vorgeschlagenen Nomenklatur (Biemann, (1989), Johnson, Martin *et al.*, (1988), Roepstorff & Fohlman, (1984), s. Abbildung 1-4, auch Spektrum in Abbildung 3-21 sowie Fragmentierschema in Abbildung 3-22 (in Abschnitt 3.4)).

Abbildung 1-4: Nomenklatur der Fragmentionen.



Verbleibt die Ladung am N-terminalen Fragment, so werden die Ionenserien mit a_i , b_i oder c_i bezeichnet, wobei „i“ die Nummer des entsprechenden AS-Restes (vom N-Terminus aus gezählt) angibt. Mit x_i , y_i und z_i werden dagegen C-terminale Fragmente benannt. In diesem Fall beginnt die Nummerierung am C-Terminus

1.1.3. Notwendigkeit und Möglichkeiten der Automatisierung

Die Zahl der bekannten Genome wächst ständig. Das staatlich geförderte Human Genomprojekt der USA plant die Identifizierung der ~80'000 menschlichen Gene bis zum Ende des Jahres 2003 (U.S.Department of Energy, (1999)). Vermutlich wird bis zum Jahre 2005 das gesamte menschliche Genom bekannt sein (~ 3 Mrd. Basenpaare). Im Januar 1998 waren bereits 15 Genome von Mikroorganismen vollständig sequenziert (U.S.Department of Energy, (1998)), derzeit sollen es etwa 18 Genome sein. Prinzipiell ist somit der Grundstein für die Analyse der Proteome gelegt. Um das komplexe Netzwerk der Genregulation, Proteinexpression, Proteinprozessierung und Proteinfaltung in überschaubaren Zeiträumen erforschen zu können ist, es sinnvoll zunächst die bekannten Techniken der Proteinidentifizierung zu automatisieren. Houthaeve *et al.* entwickelten bereits ein Gerät für den automatischen Verdau gelelektrophoretisch getrennter Proteine, mit dem 32 Proteinproben simultan prozessiert werden können (Houthaeve, Gausepohl *et al.*, (1995)). Um die Abbaupeptide aus mehreren 1000 Spots eines 2-DE Gels massenspektrometrisch zu analysieren ist die ESI-MS wenig geeignet. Zwar existieren bereits diverse Möglichkeiten zur automatischen HPLC / ESI-MS Kopplung, andererseits erscheint keines der bekannten Verfahren für einen hohen Durchsatz geeignet. Der Schwerpunkt liegt somit auf der Entwicklung automatischer MALDI-MS Verfahren. Die Datenaquisition der MALDI-MS dauert nur wenige Sekunden pro Probe, je nachdem wie viele Einzelmessungen erforderlich sind, um Spektren mit akzeptablen Signal- / Rausch-Verhältnis aufzunehmen. Ein hoher Entwicklungsbedarf besteht daher in der automatischen Herstellung geeigneter Proben und in der Auswertung massenspektrometrischer Daten.

Werden Matrix und Probenlösungen mit Volumina im unteren Mikroliterbereich (0,3 bis 0,5 μ l) auf die Edelstahloberflächen von MALDI-MS Targets übertragen und eingetrocknet, verbleiben runde Kristallflächen mit ca. 1,5 bis 3,0 mm Durchmesser (s. Abschnitt 2.5.2). Der Fokus des für die Desorption der Probe verwendeten Lasers beträgt hingegen 50 bis 70 μ m. Entsprechend wird auch nur ein kleiner Teil der Probe für die eigentliche Analyse aufgebraucht. Mit den herkömmlichen manuellen Pipettier-techniken ist es nur schwer möglich Volumina wässriger Lösungen unter 0,2 μ l zu transferieren. Bei organischen Lösungsmitteln wirkt sich der erhöhte Dampfdruck zudem erschwerend auf den Behalt der Flüssigkeit in der Pipettenspitze aus.

Der Transfer kleiner Volumina im Nanoliterbereich wurde mit der Einführung der Tintenstrahl-drucktechnik (*ink jet* oder *drop-on-demand technology*) technisch realisiert (*review* s. Wehl, (1994), (1995a) und (1995b)). Mit einem nach der *Drop-On-Demand*-Technik arbeitenden Mikrodispensiergerät lassen sich kleinste Mengen einer Substanz oder einer Lösung in Form kleiner Tröpfchen beliebig positionieren und pipettieren (Schober, Günther *et al.*, (1993)). Die Verwendung eines aus mehreren Mikropipetten bestehenden Dispensierkopfes setzt insbesondere voraus, daß die Zielpositionen der abgegebenen Flüssigkeitsmengen genauestens bekannt sind und notfalls aufeinander abgestimmt werden können. Andernfalls ist beispielsweise die Erzeugung definierter Mikrosubstanzzraster nicht möglich.

Zur Bestimmung der Zielpositionen der dispensierten Flüssigkeiten werden derzeit optische Systeme eingesetzt. Hierbei wird zunächst mit einem stroboskopischen Verfahren über eine Videokamera der Tropfenabriß überprüft. Mit einer weiteren Videokamera lassen sich die Zielpositionen der von den einzelnen Dispensern erzeugten Tröpfchen ermitteln. Dies gelingt z.B. auch durch die Analyse von bereits erzeugten Substanzzrastern. Zur Automatisierung eines derartigen Verfahrens sind technisch aufwendige Bilderkennungsverfahren notwendig. Neben der Kostenintensität dieser Systeme birgt deren Kompliziertheit einen hohen Grad an Fehleranfälligkeit. Die hieraus resultierende Unzuverlässigkeit zeitgemäßer Mikrodispensiersysteme steht Ihrem hohen technologischen Einsatzpotential bei der Bearbeitung von hochparallelen analytischen Aufgaben oder Problemstellungen aus der kombinatorischen Chemie bislang kontraproduktiv gegenüber.

1.2. Horizontaler Gentransfer durch bakterielle Konjugation

Ein aktuelles Beispiel für die Notwendigkeit Proteome gezielt zu untersuchen, stellt das Problem der zunehmenden Verbreitung antibiotikaresistenter Krankheitserreger. Immer häufiger wird hierüber in der einschlägigen Tagespresse berichtet (Förger, (1999), Szlezák, (1998)). Das wissenschaftliche Interesse richtet sich insbesondere auf die Art und Weise, wie Bakterien die genetische Information, welche zur Antibiotikaresistenz führt, weitergeben. Hieran ist ein komplexes System aus mehreren Proteinen beteiligt. Es handelt sich um den Apparat des „horizontalen Gentransfers“ der „bakteriellen Konjugation“. Dieser genetische Informationsaustausch wurde erstmals von Lederberg und Tatum 1946 für *E. coli* Stämme beschrieben (Lederberg & Tatum, (1946)). Wie man heute weiß, liegt die Ursache des Phänomens im sog. F- (*fertility*) Faktor begründet (Hayes, (1953), Cavalli-Sforza, Lederberg *et al.*, (1953)), welcher gleichsam das erste entdeckte „Plasmid“, ein extrachromosomales, zyklisches DNA-Molekül, darstellt. Dieses etwa 100 kb große Plasmid kodiert für eine Gruppe von Genprodukten, die u.a. zum Aufbau des sog. „Sex-Pilus“ führen (s. Abbildung 1-5).

Pili sind extrazelluläre, filamentöse Gebilde (s. Abbildung 1-7), welche vermutlich die Ausbildung eines engen, stabilen, Kontaktes zwischen Donor- und Rezipienten-Zelle vermitteln und sich dabei zurückbilden sollen. Somit stellen sie einen Teil des sog. *mating pair formation* (Mpf)-Systems dar, dessen Proteine mehrheitlich in den bakteriellen Zellmembranen verankert sind (Firth, Ippen-Ihler *et al.*, (1996), Lessl, Balzer *et al.*, (1992)). Weitere Faktoren sind vermutlich an der Stabilisierung des *mating pairs* beteiligt (Achtman, (1975), Achtman, Morelli *et al.*, (1978)).

Nach der Ausbildung eines derartigen stabilen Zell-Zell-Aggregates (*mating aggregate*), wird das konjugative Plasmid als Einzelstrang DNA in die Rezipientenzelle übertragen. Der DNA-Einzelstrang wird dabei über einen Proteinkomplex aus überspiralisierter Plasmid-DNA (*super helical DNA*) erzeugt. Die hierfür erforderlichen Proteine werden ebenfalls vom konjugativen Plasmid kodiert und dem *DNA-transfer and replication* (Dtr)-System zugeschrieben. Zum Dtr-System gehören Proteine, die spezifisch den sog. Transferursprung (*transferorigin*, *oriT*) des

Plasmids komplexieren und die Bindung der Relaxase ermöglichen. Diese bewirkt den DNA-Einzelstrangbruch (*nick*) an einer spezifischen Stelle innerhalb der „*nick*-Region“ des *oriT* und bindet dabei selbst kovalent an das 5'-Ende des erzeugten Einzelstranges. Der resultierende Komplex aus DNA und Proteinen liegt *in vivo* im Gleichgewicht mit intakter Plasmid-DNA vor und wird als Relaxosom bezeichnet (s. Abbildung 1-5; Clewell & Helinski, (1970), Lanka & Wilkins, (1995)).

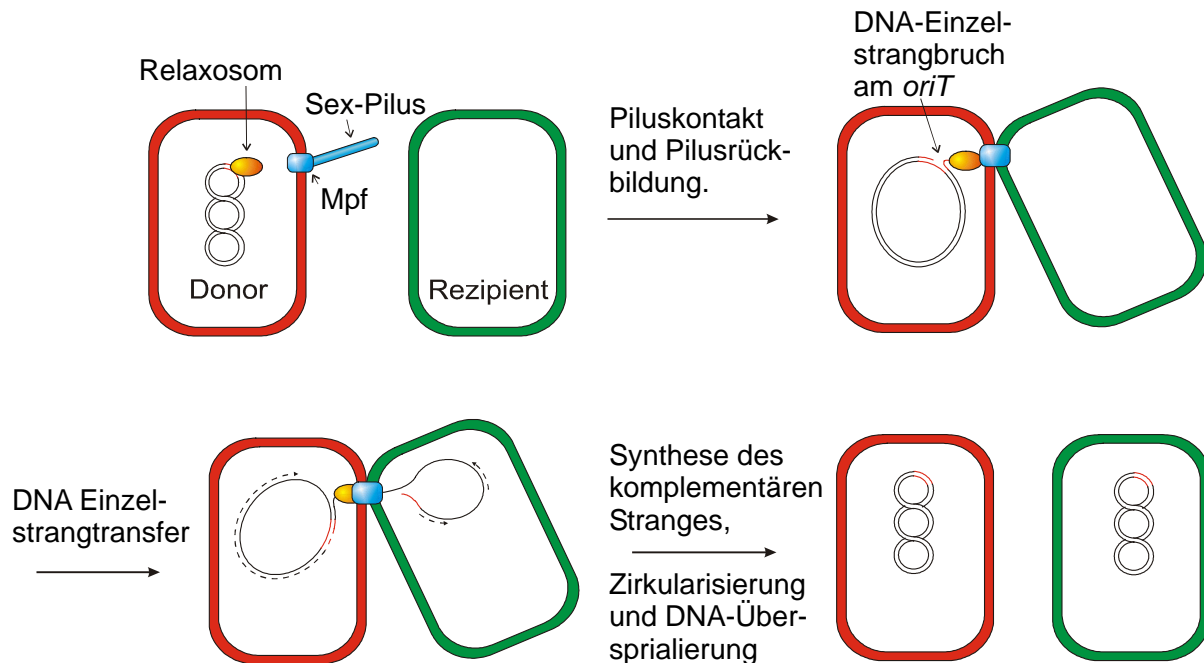


Abbildung 1-5: Der konjugative DNA-Transfer in der Modellvorstellung (Erläuterungen im Text).

Die Entwindung der DNA erfolgt durch eine Helicase nach der Kontaktaufnahme von Donor- und Rezipienten-Zelle aufgrund eines bislang nur postulierten *trigger*-Signals (*mating signal*; Firth, Ippen-Ihler *et al.*, (1996)). Wie der Einzelstrang letztendlich in die Rezipienten Zelle übertragen wird, ist unklar. Eine vorausgehende Assoziation des Relaxosoms an den Mpf-Komplex darf als wahrscheinlich angenommen werden. In der Rezipientenzelle wird der DNA-Einzelstrang zyklisiert und vermutlich nach dem *lagging strand*-Mechanismus repliziert. An der Komplementärstrang-Synthese sind wirtskodierte Funktionen wie das DNA-Polymerase-III-Holoenzym und Primasen beteiligt. Für das nachfolgend beschriebene konjugative IncP α -Plasmid (RP4) kodiert jedoch selbst eine Primasefunktion (TraC). Es wurde postuliert, dass diese zusammen mit dem Plasmid in die Rezipientenzelle übertragen wird (Rees & Wilkins, (1990)). Für das F-Plasmid wurde bisher keine Primasefunktion entdeckt.

Mittlerweile wurde eine große Anzahl weiterer konjugativer Faktoren beschrieben (*vide infra*). Der DNA-Transfer wird in den meisten Fällen durch Plasmide vermittelt, daneben sind allerdings auch konjugative *transposons* bekannt, die weniger gut untersucht sind. Diese Arbeit bezieht sich daher auf den Plasmid-vermittelten DNA-Transfer. Die meisten der bisher untersuchten konjugativen Plasmide kodieren sog. *entry exclusion* (Eex)-Funktionen, die bewirken, daß das gleiche konjugative Plasmid nur wenige Male (im „besten“ Falle einmal) in denselben Rezipienten übertragen wird (Novick, (1969)). Die Eex-Funktion verhindert den Gentransfer bereits auf der Stufe der bakteriellen Konjugation (Haase, Kalkum *et al.*, (1996)). Plasmide werden allgemein über sog. Inkompatibilitäts- (Inc)-Gruppen voneinander unterschieden. Deren Funktionsweise beruht im Gegensatz zur *entry exclusion* auf der Unverträglichkeit gleicher *partitioning* (*par*)- oder Replikations (*rep*)-Systeme mehrerer Plasmide innerhalb einer Zelle (*review s.*(Novick, (1987))).

Die dem F-Faktor ähnlichen IncI-Plasmide ColIb, R114, R64, das IncN-Plasmid pKM101, die IncP-Plasmide RP4 (IncP α), R751 (IncP β), und das IncW-Plasmid R388 kodieren Transfersysteme Gram-negativer Bakterien (Lanka & Wilkins, (1995)). Diese sind derzeit am besten untersucht.

Viele der selbst nicht konjugativ wirksamen (nicht selbst-transmissiblen), natürlich vorkommenden, kleinen Plasmide wie z.B. ColE1, pSC101 und RSF1010 lassen sich in Gegenwart konjugativer Plasmide (oder Helferplasmide) auf geeignete Rezipienten übertragen (Willetts & Wilkins, (1984), Haase, Lurz *et al.*, (1995)).

Eine interessante Variante der bakteriellen Konjugation stellt der Transfer von T-DNA aus *Agrobakterien* in Pflanzenzellen dar. Zellen verletzter Pflanzen sondern Phenolderivate wie z.B. Acetosyringon* ab. Diese Substanz wirkt als Induktor für das Bodenbakterium *Agrobacterium tumefaciens*, welches das *tumor inducing-* (Ti)-Plasmid trägt (Stachel, Messens *et al.*, (1985), Engstrom, Zambryski *et al.*, (1987)). Ähnlich wie beim F-Faktor oder den IncP-Plasmiden in *E. coli* bildet das Ti-Plasmid haltige *Agrobacterium tumefaciens* Pili aus, mit denen vermutlich die Bildung eines Mpf-artigen Zellkontaktes initiiert wird (Fullner, Lara *et al.*, (1996)). Plasmidkodierte Virulenzfaktoren (*vir*) bewirken die Übertragung, eines Teils des Ti-Plasmids, der sog. T-DNA in die Pflanzenzelle. Im Transferprozeß liegt die T-DNA als Einzelstrang und kovalent an das VirD2-Protein gebunden vor. Anschließend erfolgt die Integration der T-DNA in das Pflanzengenom, wodurch die Pflanze veranlaßt wird sogenannte Opine zu synthetisieren. Die Opine (z.B. Octopin**, Nopalين***) dienen den *Agrobakterien* wiederum als Kohlenstoff- und Stickstoffquelle.

Das Ti-Plasmid beinhaltet auch Gene einer Transfer (Tra)-Region. Diese sind ebenfalls homolog zu den Genen der Tra-Regionen der IncP-Plasmide und des F-Faktors (Altmörbe, Stryker *et al.*, (1996), Farrand, Hwang *et al.*, (1996), Li, Everhart *et al.*, (1998)). Die Tra-Region von Ti kodiert für Proteinfunktionen, die für den Austausch genetischer Information der *Agrobakterien* untereinander verantwortlich sind. Darüber hinaus bestehen hier regulatorische Verflechtungen zwischen dem T-DNA Transfer in Pflanzenzellen und dem interbakteriellen, horizontalen Gentransfer (Oger, Kim *et al.*, (1998), Cook, Li *et al.*, (1997)).

1.2.1. Die Grundlage für ein „kleines“ Proteomprojekt:

Der auf ~27 Genen basierende DNA-Transferapparat der IncP-Plasmide.

Das zur Inkompatibilitätsgruppe IncP α gehörende Plasmid RP4 verleiht seinen Wirtszellen Resistenzen gegen die drei Antibiotika Ampicillin^{4*}, Kanamycin^{5*} und Tetracyclin^{6*}, darüber hinaus besteht eine Tellurit-Resistenz. Es wurde 1969 zusammen mit den als identisch anzusehenden Plasmiden R18, R68, RK2 und RP1 aus *Pseudomonas*-Stämmen in der „burns unit“ eines Birminghamer Krankenhaus isoliert. Daher stammen auch die Bezeichnungen: Birmingham-IncP α -Plasmide und R-(*resistance*)-Faktoren (Datta, Hedges *et al.*, (1971), Burkhardt, Riess *et al.*, (1979)). Mittlerweile sind die DNA-Sequenzen von RP4 und die des nahe verwandten Schwesterplasmids R751, der IncP β -Gruppe, vollständig bekannt (Pansegrau, Lanka *et al.*, (1994), Thorsted, Macartney *et al.*, (1998)). Das RP4 Plasmid besteht aus 60'099 Basenpaaren und kodiert für mindestens 74 Gene.

Die IncP-Plasmide zeichnen sich durch einen erstaunlich breiten Wirtsbereich aus, so daß diese Plasmide in allen bislang getesteten Gram-negativen Bakterien repliziert werden können. Dieses „promiskuitive“ Transfersystem kann darüber hinaus sogar eine DNA-Übertragung in Gram-positive Bakterien (Trieu-Cuot, Carlier *et al.*, (1987)), Streptomyceten (Mazodier, Petter *et al.*,

* Acetosyringon, Trivialname für 3,5-Dimethoxy-4-hydroxyacetophenon

** Octopin \equiv (+)-N α -(1-Carboxyethyl)-L-arginin

*** Nopalين \equiv (+)-N α -(1,3-Dicarboxypropyl)-L-arginin

4* Ampicillin \equiv 6-[D(-)- α -Amino- α -phenylacetamido]-penicillansäure

5* Kanamycin, ein Aminoglycosid-Antibiotikum

6* Tetracyclin \equiv 4-Dimethylamino-1,4,4a,5,5a,6,11,12a-octa-hydro-3,6,10,12,12a-pentahydroxy-6-methyl-1,11-dioxo-naphthacen-2-carboxamid

(1989)) und sogar in Hefezellen (Heinemann & Sprague, Jr., (1989)) ermöglichen. In diesen Organismen werden die IncP-Plasmide allerdings selbst nicht repliziert. Der DNA-Transfer läßt sich jedoch durch die Integration übertragener genetischer Information in das Chromosom des Rezipienten nachweisen.

Die Transferregionen von RP4 und R751 sind auf ihren Plasmiden in jeweils zwei voneinander getrennten Regionen kodiert, die mit Tra1 und Tra2 bezeichnet werden (Lessl, Balzer *et al.*, (1992)). Die Gene des Mpf-Systems kartieren bis auf *traF* in die Tra2 Region. Gene, die an den DNA-Prozessierungsreaktionen beteiligt sind und daher zum Dtr-System gehören, kartieren in die Tra1 Region. Gene der Tra1-Region werden mit dem Präfix „tra“, die der Tra2-Region mit „trb“ bezeichnet. Insgesamt kodiert die Transferregion von RP4 für 27 Genprodukte. Für den Gentransfer in *E. coli* sind nur 19 RP4-Gene essentiell, die in die *core*-Region von Tra1 und Tra2 kartieren (s. Abbildung 1-6).

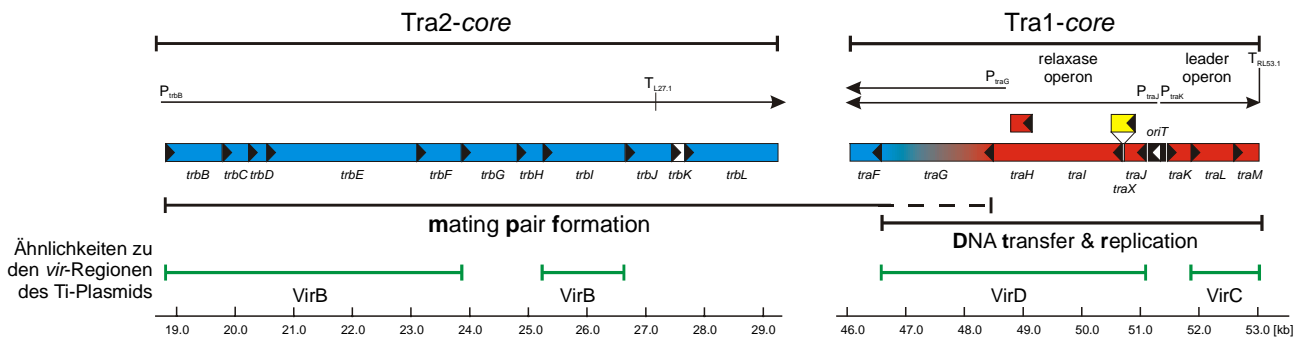


Abbildung 1-6: Physische und genetische Karte der core-Transferregion von RP4 (Koordinaten: 18,2-30,0 kb \Rightarrow Tra2 und 45,8-53,0 kb \Rightarrow Tra1). Die Pfeile in den Balken symbolisieren die Genstarts, sowie den Transferorigin (weiß, *oriT*) und die Leserichtung. Promotoren, die funktionellen Bereiche des Mpf- und des Dtr-Systems sowie Ähnlichkeiten zur Virulenzregion (*vir*) des Ti-Plasmids sind gekennzeichnet (in Anlehnung an Pansegrau, Lanka *et al.*, (1994), Lessl, Balzer *et al.*, (1992) und Haase, (1996)).

Das Relaxosom des RP4-Transferapparates besteht TraH, TraI, TraJ. In gewisser Weise zählt auch TraK dazu, welches an den *oriT* bindet. TraJ erkennt spezifisch den *oriT* und bindet dann zusammen mit der Relaxase, TraI, an das Plasmid. Nach dem „cleavage“ der DNA verbleibt TraI kovalent das 5'-Ende des resultierenden DNA-Einzelstrangs gebunden. TraH scheint diesen Protein-DNA-Komplex zu stabilisieren. Es besitzt eine auffällig braune Farbe, dessen Ursache der Klärung bedarf. (Pansegrau, Balzer *et al.*, (1990), Pansegrau, (1991), Pansegrau, Schröder *et al.*, (1993)).

Komplizierter als das Relaxosom (Dtr-System) ist das Mpf-System der Tra2-Region aufgebaut. Anhand von Sequenzähnlichkeiten zu *virB*-Region von Ti und Lokalisierungs-Experimenten können nur Vermutungen über die Funktion der meisten Tra2-Proteine aufgestellt werden.

Eindeutig geklärt werden konnte bis dato nur die Funktion des Genproduktes von *trbK*, welches die „entry exclusion“ vermittelt und in der Zytoplasmamembran verankert ist (Haase, (1996)). Nach wie vor unbekannt ist die Zusammensetzung der filamentösen Pili, für deren Biosynthese die 10 Gene *trbB*, *-C*, *-D*, *-E*, *-F*, *-G*, *-H*, *-I*, *-J*, *-L* der Tra2-Region und ein Gen, das *traF*, der Tra1-Region erforderlich sind (Haase, Lurz *et al.*, (1995)). Letzteres ist aus Untersuchungen mit den piluspezifischen Bakteriophagen Pf3 und PRR1 sowie aus elektronenmikroskopischen Studien bekannt. Die genannten Phagen docken an die IncP-Pili an (Barth, Grinter *et al.*, (1978)). Beim F-Faktor wurde die Hauptkomponente des F-Pilus bereits als TraA von F identifiziert (Maneewannakul, Maneewannakul *et al.*, (1993)). Dieses Protein ist N-terminal acetyliert (Moore, Hamilton *et al.*, (1993), Maneewannakul, Maneewannakul *et al.*, (1995)). Serologische Studien und Sequenzvergleiche legen nahe, das evtl. das Genprodukt von *trbC* von RP4 am IncP-Pilusaufbau

beteiligt sein könnte. TrbC wird allerdings in nicht völlig bekannter Weise mehrfach *in vivo* prozessiert (s. Abbildung 3-31 in Abschnitt 3.5.3 und Haase & Lanka, (1997)).

Elektronenmikroskopische Aufnahmen der RP4 kodierten IncP-Pili zeigen, daß diese faserige, zerbrechliche Strukturen bilden, die sich in Aggregaten zusammen lagern können (s. Abbildung 1-7). Die Pili von F und Ti scheinen dagegen flexibel zu sein (Ti Pili sind in Abbildung 1-7 gezeigt). Parallel dazu sollte erwähnt werden, das die bakterielle Konjugation RP4-haltiger *E. coli* Bakterien nur auf festen Medien stattfindet. Dagegen konjugieren z.B. F-Plasmid-haltige Donorzellen auch in flüssigen Medien mit meßbarer Effizienz (Bradley, Taylor *et al.*, (1980)). Der Durchmesser der Pili von F, Ti und RP4 beträgt ~10 nm. Ihre Länge kann mehrere mm betragen und somit die tausendfache Länge eines Bakteriums übersteigen. Interessanter Weise finden sich häufig blasenförmige Ausbuchtungen an einem Pilusende. Für den F-Pilus wurde postuliert, daß die Pilusspitze ein „Adhäsion“ exponieren könnte, welches den Zell-Kontakt initiiert (Anthony, Sherburne *et al.*, (1994)). Dieses wurde bisher jedoch nicht experimentell bestätigt. Für die Typ-IV Pili der pathogenen Bakterien *Neisseria gonorrhoeae* und *Neisseria meningitidis*, die eine Schlüsselrolle bei neisserialen Infektionen spielen sollen, wurde PilC als Pilusspitzen-gebundenes Adhäsion identifiziert (Rudel, Scheurerpflug *et al.*, (1995)).

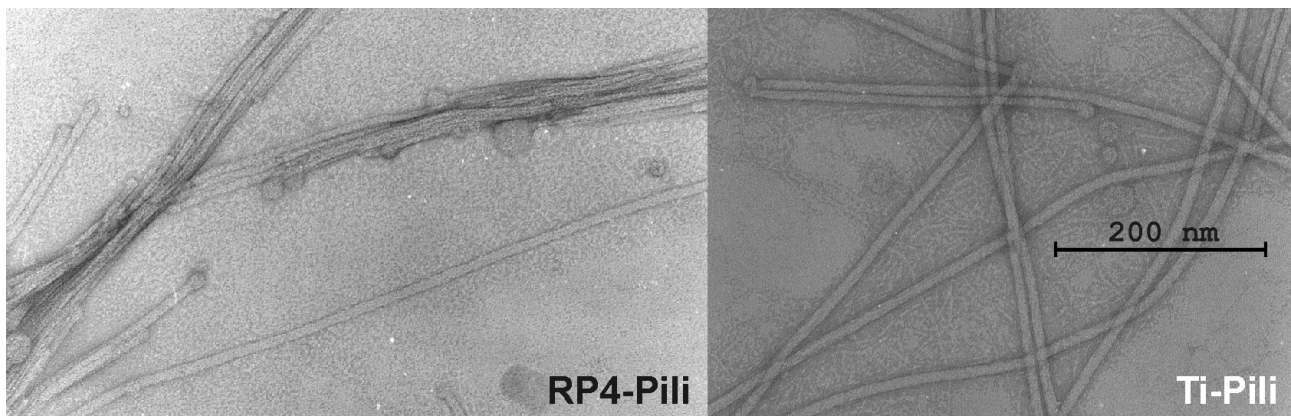


Abbildung 1-7: Elektronenmikrographie von Pili aus RP4 in E. coli (links) und Ti-Pili aus Agrobacterium tumefaciens (rechts). Blasenförmige Ausbuchtungen an einigen Pilusenden sind in beiden Mikrographien zu sehen.*

* Für die Zurverfügungstellung des Bildmaterial danke ich Herrn Dipl.-Ing. für Biotechnologie Ralf Eisenbrandt und Dr. Rudi Lurz (MPI für molekulare Genetik, Berlin).

1.2.2. Aufgabenstellung und Zielsetzung

Das Ziel dieser Arbeit bestand darin, geeignete chemische, biologische und technische Methoden zu entwickeln, welche a) die Einsetzbarkeit massenspektrometrischer Analyseverfahren für die allgemeine Durchführung umfangreicher Proteomprojekte ermöglicht und die sich, b) in Modellversuchen nutzen lassen, um spezielle Proteinfunktionen des von den IncP-Plasmiden kodierten DNA-Transferapparates der bakteriellen Konjugation zu entschlüsseln.

Im Einzelnen ergeben sich die folgenden Aufgaben und Fragestellungen:

Wie kann die Präparation von MALDI-MS-Proben durch Anwendung der Tintenstrahldrucktechnik miniaturisiert werden, um ihre hohe Empfindlichkeit und ihren minimalen Probenbedarf für die Peptidkartierung mit geringsten Proteinmengen nutzen zu können?

Miniaturisierte MALDI-MS Proben sollen rasterförmig angeordnet werden und automatisch gemessen werden. Wie läßt sich die präzise Anordnung der Mikroproben realisieren? An Modellbeispielen mit *E. coli*- und RP4- kodierten Proteinen soll die Eignung der miniaturisierter MALDI-MS-Proben ermittelt werden.

Die Auswertung einer großen Anzahl von Massenspektren ist manuell nicht zu bewerkstelligen. Wie können massenspektrometrische Daten wie m/z des monoisotopischen Signals, durchschnittliches m/z , Auflösung und Signalgüte automatisch extrahiert werden? Wie lassen sich die bisher ungenutzten Intensitätsverteilungen der Isotopen-*Peaks* sinnvoll für die Interpretation von Spektren nutzen? Ist die vollautomatische Identifizierung von Proteinen anhand der Massen entsprechender Abbaupeptide realisierbar? Lassen sich posttranslationale Modifizierungen auf diese Weise erkennen? Insbesondere soll untersucht werden, welche chromophore Gruppe die braune Farbe des *traH*-Genproduktes verursacht, und in wie fern vorhergesagte Signalpeptide von Tra2-Proteinen tatsächlich abgespalten werden. Beispielhaft soll die Prozessierung des *trbK*-Genprodukts ermittelt werden.

Die Anwendung der „klassische“ Methode der Proteomanalyse: 2-DE-Trennung der Proteine, Proteolyse und massenspektrometrische Identifizierung, soll zur Klärung der Frage dienen: „Welche RP4-kodierten Proteine werden in *E. coli* am stärksten exprimiert?“ Möglichkeiten und Limitierung der Methodik sind zu evaluieren.

Per MALDI-MS können auch Proteine ohne vorangehende Trennung direkt aus Zellextrakten charakterisiert werden. Diese Methode soll optimiert werden, um die Untereinheiten der extrazellulären Pili untersuchen zu können. Es wird erhofft, die Frage zu klären ob das Genprodukt von *trbC* der Pilusuntereinheit entspricht, ob und wie dieses posttranslational modifiziert wird.