

I. Einleitung

1.1 Das Konzept einer Large-Scale Analyse in der Biologie

In der Biologie breiten sich heute zunehmend Begriffe wie die 'systematische Analyse komplexer Systeme' oder 'globale Betrachtung' (global view) zusammen mit Forschung im 'Grossen Massstab' ('large-scale approaches') und 'Quantitative Analyse' aus. Sie treten in Verbindung mit den Forschungsprojekten des 'Genoms' und seines Zwillings des 'Proteoms' (Wilkins et al. 1996a) auf. Geschichtlich gesehen hat der Mensch das Bedürfnis schon immer gehabt, aus den Einzelbeobachtungen übergeordnete Regeln abzuleiten (Induktion) und gedankliche Systeme zu erstellen, die eine möglichst umfassende Erklärung für alle Einzelbeobachtungen liefern. Er pflegt religiöse Systeme, erstellt geographische Karten, ordnet das Tier- und Pflanzenreich in Stämme, Klassen, Ordnungen, Familien, Gattungen und Arten und versucht die Beziehungen der Teilnehmer seines Systems zu erfassen. Anfangs bildete die detaillierte Beschreibung der äusseren Morphologie von Tieren und Pflanzen die Grundlage für ihre Ordnung. Dann konnten physiologische Merkmale und schliesslich biochemische Merkmale zum Erkenntnisgewinn herangezogen werden. Mit dem Aufkommen der Computertechnik erreichte diese Entwicklung jedoch eine völlig neue Dimension. Während noch vor 20 Jahren die Analyse einzelner Proteine (z.B. Hämoglobin, Lactatdehydrogenase) oder einzelner Gene (Histongene) bis in alle Details im Mittelpunkt des Interesses standen, geht es heute um die parallele Analyse von tausenden von Genen und Proteinen.

Als 1975 die Restriktionsenzyme entdeckt wurden (Nathans und Smith 1975) und damit die Möglichkeit, DNA sequenzspezifisch zu schneiden, konnte in bedingtem Maße die DNA-Sequenz von Individuen verglichen werden. Mit der Entdeckung der Taq-Polymerase, der Entwicklung der PCR-Technik (Mullis et al. 1986) und der DNA-Sequenzierung (Sanger und Coulson 1975; Maxam und Gilbert 1977) wurde die gesamte Erbsubstanz zugänglich und folgenden Fragen unterworfen: welche Sequenzanteile sind artspezifisch, welche Gene kommen gemeinsam in allen lebenden Organismen vom Bakterium bis zum Menschen vor und was unterscheidet den Menschen – bei einer Vielzahl homologer Gene – von Affe und Maus. Diese Techniken wären nicht entwickelt worden, wenn die Computertechnologie nicht die Möglichkeit geschaffen hätte, Datenmengen in riesigem Ausmaß zu verarbeiten. Damit ist jedoch eine völlig neue Strategie in der biologischen Forschung entstanden, die man als Large-scale-Biology bezeichnen könnte. Das wesentliche Merkmal dieser Strategie ist, dass wissenschaftliche Daten erzeugt werden, die eher deskriptiver Natur sind, als dass sie der Verifizierung oder Falsifizierung einer wissenschaftlichen Hypothese dienen. Erst nach Erstellung riesiger Datenmengen wird mit Hilfe der Bioinformatik nach den biologischen Zusammenhängen gefragt.

Das bisher grösste konzertante Forschungsprojekt – die Sequenzierung des gesamten Genoms des Menschen, wurde Mitte der 80er Jahre gestartet (Dulbecco 1986; Cook-Deegan 1989). Jedoch schon Jahre zuvor wurde an Konzepten gearbeitet, alle Proteine des Menschen aufzutrennen und zu

katalogisieren (Anderson und Anderson 1982; Taylor et al. 1982). Dieses Vorhaben nahm seinen Ausgang von Klose (Klose 1975) und O'Farrell (O'Farrell 1975), die 1975 unabhängig voneinander die 2-dimensionale Elektrophorese von Proteinen entwickelten. Mit dieser Methode konnten Proteinextrakte von gesamten Organen und Geweben aufgetrennt und auf der Ebene einzelner Proteine charakterisiert werden. Zu dieser Zeit - vor der Entwicklung der Methoden zur DNA-Analyse (Sequenzierung, Klonierung, PCR, rekombinante Techniken) erschien die 2-DE als die einzig praktikable Möglichkeit, einen Organismus, z.B. ein Bakterium oder ein Organ (Gehirn, Leber) auf molekularem Niveau zu beschreiben und einen Zugang zum komplexen molekularen Geschehen der Zelle zu finden. 1976 begannen Anderson und Anderson mit der Entwicklung von Technologie und Software für eine systematische Erfassung menschlicher Proteine von 2DE Gelen - dem Human Protein Index 'HPI'-Projekt (Anderson und Anderson 1982) – und bezeichneten es als das biologische Äquivalent zum Periodensystem der Elemente in der Chemie. Unter der Koordination von N.G. Anderson wurden staatliche Förderprojekte aufgesetzt, mit deren Hilfe das HPI-Projekt umgesetzt werden sollte. Die rasante Entwicklung der DNA-Technologie zu dieser Zeit lenkte das allgemeine Interesse jedoch mehr und mehr auf die Genomanalyse anstatt auf die Analyse der Proteine. Die Genomanalyse richtete sich zunächst auf die Erstellung genetischer Karten. Die erste genetische Karte des menschlichen Genoms, die auf DNA-Polymorphismen beruhte, wurde 1980 von Botstein (Botstein et al. 1980) entwickelt und beruht auf Restriktionsfragmentlängenpolymorphismen (RFLPs). RFLPs entstehen durch Unterschiede in der DNA-Sequenz von Individuen, die von sequenzspezifisch schneidenden Restriktionsenzymen erkannt werden. Ein Verdau genomischer DNA zweier Individuen resultiert daher in unterschiedlich langen DNA-Fragmenten. Mit Auffinden hochpolymorpher Repeatsequenzen (Dinukleotid-wiederholungen) in Genomen von Mensch und Maus stand eine weitere Quelle von genetischen Markern, die Mikrosatelliten, zur Verfügung. Diese können mit sequenzspezifischen Primern mittels PCR von einem Locus des Genoms amplifiziert werden und den Locus 'sichtbar' machen oder markieren. Aus der Vererbung von polymorphen Markern und einem varianten Phänotyp, z.B. verschiedener Fellfarben der Maus oder einer erblichen Krankheit des Menschen konnte durch Segregationsanalyse die chromosomale Region ermittelt werden, die für den Phänotyp verantwortlich war.

Einen weiteren Meilenstein stellte die Entwicklung von YAC-Vektoren dar (Burke et al. 1987): in YACs (yeast artificial chromosomes, künstliches Hefechromosomen) können sehr grosse genomische DNA-Fragmente (500 - 1000 kb) verpackt und vermehrt (kloniert) werden. YAC-Klonbanken wurden seither verwendet, verschiedene Genome in Form von überlappenden YAC-Klonen nachzubauen. Die Erstellung solcher physikalischer Karten des Genoms dient dazu, letztendlich die gesamte Nukleotidsequenz des Genoms zu ermitteln. Mit den gleichen Worten wie sie Anderson für das Human Protein Index-Projekt verwendet, bewertet Eric Lander in einem Artikel in der Zeitschrift Science das Human Genom Projekt als " ... die Entdeckung des Periodensystems der Elemente des 20 Jahrhunderts... . Ziel ist es, nicht nur 100 chemische Elemente, sondern 100000 Gene zu finden ... " (Lander 1996). Mitte der 80er Jahre wurde das Human Genom Projekt (HUGO) gestartet und zog während der folgenden 15 Jahre die Aufmerksamkeit der Wissenschaftsgemeinde und Förderer auf

sich. Diese Bestrebungen nähern sich nun dem Ende. Die Sequenz des menschlichen Genoms liegt in einer Arbeitsversion vor [Lander, 2001 und die Genome der wichtigsten Modellorganismen (Hefe (Goffeau et al. 1996), *Caenorhabditis elegans*, *Drosophila* (Adams et al. 2000), *Arabidopsis* (2000a) sind bereits sequenziert oder werden in absehbarer Zeit vorliegen (Maus, Ratte, Zebrafisch). Das lenkt die Aufmerksamkeit nun verstärkt auf die Frage nach der Funktion der neuentdeckten Gene und Genprodukte.

1.2 Funktionelle Genomanalyse

Einen immer grösseren Raum nehmen Methoden ein, die auf die örtliche, zeitliche und quantitative Analyse der Genexpression zielen. Sie werden als wichtige Methoden der 'Postgenomära' beschrieben (Nowak 1995). Einen Zelltyp oder ein Entwicklungsstadium anhand von Expressionsprofilen der RNA zu beschreiben sowie Kataloge aller exprimierten Sequenzen eines Organismus aufzustellen wird mit dem Begriff 'functional genomics' umschrieben.

Um die exprimierten Sequenzen des Genoms zu erfassen, wurden grosse EST-Datenbanken angelegt. ESTs - expressed sequence tags - sind cDNA-Klone, deren Enden ansequenziert wurden und die damit 'sequenzmarkiert' sind. Derzeit enthält die allgemein zugängliche dbEST (<http://www.ncbi.nlm.nih.gov/dbEST/>), die „database of ‚Expressed sequence tags‘ “ ca. 2.6 Millionen menschliche EST-Sequenzen und repräsentiert wahrscheinlich den grössten Teil der menschlichen Gene, deren Gesamtzahl auf 30 000 – 100 000 geschätzt wird (Hattori et al. 2000).

Die Entwicklung der SAGE Technik („Serielle Analyse der Genexpression“) ermöglicht es, alle exprimierten RNAs eines Gewebes zu quantifizieren. Hierzu werden die RNAs in cDNA umgeschrieben. Von jeder cDNA wird eine definierte 9 Basenpaar lange Sequenz ausgeschnitten, die die cDNA repräsentiert. Diese sogenannten 'Tags' (Markierungen) werden ligiert und sequenziert. Die Sequenzanalyse ermittelt zum einen die Identität jeder cDNA anhand des Vergleichs der 9bp langen Sequenzen mit EST-Datenbanken und zum anderen die Häufigkeit, mit der die cDNA vertreten ist (Velculescu et al. 1995).

Eine andere Methode basiert auf der Erstellung von 'Microarrays' von cDNAs (Skena et al. 1995; Lockhart et al. 1996). Hier werden bekannte cDNAs eines Gewebes auf einen kleinen Glasträger (1x1cm) aufgebracht. Die Analyse der differentiellen Genexpression zweier Gewebe wird mittels Hybridisierung durchgeführt: die RNA beider Gewebe wird in cDNA umgeschrieben und mit jeweils verschiedenen Fluoreszenzfarbstoffen markiert. Der Microarray wird entweder nacheinander mit den markierten cDNA-Proben hybridisiert oder die cDNA-Proben miteinander gemischt und hybridisiert. Dem gleichen Ziel - der quantitativen und qualitativen Analyse aller exprimierten Gene eines Gewebes - dienen die von Hans Lehrach entwickelten gewebespezifischen cDNA Klonbanken, die auf Nylonfilter aufgebracht und durch Hybridisierung mit Oligonukleotiden charakterisiert werden (Lehrach et al. 1990; Maier et al. 1997), sowie das 'Differential Display' (Liang und Pardee 1992; Bauer et al. 1993). Die Untersuchung der RNA-Expression reicht jedoch für eine funktionelle Genomanalyse nicht aus. In letzter Zeit wurde gezeigt, dass sich die bei der RNA-Mengenanalyse gefundenen Quantitäten nicht mit

der real existierenden Menge an Protein korrelieren lassen. Anderson und Seilhamer haben die Häufigkeit vieler RNAs (ESTs) und Proteine eines 2D-Gels gemessen und einen Korrelationskoeffizienten von 0,48 gefunden (Anderson und Seilhamer 1997). Tew untersuchte die Korrelation des Proteins Gluthathione-S-Transferase mit Immunoaffinitäts-chromatographie in 60 Tumorzelllinien im Vergleich zur exprimierten RNA (quantitative Northern Blot Analyse) und fand einen ähnlichen Korrelationskoeffizienten von 0,43 (Tew et al. 1996). Für das Verhältnis der exprimierten RNA im Vergleich zur Menge Protein muss eine grosse bis zu 10-fache Variationsbreite angenommen werden. Die Menge an Protein scheint also durch zusätzliche Faktoren reguliert zu werden. Dies können Protein-Protein-Interaktionen sein, die die Synthese und den Abbau regulieren und die selbst durch Umwelteinflüsse beeinflussbar sind oder Regulationsmechanismen der Translation, sowie posttranslationale Modifikationen, die die Proteinstabilität und die Degradation modifizieren.

Eine weitere Beobachtung stellt differentielles Splicing einer RNA dar: von einem RNA-Strang können verschiedene Proteine translatiert werden. Ferner treten an Proteinen in grosser Zahl Modifikationen auf, die nicht über die RNA zu erkennen sind. Als Beispiel kann das tau-Protein dienen, das bei der Entstehung von Alzheimer wichtige Rolle spielt: aus dem transkribierten RNA-Strang können durch unterschiedliche Prozessierung (differentielles Splicing) sechs verschiedene RNA-Stränge entstehen. Diese werden in sechs verschiedene Proteine translatiert. Die translatierten Proteine erfahren eine vielfache Phosphorylierung, so dass das tau-Gen sich in einem 2D-Gel in ca. 80 Proteinspots darstellt (Janke et al. 1996). Es sind hunderte verschiedene posttranslationale Modifikationen für Proteine beschrieben worden. Für einige wurde die Bedeutung bei der Entstehung von Krankheiten nachgewiesen (s.u.).

1.3 Proteomanalyse – die Wiederentdeckung der Proteine

Proteine stellen das aktive Leben dar, üben Funktionen aus und gestalten andere Proteine, während das Genom lediglich den Bauplan der Proteine beinhaltet und die RNA ein Informationsträger, eine Vorlage für die Synthese eines Proteins ist. Völlig unabhängig von jeglicher RNA-Expression verändern Protein-Interaktionen innerhalb der Zelle den ‚Proteomzustand‘ der Zelle: Die Reifung sowie der Abbau von Proteinen sind Prozesse, die auf Proteinebene geregelt werden (Hargrove und Schmidt 1989). Anschliessend an die Synthese können eukaryontische Proteine mit einer Vielzahl posttranslatinaler Modifikationen versehen werden, die Enzymaktivitäten verändern oder durch Abspalten von Signalsequenzen den Ort des reifen Proteins innerhalb der Zelle bestimmen. Signaltransduktion aus dem extrazellulären Raum über die Zellmembran in die Zelle hinein ereignet sich, indem G-Protein-gekoppelte Rezeptoren (7-TM GPCRs) als Folge der Bindung eines extrazellulären Liganden einen intrazellulären Liganden phosphorylieren oder dephosphorylieren und damit seine Aktivität regulieren (Wess 1997). Die Freisetzung membrangebundener Proteine als extrazelluläre Signale, wie z.B. des Alzheimer'schen beta-Amyloids aus dem

Amyloidprecursorproteins APP, geschieht durch geregelte Proteolyse (desgleichen Angiotensin, Tumornekrosefaktor, Interleukine) (Hooper et al. 1997). Das wirkliche Leben, Funktionieren und Reagieren einer Zelle oder eines Organismus findet also auf Proteinebene statt. Die Technik, mit der sich hunderte und, heutzutage weiterentwickelt, tausende Proteine eines Zelltyps in einem einzigen Experiment analysieren lassen – die 2-dimensionale Elektrophorese, wurde gleichsam mit den Proteinen nach 2 Jahrzehnten wiederentdeckt. Gleichzeitig wurden aber auch die Limitierungen dieser Technik deutlich. Sie begründen sich zum einen darauf, dass Proteinspots aus 2-D Gelen einen sehr geringen Proteingehalt haben, der für analytische Techniken zur Identifizierung des Proteins wie Edman-Sequenzierung oder Aminosäureanalyse meist nicht ausreichte. Ein weiterer Grund bestand in der geringen Reproduzierbarkeit eines 2D-Musters, wenn eine Proteinprobe in verschiedenen Laboratorien aufgetrennt wurde. Aber selbst mit ein und demselben Gelsystem waren Unterschiede der Muster nicht völlig zu vermeiden. Die Analyse, der Vergleich der Muster, enthält Elemente qualitativer ja/nein Information, der Grossteil der erzielten Daten ist jedoch von quantitativer Natur. Dies steht ganz im Gegensatz zu den genomischen Datenbanken, die sich mit digitalisierbarer Information füllten. Der Vergleich von 2D-Mustern zur Ermittlung von Spotvolumina erfordert computergestützte Spoterkennung mit zeitaufwendiger Editierung einzelner Spots, besonders in Regionen hoher Spotdichte, sowie zeitaufwendigem Zuordnen der entsprechenden Spots verschiedener Gele (Spotmatching).

Zur Lösung des Hauptproblems, der Identifizierung der 2D-Spots, wurden Fortschritte in den 80er Jahren gemacht: Mit der Entwicklung von Techniken zum Transfer der Proteinspots aus dem 2D-Gel auf eine Membran (Westernblot) konnten Proteinspots durch Bindung eines spezifischen Antikörpers identifiziert werden. Die auf diesem Wege gleichzeitig erreichte Reinigung von Proteinen kleinster Menge war dann auch die Grundlage für die Entwicklung von Mikroverfahren in der Aminosäuresequenzierung (Aebersold et al. 1986; Lopez et al. 1994). Den entscheidenden Fortschritt in dieser Richtung brachten jedoch Weiterentwicklungen der Massenspektrometrie zur Bestimmung von Molekülmassen grosser Biomoleküle, insbesondere zur Bestimmung von Peptidmassen fragmentierter Proteine: die Laser Desorptions Massenspektrometrie (Karas and Hillenkamp 1988; Hillenkamp and Karas 1990), sowie die Elektrospray Massenspektrometrie (Fenn et al. 1989) ermöglichten die Identifizierung der bis dahin weitgehend anonymen Proteinspots eines 2-D Gels. Identifizierung meint: die Zuordnung des Proteins zu einer Aminosäuresequenz in einer Proteindatenbank oder einer theoretisch translatierten DNA-Sequenz einer DNA-oder EST-Datenbank. Grundlage hierfür ist die Massengenauigkeit der Massenspektrometrie sowie die Exaktheit der Suchalgorithmen, die Datenbanksequenzen nach den gemessenen Massen durchforsten. Suchalgorithmen wurden entwickelt von Daryl Pappin (Pappin 1997), Matthias Mann (Mann und Wilm 1994), R. Aebersold (Figeys et al. 1999) und J. Yates (Yates et al. 1993).

Nach Blackstock kann man in der Proteomforschung zwei Richtungen unterscheiden, die mit den Begriffen ‚Expression Proteomics‘ und ‚Functional Proteomics‘ umrissen werden können (Blackstock und Weir 1999). Unter Expression Proteomics versteht man diejenigen Experimente, die darauf

abzielen, alle exprimierten Proteine eines Organismus – das Proteinkomplement seines Genoms - systematisch zu erfassen, und zwar qualitativ wie quantitativ. Die Methode der Wahl hierfür ist die Auftrennung der komplexen Proteinmischung mittels 2-dimensionaler Elektrophorese. Die grösste Zahl trennbarer Proteinspots und damit höchste Auflösung einer Probe lässt sich gewinnen, wenn eine hochkonzentrierte, gut solubilisierete Proteinprobe mit einer grossformatigen 2D-Technik kombiniert wird (Klose und Kobalz 1995). Es wurde gezeigt, dass mit der von Klose beschriebenen 40x30cm grossen Trennfläche Proteinextrakte aus Säugetiergeweben in 5000 - 10000 Proteinspots aufgetrennt und mit Silberfärbung nachgewiesen werden können. Durch fraktionierte Extraktion von Geweben der Maus in eine lösliche, eine Membran- und eine Kernfraktion wurde eine Auflösung von über 11000 unterschiedlichen Spots pro Gewebe erreicht. Zu Expression Proteomics zählen 2-D Gelanalysen von Einzellerkulturen in verschiedenen Entwicklungsstadien sowie Analysen von verschiedenen Organen und Zelltypen eines Organismus, die zum Ziel haben, organspezifische Proteine ausfindig zu machen. Ebenso gehören dazu Experimente, in denen ein gesundes Organ mit einem erkrankten oder ein junges Gewebe mit einem alten verglichen werden (subtraktive Analyse). Kennzeichnend für diese Experimente ist der Einsatz einer möglichst gering manipulierten Ausgangsprobe hoher Komplexität. Mit der Analyse ganzer Proteome wurde bei Mikroorganismen begonnen. Die Analyse des Proteoms der Hefe *Saccharomyces cerevisiae* wird im Rahmen des ‚European Functional Analysis Network‘ (EUROFAN) durchgeführt (Dujon 1998). Sie erbrachte bis 1994 eine Auftrennung von 1500 Proteinspots von denen 40 – 60 durch Mikrosequenzierung und Aminosäureanalyse identifiziert wurden (Garrels et al. 1994). Fortschritte in der MS haben bis 1997 zur weiteren Identifizierung der doppelten Anzahl von Proteinspots geführt hat (Shevchenko et al. 1996a). Bis August 1999 wurden 401 Proteinspots des Proteoms von *Saccharomyces cerevisiae* identifiziert. Sie gehörten 279 verschiedenen Genen an (Perrot et al. 1999).

Functional Proteomics richtet – nach der genannten Definition – sein Interesse auf die Reaktionen, die das Proteom eines Organismus oder Organs nach Manipulation mit Pharmaka, nach Induzierung von Mutationen oder auf Umwelteinflüsse hin zeigt: Das Ziel besteht darin, all diejenigen Proteine zu entdecken, die sich aufgrund natürlicher oder manipulierter Einflüsse hin verändert haben. In diesen Bereich fallen toxikologische, teratologische und Mutagenitäts-Untersuchungen. Interessante Untersuchungen hierzu wurden von Anderson durchgeführt. Sie zeigten, dass der Grad, mit dem die Zelle versucht, die Aktivität von zwei Genprodukten zu koordinieren, ein Mass ist für die funktionelle Verwandtschaft dieser Proteine. Anderson belegt sein Konzept, Proteine auf der Basis ähnlicher Regulation zu gruppieren, mit dem Begriff der ‚Regulatorischen Homologie‘, angelehnt an den Begriff der Sequenzhomologie von DNA oder Proteinen (Anderson und Anderson 1998). Basierend auf der Manipulation durch Pharmaka - sozusagen chemischen knockouts von Proteinen - will Anderson systematisch Funktionen uncharakterisierter Proteine von ihrer Koregulation mit bereits bekannten Proteinen ableiten und Regulations-Netzwerke aufdecken. Er hat dazu die Datenbank ‚Molecular Effects of drugs‘ (MEDTM) etabliert. Sie enthält Daten, die an Ratten *in vivo* gewonnen wurden. Diese stellen Reaktionen des Leberproteoms auf 51 pharmazeutische Chemikalien dar, die derzeit in

Benutzung sind. Den grössten Teil der Daten stellen quantitative Veränderungen von Proteinspots in 2D-Mustern dar: die wichtigsten Entdeckungen werden, so Anderson, durch Quantifizierung einer grossen Zahl von Proteinen gemacht. Ein anderer Ansatz der Funktionellen Proteomanalyse zielt hingegen nur auf eine Auswahl der exprimierten Proteine eines Gewebes ab; auf solche nämlich, die einen Komplex bilden und damit im Verdacht stehen, innerhalb der Zelle miteinander zu interagieren (Neubauer et al. 1998; Blackstock und Weir 1999; Rigaut et al. 1999). Es geht auch hier um die Aufklärung von Proteinfunktion durch die Analyse derjenigen Proteine, die eine gemeinsame Funktionseinheit - einen Proteinkomplex - bilden.

1.4 Proteomanalyse – Stand der Technik

Die 2D-Elektrophorese als Methode zur Auftrennung komplexer Proteinmischungen lässt sich mit dem Klonieren von Genen vergleichen. Die Sequenzierung des menschlichen Genoms wird eine lange Zeit brauchen, um die letzten Lücken der DNA-Sequenz zu schliessen und damit die letzten 5% aller Gene zu identifizieren. Ebenso sind der Auftrennung einer Proteinprobe Grenzen gesetzt. Bei maximaler Auflösung liegt die Anzahl von 10000 Proteinspots in der Grössenordnung der Zahl unterschiedlicher, exprimierter RNA-Spezies pro Zelle. Allerdings hat die Identifizierung von 250 Proteinspots eines 2D-Gels von Mausgewebe ergeben, dass ein Protein durchschnittlich durch drei Spots eines 2D-Gels vertreten wird (Gauss et al. 1999).

1.4.1 Gelsysteme

Für die Durchführung der 2D-Elektrophorese sind heute zwei unterschiedliche technische Systeme in Gebrauch. Sie unterscheiden sich in der Durchführung der isoelektrischen Fokussierung, d.h. der Trennung der Proteine aufgrund ihres Isoelektrischen Punktes in der ersten Dimension. Nach dem ursprünglichen Verfahren (Klose 1975; O'Farrell 1975) werden für den Aufbau des pH-Gradienten freibewegliche Trägerampholyte (carrier ampholytes, CA-Technik) verwendet, nach dem neueren Verfahren acrylamidgebundene Trägerampholyte (Immobiline, IPG-Technik) (Bjellqvist et al. 1982; Gorg et al. 2000). Bei der Isoelektrischen Fokussierung mit freibeweglichen Trägerampholyten wird der pH-Gradient während des Gellaufs aufgebaut, während die IPG-Technik Ampholyte benutzt, die an Acrylamid, die Gelmatrix, gebunden sind. IPG-Gele werden aus einer Acrylamidlösung basischer Immobiline und eine Acrylamidlösung saurer Immobiline mit Hilfe eines Gradientenmischers hergestellt. Immobiline bilden einen stabilen ins Gel eingegossenen pH-Gradienten. Beide Techniken haben Vor- und Nachteile. Bei der IPG-Technik kann sich der pH-Gradient nicht durch Kathodendrift verschieben, was einen Vorteil im Hinblick auf die Reproduzierbarkeit der Trennung darstellt. Der Transfer der isoelektrisch fokussierten Proteine aus dem 1D-Gel in das 2D-Gel ist jedoch dadurch erschwert, dass die Immobiline nicht mit den Proteinen aus dem 1D-Gel herauswandern können. Das führt zu unsauberer Trennung der Proteinspots in der zweiten Dimension (vertikale Streaks). Der

grosser Vorteil CA-basierter Gele besteht in der hohen Auflösung der Trennung. Der pH-Gradient lässt sich auf eine Gellänge von 40 cm strecken (Klose und Kobalz 1995). Die Entwicklung einer CA-basierten Isoelektrischen Fokussierung für eine weiter vergrösserte Trennstrecke (>40cm) ist durch die begrenzte Verfügbarkeit unterschiedlicher pH-spezifischer Trägerampholyte limitiert. Ebenso würden längere Laufzeiten für längere Gele zu Diffusion der Proteine führen und damit zu schlecht fokussierten Spots. Mit Immobilinen konnte keine bessere Auflösung in 40cm langen IEF-Gelen erzielt werden, da die Proteinspots eine längliche, bandenartige Form annehmen. Um die Trennstrecke über den pH-Gradient zu erhöhen, haben Bjellqvist IPG-Strips mit engen pH-Gradienten, z.B. pH 4,5 – pH 5,5 entwickelt (AmershamPharmaciaBiotech; Bjellqvist et al. 1993b; Wildgruber et al. 2000). Der Vorteil scheint tatsächlich in einer höheren Auflösung zu bestehen. Der Nachteil besteht darin, dass ein Vielfaches an Proteinprobe nötig ist, um diese Auflösung zu erreichen und dass mehrere überlappende Proteilmuster für die Analyse korrekt zusammengesetzt werden müssen. Für die Reproduzierbarkeit von 2-DE Mustern ist die Reproduzierbarkeit des pH-Gradienten im IEF-Gel ausschlaggebend. Schwierig ist es, mit freibeweglichen Trägerampholyten einen stabilen pH-Gradienten im basischen Bereich zu erhalten. Da die TA-basierten Gele in ein Glasrohr hineingegossen werden und während der Separation schrumpfen, kann sich Elektrodenlösung zwischen Gel und Innenwand des Glasröhrchens hochziehen und den pH-Gradienten zerstören. Eine Lösung bietet die Einführung eines 12% ‚cap‘ Gels zur Versiegelung des basischen Endes des IEF-Gels (Klose 95). Weiterhin wird die Fokussierung nach einem festgelegten Protokoll durchgeführt: die Fokussierung wird gestoppt bevor die basischen Proteine ihren Isoelektrischen Punkt erreichen. Dadurch wird verhindert, dass sie aus dem Gel hinauswandern. Immobilinglee weisen einen stabilen pH-Gradienten auch im basischen Bereich auf (Gorg 1999; Gorg et al. 1999). Die IPG-Technik ist heute weiter verbreitet als CA-Techniken, da sie benutzerfreundlicher angeboten wird: fertige IPG-Gele und zugehörige Puffer können von verschiedenen Firmen (AmershamPharmaciaBiotech, Biorad) bezogen werden. Immobilinglee werden als besser reproduzierbar angesehen – Gele verschiedener Labors können angeblich computergestützt ausgewertet und verglichen werden (Corbett et al. 1994; Blomberg et al. 1995) - Trennungsergebnisse sind jedoch genau wie bei der CA-Technik von der verwendeten Charge der Ampholyte abhängig. IPG-basierte Referenzkarten verschiedener Gewebe und Organismen werden im Internet veröffentlicht (z. B. SWISS-2DPAGE (Sanchez et al. 1995), (Boutell et al. 1994). Mit der Zeit wurde jedoch erkennbar, dass sich die Idee der Referenzkarte für ein Gewebe oder einen Organismus, an dem sich alle Labors weltweit orientieren, sich nicht realisieren lässt. Es stellte sich heraus, dass Referenzkarten /-muster nur für eine exakt definierte Technik innerhalb eines Labors geeignet sind. Die Reproduzierbarkeit der CA- sowie der IPG-Technik hängt zum grössten Teil von den Ampholyten ab, die für die Isoelektrische Fokussierung eingesetzt werden. Die Synthese der Ampholyte – der Carrierampholyte sowie der Immobiline – ergibt von Charge zu Charge variierende Ergebnisse. Für Experimente mit käuflich erworbenen Immobilingleen besteht daher immer das Risiko, von Bestellung zu Bestellung Gele mit leicht unterschiedlichen pH-Gradienten zu erhalten. Anwender der Carrierampholyttechnik profitieren davon, dass sie ihre Ampholyte selbst zusammenstellen und Aliquote einfrieren und damit die eigene Ampholyt-Charge für eine grosse Zahl

von Experimenten über Jahre hinweg verwenden können (Klose und Kobalz 1995). Der weitere grosse Vorteil der CA-Technik liegt in der höheren Auflösung: sie liegt mit etwa der doppelten Anzahl von Spots weit über derjenigen von IPG-Gelen.

1.4.2 Proteinextraktion

Das Ziel der Proteinextraktion ist es, alle vorhandenen Proteinspezies eines Gewebes quantitativ zu extrahieren und zu solubilisieren. Für die Extraktion von Organen hat sich eine Fraktionierung in drei Fraktionen bewährt. Diese zeigen auf insgesamt drei 2D-Gelen ein Maximum an verschiedenen Spots. Zellfraktionierung hat normalerweise zum Ziel, ein Zellkompartiment wie z.B. Mitochondrien oder Membranen in mehreren Aufarbeitungsschritten möglichst gut zu reinigen und davon Proteine zu extrahieren. Das kann aber den Verlust von bestimmten Klassen von Proteinen durch Fällung, Waschschriffe, Lyophilisieren und Dialysieren zur Folge haben und stellt nicht das Ziel der Proteinextraktion für 2-DE dar. Die von Klose beschriebene Fraktionierung von Geweben (Klose 1999a) zielt darauf ab, alle in einem Gewebe vorhandenen Proteine quantitativ zu extrahieren. Gewebe wie Leber, Herz oder Gehirn von Maus (oder Mensch) werden daher in folgende drei Extrakte fraktioniert:

Den Überstand I+II, der Tris-Pufferlösliche Proteine des Cytoplasmas und der Zellorganellen enthält; das Pelletextrakt: es enthält Proteine, die in 9M Harnstoff und CHAPS löslich sind, das sind Membranproteine und andere Strukturproteine und die Pelletsuspension, die Proteine enthält, die nach DNA-Verdau von der DNA freigesetzt werden, wie Histone und andere chromosomale Proteine.

Ein weiteres wichtiges Ziel, nämlich die Proteolyse der Proteine während der Aufarbeitung zu minimieren, wird dadurch erreicht, dass das Gewebe zusammen mit einem definierten Volumen Proteasehemmercocktails in flüssigem Stickstoff homogenisiert wird und so die Hemmer beim Auftauen gleich überall im Homogenat wirksam werden. Das Ziel einer maximalen Solubilisierung der Proteine wird dadurch sichergestellt, dass Proteine unter möglichst natürlichen Bedingungen extrahiert werden: dies sind Bedingungen wie sie in der lebenden Zelle vorgefunden werden. Die Ionenstärke des Gewebehomogenats wird zwischen 150 und 200mM gehalten und der pH-Wert zwischen 7.0 und 7.5, die Proteine werden möglichst bei ihrer natürlichen Konzentration belassen, der Extraktionspuffer enthält Glycerol zum Schutz der Proteine vor Wasser. Die optimalen Bedingungen wurden von Klose empirisch ermittelt. Eine zu hohe Proteinkonzentration des Extraktes verhindert eine effektive Ausschaltung von Proteasen, was in einer erhöhten Zahl von Spots kleinen Molekulargewichtes sichtbar wird. Je stärker die Proteinprobe jedoch verdünnt wird, desto weniger der niedermolekularen Proteine bleiben löslich, während die hochmolekularen sich besser zu lösen scheinen und an Intensität und Zahl zunehmen (Klose 1999a). Die optimale Proteinkonzentration einer Probe hängt demnach von dem jeweiligen Gewebe ab.

1.4.3 Quantitative Analyse von 2D-Mustern

Der grösste Teil der Unterschiede zwischen 2D-Mustern eines kranken und gesunden oder jungen und alten Gewebes besteht in quantitativen Veränderungen von Proteinspots, d.h. ein Spot eines 2D-Gels färbt sich stärker als sein korrespondierender Spot im Referenzgel. Gegenwärtige Färbemethoden lösen das Problem der Quantifizierung nicht befriedigend. Die beiden wichtigsten Proteinfärbungen für die Quantifizierung von Spots sind die Coomassiefärbung und die Färbung mit den Fluoreszenzfarbstoffen SYPRO Orange, SYPRO Red und SYPRO Rubi. Die Silberfärbung zeigt nur über einen kleinen Bereich von 1 – 50ng Protein eine lineare Zunahme der Optischen Dichte mit der Proteinkonzentration. Bei grösseren Proteinmengen erreichen die Spots schnell die Sättigung. Die weniger sensitiven Coomassiefärbungen R-250 und G-250 weisen einen weiteren linearen Färbungsbereich auf als die Silberfärbung. Besonders das colloidale Coomassie G-250, die dimethylierte Form von CBB R-250, detektiert Proteinmengen von 8-10ng und erlaubt damit die Quantifizierung von Proteinmengen innerhalb des Bereichs von 10ng bis einigen 100ng Protein pro Spot (Neuhoff et al. 1988; Neuhoff et al. 1990). Eine neue, genauso einfach wie die Coomassiefärbung durchzuführende Färbung mit dem Fluoreszenzfarbstoff SYPRO Rubi, soll quantitativ über einen Bereich von 3 Grössenordnungen (1ng – 1000ng Protein) färben (Berggren et al. 2000; Patton 2000; Valdes et al. 2000). Sie entspricht in der Sensitivität jedoch nur der Coomassiefärbung und stellt damit keine Alternative zur empfindlichen Silberfärbung dar. Die beste Wahl ist daher immer noch die colloidale Coomassiefärbung. Ein grosses Problem besteht darin, dass die Proteinmenge der einzelnen Proteine einer Probe über 7-8 Grössenordnungen variiert und dass häufige Proteine, die grossen, stark gefärbten Spots eines 2D-Gels, kleine Proteinspots überlagern können. Durch Fraktionierung der Proteinprobe mit chromatographischen oder elektrophoretischen Methoden (Hochstrasser et al. 1991) oder durch Zellfraktionierung können Proteine, die in wenigen Molekülen pro Zelle vorliegen, angereichert werden, allerdings unter Gefahr der Veränderung der quantitativen Verhältnisse in der Probe und mit einem zusätzlichen Risikofaktor für die Reproduzierbarkeit des Experiments.

1.4.4 Bildanalyse von 2D-Mustern

2D-Gele können tausende Proteinspots enthalten. Der Vergleich mehrerer Gele auf einem Lichtkasten ohne technische Hilfsmittel ist subjektiv und sehr mühsam. Daher wurden verschiedene Auswertungsprogramme entwickelt, die die Bildanalyse von 2D-Mustern erleichtern sollen und eine teilweise Automatisierung der erlauben. Am bekanntesten sind die folgenden Computerprogramme: Phoretix-2D-Analysis, Melanie II (Appel et al. 1997) sowie PDQuest (Garrels 1989). Das Vorgehen stellt sich folgendermassen dar: die 2D-Muster werden mittels Laserscanner oder CCD-Kamera aufgenommen und die Messwerte digitalisiert. Das entstandene Computermuster wird mittels 2D-Software analysiert mit dem Ziel der Spoterkennung, des Spotmatching (Zuordnen von Spots

verschiedener Gele), der Hintergrundkorrektur und der Normalisierung der Muster. Als Ergebnis dieses Prozesses erhält man quantitative Messwerte des Spotvolumens der einzelnen Spots. Die computergestützte Quantifizierung bestimmt das Spotvolumen eines Spots aus Spotfläche und optischer Dichte. Die grösste Herausforderung stellt die korrekte Spoterkennung dar. Dicht beieinanderliegende Spots werden oft als ein einziger Spot markiert, schwache Spots werden nicht erkannt und Streifen (Streaks) sowie Unreinheiten werden fälschlich als Spots eingekreist. Mit grossem Aufwand muss daher nach der computergestützten Spoterkennung das manuelle Editieren der Muster durchgeführt werden. Auch der nächste Schritt der Analyse, das Zuordnen (Matchen) von Spots mehrerer 2D-Gele ist zeitintensiv und kann nur teilweise computergestützt geschehen: sowohl während der Isoelektrischen Fokussierung als auch im SDS-Gellauf können Verzerrungen des Spotmusters auftreten. Dies kann an der leicht unterschiedlichen Beladung der IEF-Gele, der unterschiedlichen Proteinkonzentration der Proben oder den Laufbedingungen (z.B. Temperatur) liegen. Die nicht 100%ige Reproduzierbarkeit der geladenen Proteinmenge sowie der Intensität der Färbung verschiedener Gele verursachen Probleme bei der Spoterkennung, während das Abgleichen von Spots verschiedener Gele (Spotmatching) durch die nie 100%ig übereinstimmenden Koordinaten der Position eines Spots auf verschiedenen Gelen erschwert wird.

2D-Gele sowie synthetisch per Computer erstellte 2D-Standardmuster - auch Master- oder Referenzgele genannt - werden heute im Internet präsentiert (Swiss 2D-Page: <http://expasy.cbr.nrc.ca/ch2d>; Keratinozyten-2-D-Gel-Datenbank: <http://biobase.dk/cgi-bin/celis>; Maus-Gehirn 2D-Muster: <http://www.ukrv.de/humangenetik>). Dahinter stand ursprünglich die Idee, Charakterisierungs- und Identifizierungsdaten für einen Proteinspot eines Gels auf Gele anderer Laboratorien zu übertragen und die Charakterisierung verschiedener Gewebe durch Zusammenarbeit vieler Laboratorien voranzutreiben. Diese Idee muss heute als gescheitert betrachtet werden, da jedes 2D-Labor eigene Protokolle für die Proteinextraktion aus Geweben, die Erstellung der 1D- und 2D-Gele sowie für die Färbung verwendet. Die Reproduzierbarkeit der Proteinmuster derselben Proteinprobe in verschiedenen Gelsystemen ist schlecht.

1.4.5 Massenspektrometrie von Proteinen

Die Massenspektrometrie nimmt für die Charakterisierung von Proteinspots denselben Stellenwert ein wie die Sequenzierung für die Charakterisierung von DNA. Für die Arbeitsschritte des Ausschneidens der Spots aus dem Gel, sowie für den proteolytischen Verdau, die Konzentrierung und Reinigung der gewonnenen Peptide sind Roboter käuflich zu erwerben (der MassPREP-Roboter der Firma Micromass, UK sowie das ProteomeWorks™ System, das Micromass zusammen mit Bio-Rad entwickelt hat). Zur Diskussion steht auch, die Proteine des 2D-Gels auf eine Membran zu transferieren, dort proteolytisch zu verdauen und die gesamte Membran mit Massenspektrometrie abzuscanen (Binz et al. 1999) oder sogar den Transferschritt auf die Membran wegzulassen. Nach Anderson (Anderson und Anderson 1998) werden in 3-10 Jahren alternative Trenntechniken wie

Kapillarelektrophorese und Flüssigchromatographie gekoppelt mit Massenspektrometrie alle quantitativen und qualitativen Analysen einer Proteinprobe erlauben und Gele überflüssig machen. Ein wichtiger Schritt in diese Richtung geschah mit der Entwicklung eines Reagens durch Gygi im Labor von Rudi Äbersold, dem ICAT-Reagenz (isotope-coded affinity tags) (Gygi et al. 1999a). Das Reagenz enthält eine Sulfhydrylgruppe, die die kovalente Bindung an die SH-Gruppen von Proteinen erlaubt. Es enthält einen sogenannten ‚linker‘, der entweder 8 Deuteriumatome ^2H oder acht normale Wasserstoffatome trägt ^1H und es enthält eine Biotin-Gruppe, die die Aufreinigung von gebundenen Peptiden über eine Affinitätssäule ermöglicht. Mit dem ICAT-Reagenz wurde es möglich, fertig extrahierte Proteinproben zu markieren – eine Probe mit dem leichten ICAT-Reagenz, die andere Proteinprobe mit dem schweren ICAT-Reagenz. Nach Mischen der Proben, Trypsinverdau und Flüssigchromatographie werden die – weil sie identische chemische Eigenschaften haben – gemeinsam eluierten Peptide (ein leichtes, ein um 8Da schwereres) gleichzeitig massenspektrometrisch analysiert. Die Peakhöhe der beiden um 8Da versetzten Peaks spiegelt das quantitative Verhältnis dieses Proteins in den beiden Proteinextrakten wider. Besonders wertvoll ist diese Methode, weil eine Quantifizierung von Proteinen zweier Proteinproben mit Massenspektrometrie alleine nicht möglich ist: die Amplitude der Peptidsignale variiert stark und korreliert nicht mit der aktuellen Peptidmenge. Als ein vollständiger Ersatz für die 2D-Elektrophorese kann die ICAT-Methode jedoch nicht betrachtet werden, denn sie wird modifizierte Proteine, Isoformen und genetisch variante Proteine nicht gesondert erfassen. Bisher ist auch nicht geklärt, wieviele verschiedene Proteine eines komplexen Proteinextraktes, z.B. eines Gewebes, wirklich erfasst werden können. Die Massenspektrometrie wird auch in Zukunft wichtig für den Vergleich von Proteinmustern über Laborgrenzen hinaus sein. Da ein weltweiter Labor-zu-Labor-Vergleich von 2D-Mustern auf Bildanalyse-Ebene nicht durchführbar ist, werden alle Hoffnungen auf die Massenspektrometrie gesetzt, diese Probleme in Zukunft lösen. Der Vergleich von Proteinspots erfolgt dann auf der Ebene von Peptidprofilen und damit über Datenbanken. Dazu müssen die verschiedenen Massenspektrometrie-techniken jedoch auf einen sehr hohen, automatisierten Durchsatz hin entwickelt werden. Die Firma Toplab erreichte Mai 1999 eine Identifizierung von 50 – 60 Spots täglich (C. Eckerskorn, persönliche Mitteilung). Hoffman-Laroche gibt eine Charakterisierungsrate von 500 Spots pro Tag an.

1.5 Proteomanalyse im Vergleich zur Genomanalyse

1.5.1 Methodische Aspekte

Zwischen den DNA-Analysetechniken und den Proteinanalysetechniken lassen sich gewisse Parallelen ziehen: Klonierung von DNA bedeutet, das Genom in handhabbare Stücke zu zerlegen. Proteine eines Proteinextraktes lassen sich in einem 2-D Gel in einzelne Proteinspots auftrennen. Ein DNA-Klon lässt sich durch Sequenzierung der Nukleotide charakterisieren, ein Proteinspot lässt sich

mit Massenspektrometrie charakterisieren und teilweise sequenzieren (Aminosäuresequenz). Jedoch ist der Proteinspot nicht amplifizierbar, weder durch PCR noch durch Amplifizieren des DNA-tragenden Plasmid- oder Phagenvektors in einem Wirtsbakterium. Während sich im Rahmen des Humangenomprojektes für die weltweite Verteilung von DNA-Klonen vier Ressourcenzentren sowie das IMAGE-Konsortium etabliert haben, arbeitet jedes 2D-Labor mit unterschiedlichen, nicht standardisierten Proteinextrakten und Elektrophoresesystemen. Zur Analyse der Genexpression werden DNA-Filter oder DNA-Chips mit Gesamt-RNA verschiedener Gewebetypen hybridisiert. Dies erfordert in gleichem Masse eine qualitative und quantitative Image-Analyse wie der Vergleich von 2-dimensionalen Gelen. Die Expressionsanalyse auf Nukleinsäureebene hat hier jedoch den Vorteil, dass die DNA-Segmente geordnet an festgelegte Positionen einer DNA-bindenden Membran (oder in kleinerem Format auf ein ‚Chip‘) gebunden werden, bevor sie mit der zu untersuchenden cDNA/RNA Probe hybridisiert werden, während beim Vergleich von Proteinspots verschiedener Gele das Zuordnen (Matchen) der Spots Zeit kostet und nur teilweise computergestützt durchgeführt werden kann.

1.5.2 Syntheserate und zelluläre Konzentration der Proteine

Die Proteine verfügen molekular über alle Eigenschaften, um die Funktion des zugehörigen Gens zu erfüllen. Alle Eigenschaften eines Proteins jedoch, die über die Sequenzhomologie von Gen und Protein hinausgehen, lassen sich nicht an der DNA ablesen. Das betrifft die Syntheserate und zelluläre Konzentration der Proteine und die strukturelle Modifizierung der Proteine durch posttranslationale Prozesse. Auf den ersten Punkt wurde bereits eingegangen. Es wurden Untersuchungen durchgeführt, denen zufolge die zelluläre Konzentration der Proteine häufig nicht konform geht mit der Konzentration der zugehörigen mRNA (Anderson und Seilhamer 1997; Gygi et al. 1999b).

1.5.3 Posttranslationale Modifikationen der Proteine

Welche posttranslationalen Modifikationen (PTM) existieren? Die translatierte Aminosäure-kette kann eine grosse Zahl von Modifikationen erfahren, um ihre Funktion in der Zelle zu erfüllen (Packer und Harrison 1998). Oft zeigen der Modifikationszustand eines Proteins und die spezifische Konstellation modifizierter Proteine – z.B. der Phosphorylierungsgrad von Proteinen einer Signaltransduktionskette – den Aktivitätszustand der Zelle an. Wichtige posttranslationale Modifikationen sind die Glykosylierung und die Phosphorylierung, die Abspaltung von Signalsequenzen, die Entfernung des Start-Methionins, die Ausbildung von Disulfidbrücken innerhalb eines Peptids oder zwischen mehreren Peptiden sowie das Anhängen von Fettsäuren (Palmitoylierung). Welche Modifikationen wirklich vorgenommen werden, lässt sich aus der reinen DNA- oder Aminosäuresequenz des Proteins nicht

ableiten und hängt von der Faltung ab, die die Proteinkette nach ihrer Synthese annimmt. Diese Faltung bestimmt die Zugänglichkeit der Aminosäuren für Modifikationen.

PTMs sind wichtig für die korrekte Funktion eines Proteins. Das Fehlen einer Modifikation kann Ursache für eine bestimmte Krankheit sein (Ohno et al. 1992; Yamashita et al. 1993). In eukaryontischen 2DE-Mustern hat man Serien von Spots – sogenannte ‚trains‘ (Hochstrasser et al. 1992) – in Verbindung mit Krankheiten gebracht (Duthel und Revol 1993; Yamashita et al. 1993; Sarto et al. 1997; Sarto et al. 1999). Zum Beispiel zeigt das tau-Protein in Gehirngewebe von Alzheimerpatienten Veränderungen seiner phosphorylierten Produkte (Janke et al. 1996). PTMs können gewebsspezifisch auftreten, wie das Protein Serotransferrin, das im menschlichen Serum 4 Spots, in der Cerebrospinalflüssigkeit des Menschen jedoch 8 Spots aufweist (Wilkins et al. 1996b). Hier schliesst sich die Frage an, ob PTMs eventuell für die Organspezifität mitverantwortlich sind. Die modifizierten Formen eines Proteins zeigen sich in der 2DE-Analyse als Spots mit Verschiebung in Richtung des Isoelektrischen Punktes oder des Molekulargewichtes. Modifikationsstellen eines Proteins können durch Mutationen verändert werden. Eine Mutation, die zur Aminosäuresubstitution führt, kann daher die Grundlage für eine Fehlfunktion bilden. Es kann aber auch das Enzym, das die Modifikation ausführt – der Modifikator (Modifier) – mutiert und in seiner Funktion verändert sein. Wo auch immer die Ursache für eine veränderte, funktionsgebende Modifikation liegt: die Analyse der normalen Heterogenität der PTMs und der krankheitsassoziierten PTMs stellt die Schnittstelle dar zwischen Genombedingtheit und Umweltbedingtheit der Proteinfunktion.

Die Strategien, die für die globale Analyse von PTMs verfolgt werden, lassen sich – wie die Proteomanalyse auch – in die beiden Kategorien des rein deskriptiven und des funktionellen Ansatzes einteilen.

Das deskriptive Konzept basiert auf dem Vergleich von 2DE Mustern mit einer 2DE-Referenzkarte (Wilkins et al. 1996b). Das 2D-Muster eines Kontrollgewebes bildet das Referenzmuster für die Analyse eines veränderten Gewebes. Wenn Proteine durch 2D-Elektrophorese aufgetrennt werden, stimmen der Isoelektrische Punkt und das Molekulargewicht von Proteinen, die keine posttranslationalen Modifikationen tragen, normalerweise gut mit den aus der Aminosäuresequenz vorhergesagten pI- und Mr- Positionen überein (Bjellqvist et al. 1993a; Bjellqvist et al. 1994). Das trifft jedoch nicht zu für Abbauprodukte oder für ein modifiziertes Protein. Wilkins verfolgte deshalb den Ansatz, sogenannte Vektorkarten posttranslationaler Modifikationen zu erstellen: er verglich die vorhergesagte und die tatsächliche Position identifizierter Spots von 2D- Gelen, um festzustellen, welche Spots Modifikationen tragen und zeichnete zwischen theoretischer und echter Position einen Verschiebungs-Vektor ein. So entstanden Vektorbibliotheken der 2DE-Spots von *Eschericia coli*. Diese enthalten nicht nur einfache Spotverschiebungen, sondern ebenfalls Vektorenfelder (arrays). Ausgehend von einem einzelnen Spot weisen Vektoren auf Spotserien (sogenannte ‚trains of spots‘), die Modifikationsprodukte dieses Proteins darstellen. Die Anwendung von Vektorkarten liegt in der Vorhersage von PTMs. Für Phosphorylierung und Sulfatierungen erwartet man eine Verschiebung des Spots hin zu niedrigerem Isoelektrischen Punkt, jedoch ohne Änderung des Molekulargewichtes, während für zusätzliche neutrale Glykane wie Galaktose, Glukose N-Acetylgalaktosamin und N-

Acetylglukosamin das Molekulargewicht steigen und der Isoelektrische Punkt konstant bleiben. Im Rahmen der Reproduzierbarkeit des IPG-Gelsystems ermöglichen es diese Vektorbibliotheken anderen 2DE-Laboratorien, das Vorhandensein oder Fehlen einer PTM sowie Art und Ausmass der PTM eines verschobenen Spots abzuschätzen.

Funktionelle Konzepte für die Untersuchung von PTMs innerhalb der Proteomanalyse, auch Schrotschuss-Verfahren (shotgun approach) genannt, sind zum Beispiel die enzymatische oder chemische Behandlung einer Proteinprobe vor der Trennung mit 2D-Elektrophorese. Mit dem Enzym PNGaseF können N-gekoppelte Glycane, mit milder saurer Behandlung kann terminale Sialinsäure von N- oder O-gebundenen Glycanen entfernt werden (Packer et al. 1997; Packer et al. 1998). Die Einführung eines Knockouts eines modifizierenden Enzyms in einen Organismus und Auftrennung der Proteine auf 2DE kann zeigen, welche Spots Substrate für das modifizierende Enzym darstellen. Bakterien-, Hefe- oder Säugerzellkulturen können mit ^3H -Mannose oder ^{32}P markiert werden, um Glykosylierungen und Phosphorylierungen nachzuweisen (Garrels und Franza 1989). Immunochemische Färbungen wie Lectinbindestudien von Proteinen, die aus dem Gel auf PVDF Membranen transferiert wurden, weisen glycanhaltige Proteinspots nach (Gravel et al. 1994). Leffers exprimierte eine cDNA unter Verstoffwechslung radioaktiven Phosphats in Säugerzellen und zeigte, dass das Proteinprodukt mehrere Phosphorylierungen trägt, das 2D- Gel zeigt eine Serie von mehreren radioaktiv markierten Spots (Leffers et al. 1996). Als Beispiel für die Vielfalt der PTMs eines Gens kann das mit der Alzheimer'schen Krankheit assoziierte Tau Protein gelten: 6 von den über 80 Tau-Spots eines 2D-Gels lassen sich auf alternatives Splicing der Tau-RNA zurückführen. Ein Teil der weiteren Isoformen verschwindet nach Behandlung mit Alkalischer Phosphatase vom 2D-Muster. Die verbleibenden Spots müssen auf andere Modifikationen zurückgeführt werden (Janke et al. 1996).

1.6 Die genetische Analyse als Bindeglied zwischen Genom und Proteom

1.6.1 Proteinpolymorphismen

Polymorphismen und Neumutationen der DNA wirken sich unmittelbar auf die Proteine aus, es sei denn, das mutierte Triplet kodiert - aufgrund des degenerierten genetischen Kodes - für dieselbe Aminosäure (stille Mutation). Nukleotidaustausche in der DNA können zu Missensemutationen und damit einer Aminosäuresubstitution im Protein führen. Ein Nukleotidaustausch kann zur Einführung eines Stopcodons und damit zu Kettenabbruch und verkürztem Protein führen. Ebenso können bestehende Stopcodons mutieren und die Proteinkette wird verlängert. Deletion oder Insertion von Nukleotiden der DNA verursachen Leserahmenmutationen und führen damit zu einer veränderten Aminosäuresequenz des Proteins. Proteinpolymorphismen, d.h. Mutationen, die sich in einer Population etabliert haben, beruhen jedoch zum grössten Teil (90%) auf Aminosäuresubstitutionen. Als Folge einer Aminosäuresubstitution können sekundäre Effekte wie der Verlust einer

Modifizierungsstelle für Phosphorylierung oder Glykosylierung oder veränderte Stabilität und Degradation des Proteins auftreten. All diese Mutationen wirken sich auf das Laufverhalten des Proteins in der 2D-Elektrophorese aus, denn sie können Ladungsänderungen, Molekulargewichtsänderungen, Konformationsänderungen oder eine Kombination dieser drei hervorrufen. Da die 2D-Elektrophorese die Proteine zunächst nach Ladung und anschließend nach Molekulargewicht (Masse) trennt, können die meisten Mutationen als Positionsverschiebungen der Proteinspots im 2D-Muster erkannt werden. Die Mutationen bringen im 2D-Muster verschiedene „Proteinphänotypen“ hervor, die man im Prinzip wie folgt einteilen kann: (1) Qualitative Varianten: a) Positionsverschiebungen alleler Proteinspots in horizontaler, vertikaler oder diagonaler Richtung, b) Aufspaltung eines Proteins in verschiedene Isoformen (posttranslationale Modifikationen), Entstehung von Spotserien und anderen Spotkomplexen (2) Quantitative Varianten: a) Variation in der relativen Konzentration eines Proteins, b) Verschwinden eines Proteinspots oder Auftreten eines zusätzlichen Proteinspots. Quantitative Variationen können sowohl auf Mutationen der Primärstruktur des Proteins als auch auf Mutationen der DNA im Promotorbereich des zugehörigen Gens beruhen. Diese wirken sich auf die Syntheserate des Proteins aus (Klose 1982, 1999b). Die Möglichkeit, Genmutationen in 2D-Mustern nachzuweisen, wurde in zahlreichen Untersuchungen benutzt, um Gendefekte zu analysieren. Es wurden menschliche Tumore und Organe untersucht, um die dem Defekt zugrunde liegenden mutierten Proteine zu finden (Jungblut et al. 1999; Celis et al. 2000). Es wurden die Rückwirkungen von Umwelteinflüssen, Mutagenen und Teratogenen auf Proteine untersucht und Rückschlüsse auf Mutationsereignisse der DNA gezogen. Das Auftreten von polymorphen Proteinspots in 2D-Mustern des Menschen macht klinische Untersuchungen jedoch schwierig. Die Grundlage für klinische Untersuchungen menschlicher Gewebe mit 2D-Elektrophorese bildet daher populationsgenetische Untersuchungen zur Feststellung der generellen genetischen Heterogenität eines Gewebes und die Erstellung eines synthetischen Referenzgels aller Variationen des gesunden Gewebes, um anschließend krankheitsassoziierte Daten aus dem Vergleich mit dem Referenzgel zu filtern (Celis et al. 1996). Bisher wurden jedoch nur wenige Versuche gemacht, die genetischen Varianten eines 2D-Musters in systematischer und umfangreicher Form zu erfassen und zu kartieren. Elliott führte eine genetische Kartierung von 8 polymorphen Proteinspots aus Leberextrakten der Maus durch und lokalisierte deren Genorte auf den Mausechromosomen (Elliott 1979). Silver hat die 2D-Elektrophorese angewendet, um in kongenen Mauslinien nach Proteinpolymorphismen in der ‚t-complex‘ Region von Chromosom 17 zu suchen (Silver et al. 1983). Polymorphe Proteinspots von *Drosophila* wurden kartiert (Spicer 1988; Zeng und Singh 1993). Einige 100 genetische Varianten von Mais (Damerval et al. 1994; Touzet et al. 1995), Pinie (Bahrman und Damerval 1989) und Erbse (de Vienne et al. 1988) wurden auf der genetischen Karte dieser Organismen eingeordnet und als genetische Marker zur Kartierung von komplexen Merkmalen und quantitativen Proteinphänotypen eingesetzt (de Vienne et al. 1996).

1.6.2 Das Netzwerk zwischen Gen- und Proteinebene

Proteinphänotypen, wie sie in 2D-Mustern zutage treten, können auf allelischer Variation des Strukturgens des betreffenden Proteins beruhen. Die Ursache der Variationen eines Proteinspots – und das ist dem Polymorphismus nicht anzusehen – kann aber auch auf Mutationen in Genen beruhen, die in Interaktion mit diesem Protein stehen und für seine Expression oder Modifikationen verantwortlich sind. (1) Regulatorgene für z.B. Transkriptions- und Translationsfaktoren haben Einfluss auf die gebildete Menge eines Proteins. (2) Gene, die Protein-modifizierende Enzyme kodieren wie z.B. Kinasen, Phosphorylasen und Proteasen – sie sollen hier „Protein-Modifizierer“ genannt werden, haben Einfluss auf den Modifizierungszustand eines Proteins. Demnach ist selbst ein einzelnes Protein das Ergebnis polygener Wirkung. Die Beziehung zwischen dem Gen und seinem Produkt, dem Protein, kann daher nicht linear gedacht werden, denn das funktionelle Protein ist als das Produkt mehrerer Gene (Polygenie) anzusehen. Ebenso übt ein Gen, z.B. eine Glykosylase, eine modifizierende Wirkung auf viele Proteine aus und versetzt sie in einen funktionellen Zustand (Pleiotropie).

In naher Zukunft werden das Genom des Menschen und aller wichtigen Modellorganismen – wie z.B. der Maus – komplett sequenziert sein. Es wird dann über die Proteomanalyse (2D-Elektrophorese und Massenspektrometrie) möglich sein, einen Proteinspot aus einem 2D-Gel sofort seinem Strukturgen zuzuordnen. Sequenzhomologien zwischen DNA und Proteinen sowie die physikalische Kartierung können jedoch nicht als Werkzeuge dienen, das Netzwerk der Interaktionen zwischen Genen und Proteinen aufzudecken. Die Sequenzanalyse des Genoms kann nur die lineare Beziehung zwischen Gen und Protein etablieren. Zur Aufklärung von Interaktionsnetzwerken kann jedoch die genetische Analyse von Proteinpolymorphismen herangezogen werden. Man kann postulieren, dass die Proteinphänotypen, wie sie in vielfältiger Form in 2D-Mustern auftreten, das Ergebnis polygener und pleiotroper Effekte sind. Diese Annahme kann an einem genetischen Modellsystem geprüft werden, das Proteinpolymorphismen in grosser Zahl hervorbringt. Ein solches Modellsystem könnte aus zwei Mauseinzuchtstämmen bestehen, die eine grosse genetische Distanz aufweisen, aber trotzdem noch verpaart werden können. Die Analyse der Vererbung eines Proteinpolymorphismus über zwei Generationen ermöglicht die genetische Kartierung des Polymorphismus.

1.7 Genetische und physikalische Kartierung von Proteinen der Maus Ansatzpunkte zur eigenen Untersuchung

1.7.1 Ein Europäisches Kollaborationsprojekt zur genetischen Kartierung der Maus (EUCIB)

EUCIB - die Abkürzung für 'European Collaborative Interspecific Backcross' - ist der Name des 1992 von Steve Brown am Sanger Center in Hinxton, UK und Phil Avner am Institut Pasteur, Paris, Frankreich, ins Leben gerufenen Projektes zur Feinkartierung des Genoms der Maus. In Paris und London wurde eine Mausekreuzung angesetzt, die mit 1000 Rückkreuzungstieren 1000 Meiosen für die Erstellung einer hochauflösenden genetischen Karte des Mausgenoms bietet [Breen, 1994 #399](Rhodes et al. 1998). Die angestrebte Auflösung der genetischen Karte sollte 0.1 cM betragen und der wissenschaftlichen Öffentlichkeit eine Ressource zur Positionalen Klonierung und Feinkartierung von Genen oder Loci zur Verfügung zu stellen.

Als Ausgangsstämme wurden die beiden Spezies *Mus musculus domesticus* C57BL/6 und *Mus spretus* gewählt. Da sie stammesgeschichtlich 3 Millionen Jahre (s. Abb.1.7.1) auseinander liegen, weisen sie im Vergleich zu Subspezieskreuzungen eine höhere DNA-Polymorphismusrate auf, die für die Erstellung einer genetischen Karte essentiell ist. Diese beiden Arten sind kreuzbar. Wenn ein weibliches *Mus musculus* Tier (*Mus musculus domesticus* C57BL/6J, abgekürzt B6) mit einem männlichen *Mus spretus* Tier (Msp) verpaart wird, entstehen fruchtbare weibliche Nachkommen und unfruchtbare männliche (F1-Generation) (Avner et al. 1988). Die weiblichen Hybriden werden mit männlichen B6 oder Msp Tieren gekreuzt, die Nachkommen werden als Rückkreuzungsgeneration bezeichnet (B1-Generation).

Für die hier beschriebene Kartierung der Proteinpolymorphismen auf der genetischen Karte der Maus wurden 64 Rückkreuzungstiere des Typs (B6 x Msp) x Msp eingesetzt. Vorarbeiten wurden im Labor Klose durchgeführt und betrafen die Präparation der Organe Gehirn, Leber, Herz, Niere und Muskel der 64 Rückkreuzungstiere. Weiterhin wurden drei Proteinextrakte des Gehirns erstellt: eine pufferlösliche Fraktion, die hauptsächlich cytoplasmatische Proteine enthält, eine Membranfraktion und eine Rückstandssuspension mit stark basischen chromosomalen und ribosomalen Proteinen. Die cytosolischen Fraktionen der Gehirne wurden mittels 2D-Elektrophorese aufgetrennt. Die Gele wurden mit einer sensitiven Silberfärbung gefärbt und getrocknet (Jungblut und Seifert 1990). Die Analyse der Polymorphismen, die zwischen den Elterntieren auftreten, erbrachte eine Polymorphismusrate von 15.1% (1324 polymorphe Spots von 8767 insgesamt vorhandenen Spots der cytosolischen Fraktion). Die Hälfte (7.62%) der Polymorphismen zeigten eine qualitative Variation (mobility variation), die Hälfte (7.49%) eine quantitative Variation (presence/absence variation, amount variation).

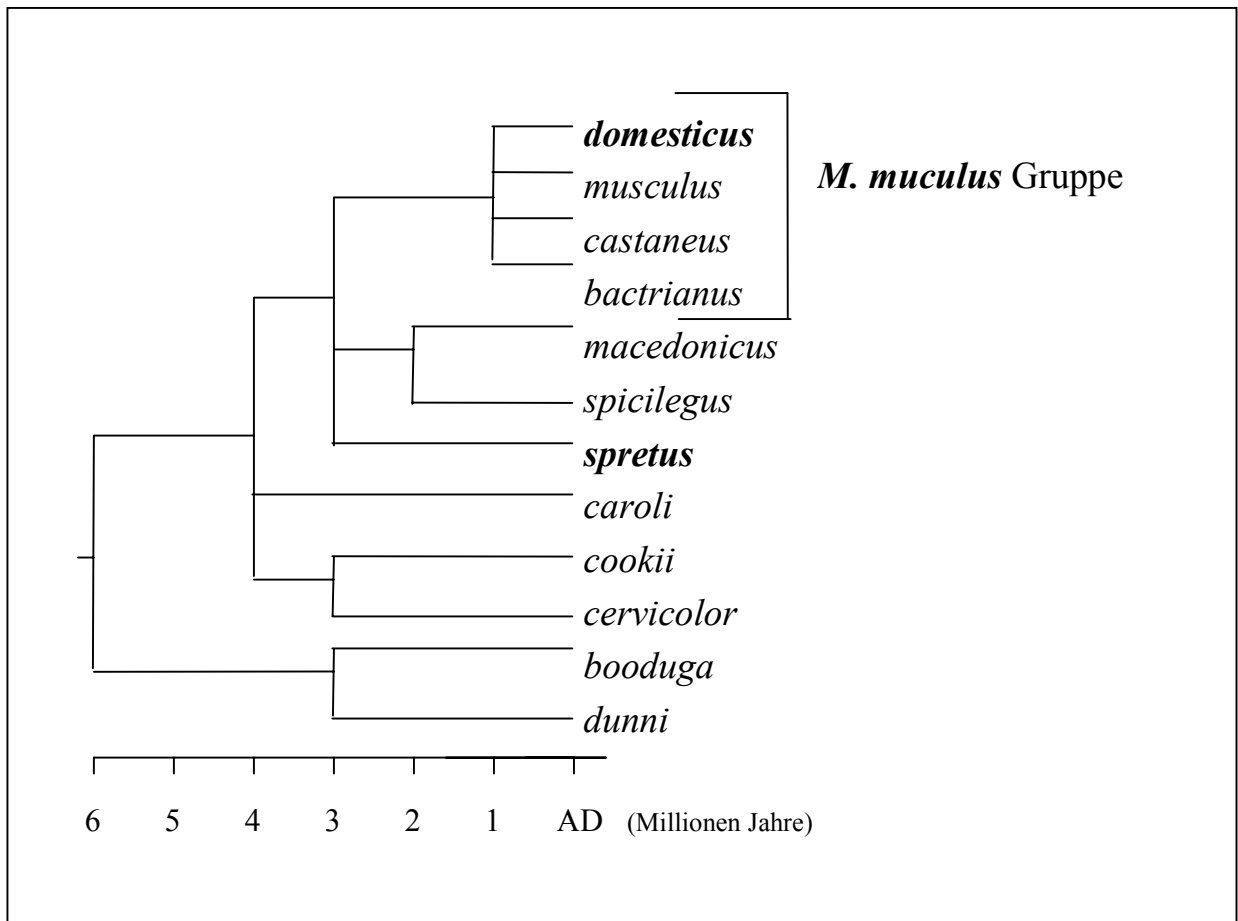


Abb. 1.7.1: Ein Stammbaum der Gattung *Mus* nach Silver (Silver 1995). *Mus musculus domesticus* und *Mus spretus* trennten sich in ihrer Entwicklung zu separaten Arten vor ca. 3 Millionen Jahren.

1.7.2 Die physikalische Karte des Mausgenoms

Physikalische Karten dienen dazu, die positionelle Klonierung von Genen zu erleichtern, die einem Phänotyp zugrunde liegen. Eine physikalische Karte besteht aus einer Reihe geordneter YAC-Klone, die die einzelnen Chromosomen repräsentieren. Eine physikalische Karte der Maus wurde Max-Planck-Institut für Molekulare Genetik, Berlin erstellt. Sie vereinigt die am Max-Planck-Institut erhobenen Daten mit den Daten, die am Whitehead Institute for Biomedical Research /MIT Center for Genome Research gewonnen wurden (Nusbaum et al. 1999). Die Karte basiert auf IRS-PCR Markern und stellt eine Karte der 2. Generation dar, die im Vergleich zur Whitehead-Karte die doppelte Anzahl an YAC-Klonen (50891) sowie genetischen Markern (21904) enthält (<http://www.molgen.mpg.de/~rodent> und Schalkwyk et al. eingereicht bei Genome Research). Diese Karte bildete die Grundlage für die exemplarische physikalische Kartierung von nichtpolymorphen Proteinspots des Mausgehirns.

1.7.3 Ansatz und Ziele der eigenen Untersuchungen

Die Untersuchung der Gehirnproteine von B6 und Msp durch 2D-Elektrophorese hatte nicht nur zur Entdeckung einer grossen Zahl von Proteinpolymorphismen geführt, sondern auch zum Nachweis eines breiten Spektrums von Proteinphänotypen (Klose 1999b). Diese Phänotypen konnten grob in drei Kategorien eingeteilt werden: Phänotypen, die auf eine strukturelle Veränderung eines Proteins hinwiesen und Phänotypen, die auf eine posttranslationale Modifizierung oder auf eine quantitative Veränderung hinwiesen.

Zielsetzung dieser Arbeit war es, Proteinpolymorphismen auf der genetischen Karte der Maus zu lokalisieren, um damit neue Gene zu kartieren. Das National Institute of Health, der Wellcome Trust und mehrere private Firmen haben am 6. Oktober 2000 die Gründung eines Maus Sequenzierungs-Konsortiums zur Totalsequenzierung des Mausgenoms angekündigt. Wenn dieses Projekt beendet ist, wird es noch geraume Zeit dauern, bis alle Gensequenzen im Genom entdeckt und gegen die nicht-kodierenden Sequenzen abgegrenzt sind. Es ist heute möglich, über Massenspektrometrie die Primärstruktur eines Proteins – z.B. von Proteinen aus 2D-Gelen – so weit zu charakterisieren, dass man das zugehörige Gen über translatierte DNA-Sequenz-Datenbanken identifizieren kann. In weiterer Zukunft wird es daher möglich sein, durch physikalische Kartierung alle Proteine im Genom der Maus zu lokalisieren. Bis dahin behält die genetische Kartierung jedoch ihre Bedeutung. Darüber hinaus kann sie auch zur Bestätigung der physikalischen Kartierung dienen. Der wichtigste Aspekt der vorliegenden Arbeit war es jedoch zu prüfen, ob über die Kartierung von elektrophoretischen Proteinphänotypen nicht nur Strukturgenorte, sondern auch Protein-modifizierende und regulierende Gene lokalisiert werden können. Die Effekte dieser Gene werden auch nach der Totalsequenzierung eines Genoms nicht über die Sequenzhomologie von Gen und Protein nachweisbar sein. Die genetische Kartierung bleibt damit für die Fragestellung ‚regulatorische Netzwerke‘ ein wichtiger Ansatz. Erst wenn alle Stoffwechselwege – und damit alle Proteininteraktionen innerhalb der Zelle – bekannt sind, wird auch die genetische Kartierung von Proteinpolymorphismen keine neuen Informationen erbringen.

Die Strategie zum Nachweis Protein-modifizierender und -regulierender Proteine, mit der in der vorliegenden Arbeit vorgegangen wurde, kann wie folgt zusammengefasst werden: 1. Genetische Kartierung einer möglichst grossen Zahl von polymorphen Proteinspots aus 2D-Gelen. 2. Identifizierung der kartierten Proteinspots durch Massenspektrometrie. 3. Proteine, die sich auf diesem Wege als bereits bekannt erweisen, werden in der genetischen Karte der Maus gesucht mit der Frage, ob ihr Strukturgen bereits kartiert ist. 4. Der Nachweis eines Modifier- oder Regulatorgens ist dann gegeben, wenn folgender Fall eintritt (Idealfall): Drei unabhängige Proteinspots werden als das gleiche Protein identifiziert. Spot 1 zeigt eine Mobilitätsvariation und kartiert an die Stelle des bekannten Strukturgens. Spot 2 zeigt eine andere Mobilitätsvariation und kartiert an eine andere Stelle. Spot 3 zeigt eine quantitative Variation und die kartiert an eine dritte Stelle im Mausgenom. Locus 2 würde als Hinweis auf ein Modifizergen gewertet werden, Locus 3 als Hinweis auf ein Regulatorgen.

Die genetische Kartierung der Proteinspots von 2D-Gelen ist nur dann möglich, wenn der Proteinspot einen Polymorphismus in zwei unterschiedlichen Mäusestämmen zeigt. Es wurde daher nach Wegen gesucht, auch nicht-polymorphe Spots auf der genetischen und physikalischen Karte der Maus zu lokalisieren. Es wurde folgende Vorgehensweise konzipiert: Im ersten Schritt wird der interessierende Proteinspot mittels Massenspektrometrie identifiziert und damit an Protein- und DNA-Sequenzdaten angeschlossen. Im zweiten Schritt können diese Sequenzdaten für den Homologievergleich mit translatierten cDNA-Datenbanken (GenBank, dbEST) verwendet werden. cDNAs und ESTs können über Verteilerstellen bezogen werden (I.M.A.G.E. Konsortium, (Lennon et al. 1996)). Eine identifizierte cDNA- oder EST-Sequenz kann dann mit verschiedenen Methoden kartiert werden: (a) durch Hybridisierung der cDNA/des ESTs gegen Strahlungshybriden (sogenannte ‚Radiation-Hybrid‘-Zelllinien), (b) durch IRS-PCR vermittelte Hybridisierung gegen YAC-Klone, die bereits zu einer physikalischen Karte zusammengesetzt worden sind (Schalkwyk et al., eingereicht bei Genome Research).

Die Kartierung von Proteinspots ist interessant und lohnend, weil 2D-Elektrophorese als Screeningtechnik dazu eingesetzt wird, Merkmale und Fehlfunktionen von Geweben bei Krankheiten und Alterung eines Organismus (Maus, Mensch) an Proteinen und damit an bestimmten Genen festzumachen. Kartiert ein veränderter Proteinspot gemeinsam mit dem Merkmal, so kann er als Kandidaten‘gen‘ für dieses Merkmal herangezogen werden. Aber auch jeder andere, veränderte Proteinspot kann, wenn er kartiert wird, einen chromosomalen Locus zu Tage fördern, der bei der Ausbildung des Merkmals beteiligt ist.

Die Kartierung von Modifier- und Regulatorgenen würde einen Weg bieten, das komplexe Netzwerk zwischen Genotyp und Phänotyp aufzuhellen und die Entstehung multifaktorieller Krankheiten besser zu verstehen. Es wird heute deutlich, dass selbst klassische monogene Erbkrankheiten im Grunde polygene Krankheiten sind (Scriver und Waters 1999).