

# MOLEKULARE NETZWERKE

## HERAUSFORDERUNG FÜR BIOLOGIE UND PHARMAFORSCHUNG

Walter Schubert

*Nach der Entschlüsselung des menschlichen Genoms ist die Entschlüsselung der molekularen Netzwerke der Zelle, die den Funktionsplan – den biologischen Code – aller Zellfunktionen beinhalten, die nächste große Herausforderung in der humanen Biotechnologie. Die genaue Kenntnis dieses Funktionsplans, den wir als Toponom bezeichnen, wird es erlauben, die zelluläre Funktion der Proteine als Elemente räumlich determinierter molekularer Netzwerke zu verstehen und gezielt molekulare Therapieformen zu entwickeln. Grundlegende Verfahren der direkten Analyse auf der Einzelzellebene wurden in Magdeburg entwickelt. Erste Erfolge der Vorhersage krankheitsrelevanter Proteine aus partiellen Toponom-Analysen sind Grundlage für die Entwicklung neuer Wirkstoffe.*

Die Kosten für die Entwicklung neuer Arzneimittel steigen international weiter an. Wie das „Start-centre for the Study of Drug Development“ im Jahr 2001 berichtete, betragen die mittleren Entwicklungskosten pro Arzneimittel im Jahr 2000 802 Millionen US-Dollar. Dennoch ist die klinische Erfolgsrate unbefriedigend. Eine Verbesserung der Erfolgsrate von derzeit 1:5 auf 1:3 pro zugelassenen Arzneimittel würde die Entwicklungskosten wahrscheinlich um 221 bis 242 Millionen US-Dollar senken /1/.

Die Ursache für diese Entwicklung sehen Lehmann Brothers & McKinsey /2/ in der schlechten Target-Selektion. In diesem Zusammenhang sind Targets molekulare Zielstrukturen (in der Regel Proteine), die eine krankmachende Wirkung haben. Sie sind Zielmoleküle, gegen die sich ein zu entwickelnder Wirkstoff richten muss, um deren krankmachende Funktion zu unterbinden. Dabei handelt es sich um ein Problem an der Basis jeglicher Wirkstoffentwicklung, denn ein guter Wirkstoff kann nur dann krankheitsspezifisch wirken, wenn er das für die jeweilige Krankheit kritische Target-Molekül adressiert. Es kommt also ganz wesentlich darauf an, das für eine Krankheit „richtige“ Target-Molekül zu finden. Wie die genannten Autoren nachweisen, sind die bisher üblichen Verfahren, um solche Proteine aus der Vielzahl möglicher Kandidaten zu identifizieren, offenbar nicht hinreichend für eine valide Wirkstoffentwicklung. Der Zusammenhang wird deutlich, wenn man die Umsätze pro Wirkstoff in den Jahren von 1995 bis 2001 verfolgt: Bis ca. 1995 wurden noch die „erfolgreichen“ klassischen pharmakologischen Verfahren eingesetzt, um Symptome zu beeinflussen, wie hohen Blutdruck, Spasmen der Atemwege (z. B. Asthma) etc. Der mittlere Umsatz pro Wirkstoff (new chemical entity, NCE) betrug pro Unternehmen im Jahr 1995 noch 263 Millionen US-Dollar. Als dann zunehmend in den darauffolgenden Jahren neue, auf genomischen Techniken beruhende Verfahren eingesetzt wurden, um

komplexe chronische Krankheiten, wie Krebs, Alzheimer u. a., ins Visier zu nehmen, sanken die Umsätze bis zum Jahr 2001 sehr deutlich auf 34 Millionen US-Dollar pro NCE ab, obwohl in demselben Zeitraum durch Einsatz der jetzt möglich gewordenen breit angelegten Targetsuche (Large-Scale-Expression Profiling Verfahren)<sup>1)</sup> die Zahl so genannter Target-Kandidaten um den Faktor 4 bis 5 gegenüber 1995 gestiegen war. Inwieweit ein Molekül (z. B. Protein) als Target-Kandidat infrage kommt, wird bei diesen Verfahren davon abhängig gemacht, ob es in größerer Menge bei einer bestimmten Krankheit vorkommt. Das Fazit, das u. a. hieraus gezogen werden kann, lautet: Die methodische Fähigkeit, eine immer größere Zahl von Proteinen gleichzeitig aus einer Gewebeprobe extrahieren und bestimmen zu können, führt nicht zwingend zu einem biologischen Informationsgewinn. Eine Folge ist, dass der Produktdruck in der Pharmaindustrie erheblich gestiegen ist: Oft entscheidet sich erst in der klinischen Studie, ob das ‚richtige‘ Target gefunden wurde oder nicht. Dann allerdings sind schon gewaltige Kosten, oft mehrere hundert Millionen US-Dollar angefallen, und ein Abbruch einer klinischen Studie wegen nachgewiesener Unwirksamkeit oder Toxizität eines entsprechenden Wirkstoffs beeinflusst nachhaltig die Forschungs- und Kostenentwicklung. Diese Zusammenhänge kennzeichnen einerseits ein signifikantes ökonomisches und gesundheitspolitisches Problem, andererseits eine große Herausforderung an die Wissenschaft, fundamentale biologische Prozesse besser analysieren und verstehen zu lernen.

Ein wesentliches Ziel besteht also in Zukunft darin, Target-Kandidaten aufgrund „besserer“ biologischer Informationen zu ermitteln, d. h. die zelluläre Funktion von Target-Kandidaten so genau wie möglich zu entschlüsseln, und zwar direkt auf subzellulärem Niveau in situ, d. h. an den Zielstellen der Krankheit im erkrankten Gewebe oder auch an isolierten krankheitsrele-

1) Diese Verfahren beruhen auf einer Zerstörung von Zellen und Geweben, um Moleküle (Proteine, Nucleinsäuren) zu isolieren und zu bestimmen.

vanten Zellen. Dazu müssen Verfahren entwickelt werden, die es erlauben, den Kontext der Proteine auf dem Einzelzellniveau messtechnisch zu erfassen und funktionell zu analysieren. Ein Quantensprung wird erwartet, wenn es gelänge, die Gesamtheit aller Proteine (das Proteom) einer Zelle in einem einzigen Experiment zu bestimmen /3/.

**DIE ZELLE – EIN NETZWERK INTERAGIERENDER MOLEKULARER MASCHINEN**

Um dieses Problem zu erläutern nehmen wir die geschriebene Sprache als Metapher: Nicht die Buchstaben, sondern erst der konkrete „topologische“ Kontext von Buchstaben und Leerzeichen, also ihre Zusammensetzung zu Wörtern, Sätzen, Texten usw. ergibt Sinn bzw. Bedeutung. Abbildung 1 illustriert das Beispiel: Die Bestimmung der Zahl der Buchstaben in zwei englischsprachi-

darum geht, ihre Rolle bei der Verschlüsselung von Zellfunktionen zu verstehen. Tatsächlich deuten zahlreiche Analysen in den letzten Jahren darauf hin, dass derartige biochemische Extraktions- und Bestimmungsverfahren, die man als „Large-scale-protein-profiling“-Verfahren bezeichnet, für sich betrachtet einen relativ geringen biologischen Informationswert besitzen. Die schlechte Target-Selektion, die Lehman Brothers & Mc Kinsey /2/ als den eigentlichen Treiber der Probleme in der Arzneimittelentwicklung identifiziert haben, scheint also darin zu bestehen, dass Target-Proteine noch nicht systematisch aus dem Kontext der Proteine in Zellen ausgelesen und als valide Kandidaten für die Wirkstoffentwicklung bestimmt werden können.

Andererseits können solche Target-Proteine auch nicht einfach aus der molekularen Funktion, die für viele Proteine schon bekannt ist, postuliert werden: Ein Protein mit der bekannten molekularen Funktion „x“ hat im Zusammenhang bzw. Wechselspiel mit bestimmten anderen Proteinen innerhalb der Zelle eine bestimmte zelluläre Funktion. Im Zusammenspiel desselben Proteins ‚x‘ mit wiederum anderen Proteinen kann diese zelluläre Funktion aber eine ganz andere sein. Man muss also die molekulare(n) Funktion(en) der Proteine von deren zellulären Funktion(en), die sich nur aus dem Kontext, also dem Zusammenspiel mit anderen Proteinen, ergeben, trennen.

Nachdem der genetische Code für die Synthese der Proteine bei zahlreichen Spezies heute praktisch vollständig bekannt ist, wird die Herausforderung für die Zukunft darin bestehen, die zellulären Funktionen der Proteine zu entschlüsseln. Dieses Ziel zwingt zur Entwicklung von Verfahren, die es ermöglichen, Proteine im Kontext ihrer Netzwerke in morphologisch intakten Zellen und Geweben auszulesen.

Ob die erwählte Sprach-Metapher mit allen semantischen Konsequenzen und Regeln auf das zelluläre System der Proteine abgebildet werden kann, ist natürlich fraglich. So ist es heute noch unklar, ob es kleine Bedeutung tragende Protein-Einheiten, also kleine Protein-Ensembles (Protein-,Morpheme) gibt, die von quasi allen Zellen als modulare Einheiten räumlich und zeitlich verschiedenartig zusammengesetzt werden können, um eine ‚Syntax‘ in größeren Zusammenhängen zu erzeugen. Es dürfte allerdings auf der Grundlage des bisherigen biologischen Wissens und gut begründeter Annahmen als sicher gelten, dass Proteine in Zellen zu mehr oder weniger fest assoziierten Proteinkomplexen, also Einheiten höherer Ordnung zusammentreten.

Mit bestimmten biochemischen Verfahren können solche Protein-Komplexe aus Homogenaten, also morphologisch zerstörten einfachen Zellen isoliert werden /4/. Die gewonnenen Daten sind allerdings z. Zt. nur schlecht reproduzierbar, und man kann Artefakte, die im Isolierungsprozess liegen, von realen Komplexen nur schwer unterscheiden. Außerdem sind zahlreiche, eher lockere Protein-Ensembles, die für viele Zell-

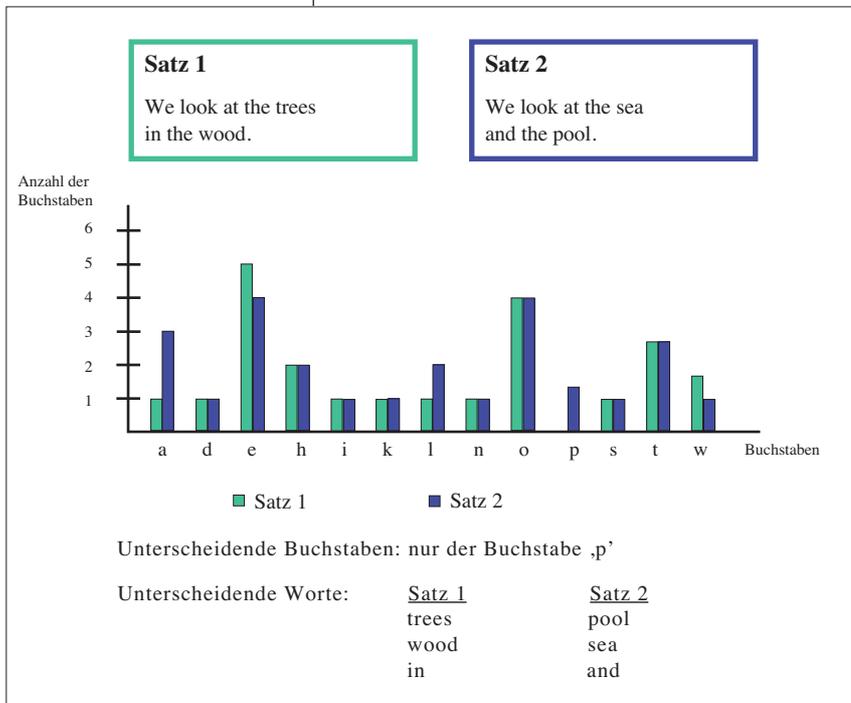


Abbildung 1 Metapher der geschriebenen Sprache: Zusammenhang zwischen Buchstaben und Syntax. Nur die Ebene der Syntax erlaubt es, die beiden Sätze mit ganz verschiedenen Bedeutungen zu unterscheiden.

gen Sätzen mit jeweils ganz verschiedener Bedeutung ergibt eine annähernd gleiche Anzahl der verwendeten Buchstaben. Man kann also allein anhand der Buchstabenprofile die beiden Sätze nicht unterscheiden. Lediglich der Buchstabe ‚p‘ ist nur in einem der beiden Sätze vorhanden. Allerdings unterscheiden sich die Sätze bereits sehr deutlich voneinander, wenn man die Wortebene betrachtet. Drei Wörter kennzeichnen spezifisch Satz 1, und drei Wörter kennzeichnen Satz 2. Mithin erweist sich erst die Ebene der Syntax als unterscheidend.

Wie aber wird in biologischen Systemen ‚Bedeutung‘ kodiert? Wenn die Sprach-Metapher zutrifft und Proteine, dieser Metapher folgend, als Buchstaben für die Zusammensetzung spezifischer syntaktischer ‚Zeichen‘, die Zellfunktionen kodieren, verstanden werden können, so wäre es nicht zielführend, Proteine durch Extraktion aus dem biologischen Kontext zu vereinzeln, wenn es

funktionen wahrscheinlich sehr wichtig sind, gar nicht isolierbar, da sie beim Isolierungsvorgang zerfallen /5/. Darüber hinaus sagen die aus Zell-Homogenaten isolierten Protein-Komplexe nichts über die Situation in individuellen Zellen aus. Diese allerdings ist von großer Bedeutung, nicht zuletzt deshalb, weil bei zahlreichen chronischen Krankheiten einzelne oder nur sehr wenige Zellen im Gewebe den Krankheitsprozess bestimmen und vorantreiben.

Unabhängig von diesen derzeit noch großen methodischen Schwierigkeiten, allein auf der Basis biochemischer Protein-Daten grundlegende Aussagen über die konkrete Realität der Zelle zu treffen, so dürften die Annahmen von Bruce Alberts zutreffen, der die Zelle als eine Art Fabrik integrierender molekularer Maschinen (Protein-Komplexe), also ein dynamisches und hoch organisiertes molekulares Netzwerk interpretiert /6/.

#### KONZEPTION DER KARTIERUNG MOLEKULARER NETZWERKE ALS MUSTER

In der theoretischen Konzeption geeigneter Verfahren für die Kartierung molekularer Netzwerke in Zellen sind die Konzepte zugrunde zu legen, welche die Zelle im Laufe ihrer evolutionsbiologischen Entwicklung selbst etabliert hat, um die Vielzahl der Zellfunktionen zu generieren: In der Zelle sind die Proteine nicht stochastisch verteilt, sondern im Gegenteil räumlich und zeitlich hoch organisiert. In jeder Zelle muss jedes Protein zum richtigen Zeitpunkt in der richtigen Konzentration am richtigen Ort vorkommen, damit es mit anderen Proteinen, für welche genau dieselben Regeln gelten, in Wechselwirkung treten kann. Erst das hieraus resultierende räumlich determinierte Funktionsnetzwerk der Proteine generiert spezifisch eine gegebene Funktion. Da mithin derartige Proteinnetzwerke in Zellen räumlich, d. h. als Muster determiniert sind, können sie mit geeigneten bildgebenden Verfahren direkt abgebildet werden, wenn es technisch gelingt, die einzelnen Proteine qualitativ und quantitativ und gleichzeitig im räumlichen Kontext zueinander zu erfassen. Wenn man bei solchen Verfahren eine geeignete Anzahl von „Datenpunkten“ pro Zelle zugrunde legt, welche die subzellulären Kompartimente (Funktionsräume der Zellen) mit hinreichender Auflösung erfassen, so können praktisch alle, primär nicht vorhersagbaren Proteinnetzwerke als „Muster“ im Zellvergleich quantifiziert werden. Der wesentliche biologische Informationsgewinn besteht in der gleichzeitigen Erfassung von subzellulären Proteinkonzentrationen und „Protein-Arrangements“. Hierin ist die Grundlage für ein systembiologisches Verständnis von Proteinen im physiologischen Kontext von Geweben zu sehen /7/. Erste Analysen in diese Richtung hatten auf der Grundlage eines neuen Markers der Muskelfaserregeneration /8/ und durch die simultane Kartierung von neun Zelloberflächenproteinen gezeigt, dass die Regeneration des Skelettmuskels offenbar teilweise auf einer Umprogrammierung von Gefäß-Endothelzellen in Muskelstammzellen

beruht /9/. Diese Beobachtung, die erstmals direkt im Gewebe möglich war, wäre mit konventionellen Verfahren der Immunzytochemie oder Verfahren der Protein-Extraktion nicht möglich gewesen. Die Bedeutung der Gefäß-Endothelzellen für die Muskelregeneration, die zunächst auf Widerstand im zellbiologischen Diskurs stieß, wurde später durch andere Verfahren bestätigt /10/.

Diese Zusammenhänge zeigten, dass komplexe zelluläre Mechanismen direkt im Gewebe aufgedeckt werden können, wenn, allgemein formuliert, eine hinreichende Zahl geeigneter zellulärer Proteine der Zelldifferenzierung simultan und quasi als Projektion auf Zellen und Strukturen in Gewebeschnitten visualisiert werden. Um hier wieder auf die Sprach-Metapher zurückzugreifen: Erst die Visualisierung der Buchstaben im korrekten Kontext zueinander auf den Seiten eines Textes zeigt die syntaktische, also Bedeutung tragende Struktur. Die methodische Konsequenz aus diesen biologischen Erkenntnissen war, ein vollautomatisiertes System für die Analyse von Proteinkontexten in morphologisch intakten biologischen Strukturen zu entwickeln, das einerseits für systembiologische Fragestellungen, andererseits auch für die Entdeckung relevanter Target-Proteine in der Arzneimittelforschung eingesetzt werden kann. Mehrjährige methodische Forschungen am Zentrum für Molekulare Biologie Heidelberg (ZMBH) und später am Institut für Medizinische Neurobiologie der Universität Magdeburg sowie Arbeiten im Magdeburger Innovationskolleg „Bildinformation“ und in der MelTec GmbH, die hieraus ausgegründet wurde, bildeten die Grundlage für die technische Realisierung.

#### ROBOTERVERFAHREN

Ein aus diesen Entwicklungsschritten resultierendes Roboterfahren (Whole Cell Protein Fingerprinting, WCPF) wurde bei der MelTec GmbH in Magdeburg etabliert /11/. Es handelt sich hierbei um Mikroskop-Roboter, deren technische Grundlage ein molekulares Scanning-Verfahren (mass molecular scanning) ist. Proteine werden in definierten subzellulären Volumina, also in den Funktionsräumen der Zelle in Form von Lichtsignalen (Photonen) optisch erfasst. Dazu bringen die Mikroskop-Roboter große Tag-Bibliotheken auf spezifisch fixierte Zellen oder Gewebeschnitte auf, um die Proteine und Protein-Ensembles in situ d. h. in ihrer natürlichen Strukturumgebung der Zelle zu markieren. Es entsteht für jedes Zellvolumen ein ‚Protein-Fingerprint‘, der in einem Datenraum ‚abgebildet‘ wird. Durch Vergleichsanalysen zwischen Mustern bei Krankheiten oder in Zell-assays werden krankheitsspezifische, zellspezifische oder mit bestimmten Zellfunktionen assoziierte Muster gefunden und dadurch funktionell kartiert. Einer der wesentlichen Fortschritte dieser Technologie besteht darin, dass die räumliche Kompartimentierung zellulärer Proteome<sup>2)</sup> erhalten bleibt und daher Zelle für Zelle im

2) Proteom: die Gesamtheit aller Proteine einer Zelle

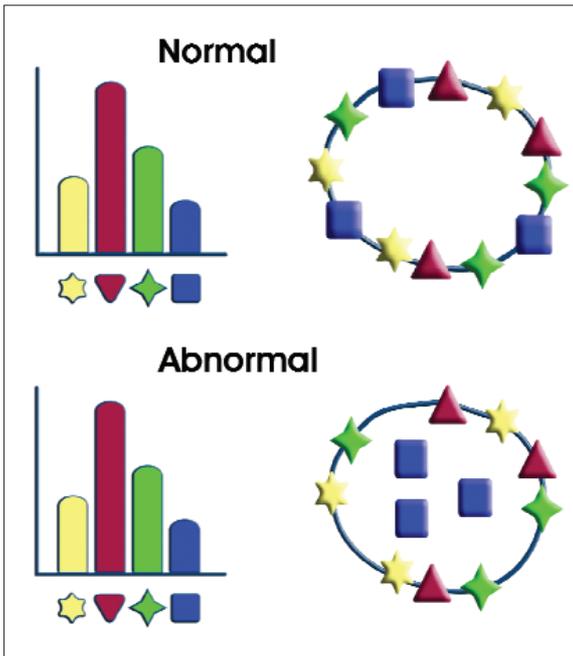
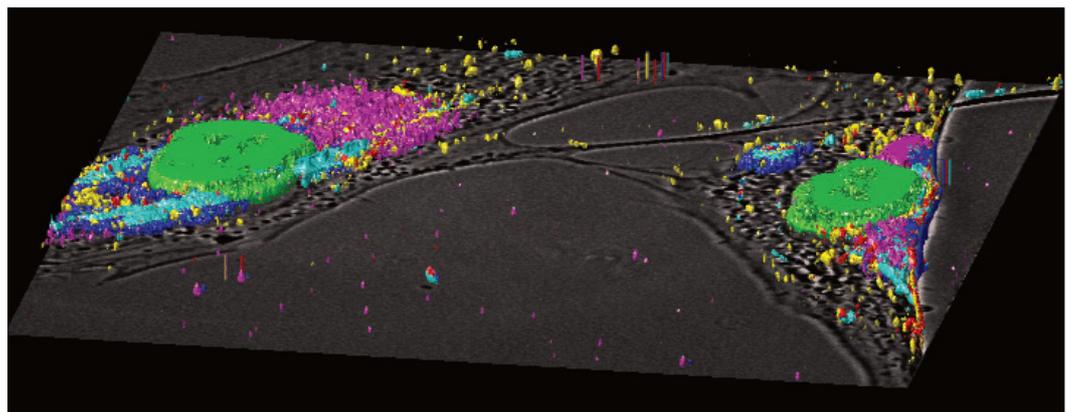


Abbildung 2  
Prinzipschema: Vergleich  
zellulärer Proteom-Muster  
und Profile

Kontext von Zellverbänden untersucht werden kann. Die dabei gewonnene kontextabhängige Proteininformation, die in Zellhomogenaten naturgemäß verloren geht, erweist sich zunehmend als unmittelbar relevant für die Identifikation spezifischer Proteinnetzwerke, die Zellfunktionen oder -dysfunktionen kodieren; diese Erfahrung basiert auf der Natur der Proteinnetzwerke selbst, denn sie sind räumlich determinierte funktionelle Einheiten in jeder einzelnen Zelle (verschiedene Funktionen = verschiedene zelluläre Proteinnetzwerke = verschiedene zelluläre Proteommuster). Das Schema (Abbildung 2) verdeutlicht, dass einzelne normale und abnorme Zellen durch die relative Anordnung von Proteinen spezifisch charakterisiert sind. Werden diese Zellen durch Homogenisierung zerstört, so ergibt die quantitative Messung der einzelnen extrahierten Proteine keinen Unterschied mehr zwischen den beiden Kategorien (Abbildung 2, links). Die klassischen Verfahren der Proteomanalyse („large-scale-protein-profiling“-Verfahren) würden in dem Beispiel zu der falschen Schlussfolgerung Anlass geben, dass die identifizierten Proteine nicht relevant für die weitere Erforschung der Krankheit sind. Die Kartierung der Proteinmuster als Netzwerke (Abbildung 2, rechts) würden zu der gegenteiligen Schlussfolgerung führen. Mit Hilfe geeigneter Computerverfahren können Zellen, die durch besondere Muster spezifisch gekennzeichnet sind, überraschend schnell ausgelesen werden /12/. Die Anwendung von WCPF bestätigt insbesondere in der Analyse invasiver Zellen (Immunzellen, Tumorzellen), dass nur ein kleiner Ausschnitt des Gesamtproteoms, nämlich die gleichzeitige kombinatorische Kar-

terierung von 20 bis 100 Proteinen der Zelloberfläche auf Einzelzellniveau bereits weitaus spezifischere Muster ergibt als die Einzel-Profile derselben Proteine. Diese Muster sind Ausdruck konkreter molekularer Netzwerke der Zelle. Abstrakt betrachtet steht jeder Zelle ein quasi unendlicher „Datenraum“ für die Generierung verschiedener Muster bzw. Funktionen zur Verfügung. Jede Zelle generiert über die Zeit jedoch nur ein limitiertes, wenn auch großes Repertoire von Mustern, das prinzipiell mittels WCPF als „Spur“ in dem Datenraum gelesen werden kann. Dieses „Tracing“ in multidimensionalen Datenräumen bedeutet, dass man gegenüber den auf Zellhomogenaten beruhenden Proteinanalysen hier einen exponentiellen Informationsgewinn hinsichtlich der Abbildung biologischer Funktionen in Proteomen erreicht: Nimmt man 20 Proteine an, die aus einer Million Zellen eines Gewebes extrahiert wurden (klassische Proteomanalysen – „large scale-Verfahren“) und in 250 verschiedenen Konzentrationsstufen gemessen werden können (0 bis 250), so enthielte ein entsprechendes Proteinprofil maximal 25 020 verschiedene Möglichkeiten der Kombination von Proteinen in verschiedenen Konzentrationen. Tatsächlich verfügt jede einzelne Zelle aber potenziell über die Möglichkeit, diese Proteine in jedem definierten subzellulären Volumen entsprechend zu kombinieren. Nimmt man an, dass für jede der 1 Million Zellen 2 000 subzelluläre Volumina (im WCPF-Verfahren) gemessen werden können, so resultiert für das Beispiel eine maximal mögliche Zahl von 1 Million x 2 000 x 25 020 verschiedene Proteinmuster. Die biologische, d. h. auf die Zelle als Funktionseinheit bezogene Information steigt also exponentiell gegenüber den klassischen Verfahren der Proteinanalyse an. WCPF stellt derartige „Leseraster“ zur Verfügung und ermöglicht dadurch einen entscheidenden Informationsgewinn für zelluläre Proteom-Analysen im Krankheitsvergleich („Walking“ durch große Proteom-Fractionen). Auf der Grundlage von Referenzkarten für bestimmte Zelltypen können durch Vergleichsanalysen von Kranken und Gesunden mittels WCPF krankheitsspezifische zelluläre Proteommuster relativ rasch, teilweise innerhalb weniger Tage identifiziert werden. Die Abbil-

Abbildung 3  
Subzellulärer Toponom-  
Fingerprint von zwei  
Muskelzellen



dung 3 zeigt ein Beispiel einer subzellulären ‚Protein-Referenz-Karte‘, eine so genannte Partielle Toponom<sup>3)</sup> Karte von zwei Muskelzellen. Man erkennt dicht gepackte Vesikel. Jedes Vesikel enthält ein jeweils spezifisches Protein-Ensemble und ist dadurch als besonderer Vesikeltyp kartierbar. Wie der Vergleich zeigt, weisen die beiden Zellen eine deutlich unterschiedliche Anordnung der verschiedenen Vesikeltypen auf. Diese kennzeichnen verschiedene Funktionszustände der Zellen: rechte Zelle (Fusionsbereitschaft), linke Zelle (Migrationsstatus).

**FORMALISIERUNG VON NETZWERK-MOTIVEN ALS GEOMETRISCHE OBJEKTE**

Wesentlich für eine exakte quantitative Beschreibung von Protein-Kontexten, die als zelluläre Netzwerke Funktionen generieren (Netzwerk-Motive), ist die Mathematisierung ihrer Muster, die mittels WCPF detektiert werden. Das sei anhand des Schemas in Abbildung 4 erläutert. Ein molekulares Netzwerk der Zelloberflächenmembran von zwei Zellen im Vergleich zeigt die Anordnung von sieben verschiedenen Proteinen zueinander als Komplexe. Die einzelnen Proteine werden mittels WCPF gleichzeitig markiert und als Photonensignal erfasst. Dabei beinhaltet jedes Photonensignal einen Grauwert, der ein relatives Maß für die Menge des jeweiligen Proteins angibt. Es hat sich im Hinblick auf die rasche Orientierung in Datensätzen als effektiv erwiesen, derartige Primärdaten zu binarisieren, d. h. jedes Protein, bezogen auf einen, nach bestimmten Kriterien gesetzten Schwellenwert als anwesend oder abwesend aufzutragen (anwesend [1], abwesend [0] = 1 Bit). Auf dieser Basis kann für jedes vorliegende Teilvolumen einer Zelle (Pixel/Voxel) ein kombinatorischer Binärcode zugeordnet werden. Die primäre Bildinformation, die in komplexen Grauwertverteilungen besteht, wird dabei durch eine relativ einfache geometrische Beschreibung in x, y, z-Koordinaten ersetzt. Wie in Abbildung 4 gezeigt ist, kann man sich in solchen Datensätzen relativ rasch darüber informieren, welche kombinatorischen Muster eine biologische Struktur aufweist und welche nicht. In dem gezeigten Beispiel erweist sich, dass die Binär-Codes über eine gegebene Strecke der Zellmembran hinweg verschieden sind. Biologisch formuliert würden wir diese Unterschiede als „supramolekulare Domänen“ der Zellmembran bezeichnen. Als sehr bedeutsam hat es sich erwiesen, durch einen Quervergleich solcher Binär-Codes nach Gemeinsamkeiten zu fahnden. In dem gezeigten Schema (Abbildung 4) erweist sich z. B. in der Zelle 1 das Protein 1 als das einzige Protein, das in allen Domänen derselben Zellmembran vorkommt. Es kann mit anderen Proteinen verschiedenartig gekoppelt sein (in Abbildung 4 sind variable Kopplungen mit einem [\*] gekennzeichnet). Das Protein 7 ist invers mit Protein 1 gekoppelt (Abbildung 4, [0] = Protein 7). Im Gegensatz hierzu ist in Zelle 2 das Protein 7 allen Domänen gemeinsam. Es hat sich gezeigt, dass die Detektion solcher Gemeinsam-

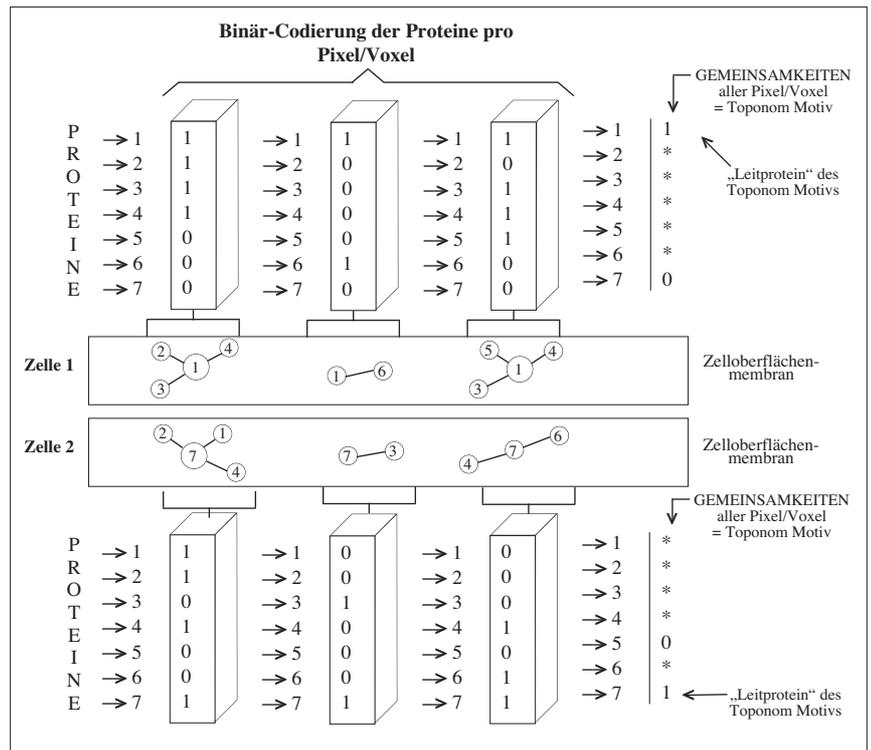


Abbildung 4 Schema der Toponom-Kartierung molekularer Netzwerke in zwei Zellen im Vergleich. Die durch Kreise gekennzeichneten Proteine 1 bis 7 sind in verschiedener Weise miteinander assoziiert. Diese Muster können mit Hilfe von Lese-Robotern als Binär-Codes ausgelesen werden. Dabei ergeben sich Gemeinsamkeiten und Unterschiede (horizontaler Vergleich in dem Schema), die gesamthaft als Toponom-Motive bezeichnet werden (rechte Seite der Abbildung). [1] = Leit Protein; [0] = Invers-gekoppelte Proteine; [\*] = variabel assoziierte Proteine. Das Schema ist stark vereinfacht. In der Realität können wesentlich mehr Proteine gleichzeitig kartiert werden (mind. 100 verschiedene Proteinspezies).

keiten und Unterschiede, die wir gesamthaft als „Toponom-Motive“ bezeichnen, bestimmte Zellfunktionen oder Zelltypen spezifisch kennzeichnen. Als besonders wichtig haben sich diejenigen Proteine erwiesen, die in allen Domänen vorkommen (Abbildung 4: Protein 1 in Zelle 1 und Protein 7 in Zelle 2). Wir bezeichnen solche Proteine als Leitproteine (Lead Proteins). Eine gezielte Blockierung solcher Proteine, nicht aber die Blockierung der variablen Proteine ([\*]), führt zu schweren Funktionsstörungen, die mit Hilfe geeigneter biologischer Modelle gemessen werden können. Wir können aus solchen Experimenten ableiten, dass Leitproteine innerhalb eines Netzwerkes von Proteinkomplexen eine dominierende hierarchische Rolle spielen. Solche, auf geometrischen Binär-Code-Daten beruhenden Experimente sind die Grundlage für eine systematische Analyse von Protein-Netzwerken in der Zelle. Die Abbildung 5 zeigt als Schema das Ergebnis eines in unseren Labors durchgeführten Experimentes an wandernden Zellen: Eine wandernde Zelle ist in den Proteinkomplexen der Zelloberfläche durch zwei Leitproteine spezifisch charakterisiert (Abbildung 5a). Die experimentelle Einzelblockierung jeweils eines dieser Proteine (Protein 1 oder 2) hat zwei Folgen: Erstens kommt es zu einer Zerstörung des Proteinmusters (Abbil-

3) Toponom: Gesamtheit der molekularen Netzwerke der Zelle, einschließlich aller Regeln (eine Art Grammatik) für die Ko-Kompartimentierung der Proteine. Der Begriff ist eine Zusammensetzung der altgriechischen Worte ‚Topos‘ (= Ort) und ‚Nomos‘ (= Gesetz).

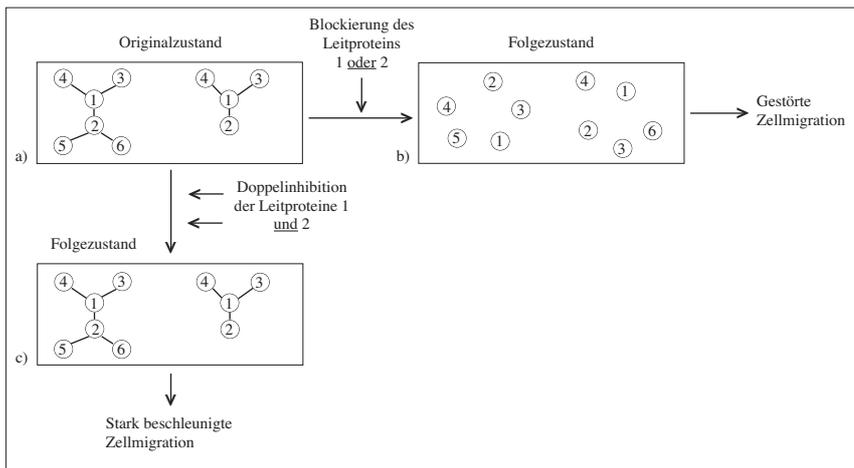


Abbildung 5  
 Schema zur Erläuterung einer experimentellen Analyse eines molekularen Netzwerkes der Zellmembran (Erläuterungen finden sich im Text)

Abbildung 5b: Die Proteine des ursprünglichen Netzwerkes verbleiben in der Membran, sind offenbar aber nicht mehr miteinander assoziiert. Zweitens resultiert offenbar wegen der Zerstörung der topologischen Muster des molekularen Netzwerkes eine starke Einschränkung der Zellwanderung. Würde man beide Leitproteine blockieren (Abbildung 5: Doppelinhibition von Protein 1 und 2 gleichzeitig) so würde man eine Addition dieses Effektes, also eine noch stärkere Einschränkung der Wanderungsfähigkeit erwarten. Es zeigt sich allerdings im Experiment das Gegenteil: Das Muster scheint weitgehend erhalten zu bleiben (Abbildung 5c), und die Zellen steigern die Wanderungsgeschwindigkeit um ein Vielfaches. Man kann mithin derartige Netzwerkstrukturen und Funktionen mit der uns gegenwärtig zur Verfügung stehenden ‚molekularen Logik‘ nicht einschätzen, geschweige denn vorhersagen. Erst die Kopplung von Toponom-Messungen und Experimenten wird es erlauben, die Kodierungsfunktionen, die in solchen Netzwerken verborgen sind, genau zu verstehen. Derartige Analysen befinden sich derzeit noch in einem frühen Stadium, jedoch wird deutlich, dass sie tiefe Einblicke in die Funktionalität molekularer Netzwerke erlauben: Kopplungen von WCPF-basierten Muster-Funktionsanalysen ermöglichen einen direkten Zugang zur hierarchischen Funktion molekularer Netzwerk-Motive. Die dabei entdeckten Leitproteine scheinen biologisch relevante Target-Kandidaten für die Wirkstoffsuche und Entwicklung zu sein. Beides ist für die biologische Grundlagenforschung und die Pharmaforschung gleichermaßen wichtig.

**FUNKTIONELLER INFORMATIONSGEWINN: PERSPEKTIVEN FÜR DIE WIRKSTOFFENTWICKLUNG**

MelTec hat ein System für die Entschlüsselung solcher Muster entwickelt. Es besteht aus den Kategorien: *Messen* (WCPF), *Filtern* (Vergleichen von zellulären Toponom-Mustern in multidimensionalen Datenräumen), *Vorhersagen* (von krank-

heitsspezifischen Targets) und *Experimentelle Überprüfung* (zelluläre Assays, mit denen Targets validiert werden). Mit Hilfe dieses „Entschlüsselungsapparates“ konnten bereits Targets und so genannte Drug Leads (Wirkstoffkandidaten) identifiziert werden, welche die Invasivität von Immun- und Tumorzellen bei bestimmten Erkrankungen spezifisch beeinflussen. Ein von WCPF vorhergesagtes Leitprotein als kritisches Targetmolekül bei der Amyotrophen Lateralsklerose (ALS) /13/, einer fatal verlaufenden Erkrankung des motorischen Nervensystems, wurde kürzlich durch ein unabhängiges Knock-Out-Maus-Modell bestätigt /14/, nachdem MelTec anlässlich einer Konferenz der amerikanischen ALS-Association (2001) die WCPF-Daten offengelegt hat. Erste Gespräche mit der amerikanischen Zulassungsbehörde (Food and Drug Administration, FDA) (2004), bei dem die pathophysiologische Bedeutung dieses Target-Proteins erörtert wurde, haben die Weichen für den nächsten Schritt auf dem Weg zu einer klinischen Studie gestellt. Dadurch und durch weitere ähnliche Protokolle, die z. Zt. der wissenschaftlichen Öffentlichkeit noch nicht zur Verfügung stehen, wurde die funktionelle Prädiktivität der Toponom-Analysen molekularer Netzwerke bestätigt. Für die Zukunft werden allerdings noch zahlreiche Untersuchungen notwendig sein, um dieses Prinzip, das eine effektivere Target-Suche und Wirkstoffentwicklung impliziert, weiter zu untermauern. WCPF kann in verschiedener Weise mit anderen proteomanalytischen Verfahren und Hochdurchsatz-Verfahren im Bereich „Drug Discovery“ gekoppelt werden. Interessant ist es insbesondere, identifizierte Wirkstoffkandidaten mit Hilfe von WCPF funktionell zu überprüfen und hierarchisch zu priorisieren. Andererseits stellt WCPF auch eine effiziente Plattform für die Validierung von Targets aus klassischen Genomics- und Proteomics-Ansätzen (Large Scale Expression Profiling, s. o.) dar. Die Bedeutung dieser Technologie dürfte jedoch vor allem in dem Potenzial für die Identifikation krankheitsspezifischer Targets und Wirkstoff-Kandidaten liegen. Wir erwarten, dass dadurch die Effizienz der Wirkstoffentwicklung deutlich gesteigert werden kann. MelTec hält heute über 50 Patente auf dem Gebiet der Toponom-Forschung, gegliedert in Technologie-, Target- und Wirkstoffrechte.

**TECHNOLOGIETRANSFER IM ZENIT**

Am Zentrum für Neurowissenschaftliche Innovation und Technologie (ZENIT) an der Medizinischen Fakultät Magdeburg sind mehrere Firmen und Forschergruppen, wie auch die MelTec GmbH & Co. KG und ein Teil des Instituts für Medizinische Neurobiologie, beheimatet. Das ZENIT bietet eine hervorragende Infrastruktur für den Technologietransfer am Standort Magdeburg. Insbesondere die enge räumliche Assoziation von Grundlagenforschung, Technologieentwicklung und Produktentwicklung erlaubt hier prinzipiell einen fast nahtlosen Fluss von Projekten durch verschiedene Entwicklungsstufen –

von der marktfernen Grundlagenforschung bis zur marktnahen Anwendung von Prototypen. Eine besondere Herausforderung ist hierbei die Auflösung der Grenzen zwischen akademischer und industrieller Forschung und Anwendung zum gegenseitigen Nutzen der heute noch weitgehend getrennten Strukturen. Dabei müssen auch insbesondere die Zusammenhänge zwischen den Forderungen des Patentrechts, der Industrie, des Erreichens einer hinreichend kritischen Masse an Kompetenz und Institutionalisierung sowie Publikationsstrategien berücksichtigt werden. Im vorliegenden Fall ist es der MelTec im ZENIT gelungen, die Entwicklung einer sehr komplizierten Technologie bis zur Anwendungsreife, und zwar durch ein Zusammenschweißen von biologischen Methoden, Ingenieurwesen, Informationstechnologie und Mathematik voranzutreiben. Die dabei entstandenen Patente mit internationalen Alleinstellungsmerkmalen stellen potenziell eine bedeutende Hebelkraft für die weitere Entwicklung, u. a. die Standortentwicklung, und die Assoziation mit der akademischen Forschung dar. Die mittels der Technologie gefundenen Wirkstoffe und potenziellen neuen Diagnostika bis zur Anwendung am Menschen zu entwickeln, wird ein nächstes Ziel sein. Ein zweites wichtiges Ziel ist die Einführung der Technologie in die systembiologische Grundlagenforschung. Schritte zu diesen Zielen wurden anlässlich der von MelTec und dem Institut für Medizinische Neurobiologie organisierten internationalen Konferenz „Linking

mathematical and biological models in cancer research“ in Magdeburg (2003) dokumentiert. Neue Kooperationen im Nationalen Genom-Forschungs-Netz (NGFN-2), die zur Zeit beginnen, sind in der Weltorganisation „Human Proteome Organisation“ verankert.

#### DISKURS ZUKÜNFTIGER ZIELE, VISIONEN

Ein sehr ehrgeiziges Ziel, das von Gene Myers<sup>4)</sup>, dem ehemaligen Chief Sales Officer der amerikanischen Firma CELERA Genomics, formuliert wurde, ist die komplette, d. h. Genomweite Entschlüsselung der Proteinnetzwerke, also des Toponomes des Menschen. Vielleicht werden aber zunächst einfache Tiermodelle, z. B. die Fruchtfliege *Drosophila*, im Vordergrund stehen müssen. Erste vorausschauende Analysen zwischen G. Myers und MelTec, die seit ca. einem Jahr erfolgen, ergaben, dass eine Verdoppelung der derzeit 15 MelTec-Roboter auf 30 Roboter in einem Zeitraum von drei bis fünf Jahren zunächst eine Kartierung von 10.000 bis 30.000 Proteinen in mehreren verschiedenen Gewebetypen erlauben würde – die Verfügbarkeit einer geeigneten Infrastruktur und Vorlaufleistungen zur Generierung großer Tag-Bibliotheken vorausgesetzt. Eine stufenweise Herangehensweise wird derzeit zwischen einigen führenden Forschern aus den Bereichen Mathematik, Biologie und Medizin diskutiert. Die treibende Kraft ist die große Vision einer umfassenden Grammatik des biologischen Codes der Zellfunktionen.

4)

Gene Myers hat durch Entwicklung und Anwendung des so genannten „Shot Gun“-Verfahrens in Rekordzeit von vier Jahren die vollständigen Genome von vier Spezies einschließlich des Menschen sequenziert. G. Myers lehrt heute an der Universität in Berkeley (USA) und erhielt vor kurzem einen hohen Preis der Max-Planck-Gesellschaft für seine Forschung.

## Literatur

- /1/ Tufts CSDD Impact Report, Vol. 4, No 5, 2002
- /2/ Lehman Brothers / Mc Kinsey & Co. The Fruits of Genomics, 2001
- /3/ Collins FS, Green ES, Guttmacher AE, Guyer MS. A vision for the future of genomics research. *Nature*, 2003; 422:835-847.
- /4/ Gavin, A.C. et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141-147 (2002)
- /5/ Hazbun, T.R. et al. Networking proteins in yeast. *Proc. Natl. Acad. Sci.* 98, 4277-4278 (2001)
- /6/ Alberts, B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* 92, 291-294 (1998)
- /7/ Schubert, W.: Topological Proteomics, Toponomics. MELK Technology. In: Hecker, M., Müllner, S. (eds). *Proteomics of Microorganisms. Fundamental Aspects and Application. Advances in Biochemical engineering/Biotechnology*, Vol. 83, Springer Verlag, pp. 189-211, 2003
- /8/ Schubert, W., Zimmermann, K., Cramer, M., Starzinski-Powitz, A.: Lymphocyte antigen Leu19 as a molecular marker of regeneration in human skeletal muscle. *Proc. Natl. Acad. Sci. USA* 86: 307-311 (1989).
- /9/ Schubert, W.: Antigenic determinants of T lymphocyte a/b receptor and other leukocyte surface proteins as differential markers of skeletal muscle regeneration: detection of spatially and timely restricted patterns by MAM microscopy. *Eur. J. Cell Biol.* 58: 395-410 (1992)
- /10/ De Angelis L. et al. Skeletal myogenic progenitors originating from embryonic dorsal aorta coexpress endothelial and myogenic markers and contribute to postnatal muscle growth and regeneration. *J Cell Biol.* 147, 869-878 (1999)
- /11/ Schubert, W: US-Patent 6,150,173 (2000); EP 1 069 431 (00 114 854.3) (2003)
- /12/ Nattkemper, T., Ritter, H., Schubert, W.: A neural classifier enabling high-throughput topological analysis of lymphocytes in tissue sections. *IEEE Trans. Inf. Techn. in Biomed.* 5, 138-149 (2001).
- /13/ Schubert, W.: USA Patents 09/367,011; 09/802,305; 09/801,414. (1999)
- /14/ Mohammed, HA et al: Immunoglobulin Fc gamma receptor promotes immunoglobulin uptake, immunoglobulin-mediated calcium increase, and neurotransmitter release in motor neurons. *J Neurosci Res* 69, 110-116 (2002)



### Hochschuldozent Dr. med. Walter Schubert

studierte Medizin an den Universitäten Aachen und Bonn, sowie Musik in Salzburg. Als Leiter des neuromuskulären Labors der neurologischen Universitätsklinik Bonn erhielt er 1986 im Rahmen einer Ausschreibung der Deutschen Forschungsgemeinschaft ein Projekt zur mikroskopischen Erforschung von Protein-Konstellationen in Zellen mit Hilfe mehrerer Laserlinien.

Theoretische Arbeiten zur kontinuierlichen Lokalisierung von Proteinen in situ und deren Anwendung in der Medizinischen Forschung führten ihn von 1988 bis 1992 als Research Associate an das Zentrum für Molekulare Biologie Heidelberg (ZMBH), wo er die entsprechenden Protokolle für zahlreiche Gewebetypen erarbeiten konnte. 1993 wurde er zum C2-Hochschuldozenten für das Fach Medizinische Neurobiologie an der Universität Magdeburg ernannt, wo hervorragende Möglichkeiten für die Etablierung der von ihm entwickelten Methoden der Proteinlokalisierung als Roboterverfahren bestanden. Hier war er von 1994-1997 Sprecher des Innovationskollegs „Bildinformation“, gründete die Arbeitsgruppe „Molekulare Mustererkennung“ am Institut für Medizinische Neurobiologie und beteiligte sich am Aufbau des Studiengangs Computervisualistik. Im Jahr 1999 gründete er die Biotech Firma Mel-Tec GmbH als Spin Off aus, die das von ihm entwickelte Verfahren der Analyse von Protein Netzwerken als Robotertechnologie etablierte. Als Geschäftsführer und CSO der Firma erhielt er im Jahr 2000 den Biochance und 2003 einen Proteomics Award des BMBF und etablierte Kooperationen mit Industriepartnern. Diese Projekte bildeten die Grundlage für ein nationales Kooperationsnetzwerk auf dem Gebiet der Protein-Systemforschung, das von der Mel-Tec GmbH kofinanziert wird (CELLECT). Im Jahr 2004 wurde er zusammen mit Vertretern des *Wellcome Trusts* in den wissenschaftlichen Aufsichtsrat der „University of Warwick“ berufen. Seit Oktober 2004 ist er mit einem Projekt im Rahmen des „Human Brain Proteome“-Projektes (NGFN-2) Mitglied der internationalen „Human Proteome Organization (HUPO)“.