

Hightech im Dienste der Infektiologie

JOACHIM BUGERT

Einleitung

Jeden Tag sterben weltweit etwa 13 Millionen Menschen an den Folgen viraler, bakterieller oder parasitärer Erkrankungen.

Zur Entwicklung neuer und hoch wirksamer Impfstoffe und Medikamente ist das Studium der Genome und Proteome dieser Mikroorganismen und ein Verständnis der darin abgebildeten biochemischen Regulationsnetzwerke unabdingbar. Die Verfolgung dieser Ziele hat sich in den letzten Jahren zu einer Multimilliarden-Euro-Industrie entwickelt, deren Produktivität sich Jahr für Jahr um Größenordnungen steigert und deren Komplexität ihrem Forschungsgegenstand alle Ehre macht.

Wer hätte sich vor 20 Jahren vorstellen können, mit welcher ungeheurer Geschwindigkeit heute vollständige mikrobielle Genome sequenziert werden und wie komplex sich die darin abgebildeten biochemischen Regulationsvorgänge darstellen würden. Dabei stellt die immer größere Zahl vollständiger Genomsequenzen wichtiger humanpathogener Bakterien, Viren, Pilze und Parasiten immer höhere Anforderungen an Rechner und Speicherkapazitäten, während die Auswertung dieser Daten die ständige Entwicklung neuer Methoden zur Beschleunigung der Identifizierung und funktionellen Analyse der Genprodukte erfordert.

Fünf Hauptverfahren stellen zurzeit die technologische Plattform großer Genomzentren dar: Genomsequenzierungsprojekte (genomics), DNA-Expressions-Chips (expression microarrays), Proteomforschung (proteomics), Dokumentation, Vorhersage und Simulation (bioinformatics/prediction-annotation/simulation of biological processes) und strukturelle Genomforschung (drug design), wobei letztere auch als postgenomische Verfahren (postgenomics) bezeichnet werden.

Die im Folgenden besprochenen Themengebiete Genomanalyse, Mikroarrays und Proteomik, Medikamentenentwicklung und Simulation biologischer Prozesse spiegeln diese Schwerpunkte wider.

1. Genomanalyse (High throughput genome analysis)

Die Sequenzierung mikrobieller Genome und des humanen Genoms sowie neuerdings einer Reihe sogenannter parasitärer „Vektorgenome“, zum Beispiel *Anopheles gambiae*, sind komplementäre Ansätze, die zusammen zur Entwicklung neuer Ansätze zur Prävention und Behandlung mikrobieller Erkrankungen beitragen sollen. Solche Genomprojekte übersteigen längst die Kapazitäten einzelner Forschungsgruppen

und sogar von Universitäten. Genomforschung wird heute weltweit im industriellen Maßstab von Regierungen [L1] großen „Non-Profit“-Organisationen [L2] und international vernetzten Sequenzierzentren durchgeführt.

Europäische und US-Sequenzierzentren

Dabei ist in den USA das „Institute for Genomic Research“ (TIGR), 1992 von J. Craig Venter nach Abschluss des humanen Genomprojekts (Celera-HGP) gegründet, führend. TIGR hat schon 1995 die erste vollständige Genomsequenz eines frei lebenden Bakteriums (*Haemophilus influenzae*) veröffentlicht. Seitdem wurden, zum Teil in Kooperation mit europäischen Zentren, weitere 188 bakterielle Genome [L3 Appendix], darunter humanpathogene Spezies wie *Mycobacterium tuberculosis*, *Neisseria meningitidis*, *Pseudomonas aeruginosa*, *Escherichia coli*, *Bacillus subtilis* und verschiedene Chlamydia-Spezies bearbeitet; mehr als 100 sind zurzeit in Bearbeitung. Seit 2004 gibt es ein Influenza Virus Genome Sequencing Project am TIGR/NIAID Microbial Sequencing Centre. Es wurden 2.028 vollständige Influenza A-Virusgenome und 144 Influenza B-Virusgenome sequenziert. Eine weitere Venter-Initiative ist das Metagenomikprojekt der Sorcerer II Global Ocean Sampling (GOS) Expedition. Dieses Projekt begann 2003 auf Venters Privatjacht Sorcerer II und erste Ergebnisse wurden 2007 von Yooseph und Mitarbeitern berichtet [20].

6,12 Millionen neuer Proteine wurden aus 7,7 Millionen GOS Clustern (Cluster: neue Art der Contig Assemblierung) abgeleitet, was auf einen Schlag die Zahl aller heute bekannten Proteine verdoppelte. Eintausendsiebenhundert bisher unbekannte große Proteinfamilien wurden entdeckt, wovon die meisten überragenderweise viralen Ursprungs sind. Die dazugehörigen Viren waren ebenfalls unbekannt. Wie von der Sammeltechnik (Filtrierung von Seewasser) zu erwarten war, wurden keine Angehörigen der CD-Familie (Kaspasen) gefunden, die nur bei multizellulären Eukaryonten und deren Viren vorkommen. Sechstausend bisher isolierte kodierende Sequenzen fanden Korrelate in der GOS-Datenbank. Es ist zu erwarten, dass die neuen Proteine der GOS-Datenbank einen wichtigen Beitrag bei der Targetselektion in der proteomischen Medikamentenentwicklung leisten werden (► unten).

Alle auf diese Weise gesammelten Sequenzen kommen über NCBIInr [L4], die größte öffentliche GenBank der Welt, sofort in die „public domain“. TIGR/J. Craig Venter Institut sind zusammen vermutlich das weltweit führende Genomsequenzierungszentrum. In Europa ist das französische nationale Genomsequenzierungsnetzwerk „Genopolis“ seit 1999 unter der Führung des Institute Pasteur [L5] umfassend organisiert und logistisch gut aufgestellt. Im Pasteur-

institut wurden die ersten vollständigen Genomsequenzen von Hepatitis B Virus (1979), dem humanen Papillomavirus (1981), von HIV Typ 1 (1985) und, entsprechend dem historischen Auftrag des Instituts, vom Tollwutvirus (1988) erstellt. Am Pasteurinstitut ist das „Laboratoire de Génomique des Microorganismes pathogènes“ (Laboratory of Genomics of Microbial Pathogens) ausgerüstet für große Sequenzierprojekte und kooperiert mit Génoscope, dem französischen Nationalen Sequenzierzentrum in Evry und weltweit mit einer großen Zahl nationaler und privater Sequenzierzentren. Zu diesen zählt auch das „Sanger Institute“, Teil des Wellcome Trust/MRC Genome Campus [L6] in Hinxton, Cambridge, Großbritannien, welches zu Ehren des Erfinders der Sanger Sequenzierungsmethode und dem Entschlüssler der ersten Aminosäuresequenz am Beispiel von Insulin, dem zweifachen Nobelpreisträger Frederick Sanger, dort etabliert wurde.

Technische Aspekte der Genomsequenzierung

Die Sanger Sequenzierungsmethode, basierend auf der Dideoxynukleotid-Kettenabbruch-Biochemie, wurde 1975 von Frederick Sanger entwickelt, erstmals zur Bestimmung der DNA-Nukleotidsequenz des Phagen Φ -X174 eingesetzt und ist die Grundlage für alle heute verwendeten Sequenzierungsverfahren. Für die Erfindung dieses Verfahrens erhielt Sanger 1980 zusammen mit Walter Gilbert seinen zweiten Nobelpreis. Mit der Etablierung großer Sequenzierzentren unter Einsatz leistungsfähiger Datenverarbeitungssoftware hat sich das sogenannte „Random shotgun sequencing“ von Genomfragmentbibliotheken mit kleinen und großen Inserten, letztere in BAC-Vektoren (BAC: bacterial artificial chromosomes with low copy number: 1–2 pro Zelle) zur Vermeidung von Genotoxizitätseffekten durchgesetzt [7]. Wichtige jüngste Verbesserungen in der Genomsequenzierungstechnologie schließen die Optimierung der Sequenzierchemie, (Dyeterminator® und Thermofidela-se®), basierend auf der Thermopolymerasereaktion, einem weiteren mit dem Nobelpreis belohnten Verfahren, und die Entwicklung leistungsfähiger Genomsequenzierungs-Software ein. Automatische Sequenziermaschinen übertragen Rohdaten direkt in Rechnernetzwerke zur weiteren Auswertung.

Aus der Vielzahl vorhandener Programme sind als jüngste Weiterentwicklungen von GENSCAN der GLIMMER Gensucher (alle basierend auf dem interpolierten MARKOV Modell) und das MUMer Alignmentprogramm (Contig-Alignment ganzer Genomsequenzen mithilfe eines „suffix-trees“) zu nennen. Speziell zur Verarbeitung großer bakterieller Genome wurden die erweiterten Gensucher ORPHEUS [L7] und Glimmer2 [L8] entwickelt. Aminosäuresequenzen vorhergesagter Proteine werden mit vorhandenen

Gendatenbankeinträgen (DDBJ/EMBL/NCBI-GenBank) mithilfe der Programme FASTA (Protein FASTA) vom EMBL-EBI am Wellcome Trust Genome Campus in Cambridgeshire und BLAST (Basic Local Alignment Search Tool; L5) des Nationalen Zentrums für Biotechnologie-Information (NCBI) in Bethesda, Maryland, USA verglichen. Als Beispiel für die zur zügigen Datenverarbeitung notwendige Rechnerleistung sei nochmals das J. Craig Venter Institut genannt. Dort werden 125 duale in einem „Storage Area Network (SAN)“ vernetzten 3,06-GHz Xeon Prozessor Systeme mit jeweils 2 Gb RAM betrieben. Für das Sorcerer II Projekt wurden auf diesem System 4.666.366 CPU-Stunden geloggt. Für 588.298 CPU-Stunden wurde ein noch leistungsfähigerer Zugang „National Energy Research Scientific Computing Center (NERSC)“ Seaborg Computer Cluster mit 380 Knoten benutzt.

Nach Beendigung der Gensuche kommen Programme zum Nachweis funktionell relevanter Proteindomänen, wie Pfam am Sanger Zentrum und SWISSPROT/ExPASy-PROSITE am Schweizer Bioinformatikinstitut SIB sowie viele weitere wie RCSB-PDB (protein sequences with solved structures), PIR (Protein Information Resource) and PRF (Protein Research Foundation, Osaka, Japan), TrEMBL, PRINTS, ProDom, Smart und TIGRFAMS zum Zug. Für die Identifizierung von alpha-Helices in Transmembran-domänen gibt es das Programm TMHMM, für subzelluläre Lokalisierungsvorhersagen TargetP, beides am dänischen Zentrum für Biologische Sequenzanalyse (CBS) der technischen Universität Dänemark. Auch zur Dokumentation (annotation) gibt es Spezialsoftware wie zum Beispiel das Annotierungsprogramm Artemis.

Weiterführende bioinformatische Auswertungen, zum Beispiel mit dem Programm InterProScan (Integrated Ressource of Protein Families, Domains and Sites), beschreiben Genorganisation, suchen nach Genfamilien oder Einzelnukleotidpolymorphismen und führen phylogenetische Studien durch. Letztere werden zunehmend in phylogenetischen Vergleichsnetzwerken organisiert, die besser die Beziehungen zwischen verschiedenen Mikroorganismen darstellen als einzelnenbezogene phylogenetische Studien. Hier wird vor allem bei bakteriellen Genomen nach Beweisen für horizontalen Gentransfer gesucht [12]. Horizontaler Gentransfer spielt eine Rolle bei der Mobilität von Antibiotikaresistenzen in Prokaryonten, findet sich aber auch bei Protozoen und vermutlich sogar bei höheren Organismen, möglicherweise vermittelt durch retrovirale Vektoren und artifiziell durch Gentherapie. Horizontaler Gentransfer wird als neues Paradigma der Biologie und als eigentlicher Motor der Evolution angesehen [19].

Vieľfach werden Ressourcen-Datenbanken mit auto-

matischer Annotation auf dem WWW veröffentlicht, um die entstehenden großen Datenmengen zu organisieren, multiple Namensgebung und andere Inkonsistenzen aufzudecken oder zu vermeiden und anderen Wissenschaftlern über multiple Anfragemenüs zugänglich zu machen. Beispiele sind die „Comprehensive Microbial Resource“ [L2] des TIGR-Wissenschaftlers Owen White und mikrobenspezifische Datenbanken wie die „Encyclopedia of Escherichia coli K-12 Genes and Metabolism-EcoCyc (E. coli K12 „literature based curation“ und E. coli Wiki) und die allgemein bakterielle Metabolismus-Enzyklopädie MetaCyc der Firma SRI international in Zusammenarbeit mit TIGR oder die Biocatalysis/Biodegradation Database der University of Minnesota. Diese Datensammlungen erlauben die Rekonstruktion mikrobieller Genome nach metabolischen Gesichtspunkten, ein Ansatzpunkt der Systembiologie [7], und bieten Zugangsmenüs, die ein unbekanntes Protein mittels multipler Alignments direkt in bekannte Genfamilien oder Stoffwechselwege einordnen.

Alle Gendatenbanken sind öffentlich und frei, obwohl bei der Diskussion von Datenmissbrauch (Bioterrorismus) der öffentliche Zugang schon in Frage gestellt wurde. Die meisten Bioinformatikprogramme sind zudem unter der „General Public Licence“ (GNU) frei herunterladbar.

Das Sequenzieren mikrobieller Genome bleibt weiterhin eine Priorität in der Infektiologie. Die Verfügbarkeit einer großen Zahl sequenzierter Genome ist die Grundlage des Studiums von Genfunktionen mithilfe umfassender Mutagenese-Projekte. Mithilfe von Transposon-Mutagenese gesteuerter Genreduktion (knock-out) gelang es einer TIGR-Forschungsgruppe um Clyde Hutchinson und Scott Peterson Minimalgenvarianten verschiedener Mycoplasmen-Spezies herzustellen [12]. *Mycoplasma sp.* benötigen lediglich 265–330 essenzielle Gene zum Wachstum in angereichertem Flüssigmedium. Jeder neue analysierte Mikroorganismus bringt demnach neue Einblicke in das biochemische Potenzial der Mikroben. Immer mehr Mikroben werden in Sequenzierprogramme aufgenommen, um immer dringenderen Fragestellungen nach Biodiversität und evolutionären Prozessen, insbesondere im Hinblick auf Antibiotikaresistenz und Medikamentenentwicklung, gerecht zu werden.

2. Mikroarrays in der Infektiologie

Bioinformatische Ansätze in der Infektiologie finden ihre Grenze, wenn es um die Frage geht, ob das unbekannte mikrobielle Gen oder das hypothetische Genprodukt in der Realität existiert. Die Frage lässt sich in mindestens 4 Teilen stellen:

1. Wird messenger-RNA produziert?
2. Was ist die Halbwertszeit der mRNA?
3. Wird Protein translatiert?

4. Was ist die Halbwertszeit des Proteins im biologischen System?

Die erste Frage wird von DNA-Mikroarrays beantwortet. DNA-Mikroarrays erlauben die gleichzeitige Untersuchung der mRNA-Expressionsmuster von Tausenden von Genen und ihrer temporären Regulation.

mRNA isoliert von Mikroorganismen verschiedener Stämme unter verschiedenen Wachstumsbedingungen in verschiedenen Wirtsorganen wird mit fluoreszenzkonjugierten Ribonukleotiden markiert und mit auf ‚Chipsets‘ von der Größe eines Mikroskop-Objektträgers immobilisierten Sonden hybridisiert. Die Auswertung durch ein Mikroarray-Lesegerät („Microarray reader“) erlaubt die Identifizierung eines spezifischen Pathogens oder ergibt einen quantitativen Überblick der relativen Genexpressionslevel der Mikrobe oder zellulärer Gene des Wirts. Mikroarrays erlauben die systematische Abfrage von Genen ausgesuchter Stoffwechselwege oder bestimmter mikrobieller Gengruppen (zum Beispiel Struktur-, Nonstrukturproteine).

Die so gewonnenen differenziellen Transkriptionsprofile erlauben theoretisch einen Einblick in die physiologischen Unterschiede zwischen mikrobiellen Populationen, abhängig von Wachstumsmedien, exponentiellen oder transitionellen Wachstumsphasen, aktiver oder persistierender bzw. latenter Infektion. Eine Zuordnung der Veränderungen zu bestimmten Stoffwechselwegen ist möglich. Mikroarray-Technologie ist sensitiv genug, um dynamische Prozesse, wie zeitliche Genregulation und mikrobielle Genomreplikation, mit hoher Auflösung zu analysieren. Solche Analysen reduzieren die Notwendigkeit zur Erstellung teurer initialer biochemischer Reaktionsprofile und bieten gleichzeitig Anhaltspunkte, welche Gene unter welchen Bedingungen Gegenstand von biochemischen Analysen sein sollten.

Mikroarrays sind in der praktischen Infektiologie noch nicht wirklich etabliert, werden zurzeit aber, ausreichende Standardisierung vorausgesetzt [16], in der mikrobiellen Diagnostik und beim Screening von Medikamentenresistenzen als Alternativen oder komplementierende Verfahren zu etablierten Methoden klinisch getestet.

Diagnostische Mikroarrays in der Infektiologie

Im Wesentlichen gibt es zwei Anwendungsformen für Mikroarrays in der mikrobiologischen Diagnostik:

1. Pathogenspezifische DNA Mikroarrays für den Nachweis und die Identifizierung von Mikroben auf Subspezies-, Spezies- und höherer phylogenetischer Ebene im Hochdurchsatzverfahren und
2. Genexpressions-Mikroarrays infizierter Zellen und Gewebe zur Charakterisierung klinisch signi-

fikanter Stoffwechselveränderungen unter den Bedingungen der Infektion.

Praktisch interessiert bei letzterem vor allem die Definition der Mindestanzahl von Gengruppen mit signifikanten Änderungen des mRNA-Expressionslevels im Hinblick auf eine Korrelation mit dem Infektionsstadium und einer eindeutigen Diagnose.

Genexpressionsstudien dieser Art stehen nicht allein, sondern müssen eigentlich immer im Kontext messbarer Proteinkonzentrationen gesehen werden. Dabei stellt sich heraus, dass mRNA und Proteinkonzentrationen oft unabhängig voneinander mit dem Krankheitsprozess korreliert sind. Mit anderen Worten, das Genprodukt eines zellulären Gens, dessen mRNA-Expression bei Infektion mit einem bestimmten Mikroorganismus signifikant und pathognomonisch hochreguliert ist, muss nicht unbedingt eine signifikante Rolle bei der Pathologie dieser Infektion spielen.

Geignete Parameter könnten kostengünstig unter den Bedingungen des mikrobiologischen Labors untersucht werden. Solche Arrays werden in der Praxis allerdings noch nicht eingesetzt. Dies liegt vor allem an den hohen Kosten der Mikroarray-Technologie, die einen Einsatz in der praktischen Diagnostik zurzeit wenig kosteneffizient machen. Obwohl die Herstellung von Mikroarraychips mittlerweile ein hoch automatisierter Prozess ist, liegen die Kosten für einen Genexpressionschip im Durchschnitt mehrerer Anbieter (Nimblegen, Affymetrix) bei umgerechnet etwa 750 € pro Chip.

Im Gegensatz dazu werden pathogenspezifische Mikroarrays bereits mit einigem Erfolg in der mikrobiologischen Diagnostik angewendet. Ein Beispiel ist der sogenannte ViroChip®.

Entwickelt von Joe DeRisi an der Universität von Kalifornien in San Franzisko in Zusammenarbeit mit NimbleGen, ist der ViroChip® der dritten Generation mit 22.000 Oligonukleotidsequenzen repräsentativ für alle publizierten Virusgenomsequenzen in der GenBank bestückt. Im Laufe der Entwicklung dieses Chip-Arrays stellte sich heraus, dass Oligonukleotidchips einfacher herzustellen sind als cDNA-Chips und bei einer Durchschnittslänge von 70 Nukleotiden (70mers) eine vergleichbare Spezifität aufweisen. Die Spezifität ist hierbei weniger von der Länge des Oligonukleotids, sondern vielmehr von einer minimalen Kreuzreaktivität abhängig, die in Sequenzvergleichs-(BLAST) Analysen mit bekannten menschlichen Genomsequenzen ständig optimiert wird.

Der ViroChip® der dritten Generation wurde 2003 von den US Centers of Disease Control (CDC, Atlanta, Georgia) zur Identifizierung des SARS-Pathogens eingesetzt und identifizierte ein Virus der Familie Coronavirus innerhalb von 24 Stunden nach Zusendung der Untersuchungsproben vom CDC und damit lange

vor den Ergebnissen der mit differenzieller PCR-Technologie arbeitenden Arbeitsgruppen. Hybridisierende virale Gensequenzen konnten vom Mikroarray wiedergewonnen und sequenziert werden, was eine genaue phylogenetische Einordnung des neuen Pathogens erlaubte.

Der ViroChip® ist im Gegensatz zum pathogenspezifischen PCR-Nachweis und ähnlich wie der kulturelle Nachweis mikrobieller Erreger ein offenes Untersuchungsverfahren. Der ViroChip® funktioniert direkt mit Patientenmaterial, wobei sich häufig klinisch interessante Mehrfachinfektionen mit verschiedenen viralen Infektionserregern herausstellen. Ein bakterieller und ein mykologischer Mikroarraychip sind in Entwicklung. Diese Ergebnisse eröffnen das Studium synergistischer Effekte von Infektionen mit mehreren Erregern als neues infektiologisches Arbeitsgebiet. Solche Untersuchungen sind der Gegenstand mehrerer klinischer Studien in den USA.

Genotypische Resistenztestung mit Mikroarrays

Ein Beispiel für den Einsatz von Mikroarraychips bei der Hochdurchsatz-Testung von Medikamentenresistenzen ist der Virochip® der Firma Biotherapix. Die Entwicklung von Medikamentenresistenzen ist ein Hauptmerkmal der HIV/AIDS-Therapie. HIV-Patienten beherbergen eine breitgefächerte Population von HI-Viren (Pseudo-Spezies), von denen einige bereits vor dem Einsatz antiviraler Medikamente spezifische Resistenzmutationen aufweisen. Die folgende Behandlung, auch mit Kombinationstherapie (Hochaktive Antiretrovirale Therapie: HAART), induziert keine Resistenzmutationen, sondern stellt lediglich einen Selektionsdruck auf die Gesamtpopulation dar, der zur vorwiegenden Vermehrung der „resistenten“ Virustypen führt und den Einsatz einer neuen Medikamentenkombination erfordert. Der Virochip® ist ein mikroarraybasiertes Prognosewerkzeug, das dem Kliniker einen Einblick in die genotypische Resistenzlage seines Patienten gibt. Der Virochip® erlaubt schnelle Probenverarbeitung, Identifizierung und Quantifizierung auch seltener Virussubpopulationen bei hoher Sensitivität und eine akurate Genotypisierung. Sein Einsatz soll die antiretrovirale Therapie optimieren und Therapiekosten sparen. Das Spektrum nachweisbarer Genotypen ist schnell an veränderte Resistenztypen anpassbar. Der Virochip® v 4.0 wird zurzeit in klinischen Studien ausgewertet.

DNA-Mikroarrays leisten bereits einen wichtigen Beitrag in der Infektiologie. Hauptproblem bei der Implementierung von Mikroarrays in der Infektiologie sind die hohen Kosten. Trotzdem werden mehrere kommerzielle Produkte zurzeit in umfassenden klinischen Studien getestet.

3. Jenseits des Genoms: Funktionelle Proteomanalysen in der Infektiologie (Proteomics)

Wie perfekt auch immer DNA-Mikroarrays und die dazugehörigen analytischen Werkzeuge werden, wichtige Fragestellungen in der Infektiologie deuten über die Sequenzierung von Genen und die Messung von mRNA-Konzentrationen hinaus. Posttranslationale Veränderungen an mikrobiellen wie auch Wirtspoteinen, sind wichtige Aspekte vieler Erkrankungsprozesse und können nur mit proteinbiochemischen Methoden untersucht werden. Hier sind vor allem komplexe Proteinadress/Verkehrssysteme (trafficking), selektive Interaktionen mit anderen Biomolekülen, Proteinen, Antikörpern, Medikamenten oder verschiedenen kleinen Liganden zu nennen [10].

Diese Fragestellungen können in Analogie zu den DNA-Mikroarrays durch parallele Untersuchung einer großen Zahl von Proteinen gleichzeitig in Protein-Mikroarrays bearbeitet werden.

Dabei sind sogenannte capture-Moleküle an der Festphase gebunden und extrahieren Liganden aus komplexen Proteingemischen in wässriger Lösung, zum Beispiel biologischen Flüssigkeiten. Letztere sind den vorher besprochenen Methoden der Genomsequenzierung und DNA-Mikroarrays vollständig unzugänglich und daher eine Spezialdomäne der Proteomik.

Das Gebiet möglicher Proteinmikroarrays ist wegen seiner Vielfalt an Bindemolekülen besonders unübersichtlich. Als Beispiele seien hier neben rekombinanten Proteinen und Peptiden, die in großem Umfang direkt auf einem Chip produziert werden können [15], auch protein- und andere wirkstoffbindende Nukleinsäuren genannt. Diese capture-Moleküle, Peptide und Nukleinsäuren, werden auch als Aptamere bezeichnet. Kommerziell vertriebene Arrays basieren oft auf der Aptamer-Technologie, wie die Produkte von SomaLogic und Archemix sowie eine Vielzahl anderweitig modifizierter Bindeproteine und Teilmoleküle von Aspira Biosystems (PMI: partial molecule imprinting) und Phyllos (TRINECTIN™ binding proteins). Ein kurzer Exkurs in einige technische Aspekte der Proteomik soll die Vielfalt der Methodik illustrieren.

Technische Aspekte moderner Proteomanalyse

Zwei klassische Verfahren der Proteinbiochemie sind die Standbeine der Proteomik: die zweidimensionale Proteinelektrophorese und verschiedene Spielarten der Massenspektroskopie [2]. Besonders in Kombination eignen sich beide auch für Hochdurchsatz-Untersuchungen. Ergänzt werden diese Standardmethoden durch „Two hybrid-Verfahren“ zur Hochdurchsatz-Evaluation möglicher Protein-Protein-Interaktionen sowie Laser gesteuerte Mikrodissektion zur Re-

duktion der Probenkomplexität. Ein „reverse-phase protein array“-Ansatz erlaubt es, das ganze Repertoire der Proteine eines Gewebes auf einem Proteinmikroarraychip zu immobilisieren. Untersuchungen des Phosphorylierungsstatus von Proteinen verschiedener Signalkaskaden in Subpopulationen menschlicher Gewebezellen werden hierdurch möglich. Phosphorylierung ist eine der möglichen posttranslationalen Proteinmodifikationen, die üblicherweise durch einfache Peptid- oder Antikörper capture-Proteinmikroarrays nicht erfasst werden. Zur Hochdurchsatz-Erfassung posttranslationaler Modifikationen sind moderne Proteinfractionierungsmethoden unabdingbar. Die gelbasierten Methoden, wie 2D PAGE, zur Fraktionierung nach Molekulargewicht, isoelektrischem Punkt und eben posttranslationalen Modifikationen sind trotz vieler Verbesserungen, wie immobilisierte pH-Gradienten (IPGs), differenzielle in-gel Elektrophorese (DIGE) und Präfraktionierung durch „zoom gels“ [4], zu ineffizient und werden zunehmend durch adaptierte zweidimensionale Varianten der Flüssigkeitschromatographie/Kapillarelektrophorese und Kombination mit Hochdurchsatz-Massenspektrometrie (LC/MS/MS and capillary electrophoresis-MS/MS) ersetzt [18]. Reduktion der Probenkomplexität vor der Analyse durch Untersuchung von Proteinen aus spezifischen subzellulären Kompartimenten ist besonders auch zum Nachweis seltener Proteine unabdingbar.

Die Entwicklung von Mikroflüssigkeitssystemen (microfluidics) ist hier wegen der minimalen Proteinmengen notwendig und solche Systeme werden integriert mit Massenspektrometrie zum Proteinverdau und zur anschließenden Identifizierung. In fraktionierungsunabhängigen Ansätzen werden auch Sektionen gefrorener Gewebe direkt auf matrixunterstützte Desorptions-/Ionisations- (MALDI) Platten aufgegeben und Massenspektren so direkt aus der klinischen Probe gewonnen [18].

Als letzte Methode sei hier das Aktivitäts-gesteuerte Screening (activity-based screening, ABS) erwähnt, das mithilfe von activity based probes (ABP) zum Beispiel die Identifizierung von Bindungsstellen für Inhibitoren mehrerer spezifischer Proteasen in einem einzigen Experiment ohne die Notwendigkeit der Herstellung rekombinanter Enzyme erlaubt. ABPs leiten über in das neue Gebiet der chemischen Genetik [18], die sich mit der Identifizierung kleiner hochselektiver Funktionsinhibitoren in zellbasierten Experimenten befasst.

ABS-Untersuchungsmethoden wurden bei der Charakterisierung der Zysteinproteasenaktivität und zur Proteaseinhibitorentwicklung bei *Plasmodium falciparum* eingesetzt. Sie führten zur Identifizierung des Falcipain-1-spezifischen Inhibitors, YA29-Eps(S,S), und der Möglichkeit einer neuen Strategie zur Mala-

riatherapie [1]. Zwei Jahre später stellte sich jedoch in einer Falcipain-1 knock-out Studie von Sijwali und Mitarbeitern heraus, dass das Fehlen von Falcipain-1 keinen Einfluss auf die Merozoiteninvasion von Erythrozyten hat. Die Inhibition von Cysteinproteasen verzögert vielmehr lediglich die Lyse der Erythrozyten durch reife Schizonten [3].

Auf diesem Hintergrund wird klar, dass die infektiologische Proteomik noch einen weiten Weg vor sich hat. Screeningergebnisse müssen durch sorgfältige Experimente überprüft werden, am besten von Mikrobiologen/Parasitologen mit moderner Methodik und genauen Kenntnissen „ihrer“ Mikroorganismen [13, L9].

Es steht aber außer Zweifel, dass die infektiologische Proteomik in der Zukunft eine wichtige Rolle in der Mikrobiologie [5,18], vor allem bei den haploiden Organismen, wie den Protozoen *Toxoplasma gondii* und *P. falciparum*, spielen wird. Die Proteomik erlaubt bei angemessener Empfindlichkeit die Anwendung von Hochdurchsatz-Verfahren. Es wird erwartet, dass die Proteomik einige der Begrenzungen konventioneller mikrobiologischer Ansätze, insbesondere bei der Untersuchung der temporären Regulation der Proteinexpression, nicht nur in ganzen Zellen oder Geweben, sondern auch in subzellulären Strukturen, in Protein-komplexen und insbesondere in biologischen Flüssigkeiten, aufhebt. Die Rolle der Proteomik bei der Untersuchung biologischer Flüssigkeiten mit dem Ziel der Medikamentenentwicklung wird im nächsten Kapitel an einem Beispiel erläutert.

Es ist auch klar, dass jede infektiologische Proteomik abhängig von den Fortschritten bei der Proteomik des menschlichen Genoms sein muss. Ein vollständiger Überblick über Proteinmodifikationen und -interaktionen sowie die subzelluläre Lokalisation menschlicher Proteine ist zurzeit nicht absehbar. Zur Förderung der Proteomik wurde deswegen die „Human Proteome Organisation“ (HUPO) gegründet. Die Ziele von HUPO sind global, transnational und international:

1. Konsolidierung nationaler und regionaler Proteom-Organisationen in eine weltweite Organisation, Ausbildungsinitiativen zur Disseminierung von Proteomtechnologien;
2. Koordinierung öffentlicher Proteom-Initiativen (Serum/Plasmaproteom-, Leberproteomstudie) zur Charakterisierung spezifischer Gewebe und Zellproteome.

Dies bestätigt eine Vorhersage von Chargaff [14]. Alle Initiativen dieser Art haben die Standardisierung der Proteominformation zur Erleichterung translationaler Forschung und hier vor allem der potenziell lukrativen Medikamenten- und Impfstoffentwicklung zum Ziel.

4. Medikamentenentwicklung (Drug design)

Das große Interesse der Pharmaindustrie an der Proteomik zeigt sich in der verstärkten Implementierung von Proteomikprogrammen bei allen größeren Pharmaherstellern. Die Anwendung der Proteomik ist tatsächlich ganz besonders für die industrielle Umgebung mit ihrer logistischen Organisationsform und dem industrietypischen Verfahren des Hochdurchsatz-Screenings geeignet. Da überdies die meisten pharmazeutischen Wirkstoffe Target-Proteine binden, verspricht die Anwendung der Proteomik, zumindest theoretisch, eine überwältigende Flut neuer Target-Proteine und die Möglichkeit, die Wirkmechanismen und mögliche Nebenwirkungen sowohl unter vorklinischen als auch unter klinischen Bedingungen zu untersuchen. Die Proteomik in der Medikamentenentwicklung ist auch Ausdruck eines Trends zur Identifizierung von Faktoren, die die Pathogenese einer mikrobiellen Infektion bestimmen. Das Ziel ist dann durch Auswahl eines geeigneten Inhibitors für das pathophysiologische Target einfach die pathologischen Konsequenzen der Infektion zu verhindern. Die Elimination des Mikroorganismus wird vermieden. Dabei ist der Grundgedanke, dass Therapien, die das Überleben der Mikrobe nicht direkt gefährden, vermutlich weniger häufig zu Medikamentenresistenzen führen. Bei aller Euphorie muss man hier jedoch immer das Beispiel des Falcipain-1-Inhibitors YA29-Eps(S,S) in Erinnerung behalten.

Nachfolgend noch 2 Beispiele erfolgreicherer Wirkstoff-Suchstrategien. Hier ging es um die Identifizierung natürlicher antimikrobieller Substanzen aus Zellkulturüberständen und menschlichen Gewebeflüssigkeiten. Dies ist ein besonders geeignetes Einsatzgebiet der Proteomik: Biologische Flüssigkeiten sind keiner anderen Hochdurchsatz-Methodik zugänglich.

Mehrere HIV-antivirale Peptide, sogenannte Defensine, wurden unter Einsatz von Proteinmikrochips und Massenspektrometrie aus CD8 Zellkulturüberständen von HIV-Longterm-Nonprogressors isoliert [8]. Bei der Suche nach körpereigenen antimikrobiellen Substanzen in einer aus dem Filtrat von 10.000 Litern menschlichen Dialyseplasmas gewonnenen Peptidbibliothek wurde VIRIP (Virus-Inhibitory Peptide), ein C-proximales Subpeptid des humanen α 1-Antitrypsin (α 1-AT) von 20-Aminosäureresten Länge mit einer HIV-spezifischen antiviralen Wirkung gefunden [9]. VIRIP ist wie T20- (Fuzeon®) ein HIV-Fusionsinhibitor. Es bindet hochspezifisch das HIV-Glykoprotein gp41 und reduzierte dosisabhängig die Infektionsrate bei 60-HIV-Isolaten, die zum Großteil resistent gegen RT- und Proteaseinhibitoren waren. Optimierung der Aminosäuresequenz erhöhte die antivirale Wirkung um mehrere Größenordnungen. Allerdings gilt für VIRIP dasselbe wie für T-20: Da der therapeutische

Einsatz von Peptiden schwierig ist, ist die genaue Charakterisierung der gp41 Bindungsstelle und die Entwicklung eines leichter verabreichbaren Wirkstoffs unabdingbar. Trotzdem ist T20 unter dem generischen Namen Fuzeon® bereits auf dem amerikanischen Markt zugelassen.

Es bleibt festzuhalten, dass Peptidbibliotheken menschlicher Körperflüssigkeiten eine Goldmine infektiologisch relevanter Information darstellen. Peptidbibliotheken enthalten über eine Million verschiedener Peptide und kleiner Proteine und repräsentieren alle zirkulierenden Proteine mit einem Molekulargewicht unter 30 kD [17], wie Chemokine, Defensine und Peptidhormone. Darunter befindet sich möglicherweise eine große Zahl von Peptiden mit antimikrobieller Aktivität, besonders gegen latent oder persistierend verlaufende bisher weitgehend unbehandelbare Virusinfektionen, wie Herpesviren, Retroviren, Hepatitis B und C sowie chronisch verlaufende bakterielle und parasitäre Infektionen.

5. Vorsicht Computersimulation: Simulation biologischer Prozesse

Abschließend soll auch noch das Fachgebiet Systembiologie und seine Auswirkungen auf die Infektiologie kurz besprochen werden. Hier handelt es sich im Wesentlichen um die Entwicklung von mathematischen und Ingenieur-Modellen biologischer Prozesse. Der Gedanke ist naheliegend, wenn man die Ähnlichkeit von Signalkaskaden biologischer Regulationsnetzwerke mit entsprechenden Schemata elektronischer oder verfahrenstechnischer Schaltkreise akzeptiert. Wesentliche Methoden stammen aus der interdisziplinären Systemtheorie [7]. Virtuelle Experimente am Computer werden durch Vergleich mit den Ergebnissen realer Laborexperimente optimiert. Der biologische Modell-Organismus (*in vivo*) wird in einem mathematischen Modell (*in silico*) emuliert. Dabei werden biologische Systeme bis hinab auf die molekulare Ebene abgebildet (virtuelle Zelle) und Reaktionen auf äußere Einflüsse sollen simulieren oder sogar vorhergesagt werden. Auch hier muss aber gesagt werden, dass diese Simulationen immer nur so gut sein können wie die experimentellen Daten, auf die sie sich stützen. Allerdings könnte man sich vorstellen, experimentell aufgestellte Hypothesen in einer Simulation zu emulieren. Wenn die Hypothese stimmt, sollten die im Modell vorhergesagten Effekte dann experimentell nachweisbar sein. Solche Untersuchungen werden zurzeit mit großem Erfolg zur Überprüfung von aus NMR-Daten abgeleiteten 3D-Proteinstrukturmodellen eingesetzt (W. v. Gunsteren, Workshop on computer simulations of soft matter and biosystems, 14.–16. März 2007, KIP, Heidelberg). Auf der Organisationsebene sind das europäische Exzellenznetzwerk ENFIN und BIOSM, das erste deut-

sche Zentrum für Modellierung und Simulation in den Biowissenschaften, Musterbeispiele in diesem Bereich. Die *in silico*-Systembiologie wird an der Universität Heidelberg im Rahmen des Forschungsnetzes „Bioquant“ besonders gefördert [L10]. Ziele der Systembiologie sind die Suche nach wirksamen Medikamenten mit geringen Nebenwirkungen, Reduktion von Tierversuchen und die schnellere Entwicklung kostengünstiger Therapien.

Zusammenfassung und Ausblick

Angesichts der schwierigen Frage bakterieller Antibiotikaresistenzen, sich zusehends entwickelnder viraler Medikamentenresistenzen und der Ratlosigkeit behandelnder Ärzte „ohne Behandlungsoption“ ist die Zahl neuer Antibiotika, Antiparasitika und antiviraler Substanzen sowie wirksamer Impfstoffe in Entwicklung immer noch erschreckend klein; bahnbrechende Alternativen sind nicht in Sicht.

Die Ursachen des infektiologischen Dilemmas, des Wettrüstens zwischen Mensch und Mikrobe, liegen offensichtlich in der Genetik: Ständige Veränderung, schier unendliche Vielfalt und „survival of the fittest“.

Genetisch und eskalierend ist deswegen auch der angebotene Ausweg aus dem Dilemma. Genomsequenzierung im industriellen Maßstab hat in den Neunzigerjahren des zwanzigsten Jahrhunderts die „Hochdurchsatz-Revolution“ (Revolution des High-throughput-screening) eingeleitet. Die neue Forschungsrichtung der Systembiologie versucht im einundzwanzigsten Jahrhundert Genomanalyse, Mikroarrays und Proteomik zu integrieren und mit Rechnernetzwerken (gigabit applications, grid-computing) und Systemtheorie der Datenflut Herr zu werden. Systembiologie soll mikrobielle Diagnostik und Medikamentenentwicklung revolutionieren. Das Ziel: Biologische Prozesse werden ganzheitlich am Computer simuliert und neue Medikamente werden *in silico* entwickelt. Die Systembiologie wird zurzeit weltweit mit über einer Milliarde Euro gefördert und ist damit das am schnellsten wachsende Forschungsgebiet der Lebenswissenschaften.

Es bleibt abzuwarten, wie sich die Konzentration gewaltiger Geldmittel in wenigen, nach industriellen Gesichtspunkten arbeitenden, „Biodiversity“-Systembiologie-Zentren auf die Vielfalt wissenschaftlicher Arbeit an den Universitäten auswirken wird, und ob diese Art der Forschung in Zeiten knapper werdender öffentlicher Mittel den infektiologischen Herausforderungen der Gegenwart und der Zukunft gerecht werden kann (14).

Weblinks

1. Department of Energy-Genomics: <http://genomics.energy.gov/>

2. TIGR Microbial Database : <http://www.tigr.org/tdb/mdb/>
3. Genome News Network Guide to genomes: http://www.genomenewsnetwork.org/resources/sequenced_genomes/genome_guide_p1.shtml
4. National Center for Biotechnology Information (BLAST): <http://www.ncbi.nlm.nih.gov/BLAST>
5. Institute Pasteur: <http://www.pasteur.fr/english.html>
6. Welcome Trust Sanger Institute: <http://www.sanger.ac.uk/>
7. ORPHEUS: <http://kulan.gsf.de/orpheus/index.html>
8. GLIMMER: www.tigr.org/software/glimmer
9. Proteome database at Max Planck Institute for Infection Biology: <http://www.mpiib-berlin.mpg.de/2D-PAGE/>
10. BIOQUANT: <http://www.bioquant.uni-hd.de/>
8. Zhang L et al. (2002) Contribution of human α -defensin 1, 2 and 3 to the anti-HIV-1 activity of CD8 antiviral factor. *Science* 298:995–1000
9. Münch J, Ständker L, Adermann K et al. (2007) Discovery and Optimization of a Natural HIV-1 Entry Inhibitor Targeting the gp41 Fusion Peptide. *Cell* 129:263–275
10. Hanash S (2003) Disease proteomics. *Nature* 422:226–232
11. Diruggiero J et al. (2000) Evidence of recent lateral gene transfer among hyperthermophilic archaea. *Mol Microbiol* 38:684–693
12. Hutchison CA et al. (1999) Global transposon mutagenesis and a minimal Mycoplasma genome. *Science* 286:2165–2169
13. Haas G et al. (2002) Immunoproteomics of Helicobacter pylori infection and relation to gastric disease. *Proteomics* 2:313–324
14. Chargaff E (1980) In praise of smallness - can we return to small science? *Perspectives in Biology and Medicine* 23:37
15. Pellois JP et al. (2002) Individually addressable parallel peptide synthesis on microchips. *Nature Biotechnol* 20:922–926
16. Opinion (2002) Microarray standards at last. *Nature* 419:323
17. Schulz-Knappe P, Schrader M, Standker L et al. (1997) Peptide bank generated by large-scale preparation of circulating human peptides. *Journal of Chromatography A* 776:125–132
18. Phillips CI, Bogyo M (2005) Proteomics meets microbiology: technical advances in the global mapping of protein expression and function. *Cell Microbiol* 7:1061–1076
19. Martin W, Dagan T, Koonin EV, Dippio JL, Gogarten JP, Lake JA (2007) The evolution of eukaryotes. *Science* 316:542–543
20. Yoosof S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, et al. (2007) The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families. *PLoS Biol* 5:e16

Schlüsselliteratur

1. Greenbaum DC, Baruch A, Grainger M, Bozdech Z, Medzihradsky KF, Engel J, DeRisi J, Holder AA, Bogyo M (2002) A role for the protease falcipain 1 in host cell invasion by the human malaria parasite. *Science* 298:2002–2006
2. Lasonder E et al. (2002) Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry. *Nature* 419:537–542
3. Sijwali PS, Kato K, Seydel KB, Gut J, Lehman J, Klemba M, Goldberg DE, Miller LH, Rosenthal PJ (2004) Plasmodium falciparum cysteine protease falcipain-1 is not essential in erythrocytic stage malaria parasites. *Proc Natl Acad Sci USA* 101:8721–8726
4. Nilsson CL (2002) Bacterial proteomics and vaccine development. *Am J Pharmacogenomics* 2:59–65
5. Petricoin EF, Zoon KC, Kohn EC, Barrett JC, Liotta LA (2002) Clinical proteomics: translating benchside promise into bedside reality. *Nature Rev Drug Discov* 1:683–695
6. Frangeul L, Nelson KE, Buchrieser C, Danchin A, Glaser P, Kunst F (1999) Cloning and assembly strategies in bacterial genome projects. *Microbiology* 145:2625–2634
7. Mauch K, Buziol S, Schmid JW, Reuss M (2002) Compu-

Hirnabszess

- ▶ Actinobacillus
- ▶ Bacteroides
- ▶ Bilophila
- ▶ Citrobacter
- ▶ Eikenella
- ▶ Nocardia
- ▶ Pasteurella multocida
- ▶ Phaeohyphomycetes
- ▶ Porphyromonas

Histoplasma capsulatum

MARIANNE KRETSCHMAR

Erreger

Synonym(e)

Cryptococcus capsulatus, *Posadasia capsulate*, *Torulopsis capsulatus*, *Histoplasma pyriforme*, *Histoplasma farciminosum*, *Histoplasma duboisii*

Erregerspezies

Zur Gattung *Histoplasma* gehört die Spezies *Histoplasma capsulatum*. *H. capsulatum* hat zwei Varietäten: *H. capsulatum* var. *capsulatum* und *H. capsulatum* var. *duboisii*.