

Verstehen, wie ein Genom funktioniert

6

Wenn Sie Kapitel 6 gelesen haben, sollten Sie folgende Aufgaben lösen können:

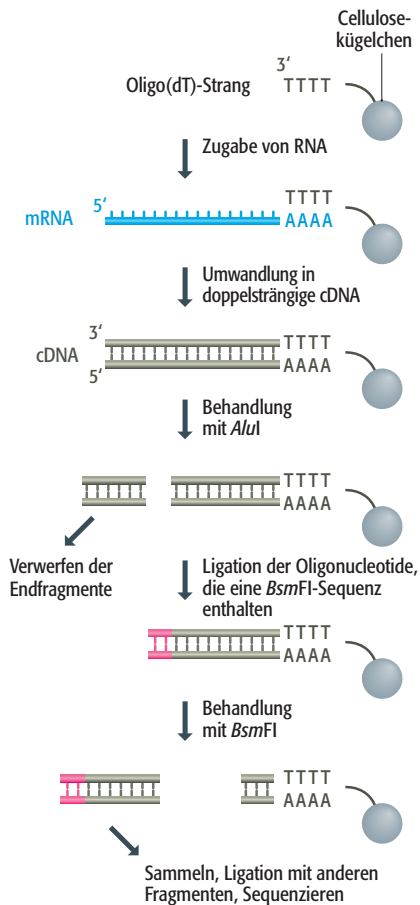
- Beschreiben Sie, wie sich ein Transkriptom durch Analyse der cDNA-Sequenzen untersuchen lässt.
- Beurteilen Sie Stärken und Schwächen der Microarray- und Chip-Technologie für die Untersuchung von Transkriptomen, und erläutern Sie, wie Expressionsmuster, die man durch diese Verfahren erhält, miteinander verglichen werden.
- Nennen Sie Beispiele dafür, wie Transkriptomanalysen zu unserem Verständnis der Hefebiologie und der Krebserkrankungen beim Menschen beigetragen haben.
- Unterscheiden Sie zwischen den unterschiedlichen Formen der Information, die man durch Transkriptom- und Proteomanalyse erhält.
- Beschreiben Sie, wie ein Protein-Profilung durchgeführt wird.
- Stellen Sie die Methoden gegenüber, mit denen man Paare und Gruppen von Proteinen, die miteinander in lebenden Zellen interagieren, ausfindig machen kann, und vergleichen Sie sie. Unterscheiden Sie insbesondere zwischen Verfahren, die direkte Interaktionen erkennen und solchen, mit denen sich funktionelle Wechselwirkungen identifizieren lassen.
- Nennen Sie Beispiele für Proteininteraktionskarten und erörtern Sie wichtige Eigenschaften.
- Erläutern Sie die Grundlagen und die Bedeutung des biochemischen Profiling.
- Beschreiben Sie in Grundzügen die Prinzipien und Ziele der Systembiologie.

In dem vorherigen Kapitel haben wir erfahren, wie man unterschiedliche computergestützte und experimentelle Verfahren einsetzen kann, um Genen, die man in einer genomischen Sequenz ausfindig gemacht hat, eine Funktion zuzuweisen. Und wir haben gelernt, wie die Anwendung dieser Methoden auf das Genom von *Saccharomyces cerevisiae* die Zahl der Hefegene, für die eine eindeutige Funktion bekannt ist, nahezu verdoppelt hat. Diese Form der Genomannotierung ist ein groß angelegtes Unterfangen, doch selbst wenn man jedes Gen eines Genoms identifizieren und ihm eine Funktion zuweisen kann, bleibt immer noch ein Problem zu bewältigen, und zwar zu verstehen, wie ein Genom als Ganzes in der Zelle agiert, wie es die Vielfalt der stattfindenden biochemischen Aktivitäten bestimmt und koordiniert. Diese allgemeinen Untersuchungen der Genomaktivität müssen sich nicht unbedingt auf das Genom selbst beschränken, sondern können sich auch auf das Transkriptom und das

[6.1 Untersuchung des Transkriptoms](#)

[6.2 Untersuchung des Proteoms](#)

[6.3 Jenseits des Proteoms](#)



6.1 SAGE. In diesem Beispiel ist *AluI* das erste verwendete Restriktionsenzym, das die 4 bp lange Sequenz 5'-AGCT-3' erkennt. Das an die cDNA ligierte Oligonucleotid enthält eine Erkennungssequenz für *BsmFI*, das 10–14 Nucleotide stromabwärts schneidet und so ein cDNA-Fragment abspaltet. Man ligiert die Fragmente unterschiedlicher cDNAs und es entsteht ein Concatemer, das anschließend sequenziert wird. Wendet man diese Methode an, dann besteht das entstehende Concatemer teilweise aus Sequenzen, die aus den *BsmFI*-Oligonucleotiden stammen. Um dieses zu vermeiden und ein Concatemer zu erhalten, das ausschließlich cDNA-Fragmente enthält, kann man die Oligonucleotide so konstruieren, dass das an die cDNA ligierte Ende eine Erkennungssequenz für ein drittes Restriktionsenzym enthält. Die Behandlung mit diesem Enzym entfernt die Oligonucleotide von dem cDNA-Fragment.

Proteom beziehen, die von dem Genom synthetisiert und aufrechterhalten werden. Ebenfalls muss geklärt werden, wie Transkriptom und Proteom die definitiv letzte Ebene der Genomexpression – das Netzwerk der miteinander verknüpften biochemischen Wege und Prozesse, die eine lebende Zelle ausmachen – etablieren und koordinieren. In diesem Kapitel werden wir uns die Methoden ansehen, die für diese allgemeinen Untersuchungen der Genomaktivität eingesetzt werden.

6.1 Untersuchung des Transkriptoms

Das Transkriptom besteht aus den mRNAs, die zu einem bestimmten Zeitpunkt in einer Zelle vorhanden sind. Die Zusammensetzung der Transkriptome kann hoch komplex sein, mit Hunderten oder Tausenden von unterschiedlichen mRNAs, von denen jede einen Teil der gesamten Population ausmacht (Abschnitt 1.2.4). Um ein Transkriptom zu analysieren muss man daher die enthaltenen mRNAs identifizieren und, im Idealfall, ihre relativen Häufigkeiten ermitteln.

6.1.1 Die Untersuchung des Transkriptoms durch Sequenzanalyse

Die direkteste Art und Weise ein Transkriptom zu charakterisieren ist, seine mRNA in cDNA zu überführen (Abb. 3.36) und dann jeden Klon der entstehenden cDNA-Bibliothek zu sequenzieren. Gene, deren mRNAs im Transkriptom vorhanden sind, werden durch Vergleiche zwischen cDNA-Sequenzen und den genomischen Sequenzen identifiziert. Dieser Ansatz ist leicht durchzuführen, doch er ist arbeitsintensiv, da viele verschiedene cDNA-Sequenzen benötigt werden, bevor ein nahezu vollständiges Bild von der Transkriptomzusammensetzung entsteht. Werden zwei oder mehrere Transkriptome miteinander verglichen, dann vergeht bis zum Abschluss des Projektes noch mehr Zeit. Lässt sich das Verfahren in irgendeiner Art und Weise abkürzen, um schneller an die entscheidende Sequenzinformation zu gelangen?

Eine Lösung ist die **serielle Analyse der Genexpression** (*serial analysis of gene expression*, **SAGE**). Anstatt vollständige cDNAs zu untersuchen, ergibt SAGE kurze Sequenzen mit einer Länge von 12 bp, von denen jede eine mRNA des Transkriptoms repräsentiert. Die Methode beruht darauf, dass diese 12-bp-Sequenzen trotz ihrer Kürze ausreichen, um das mRNA-codierende Gen zu identifizieren. Die Begründung ist, dass jede einzelne Sequenz aus 12 bp in dem Genom einmal alle $4^{12} = 16\,777\,216$ bp vorkommen sollte. Die durchschnittliche Länge einer eukaryotischen mRNA ist etwa 1 500 bp; so entsprechen 4^{12} bp einer aufsummierten Länge von über 11 000 Transkripten. Diese Zahl ist größer als die Zahl der Transkripte, die man für nahezu alle komplexen Genome erwartet. Daher sollte es möglich sein, über das 12 bp lange, so genannte Sequenz-*tag* die Gene zu identifizieren, die die vorhandenen mRNAs codieren.

Das Verfahren, mit dem die 12-bp-*tags* hergestellt werden, ist in Abbildung 6.1 dargestellt. Zuerst wird die mRNA in einer Chromatographiesäule immobilisiert, indem man den Poly(A)-Schwanz vom 3'-Ende dieser Moleküle an Oligo(dT)-Stränge binden lässt, die wiederum an Cellulosekügelchen fixiert sind. Die mRNA wird in doppelsträngige cDNA umgeschrieben und dann mit einem Restriktionsenzym behandelt, das eine 4 bp lange Erkennungssequenz besitzt und daher jede cDNA häufig schneidet. Das terminale Restriktionsfrag-

ment jeder cDNA bleibt an die Cellulosekügelchen gebunden, sodass alle anderen Fragmente von der Säule gewaschen und verworfen werden können. Nun wird ein kurzes Oligonucleotid, das eine Erkennungssequenz für *BsmFI* besitzt, an das freie Ende jeder cDNA gebunden. *BsmFI* ist ein unübliches Restriktionsenzym, das die DNA nicht innerhalb der Erkennungssequenz spaltet, sondern 10–14 Nucleotide stromabwärts davon. Die Behandlung mit *BsmFI* entfernt daher ein Fragment mit einer durchschnittlichen Länge von 12 bp vom Ende jeder cDNA. Die Fragmente werden gesammelt, miteinander „Kopf an Schwanz“ zu einem so genannten Concatemer ligiert und sequenziert. Die einzelnen *tag*-Sequenzen in einem Concatemer werden ermittelt und mit den Gensequenzen des Genoms verglichen.

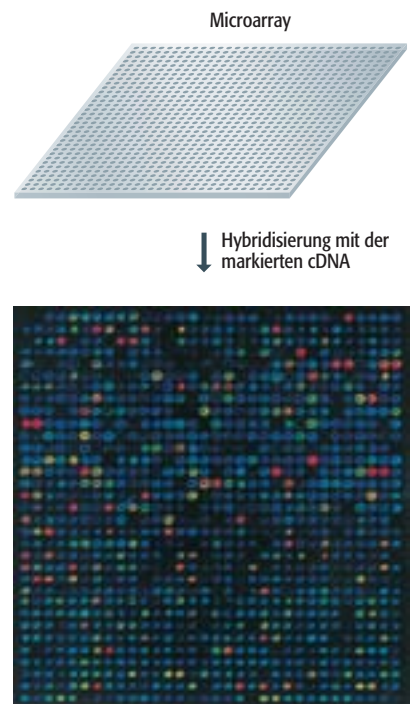
6.1.2 Untersuchung eines Transkriptoms durch Microarray- oder Chip-Analyse

DNA-Chips und Microarrays (Methoden 3.1) können auch für Transkriptomanalysen eingesetzt werden. Es sei daran erinnert, dass der Unterschied zwischen beiden darin besteht, dass ein Chip eine Sammlung von immobilisierten Oligonucleotiden trägt, die *in situ* auf der Oberfläche des Glas- oder Silikonträgers synthetisiert werden, und dass Microarrays aus DNA-Molekülen – in der Regel PCR-Produkte oder cDNAs – bestehen, die auf die Oberfläche des Glasobjektträgers oder der Nylonmembran aufgetragen werden. Microarrays und Chips werden beide in derselben Art und Weise verwendet (Abb. 6.2). Die Population von mRNAs, die ein Transkriptom ausmacht, wird in ein Gemisch von cDNAs umgewandelt, die dann (in der Regel mit einem Fluoreszenzmarker) markiert und mit dem Microarray oder Chip hybridisiert werden. Anschließend ermittelt man die Positionen, an denen eine Hybridisierung stattgefunden hat. Verglichen mit SAGE hat dieser Ansatz den Vorteil, dass die Unterschiede zwischen zwei oder mehreren Transkriptomen schnell beurteilt werden können, indem man verschiedene cDNA-Präparationen mit identischen Arrays hybridisiert und die Hybridisierungsmuster miteinander vergleicht. Eine Verfeinerung lässt sich erzielen, wenn man den Array mit cDNA hybridisiert, die von der mRNA-Fraktion stammt, die in den untersuchten Zellen an die Ribosomen gebunden ist, und nicht nur mit cDNA aus Gesamt-mRNA. Die gebundenen mRNAs entsprechen dem Teil des Transkriptoms, der aktiv in Protein translatiert wird, wodurch ein etwas anderes Bild der Genomaktivität entsteht.

Als Erstes werden wir die technischen Fragen behandeln, die bei Microarray- und Chip-Analysen auftreten, und anschließend einige Anwendungen dieser Analyseverfahren kennen lernen.

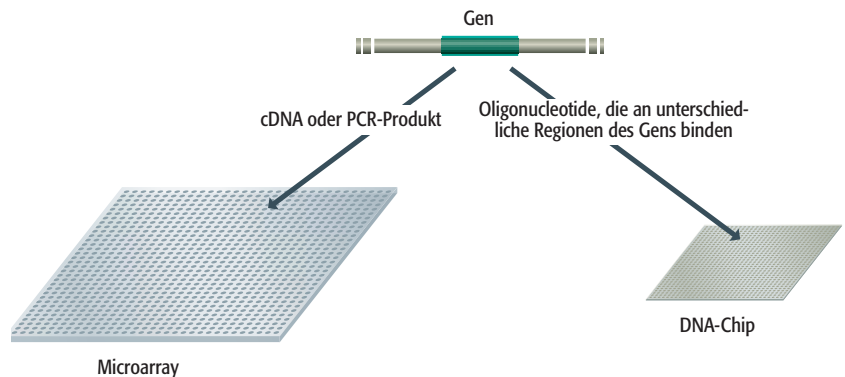
Der Einsatz von Microarray oder Chip für die Analyse von einem oder mehreren Transkriptomen

Bei der Transkriptomanalyse sind die beiden Hauptziele die Identifikation von Genen, deren mRNA vorhanden ist, und die Bestimmung der relativen Mengen dieser unterschiedlichen mRNAs. Um das erste Ziel erreichen zu können, muss jedes relevante Gen durch mindestens eine Probe im Array vertreten sein. Bei einem Microarray erreicht man dies durch die Verwendung von PCR-Produkten oder cDNAs, die sich von den gewünschten Genen ableiten, und bei einem DNA-Chip wird an jeder Position ein Gemisch von Oligonucleotiden synthetisiert, insgesamt etwa 20 unterschiedliche, deren Sequenzen unterschiedlichen Regionen des relevanten Gens entsprechen (Abb. 6.3). Das zweite



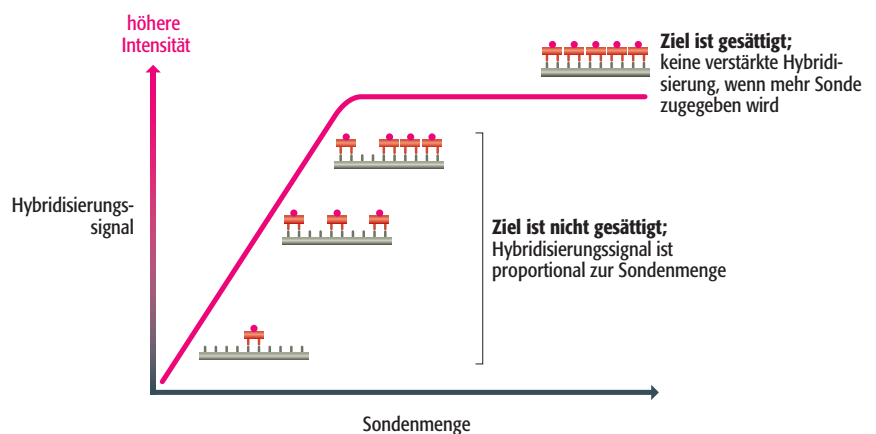
6.2 Microarray-Analyse. Eine cDNA-Präparation wird mit einem Fluoreszenzmarker markiert und mit einem Microarray hybridisiert. Die Markierung wird mit konfokaler Laser-Scanning-Mikroskopie nachgewiesen und die Signalintensität in ein Falschfarbenbild umgewandelt. Bei dieser Art der Darstellung gibt Rot die stärkste Hybridisierung an, gefolgt von Orange, Gelb, Grün, Blau, Indigo und Violett, welches das Hintergrundsignal der Hybridisierung anzeigt. Für mehr Information über die Herstellung und die Anwendung von Microarrays siehe Methoden 3.1. Mit freundlicher Genehmigung von Macmillan Publishers Ltd[©] Nature.

6.3 Microarrays und DNA-Chips. Jede Position in einem Microarray enthält eine cDNA oder ein PCR-Produkt eines gewünschten Gens. Dagegen enthält jede Position auf einem DNA-Chip ein Gemisch von Oligonucleotiden, deren Sequenzen zu unterschiedlichen Segmenten der relevanten Gene passen.

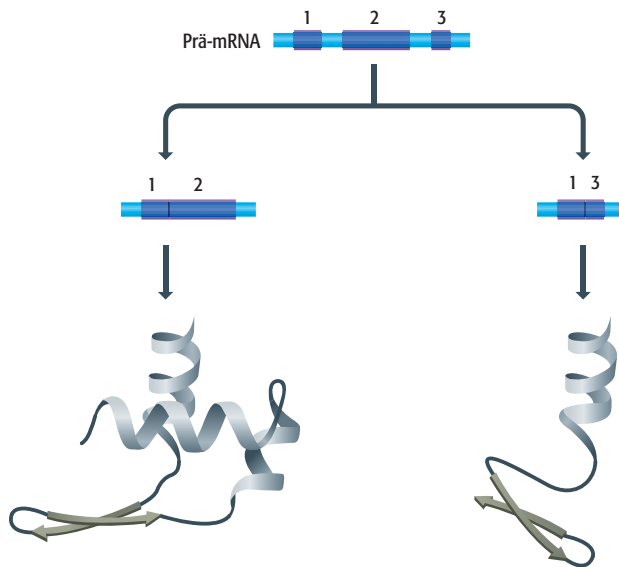


Ziel – die Bestimmung der relativen Mengen der einzelnen mRNAs im Transkriptom – erreicht man, weil jede Position auf dem Microarray oder Chip bis zu 10^9 Kopien des Zielmoleküls enthält. Diese Zahl ist höher als die erwartete Kopienzahl für jede mRNA, die in der geringen Transkriptommenge enthalten ist, mit der man den Array hybridisiert. Daher wird keine Position jemals abgesättigt, sondern jedes Sondenmolekül kann mit seinem Zielmolekül eine Basenpaarung eingehen. Das Ausmaß der Hybridisierung ist daher variabel und die Signalintensität hängt an jeder Position von der Menge jeder einzelnen mRNA im Transkriptom ab (Abb. 6.4).

Der obige Abschnitt vermittelt den Eindruck, Microarray- und Chip-Analyse seien einfache Methoden. In der Praxis existieren jedoch einige Schwierigkeiten. Die erste ist, dass die Hybridisierungsanalyse bei fast allen Transkriptomen, bis auf die einfachsten, nicht ausreichend spezifisch ist, um zwischen allen vorhandenen mRNAs zu unterscheiden. Der Grund liegt darin, dass zwei verschiedene mRNAs ähnliche Sequenzen besitzen und daher mit der spezifischen Zielsequenz der jeweils anderen mRNA auf dem Array kreuzhybridisieren können. Bei zwei oder mehreren paralogen Genen (Abschnitt 5.2.1), die in demselben Gewebe aktiv sind, geschieht dies relativ häufig. Das Transkriptom enthält dann eine Gruppe von verwandten mRNAs, von denen jede bis zu einem gewissen Ausmaß mit den Mitgliedern der Genfamilie hybridisiert. Die Unterscheidung der relativen Mengen jeder mRNA oder selbst der sichere Nachweis bestimmter mRNAs, kann in einem solchen Fall sehr schwie-



6.4 Die Beziehung zwischen Hybridisierungsstärke und der Menge des jeweiligen Sondenmoleküls.

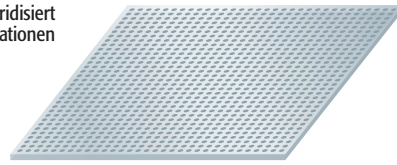


6.5 Alternatives Spleißen. Beim alternativen Spleißen werden Exons in unterschiedlichen Kombinationen miteinander verbunden. Daher werden von der gleichen Prä-mRNA unterschiedliche Proteine synthetisiert.

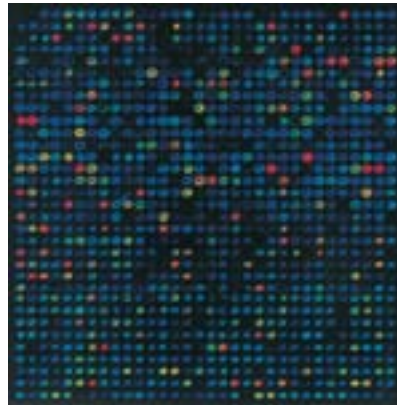
rig sein. Ein ähnliches Problem tritt auf, wenn zwei oder mehrere unterschiedliche mRNAs von demselben Gen abstammen. Dieses ist Vertebraten relativ verbreitet, weil **alternatives Spleißen** stattfindet. Bei diesem Prozess werden Exons von einer Prä-mRNA in unterschiedlichen Kombinationen zusammengesetzt und ergeben dann eine Reihe von verwandten, aber dennoch unterschiedlichen mRNAs (Abb. 6.5). Der Array muss sorgfältig angelegt werden, wenn diese Varianten nachgewiesen und genau quantifiziert werden sollen.

Der Vergleich von zwei oder mehreren Transkriptomen – wie wir unten sehen werden eine häufige Fragestellung – wirft weitere Probleme auf. Damit ein Vergleich auch aussagekräftig ist, müssen die Hybridisierungsintensitäten für dasselbe Gen bei zwei unterschiedlichen Microarrays oder Chips die tatsächlichen Unterschiede in der mRNA-Menge repräsentieren und dürfen nicht von experimentellen Faktoren wie der Menge der Ziel-DNA auf dem Array, der Effizienz der Sondenmarkierung oder der Effektivität des Hybridisierungsprozesses abhängen. Doch selbst in nur einem einzigen Labor können diese Faktoren kaum mit absoluter Genauigkeit kontrolliert werden, und die exakte Reproduzierbarkeit zwischen unterschiedlichen Labors ist nahezu unmöglich. Daher sind bei der Datenanalyse Normalisierungsverfahren notwendig, damit man die Ergebnisse unterschiedlicher Array-Experimente genau miteinander vergleichen kann. Die Arrays enthalten daher sowohl Negativkontrollen, mit deren Hilfe man in jedem Experiment die Hintergrundhybridisierung bestimmen kann, als auch Positivkontrollen, die stets identische Signale geben sollten. Für die Transkriptome von Vertebraten wird häufig das Actin als Positivkontrolle eingesetzt, da seine Expressionsstärke in einem bestimmten Gewebe, unabhängig von Entwicklungsstadium oder Krankheitszustand, relativ konstant ist. Eine etwas zufriedenstellendere Alternative ist die Durchführung eines Experiments, mit dem die beiden Transkriptome direkt, in einer einzigen Analyse und mit einem einzigen Array, miteinander verglichen werden können. Dazu kennzeichnet man die cDNA-Präparationen mit unterschiedlich fluoreszierenden Markern und analysiert den Array anschließend bei den entsprechenden Wellenlängen, um die relativen Intensitäten der beiden Fluo-

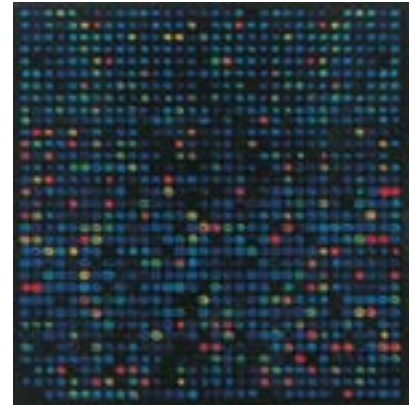
Microarray, hybridisiert mit zwei cDNA-Präparationen



Scan mit unterschiedlichen Wellenlängen



Transkriptom 1



Transkriptom 2

6.6 Vergleich von zwei Transkriptomen in einem einzigen Experiment. Mit freundlicher Genehmigung von Macmillan Publishers Ltd[®] *Nature*.

reszenzsignale an jeder Position und somit die Unterschiede zwischen den mRNA-Gehalten in den beiden Transkriptomen zu bestimmen (Abb. 6.6).

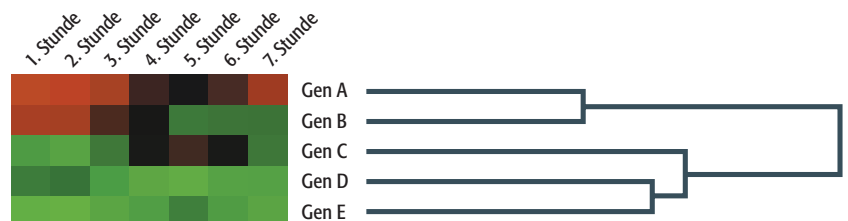
Vorausgesetzt zwei oder mehrere Transkriptome können genau miteinander verglichen werden, dann sind die festgestellten Unterschiede im Genexpressionsmuster relativ komplex. Gene mit ähnlichen Expressionsprofilen haben vermutlich miteinander in Beziehung stehende Funktionen und es sind stringente Verfahren notwendig, um diese Gruppen zu unterscheiden. Das Standardverfahren wird als hierarchische Cluster-Bildung bezeichnet und umfasst einen Vergleich der Expressionsstärken von jedem Genpaar in jedem der analysierten Transkriptome. Man weist jedem Paar einen Wert zu, der das Ausmaß der Beziehung zwischen diesen Expressionsniveaus beschreibt. Diese Daten können in einem Dendrogramm dargestellt werden, in dem Gene mit verwandten Expressionsprofilen zu Clustern zusammengefasst werden (Abb. 6.7). Das Dendrogramm liefert deutlich sichtbare Hinweise für eine funktionelle Beziehung zwischen Genen.

Untersuchungen des Hefetranskriptoms

Mit etwas über 6 000 Genen ist die Hefe *Saccharomyces cerevisiae* für Untersuchungen des Transkriptoms sehr gut geeignet und an diesem Organismus wurde viel Pionierarbeit geleistet. Eine der ersten Entdeckungen war, dass sich

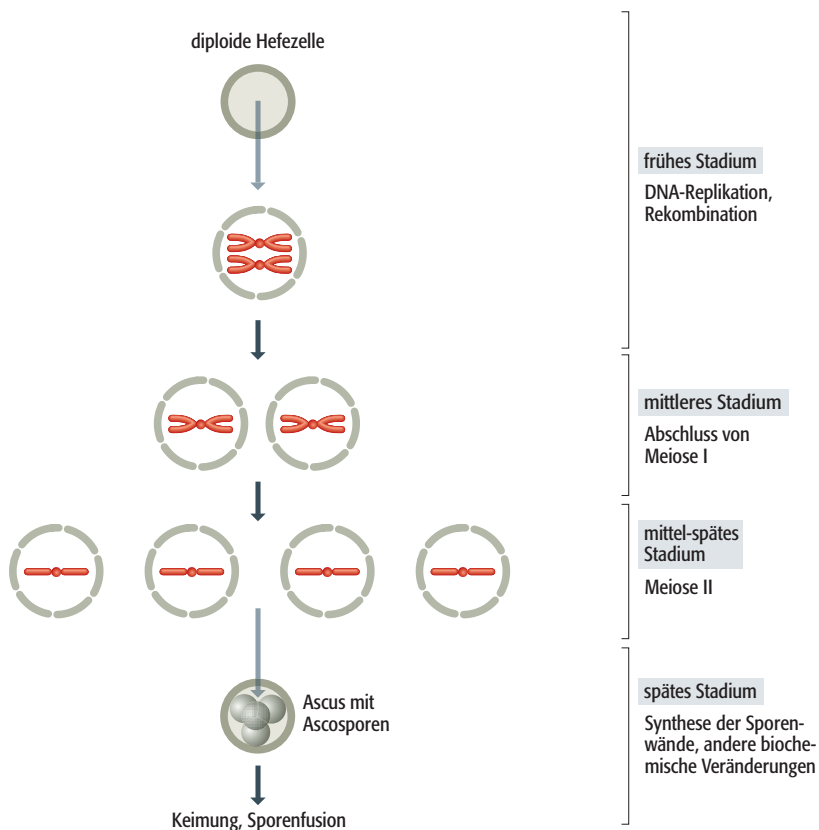
6.7 Vergleich der Expressionsprofile von fünf Genen in sieben Transkriptomen.

Aus Zellen wurden zu unterschiedlichen Zeitpunkten nach Zugabe von energiereichen Nährstoffen zum Wachstumsmedium sieben Transkriptome präpariert. Nach der Datenanalyse durch hierarchische Cluster-Bildung erstellte man ein Dendrogramm, das das Ausmaß der Beziehung zwischen den Expressionsprofilen von fünf Genen darstellt.



die Zusammensetzung des Hefetranskriptoms unter konstanten biochemischen Bedingungen relativ wenig verändert, obwohl mRNAs ständig abgebaut und neu synthetisiert werden. Wächst Hefe in glucosehaltigem Medium und teilen sich die Zellen dadurch mit einer maximalen Rate, dann ist das Transkriptom fast völlig stabil. Nur die Häufigkeit von 19 mRNAs nimmt über einen Zeitraum von zwei Stunden um mehr als das Zweifache zu. Das Transkriptom verändert sich nur dann signifikant, wenn die Glucose im Wachstumsmedium aufgebraucht ist, wodurch die Zellen gezwungen werden zwischen aerober Atmung und anaerober Gärung zu wechseln und zu einer anderen Kohlenstoffquelle überzugehen. Während dieser Veränderung steigt die Menge von über 700 mRNAs um den Faktor zwei oder mehr, und die Menge von noch einmal 1 000 mRNAs nimmt auf weniger als die Hälfte ihres ursprünglichen Niveaus ab. Die veränderten äußeren Bedingungen führen zu einer Umstrukturierung des Transkriptoms und so zu einer Anpassung an die neuen biochemischen Bedürfnisse der Zellen.

Das Hefetranskriptom durchläuft auch während der Zelldifferenzierung eine Umstrukturierung. Dieses wurde bei Untersuchungen der Sporulation (Sporenbildung) entdeckt, die durch Nährstoffmangel und andere umweltbedingte Stressfaktoren eingeleitet wird. Der Ablauf der Sporulation wird anhand der morphologischen und biochemischen Vorgänge in vier Phasen unterteilt – die frühe, mittlere, mittelspäte und späte Phase (Abb. 6.8). Frühere Untersuchungen haben, wie erwartet, gezeigt, dass jede Phase durch die Expression unterschiedlicher Gruppen von Genen charakterisiert ist. Transkriptomanalysen ergänzten unser Verständnis der Sporulation auf unterschiedliche Art



6.8 Die Sporulation von *Saccharomyces cerevisiae*. Die mittleren drei Abbildungen zeigen die Kernteilungen, die während der Sporulation stattfinden. Siehe Abbildung 3.16 für weitere Einzelheiten der Vorgänge bei Meiose I und Meiose II.

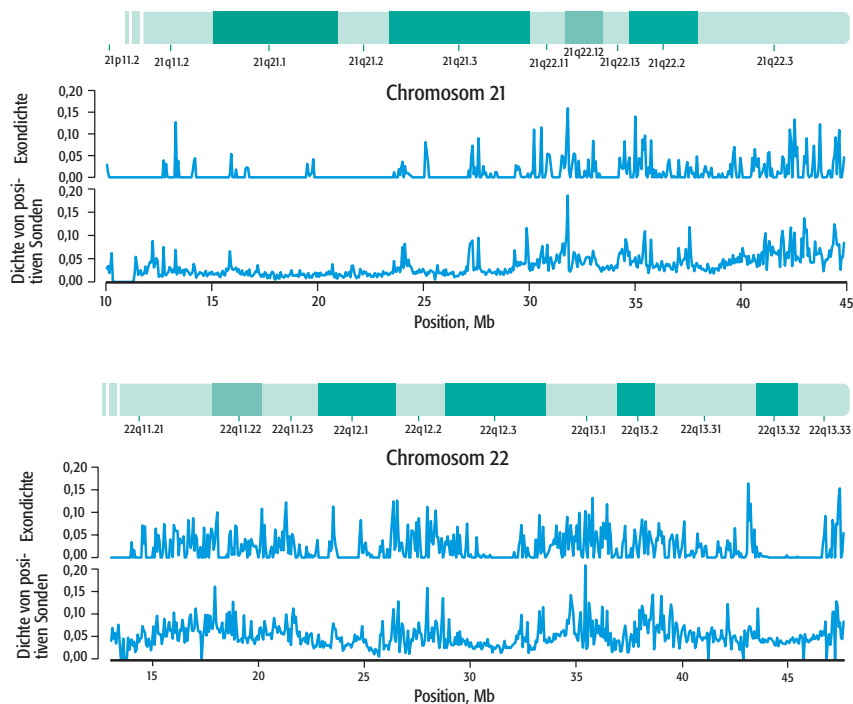
und Weise. Höchst bemerkenswert ist, dass die Veränderungen der Transkriptomzusammensetzung eine Einteilung des frühen Stadiums der Sporulation in drei unterschiedliche Phasen zeigen, die als frühe (I), frühe (II) und früh-mittlere Phase bezeichnet werden. Während der frühen Sporulation steigt die Menge von über 250 mRNAs signifikant an und die von weiteren 158 mRNAs nimmt spezifisch während der mittleren Phase zu. Bei noch einmal 61 mRNAs erhöht sich die Häufigkeit während der mittel-späten Phase und weitere 5 mRNAs kommen in der späten hinzu. Außerdem nimmt die Häufigkeit von 600 mRNAs, die vermutlich für das vegetative Wachstum notwendige Proteine codieren, deren Synthese jedoch während der Sporenbildung abgeschaltet wird, während der Sporulation ab.

Die Erforschung der Hefesporulation ist aus zwei Gründen wichtig. Erstens ebnet Transkriptomanalysen den Weg für Untersuchungen der Wechselwirkungen zwischen Genom und Umweltsignalen, die die Sporulation auslösen, indem sich mit solchen Analysen Veränderungen der Genomexpression während der Sporulation beschreiben lassen. Untersuchungen dieser Art, durchgeführt an einem relativ einfachen Organismus wie der Hefe, dienen als wichtiges Modell für komplexere Entwicklungsprozesse in höheren Eukaryoten, einschließlich dem Menschen. Zweitens sind einige der mRNAs, deren Häufigkeit während der Sporulation signifikant ansteigt, Transkripte von zuvor unbekannt Genen. Die Transkriptomanalysen sind daher für die Annotierung einer genomischen Sequenz hilfreich und unterstützen die Identifizierung von Genen, deren Funktion im Genom bisher durch andere Verfahren nicht ermittelt werden konnte.

Das Transkriptom des Menschen

Mit fünfmal so vielen Genen ist das menschliche Genom erheblich komplexer als das der Hefe, und Untersuchungen seiner Zusammensetzung stecken immer noch in den Kinderschuhen. Dennoch hat man bereits einige interessante Ergebnisse zusammentragen können. Zum Beispiel wurden die Transkriptome unterschiedlicher Zelltypen in der Genomsequenz des Menschen kartiert. Das Ergebnis ist ein Überblick über das Genexpressionsmuster entlang eines vollständigen Chromosoms. Diese Erkenntnisse führten zu der wichtigen Entdeckung, dass Transkripte von Chromosomenregionen synthetisiert werden, in denen bislang keine Gene bekannt waren. Für die Untersuchungen wurden zum Beispiel DNA-Chips hergestellt, deren einzelne Oligonucleotidsonden auf Bereiche abzielten, die auf den Chromosomen 21 und 22 einen Abstand von durchschnittlich 35 Nucleotiden haben. Diese so genannten **tiling-Arrays** (*tiling* deckend) enthalten über eine Million Zielsequenzen, doch liegen davon nur 26 000 in Exons dieser Chromosomen. Allerdings konnte man über 350 000 Zielsequenzen eine mRNA in mindestens einem von 11 menschlichen Transkriptomen zuordnen, die aus den unterschiedlichen untersuchten Zelllinien stammten (Abb. 6.9). Genomweit wurden etwa 10 500 transkribierte Sequenzen ausfindig gemacht, die aus Regionen des Genoms stammen, von denen man zuvor angenommen hatte, dass sie keine Gene einhalten. Diese Arbeit zeigt, dass die Transkriptomanalyse bei der Genomannotierung eine wichtige Rolle spielt.

Die Transkriptomanalyse hat auch auf die Untersuchungen von Krankheiten des Menschen einen großen Einfluss. Im Jahr 1997 entdeckte man die Umstrukturierung des Transkriptoms als Folge einer Krebserkrankung. Ein Vergleich von normalen Epithelzellen des Dickdarms mit Krebszellen ergab signifi-



6.9 Transkriptomanalyse der Chromosomen 21 und 22 des Menschen. Ein Teil jedes Chromosoms ist mit seinem G-Bandenmuster (Abschnitt 7.1.2) und mit den Kartenpositionen (21p11.2 usw.) dargestellt. Bei jedem Chromosom zeigt die obere Grafik die Lokalisierung der bekannten Exons, ausgedrückt als Exondichte für ein 5,7-Mb großes „Fenster“ der DNA. Die untere Abbildung zeigt die Positionen, an denen in 11 untersuchten Transkriptomen mRNAs nachgewiesen werden konnten, ebenfalls ausgedrückt als Dichte pro 5,7-Mb-Fenster. Nachdruck mit Genehmigung von Kapranov et al (2002) *Science* 296: 916 – 919, © AAAS.

kante Unterschiede in der Häufigkeit von 289 mRNAs. Ungefähr die Hälfte dieser mRNAs zeigte diese Tendenz auch in Krebszellen der Bauchspeicheldrüse. Durch diese wichtige Beobachtung von Unterschieden zwischen den Transkriptomen von normalen Zellen und Krebszellen wird die Etablierung neuer Therapieformen für Krebserkrankungen unterstützt. Transkriptomanalysen werden aber auch in der Krebsdiagnostik eingesetzt. Der Durchbruch in diesem Bereich fand im Jahr 1999 statt, als man zeigen konnte, dass sich das Transkriptom aus entarteten Zellen der akuten lymphatischen Leukämie von dem aus Zellen der akuten myeloischen Leukämie unterscheidet. Die Untersuchung von 27 lymphatischen und 11 myeloischen Krebserkrankungen ergab, dass trotz leichter Unterschiede zwischen allen Transkriptomen die Differenzen zwischen diesen beiden Formen ausreichen, um die Krebserkrankungen eindeutig voneinander abgrenzen zu können. Die Bedeutung dieser Arbeit liegt in den verbesserten Remissionsraten, die man durch eine frühzeitige Erkennung einer Krebserkrankung erreichen kann, noch bevor klare morphologische Anzeichen zu sehen sind. Bei diesen beiden Formen der Leukämie ist diese Erkenntnis zwar nicht relevant, weil sie durch nicht genetische Verfahren voneinander unterschieden werden können, doch bei anderen Krebsformen wie dem Non-Hodgkin-Lymphom ist ein früher eindeutiger Befund durchaus von Bedeutung. Die am weitesten verbreitete Form dieser Erkrankung wird als diffuses großzelliges B-Zell-Lymphom bezeichnet. Viele Jahre nahm man an, dass alle Tumoren dieses Typs gleich sind. Nach Transkriptomanalysen musste man diese Auffassung jedoch revidieren, denn es zeigte sich, dass B-Zell-Lymphome in zwei Subtypen unterteilt werden können. Durch die Unterscheidung zwischen den Transkriptomen dieser Subtypen konnte jeder einer bestimmten Klasse von B-Zellen zugeordnet werden, was die gezielte Suche nach spezifischen Behandlungsformen ermöglichte, die auf das jeweilige Lymphom zugeschnitten sind.

6.2 Untersuchung des Proteoms

Untersuchungen des Proteoms sind wichtig, weil es die zentrale Rolle des Proteoms ist, die Verbindung zwischen Genom und biochemischen Fähigkeiten einer Zelle herzustellen (Abschnitt 1.3.2). Die Charakterisierung der Proteome unterschiedlicher Zellen ist der Schlüssel zum Verständnis, wie ein Genom agiert und wie dysfunktionelle Genomaktivitäten zu Krankheiten führen können. Transkriptomanalysen können diese Fragestellungen nur zu einem Teil beantworten. Die Untersuchung des Transkriptoms gibt einen genauen Hinweis auf die aktiven Gene in einer Zelle, doch was die in ihr enthaltenen Proteine betrifft, sind die Angaben weniger genau. Der Grund dafür ist, dass nicht nur die vorhandene Menge an mRNA den Proteingehalt beeinflusst, sondern auch die Geschwindigkeit, mit der die mRNA in Proteine translatiert wird und mit der diese abgebaut werden. Außerdem ist das ursprüngliche Proteinprodukt der Translation unter Umständen nicht aktiv, sodass manche Proteine eine physikalische und/oder chemische Modifikation durchlaufen müssen, bevor sie funktionell aktiv sind (Abschnitt 13.3). Die Bestimmung der Menge der *aktiven* Form eines Proteins ist daher für das Verständnis der Biochemie einer Zelle oder eines Gewebes entscheidend.

Die Methodik, die für die Analyse von Proteomen angewendet wird, bezeichnet man als Proteomik. Genau genommen ist die **Proteomik** eine Sammlung von verschiedenen Techniken, die über ihre Eigenschaft Informationen über das Proteom zu liefern, miteinander in Beziehung stehen. Diese Informationen umfassen nicht nur die Identität der vorhandenen Proteine, sondern auch Faktoren wie die Funktionen einzelner Proteine und ihre Lokalisierung innerhalb der Zelle. Die besondere Methodik zur Ermittlung der Zusammensetzung des Proteoms wird als **Protein-Profiling** oder **Expressionsproteomik** bezeichnet.

6.2.1 Protein-Profiling – Methodik zur Identifizierung von Proteinen in einem Proteom

Das Protein-Profiling beruht auf zwei Techniken – **Proteinelektrophorese** und **Massenspektrometrie** –, die beide schon seit langem etabliert sind, doch in der prägenomischen Ära selten zusammen angewendet wurden. Heutzutage werden sie in einem der am schnellsten wachsenden Bereiche der modernen Forschung miteinander kombiniert.

Die Trennung der Proteine eines Proteoms

Um ein Proteom zu charakterisieren ist es zunächst notwendig, reine Proben der vorhandenen Proteine herzustellen. Dies ist im Hinblick auf die Komplexität eines durchschnittlichen Proteoms kein triviales Unterfangen: Erinnern Sie sich daran, dass eine Säugerzelle 10 000–20 000 unterschiedliche Proteine enthalten kann (Abschnitt 1.3.2).

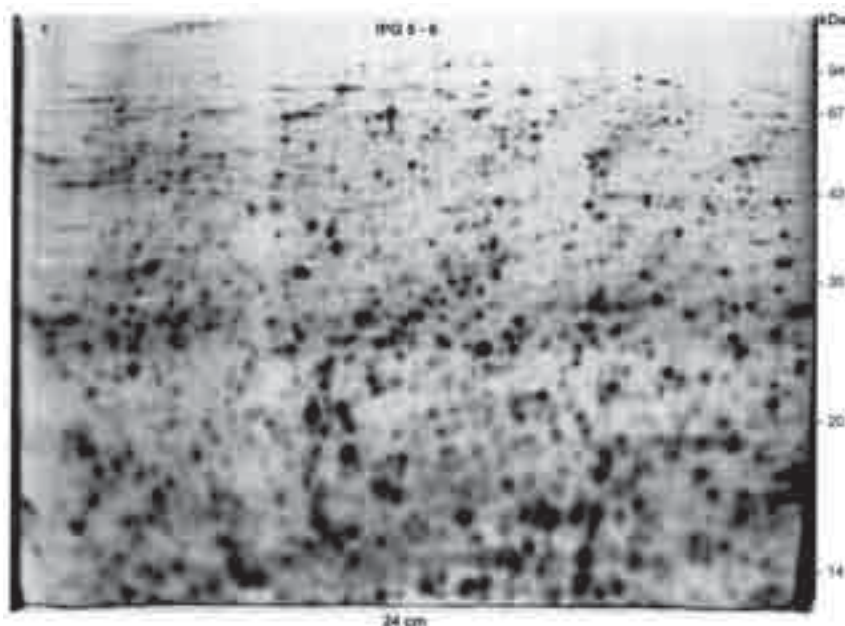
Die Standardmethode für die Trennung eines Proteingemisches ist die Polyacrylamidgelelektrophorese (Methoden 4.1). Abhängig von der Gelzusammensetzung und den Bedingungen, unter denen die Gelelektrophorese durchgeführt wird, macht man sich für die Trennung unterschiedliche chemische und physikalische Eigenschaften der Proteine zunutze. Das am häufigsten angewendete Verfahren macht von dem Detergens Natriumdodecylsulfat (*sodium dodecyl sulfate*, SDS) Gebrauch, das Proteine denaturiert und ihnen eine nega-



tive Ladung vermittelt, die ungefähr der Länge des nichtgefalteten Polypeptids entspricht. Unter diesen Bedingungen trennen sich die Proteine entsprechend ihrer molekularen Masse, wobei die kleinsten Proteine am schnellsten zur positiv geladenen Elektrode wandern. Alternativ können die Proteine auch durch **isoelektrische Fokussierung** in einem Gel getrennt werden, das Chemikalien enthält, die unter elektrischer Spannung einen pH-Gradienten aufbauen. In dieser Art von Gel wandert ein Protein bis zu seinem **isoelektrischen Punkt**, also der Position, an der seine Nettoladung gleich null ist. Beim Protein-Profiling werden diese Methoden zu einer **zweidimensionalen Gelelektrophorese** miteinander kombiniert. In der ersten Dimension werden die Proteine durch isoelektrische Fokussierung getrennt. Das Gel wird anschließend in eine SDS-Lösung gelegt, danach um 90 Grad gedreht und dann einer zweiten Elektrophorese unterzogen, bei der die Proteine entsprechend ihrer Größe im rechten Winkel zu der ersten Elektrophorese voneinander separiert werden (Abb. 6.10). Mithilfe dieses Ansatzes lassen sich in einem einzigen Gel Tausende von Proteinen trennen.

6.10 Zweidimensionale Gelelektrophorese.

Nach der Elektrophorese zeigt die Färbung des Gels ein komplexes Muster von Spots, von denen jeder ein anderes Protein enthält (Abb. 6.11). Bei einem Vergleich von zwei Gelen weisen Unterschiede im Muster und in der Intensität der Spots auf Unterschiede in der Identität und den relativen Mengen der einzel-



6.11 Das Ergebnis einer zweidimensionalen Gelelektrophorese. Proteine aus der Mausleber wurden durch isoelektrische Fokussierung mit einem pH-Bereich von 5–6 in der ersten Dimension und entsprechend ihrer molekularen Masse in der zweiten Dimension aufgetrennt. Die Protein-Spots wurden durch Silberfärbung sichtbar gemacht. Nachdruck mit Genehmigung von Görg et al (2000) *Electrophoresis* 21: 1037 – 1053 Wiley-VCH Verlag.

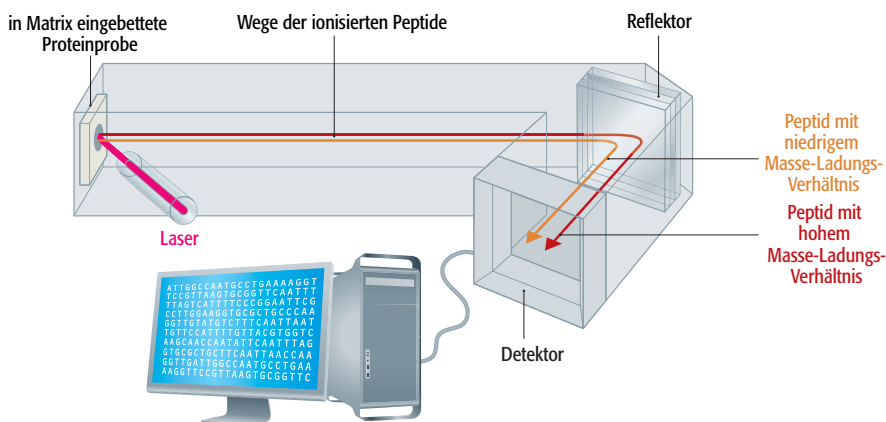
nen Proteine in den beiden untersuchten Proteomen hin. Auf diese Weise können für die zweite Phase der Proteomanalyse, in der die tatsächlich vorhandenen Proteine wie unten beschrieben identifiziert werden, interessante Spots ausgewählt werden. Doch bevor wir zu dieser Phase übergehen, müssen wir die Grenzen der zweidimensionalen Gelelektrophorese kennen lernen, die einen entscheidenden Einfluss auf die Anwendbarkeit des Protein-Profilings als Hilfsmittel für die Untersuchung von Proteomen hat. Das größte Problem ist, dass nicht alle Proteine des Proteoms in dem Gel sichtbar sind. Insbesondere fehlen die Proteine, die sich nicht in einem wässrigen Puffer lösen, wie viele der in der Zellmembran vorhandenen Proteine. Um diese Bestandteile des Proteoms untersuchen zu können, muss man spezielle Puffer und Gelzusammensetzungen verwenden. Das bedeutet wiederum, dass etliche Experimente erforderlich sind, wenn man das Proteom in seiner Gesamtheit untersuchen möchte. Außerdem gibt es Probleme mit der Reproduzierbarkeit der zweidimensionalen Gelelektrophorese, und es ist schwierig, entsprechende Kontrollen zu etablieren, um die Daten eines solchen Gels so zu standardisieren, dass zwei Proteome miteinander verglichen werden können. Aus diesen Gründen wurden alternative Trennmethode entwickelt, von denen zurzeit die Hochleistungsflüssigkeitschromatographie (*high performance liquid chromatography*, HPLC) und die *free flow*-isoelektrische Fokussierung eine zentrale Rolle spielen.

Identifizierung der Proteine eines Proteoms

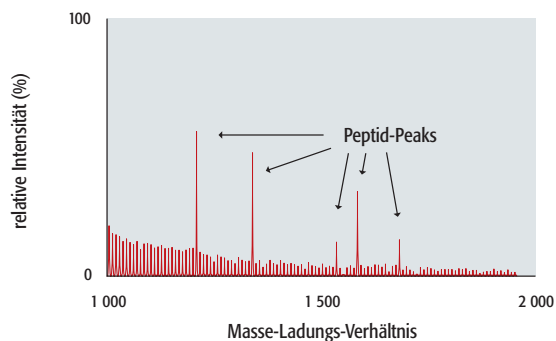
Zweidimensionale Gelelektrophorese führt zu einem komplexen Muster von Spots, von denen jeder ein anderes Protein enthält. Wie können wir ein Protein in einem Spot identifizieren? Dieses war stets eine schwierige Aufgabe doch durch Fortschritte in der Massenspektrometrie steht nun ein schnelles und genaues Erkennungsverfahren zur Verfügung, das den Anforderungen der Genomanalysen genügt. Die Massenspektrometrie wurde ursprünglich als Hilfsmittel für die Identifizierung einer Verbindung durch das Masse-Ladungs-Verhältnis ihrer ionisierten Form entwickelt, die man erhält, wenn man Moleküle der Verbindung einem starken elektrischen Feld aussetzt. Das Standardverfahren konnte bei Proteinen nicht eingesetzt werden, weil sie für eine effektive Ionisierung zu groß sind. Doch eine neue Methode, die als **MALDI-TOF** (*matrix-assisted laser desorption ionization time-of-flight*) bezeichnet wird, umging dieses Problem zumindest bei Peptiden bis zu einer Länge von 50 Aminosäuren. Tatsächlich sind die meisten Proteine viel länger als 50 Aminosäuren, wodurch es notwendig ist, sie vor der Untersuchung mit MALDI-TOF in Fragmente zu zerteilen. Üblicherweise wird das Protein aus einem Spot isoliert und dann mit einer sequenzspezifischen Protease wie Trypsin, die das Protein unmittelbar hinter Arginin- oder Lysinresten spaltet, abgebaut. Bei den meisten Proteinen führt diese Behandlung zu einer Serie von Peptiden mit einer Länge zwischen 5 und 75 Aminosäuren.

Ist das Peptid ionisiert, dann wird das Masse-Ladungs-Verhältnis durch seine „Flugzeit“ („*time of flight*“) ermittelt, die es innerhalb des Massenspektrometers für die Strecke zwischen Ionierungsquelle und Detektor braucht (Abb. 6.12). Ist das Masse-Ladungs-Verhältnis bekannt, lässt sich die molekulare Masse bestimmen, aus der man die Aminosäurezusammensetzung des Peptids ableiten kann. Wurde eine Anzahl von Peptiden eines einzelnen Proteins aus einem Spot des zweidimensionalen Gels untersucht, dann können die resultierenden Informationen über die Zusammensetzung mit der genomischen Sequenz abgeglichen werden, um die Identität des proteincodierenden Gens zu ermitteln. Die Aminosäurezusammensetzungen der von einem einzelnen

a) MALDI-TOF-Massenspektrometrie



b) MALDI-TOF-Spektrum

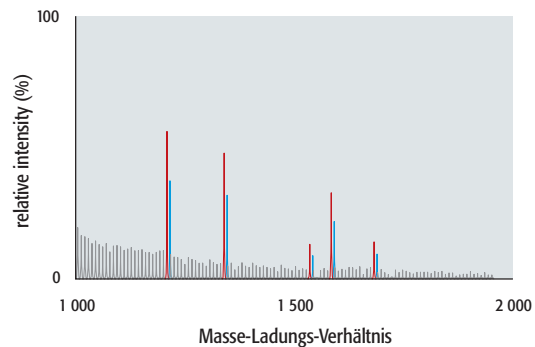


6.12 Die Verwendung von MALDI-TOF beim Protein-Profilung. Nach der zweidimensionalen Gelelektrophorese wird das gewünschte Protein aus dem Gel geschnitten und mit einer Protease wie Trypsin gespalten. Diese schneidet das Protein in eine Reihe von Peptiden, die man mittels MALDI-TOF analysieren kann (a). In dem Massenspektrometer werden die Peptide durch einen Puls mit einem Laserstrahl ionisiert, sie durchfliegen ein Rohr bis zu einem Reflektor und landen anschließend auf einen Detektor. Die Flugzeit jedes Peptids hängt von seinem Masse-Ladungs-Verhältnis ab. Die Daten werden in Form eines Spektrums sichtbar gemacht (b). Der Computer enthält eine Datenbank mit den erwarteten molekularen Massen von allen Trypsinfragmenten, die aus den Proteinen entstehen, die von dem Genom des untersuchten Organismus codiert werden. Der Computer vergleicht die Massen der nachgewiesenen Peptide mit der Datenbank und identifiziert das wahrscheinlichste Ursprungsprotein.

Protein abstammenden Peptide können auch für die Überprüfung der Gensequenz und der korrekten Lokalisierung der Exon-Intron-Grenzen eingesetzt werden. Dies ist nicht nur hilfreich, um die Position eines Gens in einem Genom zu beschreiben (Abschnitt 5.1.1), sondern es können auch alternative Spleißvorgänge in den Fällen nachgewiesen werden, in denen zwei oder mehr Proteine von demselben Gen abstammen.

Werden zwei Proteome miteinander verglichen, dann ist es in erster Linie notwendig, Proteine zu identifizieren, die in unterschiedlichen Mengen vorhanden sind. Relativ große Unterschiede sind durch einfache Betrachtung der gefärbten Gele nach der zweidimensionalen Gelelektrophorese zu erkennen. Jedoch können wichtige Veränderungen der biochemischen Kapazität eines Proteoms auch von relativ geringen Veränderungen in der Menge einzelner Proteine ausgehen, wodurch entsprechende Nachweisverfahren für die Messung kleiner Änderungen erforderlich sind. Eine Möglichkeit, dies zu bewerkstelligen, ist die Markierung der Bestandteile der beiden Proteome mit unterschiedlich fluoreszierenden Markern und die anschließende Auftrennung der Proteome in einem einzigen zweidimensionalen Gel. Hierbei handelt es sich um die gleiche Strategie wie etwa bei dem Vergleich von Transkriptomen (Abb. 6.6). Wird das zweidimensionale Gel unterschiedlichen Wellenlängen ausgesetzt, kann man die Intensitäten der entsprechenden Spots genauer bewerten, als es bei zwei getrennten Gelen möglich wäre. Eine noch genauere Alternative ist, jedes der beiden Proteome mit einem **isotopencodierten Affinitäts-**

6.13 Die Analyse von zwei Proteomen durch ICAT. In dem MALDI-TOF-Spektrum werden Peaks, die von Peptiden mit normalen Wasserstoffatomen stammen, blau dargestellt, diejenigen, die von Peptiden mit Deuterium stammen, dagegen rot. Das untersuchte Protein ist etwa 1,5-mal häufiger in dem Genom enthalten, das mit Deuterium markiert wurde.



marker (*isotope coded affinity tag, ICAT*) zu kennzeichnen. Diese Marker liegen in zwei Formen vor, die eine enthält normale Wasserstoffatome und die andere Deuterium, das schwere Isotop des Wasserstoffs. Die normale und die schwere Form können durch Massenspektrometrie unterschieden werden. Dadurch ist eine Bestimmung der relativen Mengen eines Proteins in den beiden miteinander gemischten Proteomen während der „MALDI-TOF-Phase“ des Protein-Profiling möglich (Abb. 6.13).

6.2.2 Die Identifizierung von Proteinen, die miteinander in Wechselwirkung treten

Wichtige Daten bezüglich der Genomaktivität erhält man auch durch die Identifizierung miteinander interagierender Paare und Gruppen von Proteinen. Im Einzelnen betrachtet ist diese Information oft sehr aufschlussreich, wenn einem neu entdeckten Gen oder Protein eine Funktion zugewiesen werden soll (Abschnitt 5.2), weil die Interaktion mit einem zweiten, gut charakterisierten Protein häufig auf die Funktion des unbekanntes Proteins hinweist. Zum Beispiel deutet eine Wechselwirkung mit einem auf der Zelloberfläche lokalisierten Protein darauf hin, dass das unbekanntes Protein in die Signalweiterleitung von Zelle zu Zelle eingebunden ist (Abschnitt 14.1). Etwas globaler betrachtet ist die Konstruktion einer **Proteininteraktionskarte** ein wichtiger Schritt bei der Verknüpfung des Proteoms mit der zellulären Biochemie.

Die Identifizierung von Paaren miteinander interagierender Proteine durch Phagen-Display und two-hybrid-Analysen

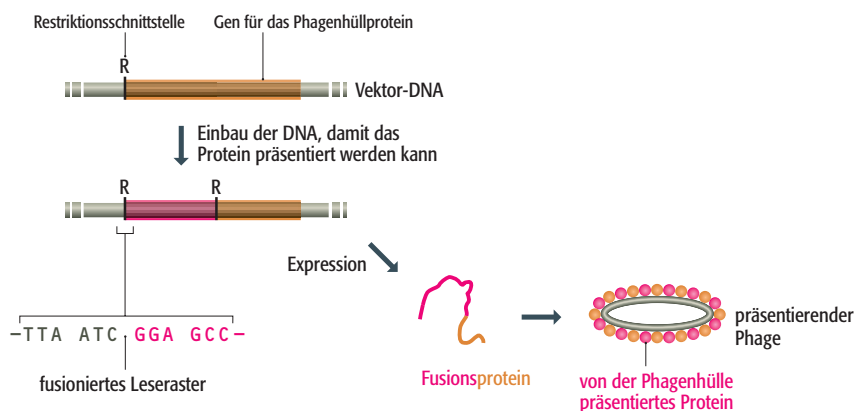
Für die Untersuchung von Protein-Protein-Wechselwirkungen existiert eine Vielzahl von Methoden, von denen die beiden nützlichsten das so genannte **Phagen-Display** und das **two-hybrid-System** in Hefe sind. Beim Phagen-Display wird ein spezieller Typ eines Klonierungsvektors eingesetzt, der auf einen λ -Bakteriophagen oder einen filamentösen Bakteriophagen wie M13 zurückgeht. Der Vektor wurde so konstruiert, dass ein in diesen Vektor kloniertes Fremd-Gen als Fusionsprodukt mit einem der Phagenhüllproteine exprimiert wird (Abb. 6.14a). Dadurch bringt das Phagenprotein das Fremdprotein in die Phagenhülle, wo es so präsentiert wird, dass es mit anderen Proteinen aus der Umgebung des Phagen interagieren kann. Es gibt viele Möglichkeiten der Analyse von Proteininteraktionen mithilfe des Phagen-Display-Systems. Bei einem der Verfahren wird das Testprotein präsentiert und man sucht nach Wechselwirkungen mit einer Reihe von gereinigten Proteinen oder Proteinfragmenten mit bekannter Funktion. Dieser Ansatz ist jedoch nur eingeschränkt verwendbar, da die Durchführung der Tests relativ zeitintensiv ist. Ein Einsatz ist

nur sinnvoll, wenn im Vorfeld Informationen über wahrscheinlich stattfindende Wechselwirkungen zur Verfügung stehen. Eine leistungsfähigeres Verfahren ist die Herstellung einer **Phagen-Display-Bibliothek**, also einer Klonsammlung, die eine Reihe von Proteinen repräsentiert. Die Aufgabe besteht in diesem Fall darin, den Klon ausfindig zu machen, der mit dem Testprotein interagiert (Abb. 6.14b).

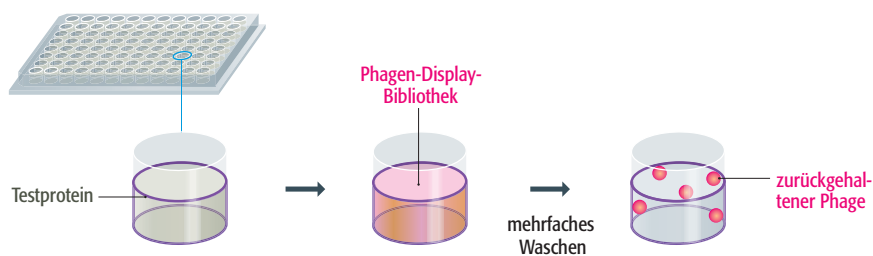
Das *two-hybrid*-System in Hefe weist Proteinwechselwirkungen auf eine komplexere Art und Weise nach. In Abschnitt 11.3.2 werden wir sehen, dass Proteine, die als **Aktivatoren** bezeichnet werden, für die Kontrolle der Genexpression in Eukaryoten notwendig sind. Um diese Funktion ausüben zu können, muss ein Aktivator an DNA-Sequenzen stromaufwärts eines Gens binden und die RNA-Polymerase dazu stimulieren, das Gen in RNA zu umzuschreiben. Diese zwei Fähigkeiten – Binden der DNA und Aktivieren der Polymerase – werden durch zwei unterschiedliche Bereiche des Aktivators vermittelt, wobei manche Aktivatoren sogar aktiv sind, wenn man sie in zwei Teile schneidet, von denen ein Segment die DNA-bindende Domäne enthält und das andere die Aktivierungsdomäne. In der Zelle interagieren beide Segmente, sodass ein funktioneller Aktivator entsteht.

Bei dem *two-hybrid*-System wird ein Stamm von *Saccharomyces cerevisiae* eingesetzt, dem ein Aktivator für ein Reporter-gen fehlt. Dieses Gen ist daher abgeschaltet. Ein künstliches Gen, das die DNAbindende Domäne des Aktivators codiert, wird mit dem Gen für das Protein ligiert, dessen Interaktionen wir untersuchen möchten. Dieses Protein kann von jedem Organismus stammen, nicht

a) Herstellung eines Phagen-Displays

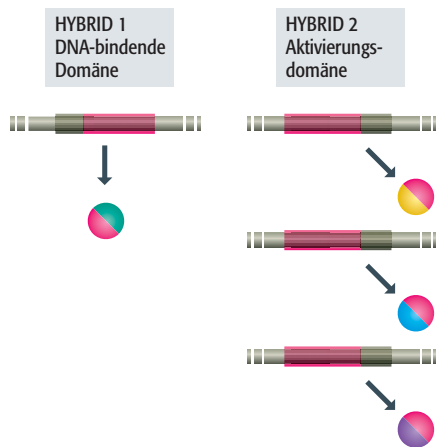


b) Verwendung einer Phagen-Display-Bibliothek

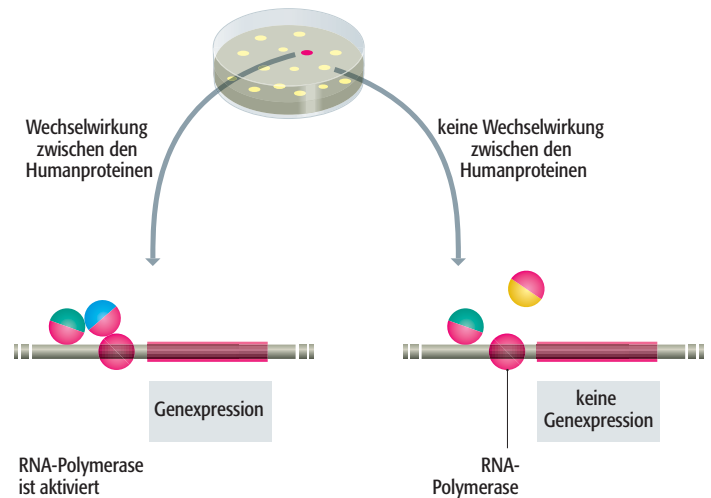


6.14 Phagen-Display. a) Der beim Phagen-Display eingesetzte Klonierungsvektor ist ein Bakteriophagen-genom mit einer nur einmal vorkommenden Restriktionsschnittstelle innerhalb eines Gens für ein Hüllprotein. Die Methode wurde ursprünglich mit dem Gen III, das ein Hüllprotein des filamentösen Phagen f1 codiert, durchgeführt, doch erweiterte man das Spektrum, sodass nun auch andere Phagen wie λ eingesetzt werden können. Für das Phagen-Display wird die DNA-Sequenz, die das Testprotein codiert, in die Restriktionsschnittstelle ligiert, sodass ein fusioniertes Leseraster entsteht – eines, bei dem die Reihe der Codons von dem Testgen bis zum Hüllproteingen bestehen bleibt. Nach Transformation in *Escherichia coli* führt das rekombinante Molekül zur Synthese eines Hybridproteins, einer Fusion des Testproteins mit dem Hüllprotein. Die von dem transformierten Bakterium produzierten Phagenpartikel präsentieren daher das Testprotein auf ihren Hüllen. b) Die Verwendung einer Phagen-Display-Bibliothek. Das Testprotein wird in einer Vertiefung einer Mikrotiterplatte immobilisiert und die Phagen-Display-Bibliothek zugegeben. Nach einem Waschschriff werden in den Vertiefungen Phagen zurückgehalten, die ein Protein präsentieren, das mit dem Testprotein interagiert.

a) *two-hybrid*-System



b) Suche nach Proteininteraktionen mit dem *two-hybrid*-System



Legende

	Hefegen		Hefedomänen		Humandomänen
	Humangen				

6.15 Das *two-hybrid*-System in Hefe.

a) Auf der linken Seite wurde ein Protein des Menschen mit dem Gen für die DNA-bindende Domäne eines Hefeaktivators ligiert. Nach Transformation der Hefe vermittelt dieses Konstrukt ein Fusionsprotein, das zum Teil aus dem Humanprotein und zum Teil aus der DNA-bindende Domäne des Hefeaktivators besteht. Auf der rechten Seite wurden verschiedene Human-DNA-Fragmente mit einem Gen für die Aktivierungsdomäne des Hefeaktivators ligiert: Diese Konstrukte codieren eine Vielzahl von Fusionsproteinen. b) Die beiden Gruppen von Konstrukten werden gemischt und Hefezellen mit dem Gemisch cotransformiert. Eine Kolonie, in der das Reportergen exprimiert wird, enthält Fusionsproteine, deren Humansegmente miteinander in Wechselwirkung treten, sodass die DNA-bindende Domäne in die Nähe der Aktivierungsdomäne gebracht und die RNA-Polymerase stimuliert wird. Siehe Abschnitt 11.3.2 für weitere Informationen über Aktivatoren.

nur aus der Hefe: In dem in Abbildung 6.15a gezeigten Beispiel ist es ein Humanprotein. Nachdem das Konstrukt in Hefe eingebracht wurde, vermittelt es die Synthese eines Fusionsproteins aus der DNAbindende Domäne des Aktivators, verbunden mit dem Humanprotein. Der rekombinante Hefestamm kann zu diesem Zeitpunkt das Reportergen nicht exprimieren, weil der modifizierte Aktivator lediglich an die DNA bindet, die RNA-Polymerase jedoch nicht beeinflusst. Eine Aktivierung findet nur statt, wenn der Hefestamm mit einem zweiten Konstrukt cotransformiert wird, das die codierende Sequenz der Aktivierungsdomäne besitzt. Diese Aktivierungsdomäne muss mit einem DNA-Fragment fusioniert sein, das ein Protein codiert, welches mit dem zu testenden Humanprotein interagiert (Abb. 6.15b). Wie beim Phagen-Display können einzelne DNA-Fragmente nacheinander mit *two-hybrid*-Systemen getestet werden, wenn mögliche Wechselwirkungen bereits im Vorfeld bekannt sind. Gewöhnlich wird die Aktivierungsdomäne jedoch mit einem Gemisch von DNA-Fragmenten ligiert, sodass viele unterschiedliche Konstrukte entstehen. Nach der Transformation werden die Zellen ausplattiert und diejenigen, die das Reportergen exprimieren, ausfärbig gemacht. Diese Zellen enthalten eine Kopie des Gens für die Aktivierungsdomäne, die mit einem DNA-Fragment fusioniert ist, das wiederum ein mit dem Testgen interagierendes Protein codiert.

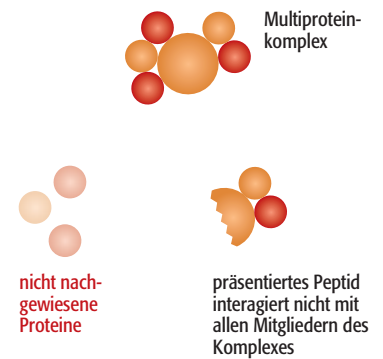
Die Identifizierung der Komponenten von Multiproteinkomplexen

Das Phagen-Display und das *two-hybrid*-System in Hefe sind effektive Methoden, um miteinander interagierende Paare von Proteinen ausfärbig zu machen, doch die Identifizierung solcher Verbindungen ist nur die Basis für die Protein-Protein-Interaktionen. Viele der zellulären Aktivitäten werden durch Multiproteinkomplexe ausgeführt, wie der Mediator-Komplex, der bei der Gentranskription

tion eine zentrale Rolle spielt (Abschnitt 11.3.2), oder das Spleißosom, das die Introns aus der Prä-mRNA entfernt (Abschnitt 12.2.2). Komplexe wie diese bestehen typischerweise aus einigen Kernproteinen (Core), die ständig anwesend sind, zusammen mit einer Vielfalt von zusätzlichen Proteinen, die nur unter bestimmten Bedingungen mit dem Komplex verbunden sind. Die Identifizierung von Core-Proteinen und zusätzlichen Proteinen ist ein entscheidender Schritt, um zu verstehen, wie diese Komplexe ihre Funktionen ausüben. Diese Proteine können, Paar für Paar, durch eine lange Reihe von *two-hybrid*-Experimenten identifiziert werden, doch wäre ein direkterer Weg zur Bestimmung der Zusammensetzung der Multiproteinkomplexe wünschenswert.

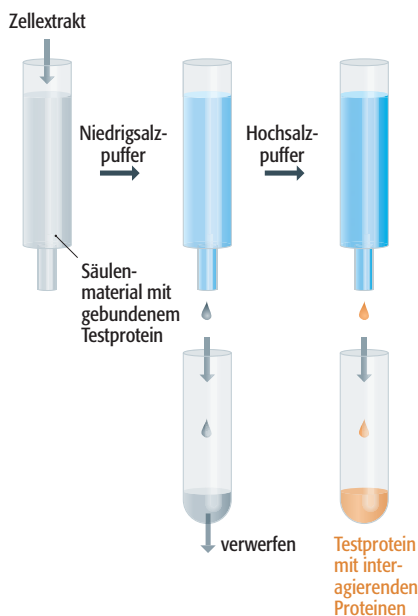
Grundsätzlich lassen sich mithilfe einer Phagen-Display-Bibliothek die Mitglieder eines Multienzymkomplexes identifizieren, da bei diesem Verfahren alle Proteine, die mit dem Testprotein interagieren, in einem einzigen Experiment ausfindig gemacht werden (Abb. 6.14). Es besteht jedoch die Schwierigkeit, dass große Proteine nicht effizient präsentiert werden, da sie den Replikationszyklus des Phagen stören. Um dieses Problem zu umgehen, ist es im Allgemeinen notwendig, statt des vollständigen Proteins ein kurzes Peptid zu präsentieren, das einen Teil des zellulären Proteins darstellt. Da dem Peptid ein in der intakten Form vorhandener Bereich für die Protein-Protein-Wechselwirkung fehlt, ist es möglich, dass das präsentierte Peptid nicht mit allen Mitgliedern des Komplexes interagiert, in dem das intakte Protein lokalisiert ist (Abb. 6.16). Ein Verfahren, das mit intakten Proteinen arbeitet und dieses Problem somit umgeht, ist die **Affinitätschromatographie**. Bei der Affinitätschromatographie wird das an Trägermaterial für die Chromatographie gebundene Testprotein in eine Säule gegeben (Methoden 2.3). Der Zellextrakt passiert die Säule in einem Niedrigsalzpuffer, der die Bildung von Wasserstoffbrücken erlaubt, die die Proteine zu einem Komplex zusammenhalten (Abb. 6.17a). Die mit dem gebundenen Testprotein interagierenden Proteine werden daher in der Säule zurückgehalten, während alle anderen von der Säule gewaschen werden. Anschließend löst (eluiert) man die interagierenden Proteine mit einem Hochsalzpuffer von der Säule. Der Nachteil dieses Verfahrens ist, dass das Testprotein gereinigt werden muss; ein für ein in großem Maßstab angelegtes Screening-Programm zeitintensiver und schwieriger Prozess. Ein Ausweg stellt die als **Tandem-Affinitätsreinigung** (*tandem affinity purification*, **TAP**) bezeichnete, technisch anspruchsvollere Methode dar, die für die Untersuchung von Proteinkomplexen in *S. cerevisiae* entwickelt wurde. Bei diesem Verfahren wird das testproteincodierende Gen so modifiziert, dass das synthetisierte Protein eine C-terminale Verlängerung trägt, die an ein zweites Protein, Calmodulin, bindet. Der Zellextrakt wird unter moderaten Bedingungen hergestellt, sodass Multiproteinkomplexe nicht auseinander brechen. Anschließend passiert der Extrakt eine Affinitätschromatographiesäule, die mit Säulenmaterial gefüllt ist, an das Calmodulinmoleküle gebunden sind. Auf diese Weise werden sowohl das Testprotein als auch die mit ihm assoziierten Proteine gebunden (Abb. 6.17b). Bei beiden Verfahren bestimmt man die Identität der gereinigten Proteine durch Massenspektrometrie. Bei der Sichtung von 1 739 Hefegenen in einem großen Ansatz ließen sich mithilfe des TAP-Systems 232 Multiproteinkomplexe identifizieren und man gewann neue Erkenntnisse über die Funktion von 344 Genen, von denen zuvor viele durch experimentelle Verfahren nicht charakterisiert werden konnten.

Ein Nachteil der Affinitätschromatographiemethoden ist, dass ein einzelnes Mitglied eines Multiproteinkomplexes als „Köder“ eingesetzt wird, um andere

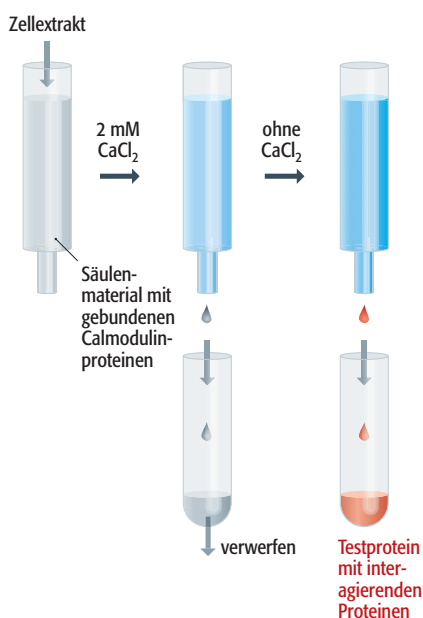


6.16 Mit dem Phagen-Display-System werden unter Umständen nicht alle Mitglieder eines Multiproteinkomplexes ausfindig gemacht. Der Komplex besteht aus einem zentralen Protein, das mit fünf kleineren Proteinen interagiert. In der unteren Darstellung wird ein Peptid des zentralen Proteins in einem Phagen-Display-Experiment eingesetzt. Mithilfe des Peptids lassen sich zwei interagierende Proteine nachweisen, die anderen drei jedoch nicht, weil ihre Bindungsstellen in einem anderen Bereich des zentralen Proteins lokalisiert sind.

a) Standardverfahren der Affinitätschromatographie



b) Tandem-Affinitätsreinigung



Proteine des Komplexes zu isolieren. Daraus folgt in der Praxis, dass sich ein Mitglied des Komplexes nicht isolieren lässt, wenn es nicht direkt mit dem „Köder“ interagiert (Abb. 6.18). Die Methoden weisen daher in einem Komplex enthaltene Gruppen von Proteinen nach, doch liefern sie nicht unbedingt das gesamte Spektrum der im Komplex vorhandenen Proteine. Die Entwicklung von Verfahren für die Reinigung intakter Komplexe ist daher ein wesentliches Ziel der aktuellen Forschung. Bei der **Coimmunpräzipitation** stellt man einen Zellextrakt unter moderaten Bedingungen her, sodass die Komplexe intakt bleiben. Anschließend gibt man einen für das Testprotein spezifischen Antikörper zu. Dadurch werden dieses Protein und alle Mitglieder des Komplexes, in dem es enthalten ist, gefällt. Eine technisch anspruchsvollere Methode für die Isolierung intakter Komplexe ist die **mehrdimensionale Proteinidentifizierungstechnologie** (*multi-dimensional protein identification technique*, **MudPIT**), die unterschiedliche Chromatographieverfahren miteinander kombiniert, zum Beispiel Umkehrphasen-Flüssigkeitschromatographie (*reversed-phase liquid chromatography*) entweder mit Kationenaustausch- oder mit Größenausschlusschromatographie. Die Komponenten eines Komplexes werden mittels Massenspektrometrie identifiziert. Diese Methode wurde zunächst für die Untersuchung der großen Untereinheit des Heferybosoms eingesetzt: Mit ihrer Hilfe wurden elf Proteine identifiziert, von denen zuvor nicht bekannt war, dass sie mit diesem Komplex assoziiert sind.

Die Identifizierung von Proteinen mit funktionellen Wechselwirkungen

Proteine müssen nicht unbedingt direkt miteinander assoziiert sein, um funktionell in Wechselwirkung zu treten. Zum Beispiel zeigen die Enzyme Lactose-Permease und β -Galactosidase in Bakterien wie *Escherichia coli* eine funktionelle Wechselwirkung, da sie beide an der Verwertung von Lactose als Kohlenstoffquelle beteiligt sind. Diese beiden Proteine interagieren jedoch nicht direkt: Die Permease ist in der Zellmembran lokalisiert und transportiert Lactose in die Zelle, die β -Galactosidase spaltet Lactose im Cytoplasma in Glucose und Galactose (Abb. 8.8a). Viele Enzyme, die in demselben Stoffwechselweg zusammenarbeiten, haben zu keinem Zeitpunkt direkten Kontakt, und wenn Untersuchungen einzig darauf abzielen, direkte Assoziationen zwischen Proteinen nachzuweisen, dann würde man viele funktionelle Wechselwirkungen übersehen.

Für die Identifizierung von Proteinen mit funktionellen Wechselwirkungen stehen diverse Methoden zur Verfügung. Die meisten von ihnen untersuchen nicht die Proteine direkt und gehören daher, streng genommen, nicht unter die allgemeine Überschrift „Proteomik“. Doch es bietet sich an, sie an dieser Stelle zu behandeln, da die Informationen, die sie liefern, zusammen mit den Ergebnissen der Proteomik in Proteininteraktionskarten einfließen. Zu diesen Verfahren gehören:

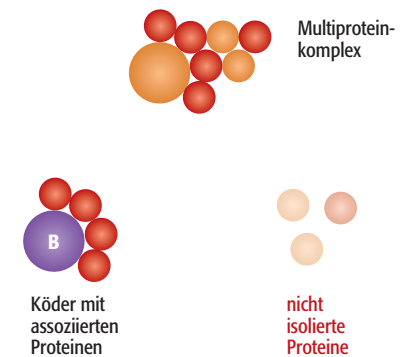
6.17 Verfahren der Affinitätschromatographie für die Reinigung von Multiproteinkomplexen. a) Bei dem Standardverfahren der Affinitätschromatographie wird das Testprotein an das Säulenmaterial gebunden. Der Zellextrakt wird in einem Niedrigsalzpuffer auf die Säule gegeben, sodass die Mitglieder des Multiproteinkomplexes an das Testprotein binden. Die Proteine werden anschließend mit einem Hochsalzpuffer von der Säule abgelöst (eluiert). b) Bei der Tandem-Affinitätsreinigung (TAP) wird der Zellextrakt in einem Puffer mit 2 mM CaCl₂ auf die Säule gegeben, also unter Bedingungen, die die Bindung des modifizierten Testproteins zusammen mit den damit assoziierten Proteinen an die säulengebundenen Calmodulinmoleküle fördern. Anschließend eluiert man die Proteine mit einem CaCl₂-freien Puffer von der Säule.

- **Vergleichende Genomik** kann auf unterschiedliche Art und Weise für die Identifizierung von Proteingruppen mit funktionellen Beziehungen eingesetzt werden. Ein Ansatz beruht auf der Beobachtung, dass Paare von Proteinen, die in manchen Organismen als getrennte Moleküle vorliegen, in anderen Organismen zu einer einzigen Polypeptidkette fusioniert sind. Ein Beispiel ist das Hefegen *HIS2*, das ein an der Histidinbiosynthese beteiligtes Enzym codiert. *E. coli* besitzt zwei zu *HIS2* homologe Gene. Eines von ihnen, als *his2* bezeichnet, hat Sequenzähnlichkeit mit der 5'-Region des Hefegens, und das zweite, *his10*, ist der 3'-Region ähnlich (Abb. 6.19). Die logische Schlussfolgerung ist, dass die von *his2* und *his10* codierten Proteine in dem *E. coli*-Proteom miteinander interagieren und dabei zur Histidinbiosyntheseaktivität beitragen. Die Analyse von Datenbanksequenzen ergibt viele Beispiele dieser Art, bei der zwei Proteine des einen Organismus in einem anderen Organismus zu einem einzigen Protein fusioniert sind. Ein ähnlicher Ansatz beruht auf der Untersuchung bakterieller Operons. Ein Operon besteht aus zwei oder mehreren Genen, die zusammen transkribiert werden und die in der Regel in funktioneller Beziehung stehen (Abschnitt 8.2). Zum Beispiel befinden sich die Gene der Lactose-Permease und der β -Galactosidase in *E. coli* in demselben Operon, zusammen mit dem Gen für ein drittes Protein, das an der Verwertung der Lactose beteiligt ist. Die Identität von Genen in bakteriellen Operons können daher verwendet werden, um funktionelle Wechselwirkungen zwischen den Proteinen abzuleiten, die in eukaryotischen Genomen durch homologe Gene codiert werden.
- Mithilfe von Transkriptomanalysen lassen sich funktionelle Wechselwirkungen zwischen Proteinen identifizieren, da die mRNAs funktionell miteinander in Beziehung stehender Proteine oft unter verschiedenen Bedingungen ähnliche Expressionsprofile zeigen.
- Genaktivierungsstudien können ebenfalls informativ sein. Wird eine Veränderung des Phänotyps schon dann beobachtet, wenn von zwei oder mehr Genen jedes einzeln inaktiviert wird, dann lässt sich daraus schließen, dass diese Gene zusammen die Ausbildung des unveränderten Phänotyps bewirken.

Proteininteraktionskarten

Proteininteraktionskarten, auch als **Interaktome** oder **Proteininteraktionsnetzwerke** bezeichnet, zeigen alle Wechselwirkungen, die zwischen den Komponenten eines Proteoms bestehen. Die ersten Karten wurden im Jahr 2001 für relativ einfache Proteome fast ausschließlich durch *two-hybrid*-Experimente erstellt. Zu diesen Karten gehörten die von *Helicobacter pylori* mit über 1 200 Wechselwirkungen, an denen nahezu die Hälfte der Proteine des Proteoms beteiligt sind, und die des *S. cerevisiae*-Proteoms mit 2 240 Interaktionen zwischen 1 870 Proteinen (Abb. 6.20a). Vor nicht allzu langer Zeit führte die Anwendung zusätzlicher Verfahren zu detaillierteren Versionen der *S. cerevisiae*-Karte. Zu diesen gehört auch eine Karte, in der auch die Interaktionen zwischen den Multiproteinkomplexen und nicht nur die innerhalb der Komplexe vermerkt sind (Abb. 6.20b). Auch erstellte man Karten von komplexeren Organismen wie *Caenorhabditis elegans*.

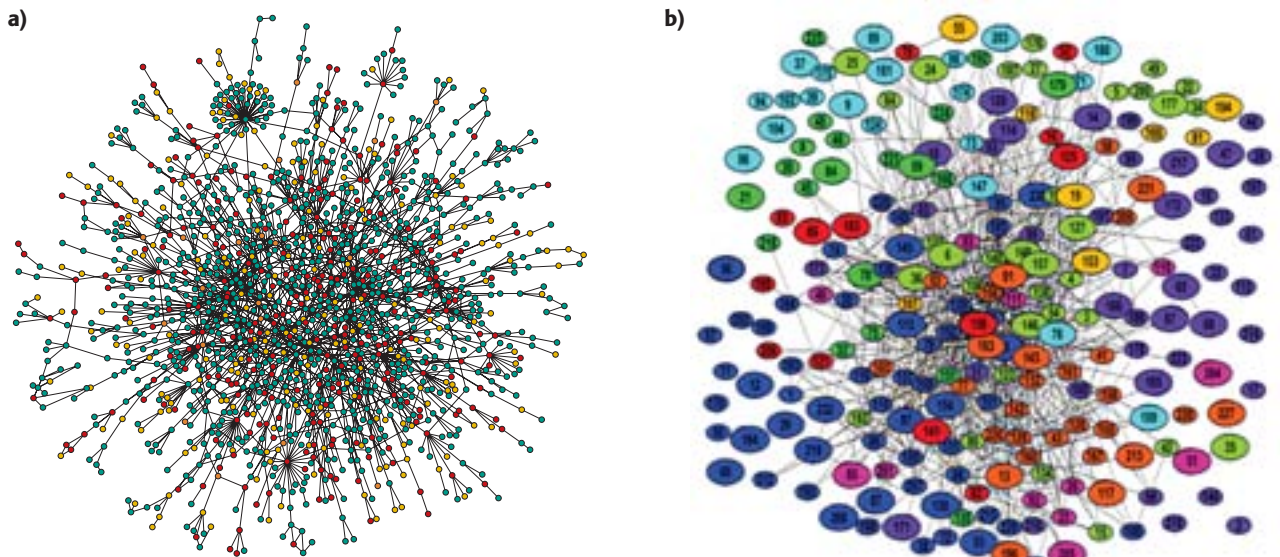
Zu welchen interessanten Erkenntnissen führten diese Proteininteraktionskarten? Die verblüffendste Entdeckung ist, dass sich jedes Netzwerk um eine kleine Anzahl von Proteinen arrangiert, die viele Interaktionen eingehen und die im



6.18 Ein Nachteil der Affinitätschromatographie. Interagiert das so genannte Köderprotein (mit einem „B“ für *bait* gekennzeichnet) nicht direkt mit einem oder mehreren Proteinen des Komplexes, dann können diese Proteine unter Umständen nicht isoliert werden.



6.19 Die Verwendung der Homologieanalyse für die Ableitung von Protein-Protein-Interaktionen. Die 5'-Region des *HIS2*-Gens der Hefe ist homolog zu *his2* aus *Escherichia coli* und die 3'-Region ist homolog zu *his10* aus *E. coli*.



6.20 Versionen von Proteininteraktionskarten von *Saccharomyces cerevisiae*.

a) Die ursprüngliche Karte wurde im Jahr 2001 veröffentlicht. Jeder Punkt repräsentiert ein Protein, wobei die Verbindungslinien Interaktionen zwischen Proteinpaaren darstellen. Rote Punkte sind essenzielle Proteine: Eine inaktivierende Mutation des Gens für dieses Protein ist letal. Mutationen in Genen für Proteine, die durch grüne Punkte dargestellt sind, sind nicht letal, Mutationen in Genen für Proteine, die in orange eingezeichnet sind, führen zu verlangsamttem Wachstum. Die Auswirkungen von Mutationen in Genen für gelb dargestellte Proteine waren zu dem Zeitpunkt, als die Karte erstellt wurde, nicht bekannt. b) Eine ausführlichere Karte wurde im Jahr 2002 veröffentlicht. In dieser Karte steht jedes Oval für einen Proteinkomplex und die Verbindungslinien zeigen Komplexe an, die mindestens ein Protein miteinander gemeinsam haben. Die Komplexe besitzen den folgenden, ihrer Funktion entsprechenden Farbcode: rot, Zellzyklus; dunkelgrün, Signalweiterleitung; dunkelblau, Transkription und Aufrechterhaltung der DNA und/oder der Chromatinstruktur; rosa, Protein- und/oder RNA-Transport; orange, RNA-Metabolismus; hellgrün, Proteinsynthese und/oder Umsatz; braun, Zellpolarität und/oder Struktur; violett, Intermediär- und/oder Energiestoffwechsel; hellblau, Membranbiogenese und/oder Membranbau. Abbildung a wurde freundlicherweise zur Verfügung gestellt von Hawond Jeong. Mit freundlicher Genehmigung von Macmillan Publishers Ltd© *Nature*. (Abb. a: Jeong et al (2001) *Nature* 411: 41 – 41; Abb. b: Gavin et al (2002) *Nature* 415: 141 – 147).

Netzwerk so genannte **hubs** (Knotenpunkte mit vielen Verbindungen) bilden, zusammen mit einer viel größeren Zahl von Proteinen mit nur wenigen Wechselwirkungen (Abb. 6.21a). Man nimmt an, dass diese Organisation die Auswirkungen von Mutationen minimiert, die einzelne Proteine inaktivieren. Nur wenn die Mutation eines der Proteine eines stark verknüpften Knotens betrifft, nimmt das Netzwerk als Ganzes Schaden. Diese Hypothese steht im Einklang mit Erkenntnissen aus Geninaktivierungsanalysen (Abschnitt 5.2.2), die zeigten, dass eine erhebliche Zahl von Hefeproteinen offenbar redundant ist. Wird die Aktivität eines solchen Proteins gestört, funktioniert das Proteom als Ganzes weiterhin normal, ohne dass eine Veränderung des Phänotyps zu erkennen wäre. Durch die Untersuchung der Expressionsprofile von *hub*-Proteinen und ihren direkten Partnern konnten *hubs* in zwei Gruppen eingeteilt werden. Die erste Gruppe von *hub*-Proteinen interagiert mit allen ihren Partnern gleichzeitig. Diese werden als „*party*“ *hubs* bezeichnet und ihre Entfernung hat nur eine geringe Wirkung auf die Gesamtstruktur des Netzwerkes (Abb. 6.21b). Im Gegensatz dazu spaltet die Entfernung der zweiten Gruppe von *hub*-Proteinen, der „*date*“- *hubs*, die zu verschiedenen Zeitpunkten mit ihren unterschiedlichen Partnern interagieren, das Netzwerk in eine Reihe kleiner Unternetzwerke (Abb. 6.21c). Die logische Schlussfolgerung ist, dass die *party hubs* innerhalb von individuellen biologischen Prozessen agieren und nur in einem geringen Umfang zur allgemeinen Organisation des Proteoms beitragen. Die *date hubs* dagegen sind die Schaltstellen für die Organisation des Proteoms, indem sie biologische Prozesse miteinander verbinden.

6.3 Jenseits des Proteoms

Das Proteom wird traditionell als Endprodukt der Genomexpression angesehen, doch diese Sicht verschleiert seine wahre Bedeutung als Teil der endgültigen Verbindung zwischen Genom und Biochemie der Zelle (Abb. 6.22). Die Erforschung der Eigenschaften dieser Verbindung erweist sich als einer der spannendsten und produktivsten Bereiche der modernen Biologie.

a) das vollständige Netzwerk



b) Entfernen der party hubs



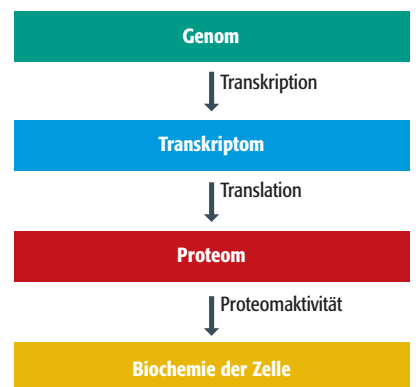
c) Entfernen der date hubs



6.21 *Hubs in der Proteininteraktionskarte von Saccharomyces cerevisiae.* Die Karte wurde im Jahr 2004 veröffentlicht. In der vollständigen Karte (a) sind die *hubs* deutlich zu erkennen. Nach Entfernen der *party hubs* bleibt das Netzwerk nahezu vollständig intakt (b). Werden jedoch die *date hubs* entfernt, spaltet sich das Netzwerk in nicht miteinander verbundene Subnetzwerke (c). Mit freundlicher Genehmigung von Macmillan Publishers Ltd© *Nature* (Han et al (2004) *Nature* 430: 88 – 93).

6.3.1 Das Metabolom

In der Biologie stammen die wichtigsten Fortschritte meist nicht von den bahnbrechenden Experimenten, sondern sie wurden möglich, weil Biologen begonnen haben, auf eine andere Art und Weise über eine Fragestellung nachzudenken. Ein Beispiel hierfür ist die Einführung des Konzepts des **Metaboloms**. Das Metabolom ist definiert als die vollständige Sammlung von Metaboliten in einer Zelle oder einem Gewebe unter bestimmten Bedingungen. Mit anderen Worten: Ein Metabolom ist eine biochemische Blaupause, und seine Untersuchung, die als **Metabolomik** oder **biochemisches Profiling** bezeichnet wird. Sie liefert eine detaillierte Beschreibung der Biochemie, die den unterschiedlichen physiologischen Zuständen, beispielsweise Krankheiten, zugrunde liegt, die von einer Zelle oder einem Gewebe angenommen werden können. Durch die Umdeutung der Biochemie einer Zelle in eine gegliederte Reihe von Metaboliten liefert die Metabolomik einen Datensatz, den man direkt mit den entsprechenden Informationen verknüpfen kann, die von der Proteomik und anderen Untersuchungen der Genomexpression zur Verfügung gestellt werden.



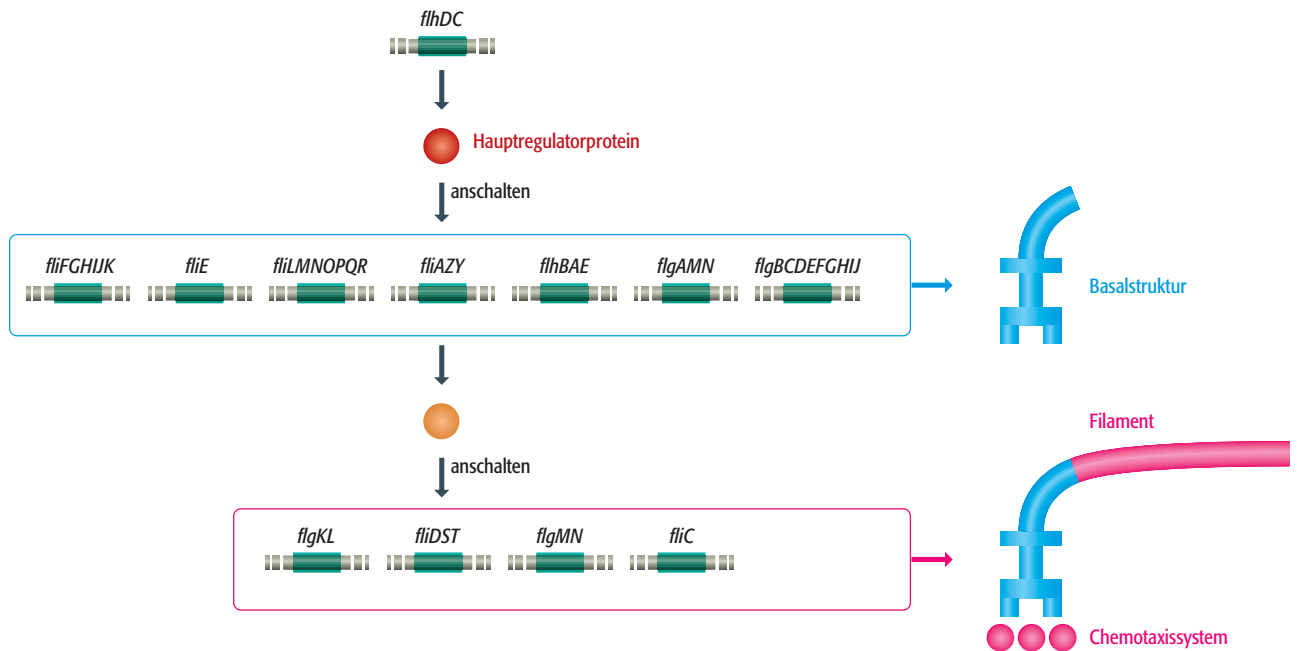
6.22 Das Proteom ist Teil der letzten Verbindung, die das Genom mit der Biochemie der Zelle verknüpft.

Ein Metabolom kann durch einzelne und in Kombination angewendete chemische Methoden charakterisiert werden, beispielweise Infrarotspektroskopie, Massenspektrometrie und Kernresonanzspektroskopie. Mit ihrer Hilfe ist die Identifizierung und Quantifizierung verschiedener kleiner Moleküle möglich, die die Metaboliten einer Zelle ausmachen. Werden diese Daten mit dem Wissen um die Reaktionsgeschwindigkeiten für die unterschiedlichen Reaktionsschritte bei gut untersuchten Stoffwechselwegen wie Glykolyse und Citratzyklus kombiniert, dann lässt sich ein Modell des **metabolischen Stoffflusses** (*metabolic flux*) aufstellen – ein Modell, das die Flussrate von Metaboliten durch das Netzwerk von Stoffwechselwegen darstellt, die die zelluläre Biochemie ausmachen. Veränderungen des Metaboloms können dann als Störungen in dem Metabolitenfluss durch einen oder mehrere Teile des Netzwerkes definiert werden, und es entsteht ein differenziertes Bild von den biochemischen Ursachen für die Veränderungen des physiologischen Zustands. Dadurch werden Möglichkeiten für das **metabolic engineering** eröffnet, bei denen man das Genom durch Mutation oder DNA-Rekombinationstechnik verändert, um die zelluläre Biochemie in einer bestimmten Art und Weise zu beeinflussen. Ein Beispiel ist die Steigerung der Syntheserate eines Antibiotikums in einem Mikroorganismus.

Zurzeit ist die Metabolomik am weitesten bei Organismen mit relativ einfacher Biochemie wie Bakterien und Hefe fortgeschritten. Ein erheblicher Forschungsaufwand zielt derzeit auf das Metabolom des Menschen ab, mit dem Ziel, metabolische Profile von gesundem und krankem Gewebe und von Gewebe aus medikamentös behandelten Patienten zu erstellen. Man hofft, dass es mithilfe der Informationen über den Stoffwechsel, die man aus diesen Studien in fortgeschrittenerem Stadium erhalten wird, möglich sein kann, Wirkstoffe zu entwickeln, die die einzelnen anormalen, im Krankheitszustand auftretenden Stoffflüsse umkehren oder abschwächen. Auch könnte biochemisches Profiling auf unerwünschte Nebenwirkungen einer medikamentösen Behandlung hinweisen, sodass die chemische Struktur des Arzneistoffes oder die Form seine Anwendung gezielt verändert werden könnten, um diese Nebenwirkungen zu minimieren.

6.3.2 Das Verstehen biologischer Systeme

Das Thema Proteininteraktionskarten und Metabolomik führt uns zu dem letzten Aspekt der Genomfunktion, den wir betrachten müssen. Es ist die Notwendigkeit, die Expression eines Genoms nicht auf Ebene der Moleküle – RNAs, Proteine und Metaboliten – zu beschreiben und zu verstehen, sondern im Zusammenhang mit dem biologischen System, das aus der koordinierten Aktivität dieser Moleküle entsteht. Es handelt sich im Wesentlichen um den Sprung von den Genen zu den Genomen, der in den vergangenen Jahren vollzogen wurde. Eines der grundlegenden Prinzipien der prägenomischen Ära der Molekularbiologie war die „ein Gen, ein Enzym“-Hypothese, die zuerst von George Beadle und Edward Tatum in den 1940er-Jahren aufgestellt wurde. Durch den Ausdruck „ein Gen, ein Enzym“ betonten Beadle und Tatum, dass ein einzelnes Gen ein einzelnes Protein codiert, welches, für den Fall, dass es ein Enzym ist, eine einzige biochemische Reaktion vermittelt. Das *trpC*-Gen aus *Escherichia coli* zum Beispiel codiert das Enzym Indol-3-glycerinphosphat-Synthase, das 1-(*o*-Carboxyphenylamine)-1'-desoxyribulose-5'-phosphat zu Indol-3-glycerinphosphat umsetzt. Allerdings arbeitet dieses Enzym nicht isoliert: Seine Aktivität ist ein Teil des Stoffwechselweges, der zur Synthese von Tryptophan



führt, wobei die anderen Enzyme dieses Weges durch die Gene *trpA*, *B*, *D* und *E* codiert werden, die zusammen mit *trpC* das Tryptophanoperon in *E. coli* bilden (Abb. 8.8b). Der Biosyntheseweg für Tryptophan ist daher ein einfaches biologisches System und das Tryptophanoperon ist die Reihe von Genen, welche diesen Weg genau bestimmen. Jedoch führt die einfache Transkription und Translation der Gene nicht zur Synthese von Tryptophan. Soll dieses System erfolgreich funktionieren, dann sind zur entsprechenden Zeit an den entsprechenden Orten in der Zelle biosynthetische Enzyme in den geeigneten relativen Mengen notwendig. Das System hängt daher von Faktoren ab, wie etwa der Syntheserate von Proteinen, die durch die Gene codiert werden, der korrekten Faltung der Proteine zu funktionellen Enzymen, der Abbaurrate der Enzymmoleküle, ihrer Lokalisierung in der Zelle und der Anwesenheit geeigneter Mengen von Metaboliten, die als Substrate und Cofaktoren für die Tryptophansynthese dienen. Dieses einfache biologische System beginnt, recht komplex zu werden. Und mit diesem System betrachten wir nur 5 der 4 405 Gene im Genom von *E. coli*.

Bis heute konnte man in der **Systembiologie** auf einer ganzen Reihe von Gebieten Fortschritte erzielen, beispielsweise bei der Untersuchung des biologischen Systems, das für die Synthese der Geißel von *E. coli* verantwortlich ist. Prägenomische Untersuchungen hatten gezeigt, dass die Geißelsynthese 51 Gene erfordert, die in 12 Operons organisiert sind, welche in drei Gruppen aktiviert werden (Abb. 6.23). Die erste zu aktivierende Gruppe enthält ein einziges Operon aus zwei Genen, die ein Protein codieren, das als Hauptregulator dient und das die Expression der zweiten Gruppe aus sieben Operons anschaltet, deren Gene zusammen die Komponenten des Basalkörpers der Geißel codieren. Eines dieser Gene codiert ein sekundäres Regulatorprotein, das die übrigen vier Operons anschaltet. Diese dirigieren die Synthese des Geißelfilaments und das biochemische System, das es dem Bakterium erlaubt, durch Geißel-

6.23 Das System, das in *Escherichia coli* an der Biosynthese der Geißel beteiligt ist.

schlag auf chemische Reize zu reagieren und auf einen Lockstoff zu zu schwimmen. Die gezielte Verwendung von Reporter genen, die mit einzelnen Operons verbunden wurden, zeigte die genaue Reihenfolge, in der die Operons jeder Gruppe aktiviert werden. Außerdem konnte man jedem Operon einen Aktivierungskoeffizient – ein Maß für die relativen Expressionsraten – zuweisen. Die erhaltene Informationen reichte aus, um das System am Computer modellieren und die genauen Funktionen der beiden regulatorischen Proteine ermitteln zu können. Diese Computermodelle bieten nun die Möglichkeit, die Auswirkung auch nur kleiner Systemveränderungen (wie die Modifikation der Eigenschaften eines der Regulatoren) vorherzusagen, die man durch weitere Experimente an dem biologischen System überprüfen kann. Das ist genau die Art von Forschung, von der wir hoffen, sie eines Tages mit menschlichen Zellen durchführen zu können, um die Ursache von Anomalitäten zu verstehen und Möglichkeiten zu entwickeln, um eine Krankheit in einen Normalzustand umzukehren. Die Übertragung der Erkenntnisse auf diese viel größeren biologischen Systeme mit dem großen Ziel, eines Tages die Funktionsweise einer bakteriellen oder eukaryotischen Zelle zu verstehen, wird die Genialität und Kreativität der Biologen in den kommenden Jahrzehnten auf die Probe stellen. Der erste Schritt wurde allerdings bereits durch eine veränderte Gewichtung vollzogen, weg von der Funktion einzelner Gene in Richtung auf die Funktionsweise des gesamten Genoms, und durch die Etablierung von Verfahren für die Untersuchung von Transkriptomen, Proteomen und Metabolomen, die zusammen die Komponenten dieser biologischen Systeme ausmachen.

Zusammenfassung

Die hauptsächliche Herausforderung der Postgenomik ist, ein Verständnis davon zu entwickeln, wie ein Genom die Vielzahl der in lebenden Zellen stattfindenden biochemischen Aktivitäten spezifiziert und koordiniert. Ein zentrales Thema in diesem Zusammenhang ist die Untersuchung des Transkriptoms und des Proteoms, die durch das Genom synthetisiert und aufrechterhalten werden. Obwohl Transkriptome durch cDNA-Sequenzierung, beispielsweise von Methoden wie SAGE, die in einem einzigen Experiment Minisequenzen von vielen cDNAs liefert, analysiert werden können, brachte die Anwendung der Microarray- und Chip-Technologie die wichtigsten Fortschritte. Die Hybridisierung unterschiedlich markierter cDNAs, die aus zwei oder mehreren Transkriptomen präpariert wurden, mit einem Microarray oder Chip, liefert Informationen über Genexpressionsmuster, die durch hierarchische Cluster-Bildung analysiert werden und so auf eine funktionelle Beziehung hinweisen können. Die Transkriptionsanalysen helfen uns, die genetische Grundlage von Entwicklungsvorgängen und Krankheiten des Menschen, beispielsweise verschiedenen Krebsformen zu verstehen. Proteomanalysen sind gleichermaßen wichtig, weil die Untersuchung des Transkriptoms lediglich Auskunft darüber gibt, welche Gene in einer bestimmten Zelle exprimiert werden und kein genaues Bild der in der Zelle enthaltenen Proteine entsteht. Für die Charakterisierung der Proteine eines Proteoms nutzt man bei einem Protein-Profiling die zweidimensionale Gelelektrophorese, gefolgt von einer MALDI-TOF der isolierten Peptidfragmente. Damit wir die Funktionsweise eines Proteoms innerhalb einer Zelle verstehen können, ist die Kenntnis der in Wechselwirkung tretenden Proteine sehr nützlich. Das Phagen-Display und *two-hybrid*-System sind die am häufigsten eingesetzten Methoden für die Identifizierung von Proteinpaaren, die direkt miteinander assoziiert sind, und man kann Methoden wie Coimmunpräzipitation einsetzen, um Multiproteinkomplexe zu isolieren. Für funktionelle Interaktionen ist es allerdings nicht immer erforderlich, dass ein Proteinpaar in direktem Kontakt steht. Solche indirekten Wechselwirkungen lassen sich aus vergleichenden genomischen Analysen, aus Untersuchungen der Genexpressionsprofile und aus Geninaktivierungsstudien ableiten. Mit den gewonnenen Informationen können Proteininteraktionskarten erstellt werden, die alle Wechselwirkungen zwischen Proteinen innerhalb eines einzigen Proteoms darstellen. Im Zentrum dieser Karten steht üblicherweise eine geringe Anzahl von Proteinen mit vielen Interaktionen, die im Netzwerk so genannte *hubs* (Knotenpunkte mit vielen Verbindungen) bilden, von denen einige einzelne biologische Prozesse repräsentieren und andere die biologischen Prozesse miteinander verbinden. Das Proteom erhält das Metabolom aufrecht – die vollständige Sammlung von Metaboliten in einer Zelle oder einem Gewebe – und aus den Untersuchungen des Metaboloms erhofft man sich Hinweise auf die genaue biochemische Grundlage von Krankheitszuständen oder unerwünschten Nebenwirkungen von Medikamenten. Die Arbeit mit Transkriptomen, Proteomen und Metabolomen führt die Biologen zur Systembiologie. Deren Anliegen ist es, die Expression eines Genoms nicht unter dem Aspekt der einzelnen Moleküle zu betrachten, deren Synthese das Genom vermittelt, sondern hinsichtlich der biologischen Systeme, die aus der koordinierten Aktivität dieser Moleküle entstehen.

Multiple-Choice-Fragen

*Antworten auf die Fragen mit den ungeraden Zahlen finden Sie im Anhang

- 6.1*** Welchen Vorteil hat die Hybridisierung eines Transkriptoms mit mRNAs, die an Ribosomen gebunden sind?
- Eukaryotische mRNA-Moleküle sind schwierig zu isolieren, wenn sie nicht im Komplex mit einem Ribosom vorliegen
 - Von Ribosomen isolierte mRNA-Moleküle werden aktiv zu Proteinen translatiert
 - Von Ribosomen isolierte mRNA-Moleküle sind stabiler als andere mRNA-Moleküle
 - mRNA-Moleküle, die nicht durch Ribosomen translatiert werden, enthalten immer noch ihre Intronsequenzen
- 6.2** Warum ist es möglich, Microarrays für die Messung der Expressionsstärke einzelner Gene einzusetzen?
- Jede Position auf einem Microarray enthält mehr Kopien der Probensequenz, als die Zahl entsprechender mRNA-Moleküle, die in dem Transkriptom vermutet wird
 - Jede Probensequenz eines Microarrays ist an vielen Positionen auf dem Array vorhanden
 - Nach der Hybridisierung werden die cDNA-Moleküle abgelöst und für jede Position des Microarrays quantifiziert
 - Die cDNA-Moleküle werden nach der Hybridisierung sequenziert und die Fluoreszenz des Sequenziersignals wird quantifiziert
- 6.3*** Warum wird Actin als Kontrolle für Transkriptionsanalysen von Vertebraten eingesetzt?
- Es wird als Negativkontrolle verwendet, da das Gen in Vertebraten nicht exprimiert wird
 - Es wird als Negativkontrolle verwendet, weil die mRNA von Actin schnell abgebaut wird
 - Es wird als Positivkontrolle eingesetzt, weil die Actinexpression in unterschiedlichen Zelltypen nahezu konstant ist
 - Es wird als Positivkontrolle verwendet, weil es in allen Zelltypen das am stärksten exprimierte Gen ist
- 6.4** Wie können zwei unterschiedliche Transkriptome mit einem einzigen Microarray analysiert werden?
- Zunächst wird ein Transkriptom hybridisiert, analysiert und seine Sequenzen entfernt, bevor das zweite Transkriptom mit demselben Microarray untersucht wird.
 - Nur eines der Transkriptome wird markiert und konkurriert mit dem zweiten, nichtmarkierten um die Bindung an die Probensequenz
 - Die Transkriptome werden vor der Microarray-Analyse miteinander hybridisiert, um die aus beiden Zelltypen stammenden cDNAs zu entfernen
 - Die beiden Transkriptome werden mit unterschiedlich fluoreszierenden Markierungen gekennzeichnet und gleichzeitig mit dem Array hybridisiert
- 6.5*** Wie werden Gene mittels hierarchischer Cluster-Bildung klassifiziert?
- Durch Expressionsmuster
 - Durch Homologie
 - Durch Sequenzübereinstimmung
 - Durch Ähnlichkeiten in den Proteindomänen
- 6.6** In welcher Weise verändert sich das Transkriptom von *Saccharomyces cerevisiae*, wenn die Zellen dauerhaft mit energiereichen Nährstoffen kultiviert werden?
- Es gibt deutliche Veränderungen in der Häufigkeit von mRNAs durch veränderte Abbau- und Syntheseraten
 - Die Häufigkeit der meisten mRNAs bleibt unverändert, doch einige schwanken während des Zellzyklus deutlich
 - Nahezu alle mRNA-Mengen bleiben konstant, nur einige wenige verändern sich
 - Die Häufigkeiten aller mRNAs bleiben unter diesen Bedingungen konstant
- 6.7*** Wie können Transkriptomanalysen die Diagnose von Krebs beim Menschen unterstützen?
- Alle Krebsformen zeigen eine verstärkte Expression einer spezifischen Reihe von Genen
 - Jeder Krebs besitzt sein eigenes einzigartiges Transkriptom
 - Die Gene, die Tumoren auslösen, werden nicht in gesunden Zellen exprimiert
 - Transkriptionsanalysen können auf die Zellteilungsrate hinweisen
- 6.8** Auf der Grundlage welches folgenden Faktors werden in der Polyacrylamidgelelektrophorese mit Natriumdodecylsulfat (SDS) Proteine getrennt?
- Verhältnis von Masse zu Ladung
 - Konformation
 - isoelektrischer Punkt
 - Größe
- 6.9*** Der isoelektrische Punkt eines Proteins ist definiert als:
- pH-Wert, bei dem ein Protein keine Nettoladung besitzt
 - pH-Wert, bei dem ein Protein seine Aktivität verliert

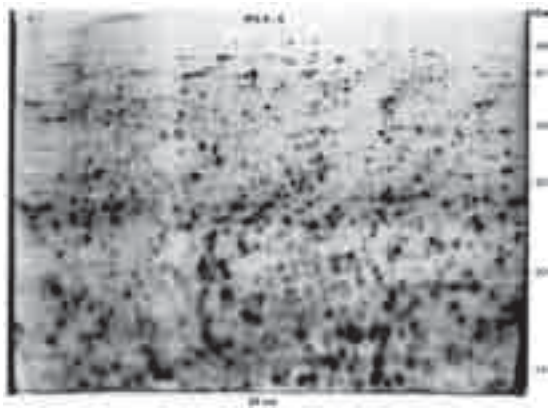
- c. pH-Wert, bei dem ein Protein seine maximale Aktivität besitzt
 - d. pH-Wert, bei dem die Aminosäuren eines Proteins alle in ionischer Form vorliegen
- 6.10** Welche folgenden Bestandteile lassen sich mit dem *two-hybrid*-System in Hefe identifizieren?
- a. alle Komponenten eines Multiproteinkomplexes
 - b. Proteine des Menschen, die für die Bindung der RNA-Polymerase erforderlich sind
 - c. zwei direkt miteinander interagierende Proteine
 - d. zwei an demselben Stoffwechselweg beteiligte Proteine
- 6.11*** Die Art der Chromatographie, bei der für die Ermittlung der bindenden Proteine ein Protein an Säulenmaterial gebunden und in einer Säule platziert wird, bezeichnet man als:
- a. Gelfiltrationschromatographie
 - b. Ionenaustauschchromatographie
 - c. Affinitätschromatographie
 - d. Isoelektrische Chromatographie
- 6.12** Was sind *hubs* in einem Proteininteraktionsnetzwerk?
- a. Es handelt sich um Proteine, die die Aktivitäten der Zelle regulieren
 - b. Es sind Proteine, die das Gerüst der Zelle bilden
 - c. Es sind Proteine, die mit vielen anderen Proteinen der Zelle interagieren
 - d. Es sind Proteine, die die Genexpression in der Zelle kontrollieren
- 6.13*** Was versteht man unter dem Metabolom einer Zelle?
- a. Alle Proteine und Nucleinsäuren einer Zelle
 - b. Alle Metaboliten einer Zelle unter bestimmten Bedingungen
 - c. Alle potenziellen Metaboliten, die von einer Zelle hergestellt werden können
 - d. Alle Makromoleküle einer Zelle

Fragen mit kurzen Antworten

*Antworten auf die Fragen mit den ungeraden Zahlen finden Sie im Anhang

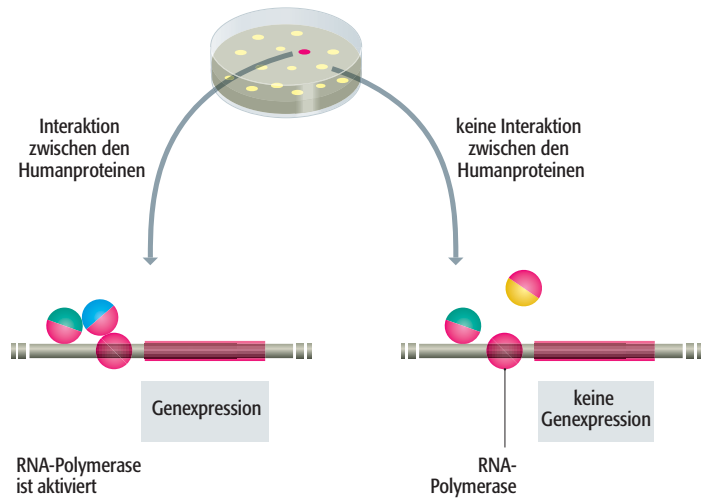
- 6.1*** Aus welchen Gründen sind Forscher an der Untersuchung von Genomen interessiert, selbst wenn allen Genen bereits eine Funktion zugewiesen worden ist?
- 6.2** Erläutern Sie, warum man für die Identifizierung eines codierenden Gens eine 12 bp kurze cDNA-Sequenz einsetzen kann.
- 6.3*** Erörtern Sie die Probleme, die paraloge Gene in Microarray-Analysen verursachen. Durch welche experimentellen Bedingungen könnte man solche Schwierigkeiten umgehen?
- 6.4** Auf welche Weise kann alternatives Spleißen bei der Charakterisierung eines Gewebetranskriptoms Probleme bereiten? Mit welchen Verfahren lassen sich die unterschiedlichen Spleißprodukte eines Gens ausfindig machen?
- 6.5*** Wie können Transkriptomanalysen Informationen über die Funktionen von Genen liefern?
- 6.6** Wie werden *tiling*-Arrays eingesetzt, um Chromosomen nach exprimierten Sequenzen abzusuchen?
- 6.7*** Warum liefert das Transkriptom keinen absolut genauen Hinweis auf das Proteom einer Zelle?
- 6.8** Wie können kleine Unterschiede in der Proteinhäufigkeit durch eine zweidimensionale Gelelektrophorese quantifiziert werden?
- 6.9*** Auf welche Weise prüfen Phagen-Display-Experimente Wechselwirkungen zwischen Proteinen?
- 6.10** Was ist der Unterschied zwischen Proteinen, die als „*party*“ *hubs* fungieren, und denjenigen, die in einem Proteininteraktionsnetzwerk die „*date*“ *hubs* darstellen?
- 6.11*** Auf welche Weise können Untersuchungen des Metaboloms die Behandlung von menschlichen Krankheiten beeinflussen?
- 6.12** Was steht im Mittelpunkt der Systembiologie und wie unterscheidet sie sich von molekularen Untersuchungen der Genregulation, die man vor der Sequenzierung von Genomen durchgeführt hat?

6.3* Erklären Sie, wie Proteinmoleküle in der zweidimensionalen Gelelektrophorese getrennt werden.



6.4 Wie prüft das *two-hybrid*-System in Hefe Interaktionen zwischen zwei unterschiedlichen Proteinen? Erläutern Sie die Aktivierung der RNA-Polymerase in diesem Experiment.

Suche nach Proteininteraktionen mit dem *two-hybrid*-System



Weiterführende Literatur

Methodik der Transkriptomanalyse

Leung YF, Cavaliere G (2003) Fundamentals of cDNA microarray data analysis. *Trends Genet* 19: 649–659

Velculescu VE, Vogelstein B, Kintzler KW (2000) Analyzing uncharted transcriptomes with SAGE. *Trends Genet* 16: 423–425

Beispiele für Transkriptomanalysen

Alizadeh AA, Eisen MB, Davis RE et al (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403: 503–511

Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, Brown PO, Herskowitz I (1988) The transcriptional program of sporulation in budding yeast. *Science* 282: 699–705

DeRisi JL, Iyer VR, Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278: 680–686 [eine der ersten Untersuchungen des Hefetranskriptoms]

Golub TR, Slonim DK, Tamayo P et al (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286: 531–537

Zhang L, Zhou W, Velculescu VE, Kern SE, Hruban RH, Hamilton SR, Vogelstein B, Kinzler KW (1997) Gene expression in normal and cancer cells. *Science* 276: 1268–1272

Protein-Profiling

Fields S (2001) Proteomics in genomeland. *Science* 291: 1221–1224 [erläutert die Bedeutung der Proteomik für das Verständnis der Humangenomsequenz]

Mann M, Hendrickson RC, Pandey A (2001) Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem* 70: 437–473

Phizicky E, Bastiaens PIH, Zhu H, Snyder M, Fields S (2003) Protein analysis on a proteomics scale. *Nature* 422: 208–215 [gibt einen Überblick über alle Aspekte der Proteomik]

Yates JR (2000) Mass spectrometry: from genomics to proteomics. *Trends Genet* 16: 5–8

Zhu H, Bilgin M, Snyder M (2003) Proteomics. *Annu Rev Biochem* 72: 783–812

Untersuchung der Wechselwirkungen zwischen Proteinen

Clackson T, Wells JA (1994) *In vitro* selection from protein and peptide libraries. *Trends Biotechnol* 12: 173–184. [Phagen-Display]

Enright AJ, Iliopoulos I, Kyripides, NC, Ouzounis CA (1999) Protein interaction maps for complete genomes based on gene fusion events. *Nature* 402: 86–90 [der Einsatz vergleichender Genomik bei der Identifizierung funktioneller Wechselwirkungen]

Fields S, Sternglanz R (2004) The two-hybrid system: an assay for protein-protein interactions. *Trends Genet* 10: 286–292

Proteininteraktionskarten

Gavin A-C, Bösch M, Krause R et al (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415: 141–147 [eine neue Hefeproteininteraktionskarte]

Han JDJ, Bertin N, Hao T et al (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* 430: 88–93 [definiert „party“ und „date“ hubs]

Jeong H, Mason SP, Barabási A-L, Oltvai ZN (2001) Lethality and centrality in protein networks. *Nature* 411: 41–42 [die erste Version der Proteininteraktionskarte der Hefe]

Lee I, Date SV, Adai AT, Marcotte EM (2004) A probabilistic functional network of yeast genes. *Science* 306: 1555–1558

Legrain P, Wojcik J, Gauthier J-M (2001) Protein-protein interaction maps: a lead towards cellular functions. *Trends Genet* 17: 346–352

Metabolomik und Systembiologie

Covert MW, Schilling CH, Famili I, Edwards JS, Goryanin II, Selkov E, Palsson BO (2001) Metabolic modelling of microbial strains *in silico*. *Trends Biochem Sci* 26: 179–186 [erläutert den Begriff des metabolischen Flusses]

Kalir S, Alon U (2004) Using a quantitative blue-print to reprogram the dynamics of the flagella gene network. *Cell* 117: 713–720

Kirchner MW (2005) The meaning of systems biology. *Cell* 121: 503–504