
PROTEINKOMPLEXE UND NETZWERKE: NEUE HERAUSFORDERUNGEN FÜR DIE PROTEOMIK

Christian von Mering und Peer Bork

*Programm für Strukturbiologie und Bioinformatik, Europäisches Molekularbiologie Labor, Heidelberg;
und Forschungsgruppe Bioinformatik, Max-Delbrück-Centrum für Molekulare Medizin, Berlin*

Die Entschlüsselung des Genoms ist zentrale Voraussetzung für ein systemweites, funktionelles Verständnis der Vorgänge innerhalb der Zelle. Funktionsträger in einem Organismus sind jedoch nicht die Gene, sondern die Gesamtheit der tatsächlich produzierten Proteine, das 'Proteom'. Die Erfassung des Proteoms ist weitaus aufwändiger als die Genomsequenzierung, da Proteine weniger gut automatisiert handhabbar sind als Nukleinsäuren. Zudem sind die anfallenden Datenmengen gewaltig: Im Unterschied zu den Genen ändert sich die Proteinausstattung eines Organismus ständig: von Gewebe zu Gewebe, während des Entwicklungszyklus, sowie in Abhängigkeit von äusseren Einflüssen.

Für eine funktionelle Charakterisierung des Proteoms sind insbesondere die Interaktionen der Proteine untereinander von Interesse – viele erfüllen ihre Funktion erst im Verbund mit anderen, und die systemweite Erfassung solcher Proteinkomplexe ist ein wichtiger Schritt hin zur präziseren Annotation von Proteinen und ihrer Wechselwirkungen. Umfassende Kenntnisse über Proteinkomplexe sind zudem oft Voraussetzung für die Identifizierung möglicher neuer Angriffspunkte für medizinische Wirkstoffe.

Im folgenden sollen erste, vielversprechende Ansätze zur Organismus-weiten Analyse von Proteinkomplexen vorgestellt werden. Zwei derartige Projekte wurden zeitgleich zu Beginn

dieses Jahres in der Zeitschrift *Nature* veröffentlicht, siehe Band 415, Seiten 141-147 und 180-183. Eine der Studien entstand in Heidelberg, in einer Kollaboration zwischen der Firma Cellzome und dem Europäischen Molekularbiologie Labor; die andere in Kanada in einer Zusammenarbeit zwischen der Firma MDS-Proteomics und Forschungseinrichtungen in Toronto. Beide Teams analysierten Proteinkomplexe in der Hefe *Saccharomyces cerevisiae*, und beide folgten dem gleichen experimentellen Grundansatz (Abb. 1).

Trotz des vergleichbaren technischen Prinzips unterscheiden sich die beiden Studien in zahlreichen methodischen Details (Abb. 2). Auch was die Auswahl der mit 'tags' versehe-

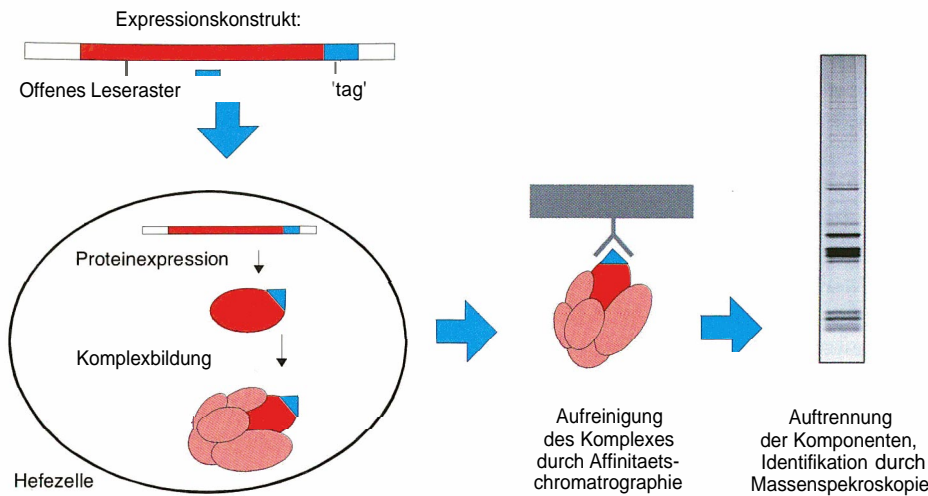


Abb. 1: Proteomweite Kartierung von Proteinkomplexen in Hefe.

Einzelne Proteine werden systematisch mittels molekularbiologischer Methoden durch kurze Ankerpeptide (sog. 'tags') markiert und in der Zelle exprimiert. Mit Hilfe der tags werden sie anschliessend aus Zellextrakten biochemisch aufgereinigt, zusammen mit evtl. gebundenen Komplexpartnern. Die gereinigten Komplexe werden durch Gelchromatographie getrennt, und die Proteine durch massenspektroskopische Analyse identifiziert. Der ganze Ablauf wird standardisiert und automatisiert, um einen hohen Durchsatz zu erreichen.

nen Proteine angeht, gibt es deutliche Unterschiede zwischen den beiden Projekten. Es wurden bei weitem nicht alle Proteine im Hefegenom analysiert – die Cellzome-Studie konzentriert sich hauptsächlich auf Proteine die in ähnlicher Form auch im Menschen vorkommen. Die MDS-Studie dagegen untersucht unter anderem eine Anzahl von Signaltransduktionsproteinen, sowie Proteine welche für die Reaktion der Zelle auf Schädigungen des Genoms essentiell sind.

Trotz der methodischen Unterschiede sind die Ergebnisse beider Projekte vergleichbar, zumindest auf den ersten Blick. Bei beiden Studien werden in den aufgereinigten Komplexen über tausend verschiedene Hefeproteine nachgewiesen (1379 bei Cellzome, 1578 bei MDS-Proteomics). Auch haben beide Projekte mit den gleichen unspezifischen Verunreinigungen zu kämpfen: Insbesondere ribosomale Proteine und andere stark exprimierte Proteine finden sich in vielen Aufreinigungen, und werden deshalb herausgefiltert. Die Grössenverteilung der gefundenen Komplexe ist ebenfalls ähnlich: Eine begrenzte Anzahl von Proteinen wird ohne assoziierte Partner gefunden (76 bei Cellzome, 110 bei MDS-Proteomics), die Mehrzahl hat fünf bis zehn Partner (im Durchschnitt 7.0 bei Cellzome, und 8.5 bei MDS), und einige haben mehr als 30 (die grösste Aufreinigung bei Cellzome enthält 54 Proteine, bei MDS 63 Proteine).

Bemerkenswerterweise gibt es auf der Ebene der einzelnen Komplexe jedoch deutliche Unterschiede – bei jenen Proteinen, die zufällig in beiden Studien mit einem 'tag' versehen wurden, ist ein relativ grosser Teil der gefunde-

nen Komplexpartner nicht identisch, 94 Proteine wurden in beiden Studien analysiert; vergleicht man deren Bindungspartner, so stimmen nur 29% der bei Cellzome gefundenen Proteine mit jenen von MDS überein, und umgekehrt nur 18% der bei MDS gefundenen mit jenen von Cellzome (MDS findet bei vergleichbaren Proteinen oft etwas mehr Komplexpartner).

Die fehlende Übereinstimmung bedeutet, dass entweder beide Studien nur einen relativ kleinen Teil der tatsächlich vorhandenen Interaktionspartner erfassen, oder dass zumindest bei einer Studie eine signifikante Anzahl an falsch-positiven Interaktionen auftritt. Möglicherweise treffen beide Erklärungen bis zu einem gewissen Grad zu. Zum einen bedingen die technischen Unterschiede zwischen beiden Projekten eine unterschiedliche Präferenz für Interaktionspartner: MDS ist aufgrund der schnelleren Aufreinigung und stärkeren Expression des markierten Proteins möglicherweise besser in der Lage, schwache und/oder transiente Bindungen zu erkennen; wogegen Cellzome durch die spezifische Aufreinigung und stochiometrische Expression eher die stabilen Bindungen erfasst. Zum anderen konnten wir im Rahmen einer vergleichenden Studie über Protein-Interaktionsdaten zeigen, dass auch ein deutlicher Anteil an falsch-positiven Interaktionen für die geringe Übereinstimmung verantwortlich ist (Nature, Band 417, Seiten 399-403). Wir untersuchten diejenigen Proteine, für welche bereits Informationen bezüglich Interaktionspartner, Funktion oder Lokalisation innerhalb der Zelle bekannt waren. Sowohl im Cellzome-Datensatz, als auch bei MDS fanden

sich innerhalb dieser Proteine Interaktionspartner, die zumindest anhand ihrer bisherigen Annotationen nicht gemeinsam in einem Komplex zu erwarten gewesen wären. Dies betraf fehlende Übereinstimmung bei zellulärer Lokalisation, funktioneller Kategorie oder bereits bekannter Komplexzugehörigkeit. Derartige Diskrepanzen tauchen besonders im MDS-Datensatz auf, was möglicherweise darauf zurückzuführen ist, dass durch Überexpression und schnellere Aufreinigung mehr Kontaminanten detektiert werden.

Trotz eines gewissen Anteils an falsch-positiven Interaktionen stellen beide Studien beeindruckende Informationsquellen zu Proteinkomplexen dar. Die vielleicht überraschendste Erkenntnis daraus ist die Vernetzung der einzelnen Komplexe. Zahlreiche Proteine werden in mehr als einem Komplex gefunden, sodass ein Netzwerk aus Protein-Protein Interaktionen entsteht. Dieses Netzwerk verdeutlicht die Komplexität der Regulierung und Verzahnung von Proteinfunktion, selbst in einem vergleichsweise 'einfachen' eukaryontischen Modellorganismus wie der Hefe. Die Vernetzung geht oft so weit, dass es schwierig wird, die einzelnen Komplexe klar voneinander abzutrennen. Auch die falsch-positiven Interaktionen erschweren dies; hier jedoch hilft die Tatsache, dass mehrere Proteine eines Komplexes unabhängig voneinander mit einem 'tag' versehen werden können – dies erlaubt die wiederholte Analyse eines Komplexes von verschiedenen Angriffspunkten aus, und damit eine interne Überprüfung der Ergebnisse (siehe z.B. Abb. 3).

Die beiden hier vorgestellten Studien repräsentieren einen Durchbruch in der systematischen Analyse von Proteinkomplexen, für sich alleine reichen sie jedoch zur angestrebten vollständigen Charakterisierung von Protein-Interaktionen nicht aus. Unterrepräsentiert sind Proteine in schlecht zugänglichen Zellkompartimenten wie z.B. Membranen oder Lysosomen, sowie Proteine die nur transient existieren oder nur unter bestimmten Bedingungen interagieren. Die Studien sind Hochdurchsatz-Studien, und man wird daher auf komplementäre Techniken wie z.B. die 'Yeast Two-Hybrid' Methode zurückgreifen, um detaillierte Einzeluntersuchungen durchzuführen. Darüberhinaus werden in Zukunft auch vermehrt miniaturisierte Technologien wie z.B. Proteinchips eingesetzt werden. Die hier vorgestellten neuen Interaktionsdaten stellen jedoch im Moment eine einzigartige Informationsquelle dar, die sich hervorragend eignet um fundamentale Erkenntnisse über Struktur und Organisation von Proteinnetzwerken zu gewinnen.

Abb. 2: Technische und konzeptionelle Unterschiede zwischen den beiden Studien zur systematischen Erfassung von Proteinkomplexen.

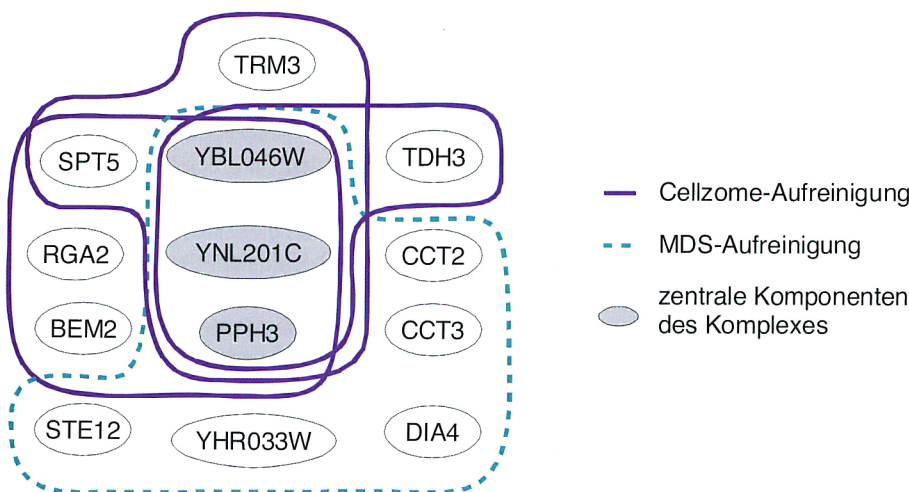


Abb. 3: Einer der zahlreichen neu entdeckten Proteinkomplexe.

Das Protein 'PPH3' ist eine Serin/Threonin-Phosphatase aus der Familie der PP2A-Phosphatasen. Das Hefegenom enthält mehrere Phosphatasen dieser Familie; PPH3 wurde bisher noch nicht eingehend charakterisiert. Durch Analyse der Komplexaufreinigungen von Cellzome und MDS-Proteomics wird ein bisher unbekannter Proteinkomplex um PPH3 sichtbar. Die beiden uncharakterisierten Proteine 'YBL046W' und 'YNL201C' bilden zusammen mit PPH3 den sicher identifizierten Kern des Komplexes, da sie in allen vier relevanten Aufreinigungen gefunden werden. Die Situation ist weniger klar für die anderen Proteine - 'SPT5' wird zwar in zwei Aufreinigungen detektiert, allerdings auch noch in anderen Komplexen (nicht gezeigt; SPT5 ist ein Beispiel für die 'Vernetzung' der Komplexe untereinander).