

# Proteomik: Proteine im (Hefe-)Kontext

P. Uetz, ITG

## Einleitung

Proteine sind die aktiven Bestandteile aller lebenden Zellen. Als Enzyme katalysieren sie fast alle Reaktionen eines Organismus. Als Strukturproteine prägen sie dessen Gestalt von der Zelle bis zum kompletten Lebewesen. Dazu kommen noch regulatorische Proteine, welche die Dynamik lebender Systeme steuern und einige andere.

Traditionell werden Proteine einzeln untersucht, oft ein Protein von mehreren Arbeitsgruppen zugleich. Die Analyse von Proteinen ist mühselig, weshalb in Einzelfällen Tausende von Wissenschaftlern an einem oder wenigen Proteinen arbeiten. Allein zum Thema HIV (= Human Immunodeficiency [= AIDS] Virus, das aus nur ca. 10 Proteinen besteht) werden pro Jahr Tausende von Arbeiten veröffentlicht. Dabei entsteht eine seltsame Diskrepanz: *wenige*, als wichtig erachtete Proteine werden von *sehr vielen* Molekularbiologen untersucht und *sehr viele* Proteine von *wenigen* oder gar keinen Wissenschaftlern. Sicher wird man mehr Ehrgeiz darauf verwenden, wichtige Proteine mit mehr Personalaufwand zu untersuchen als vergleichsweise unwichtige. Nur: woher weiß man, ob ein Protein wichtig ist, solange man es noch nicht untersucht hat? Selbst wenn man weiß, dass ein bestimmtes Protein wichtig ist, reicht es meistens trotzdem nicht aus, nur das eine Protein zu untersuchen. Illustrieren lässt sich das gut an einem Protein namens „Myc“, das u.a. eine Form des Knochenmarkkrebses verursacht

(und schon dadurch „wichtig“ ist). Der Name „Myc“ stammt von den im Knochenmark vorkommenden „Myelocyten“, aus denen sich bestimmte weiße Blutzellen bilden.

## Ein Beispiel: das Krebsprotein Myc

Das Myc Protein ist eines der meiststudierten Ursachen von Krebs. Allerdings war es ein langer Weg bis man die Funktion des Proteins geklärt hatte [1]. Myc bindet als Transkriptionsfaktor an DNA und schaltet dabei andere Gene an, die wiederum eine wichtige Rolle bei der Zellteilung spielen. Als man diese DNA-Bindungsaktivität 1991 entdeckte [2], waren bereits mehr als 2000 wis-

senschaftliche Artikel über Myc veröffentlicht worden! Die wichtigste Funktion wurde in diesen 2000 Studien nicht erkannt: Myc braucht einen Partner, um an DNA zu binden, nämlich ein anderes Protein namens Max. Beide Proteine bilden einen Komplex, der im Gegensatz zu den einzelnen Proteinen an DNA binden kann. Damit ist die Geschichte aber noch nicht zu Ende: die Aktivität von Max wird durch ein drittes Protein reguliert, das vielleicht nicht unpassend Mad getauft wurde und das wiederum eine Bindung mit Max eingeht um den Komplex Mad-Max zu bilden. Die Geschichte ist damit aber immer noch nicht zu Ende – weitere Proteine binden an Myc und Max

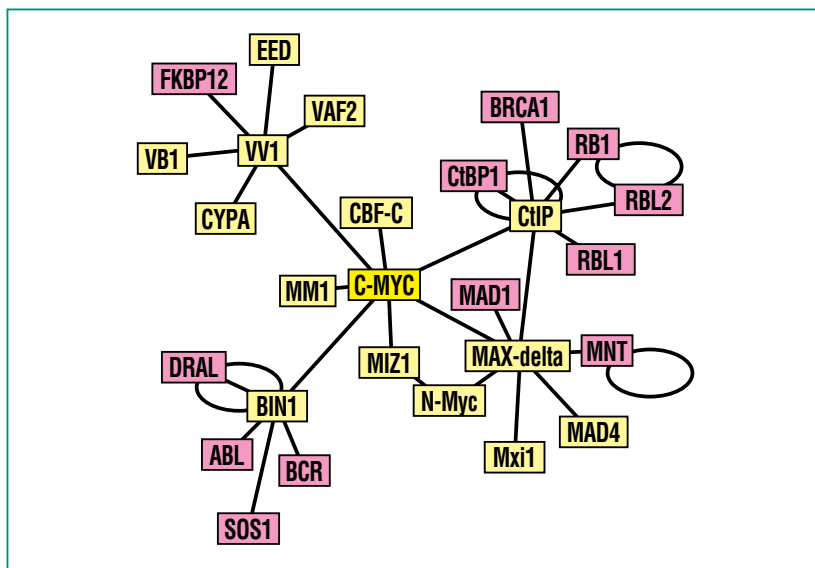


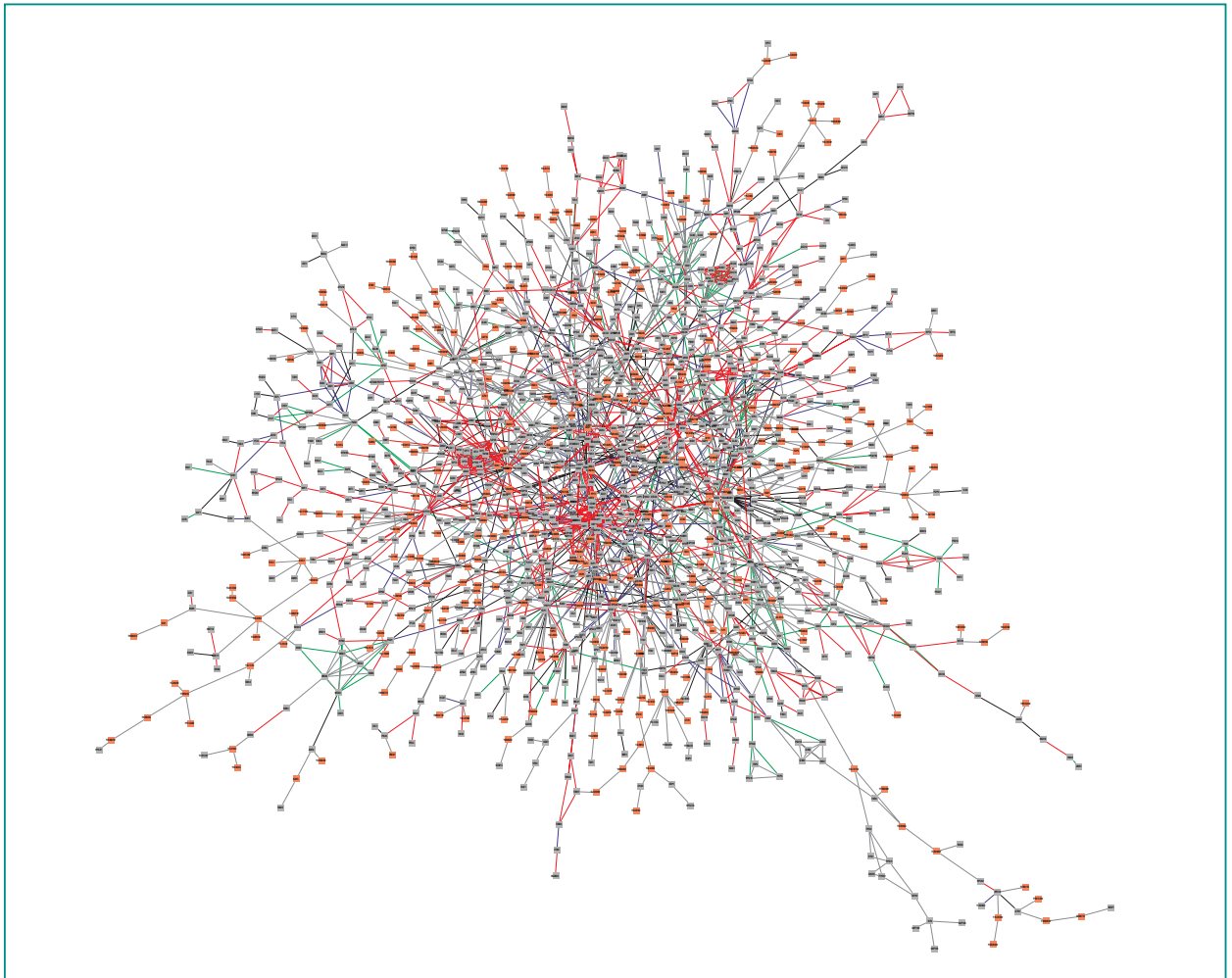
Abb. 1: Das c-Myc-Interaktions-Netzwerk. c-Myc (gelb) und seine Interaktionspartner (das „c“ steht für „cellulär“ im Gegensatz zu „v-Myc“ oder „N-Myc“, die in Viren bzw. Nervenzellen gefunden wurden). Beachte, dass c-Myc nur mit einem Interaktionspartner wie Max an DNA binden kann (Details im Text). Protein-Protein-Interaktionen sind hier als gerade Verbindungslinien dargestellt. Ovale deuten an, dass das Protein links oben am Oval mit sich selbst interagieren kann (also ein Homodimer bildet). Alle hier rot gezeigten Proteine haben wiederum weitere Interaktionspartner etc. (Daten und Diagramm aus [8]).

und an diese binden wiederum andere (Abb. 1). Wenn aber Myc für seine Aufgabe Max braucht und Max braucht Mad, und Mad wiederum andere Proteine, dann kann man Proteine wie Myc (oder Max oder Mad oder...) auch nur im Kontext ihrer Interaktionen verstehen. Tatsächlich stellte es sich erst vor 1-2 Jahren immer mehr heraus, dass die meisten Proteine einer Zelle über solche Protein-Protein-Interaktionen in komplexen Netzwerken miteinander verbunden sind (s. Abb. 2).

### Myc-Mad-Max: Krebsproteine ... und Hefe?

Offensichtlich reicht es nicht, wenn man nur einzelne Proteine untersucht, seien sie auch noch so wichtig (obwohl viele Krebsproteine *per definitionem* wichtig sind, lösen ihre ebenso wichtigen Bindungspartner oft keinen Krebs aus). Aus diesem Grund haben wir beschlossen, Protein-Protein-Interaktionen im „Gesamtkontext“ zu untersuchen, also am besten gleich Proteom-weit, d.h. *alle*

Proteine eines Organismus umfassend. Begonnen habe ich damit bereits als Postdoc in Seattle, wo wir 1997 angefangen haben alle Proteine der Bäckerhefe auf Ihre Interaktionen zu testen [3]. Aber warum gerade Hefe? Es gibt verschiedene gute Gründe und nur einer davon ist: Die Hefe war und ist *der* Modellorganismus schlechthin: die Hefe war der erste höhere Organismus, dessen Gene alle bekannt wurden, der erste Organismus, für den man DNA-Chips hergestellt hat, der



**Abb. 2: Proteinnetzwerk in Hefe. Ein Netzwerk verbindet 1548 Proteine über 2358 Interaktionen. Proteine bekannter Funktion sind hier in Grau, Proteine unbekannter Funktion in Rot dargestellt. Beachte, dass hier nur ein Teil aller Hefeproteine dargestellt ist! (nach [7]).**

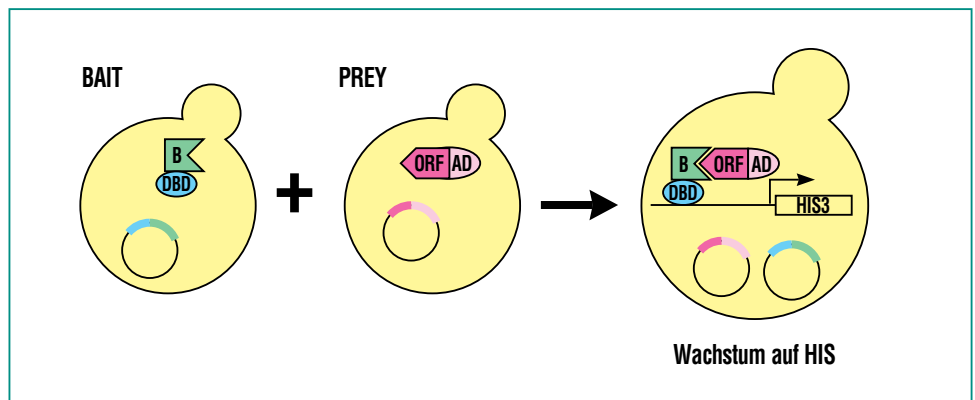
erste bei dem systematische Studien an allen Proteinen durchgeführt wurden und so weiter. Und immerhin rund ein Viertel der 6000 Hefegene haben verwandte Gene beim Menschen und sind somit auch ein sehr gutes Modell für zahlreiche Prozesse, wie sie sich auch in menschlichen Zellen abspielen. Man kann mit Fug und Recht sagen, dass alle wichtigen Prozesse bei Hefezellen und menschlichen Zellen gleich oder sehr ähnlich ablaufen. Die gestörte Zellteilung, wie sie bei menschlichen Tumoren auftritt, wurde beispielsweise zunächst bei der Hefe untersucht. Drei Pioniere der *Hefegenetik*, Leland Hartwell, Tim Hunt und Paul Nurse, wurden dafür 2001 mit dem Medizin-Nobelpreis ausgezeichnet!

ren und dann ihre Zusammensetzung mit Hilfe der Massenspektrometrie zu bestimmen. Aus der Genomsequenz kann man ja die genauen Massen aller darin kodierten Proteine vorhersagen und mit den experimentell ermittelten Massen vergleichen. Seit kurzem gibt es Projekte, die solche Analysen von Proteinkomplexen systematisch im Hochdurchsatzverfahren durchführen. Auch hier war die Hefe wieder der wichtigste Modellorganismus, bei dem diese Methoden entwickelt und dadurch auch zuerst angewandt wurden [4]. Leider liefert die Massenspektrometrie keine Information über die Anordnung der Proteine in einem Proteinkomplex. Es gäbe natürlich die Möglichkeit, die Struktur der Komplexe mit kristallographischen Methoden zu

untersuchen; auf Grund ihrer Größe ist dies bei vielen Proteinaggregaten aber nicht möglich. Wir verwenden daher das sogenannte *Two-hybrid-System*, eine Methode, die auf einem genetischen Trick basiert: man kann eine Zelle dazu veranlassen nur dann zu wachsen, wenn zwei bestimmte Proteine interagieren (Abb. 3). Damit kann man eine Proteininteraktion an einer simplen Hefekolonie gleichsam „ablesen“. Da man alle Proteine in der Hefe kennt, kann man auch alle Proteine systematisch paarweise auf solche Interaktionen testen – bei der Hefe sind dies rund 6000 mal 6000 = 36 Millionen Paare (wenn man alle Kombinationen testen will)! Freilich ist dies mit einem ziemlich großen Aufwand verbunden, aber glückli-

### Wie man komplexe Proteinnetzwerke untersucht

Nachdem das Hefegenom vollständig sequenziert war, kannte man zwar die darin kodierten 6000 Proteine, aber nicht deren Funktion und Anordnung in der Zelle. Oft gibt die Sequenz eines Proteins bereits einen Hinweis auf die Funktion oder den Ort innerhalb der Zelle. Beispielsweise enthalten mehr als 1000 Hefeproteine hydrophobe Aminosäure-Abschnitte und deuten damit an, dass sie in einer der ebenso hydrophoben Membranen stecken. Aber in welchen Membranen und in welcher Anordnung ist daraus meistens nicht ersichtlich. Eine Möglichkeit die Umgebung eines Proteins aufzuklären, besteht z.B. darin, alle damit zusammenhängenden Proteine in Form eines „Proteinkomplexes“ zu isolie-



**Abb. 3: Prinzip des Two-hybrid-Systems.** Das Two-hybrid-System beruht auf der Expression zweier Fusionsproteine in einer Zelle, die in diesem Fall durch die Verpaarung von zwei haploiden Hefezellen erreicht wird, von denen jede ein anderes Fusionsprotein exprimiert. Eines der Fusionsproteine besteht aus einer DNA-bindenden Domäne (DBD), die an den Promoter eines Reportergens (hier: His3) binden kann, und einem Protein „B“ (für „bait“ = Köder). Das zweite Fusionsprotein besteht aus einer Transkriptionsaktivierungsdomäne (AD) und einem Protein ORF (für *open reading frame* = offenes Leseraster, also ein beliebiges Protein. *Prey* = Beute). Interagiert B mit ORF wird ein Transkriptionsfaktor gebildet, der das Reportergen anschalten kann. Dadurch kann die Zelle auf Histidin-freiem Medium wachsen. Eine wachsende Hefekolonie zeigt damit an, dass die beiden Proteine interagieren (Details auf unserer Webseite [9]).



Abb. 4: Roboter, wie er von uns für automatisierte Two-hybrid-Screens verwendet wird.

cherweise lassen sich solche Tests automatisieren und damit für recht große Proteinzahlen realisieren (Abb. 4).

### Was lernt man nun daraus ... und was nicht?

Die Analyse von Proteininteraktionen hat natürlich zum Ziel, die Funktion einzelner Proteine zu verstehen. Da ein Protein seine Aufgaben aber fast immer in Zusammenarbeit mit anderen Proteinen erfüllen kann, sind seine Interaktionen von zentraler Bedeutung für dessen Funktion – genauso wie die Teile eines Automotors nicht alleine arbeiten können um einen Wagen zu bewegen. Wie bei einem Auto ist das Schlüsselwort aber auch hier „bewegen“. Proteininteraktionen werden meistens zuerst als *Zustände* beschrieben und nicht als *Aktionen* (auch wenn das Wort

*Aktion* in *Interaktion* drin steckt!). Und eine solche Zustandsbeschreibung sagt meistens nicht viel über die Konsequenzen einer Interaktion aus, die man deshalb im nachhinein noch gezielter untersuchen muss! In unserem Beispiel mit Myc bewirkt die Bindung von Max, dass der Komplex Myc-Max *an DNA binden* kann, was wiederum eine weitere *Aktion* nach sich zieht, nämlich das *Anschalten* eines Gens (also dessen Transkription), was wiederum weitere Effekte zur Folge hat. Umgekehrt verhindert die Bindung von Mad an Max, dass Max an Myc binden kann, womit die Aktivität von Myc *reguliert* werden kann.

Man kann also über Proteininteraktionen viel über die Funktion einzelner Proteine lernen, wobei die Interaktion aber nur der erste Schritt zum Verständnis eines Proteins sein kann. Nun erfordern

Experimente, welche die funktionellen Konsequenzen solcher Interaktionen erforschen, einen weit größeren Aufwand als die Identifikation der Interaktion an sich. Damit wird man vor das Dilemma gestellt, entweder viele Interaktionen zu identifizieren (ohne jedoch viel über die Funktionen zu lernen) oder wenige einzelne Interaktionen genauer zu studieren – mit der Hoffnung, ihre Funktion genau zu verstehen. Im Moment konzentrieren wir uns noch darauf, möglichst viele Interaktionen zu finden, schlicht weil bisher nur ein Bruchteil aller Proteininteraktionen bekannt ist (in der Hefe ca. 8000 von bis zu 30.000 [5,6] – beim Menschen ist das Verhältnis zwischen bekannten und allen Interaktionen sicher noch viel kleiner). Je mehr Interaktionen aber bekannt sind, desto klarer wird auch der Kontext, in dem ein Protein seine Funktion entfaltet und desto gezielter lassen sich wiederum Experimente zu seiner genauen Funktion anstellen. Ein weiteres Ziel unserer Arbeit ist es nicht zuletzt, solche „gezielten“ Experimente für eine größere Anzahl von Proteinen parallel durchzuführen, also auch schwierigere Experimente zu automatisieren.

Aus größeren Datensätzen lassen sich aber nicht nur Einsichten in die Funktion einzelner Proteine gewinnen, sondern auch ein Verständnis der Zusammenhänge auf „höherer“ Ebene. So konnten wir mit einer theoretischen Analyse aller bekannten Proteininteraktionen in der Hefe zeigen, wie verschiedene Funktionen in einer Zelle vernetzt sind [7]. Beispielsweise interagieren Proteine des

Zellzyklus mit vielen anderen Proteinklassen – was darauf hindeutet, dass letztere im Lauf des Zellzyklus (also während der Teilung) gezielt reguliert werden. Andere Prozesse wie der Proteintransport zwischen einzelnen Organellen werden offensichtlich nicht vom Zellzyklus beeinflusst. Man kann daraus z.B. schließen, dass solche Prozesse für die Krebsentstehung keine maßgebliche Rolle spielen sollten, wo Zellzyklusproteine ja *immer* beteiligt sein *müssen*.

### Epistemologischer Epilog – Prinzipien versus Details in der Biologie

Das vorher genannte Dilemma, entweder einzelne Details sehr gründlich oder größere Zusammenhänge eher oberflächlich zu studieren, führt zu einer fundamentalen Frage: reicht es für das Verständnis eines biologischen Organismus aus, anhand einer kleinen Anzahl von Proteinen dessen Funktions*prinzipien* zu verstehen?

Die Biologie und besonders die Molekularbiologie haben sich in den letzten 40 Jahren vor allem um die Aufklärung von Mechanismen gekümmert, also zum Beispiel „Wie funktioniert der genetische Code?“ oder „Wie werden Nervensignale übertragen?“. Ernüchternd ist, dass selbst ein gutes Verständnis dieser Mechanismen die realen Organismen nicht wirklich erklärt. Es gibt zwei Gründe für diese Diskrepanz: erstens berücksichtigen Prinzipien nicht die Komplexität eines Lebewesens. Beispielsweise sind die

Mechanismen der Signalübertragung im Nervensystem relativ gut verstanden, aber zur Informationsspeicherung reicht eben eine Synapse nicht aus – man braucht davon sehr viele, komplex verschaltete Synapsen. Zweitens unterliegen Prinzipien zahlreichen Variationen, so wie die Synapsen im Gehirn viele verschiedene Transmitter, Signaltransduktionswege, und physiologische Zustände besitzen. Diese mehr oder weniger großen Variationen erklären wiederum *im Prinzip* die Unterschiede zwischen Arten, also z.B. zwischen Rattenhirn und Menschenhirn, aber eben nur im Prinzip. Man ist sich dieser Variationen gewiss, ohne jedoch viele konkrete Details zu kennen.

Die Schlussfolgerung daraus ist, dass wir nicht nur die Mechanismen verstehen, sondern die *Details* kennen müssen, die Ratten- und Menschenhirn unterscheiden. Man muss aber nicht einmal bis zur Hirnforschung gehen: selbst um eine einzelne Hefezelle zu verstehen, reicht es nicht, wenn man die Details der Transkription, der Replikation etc. kennt. Man müsste wissen, wie *jedes einzelne* Gen transkribiert wird, wie *jedes einzelne* Protein gefaltet wird, wie stabil *jede einzelne* mRNA und *jedes einzelne* Protein ist, ihre Funktion, Struktur *et cetera*. Prinzipien erklären zwar *virtuelle* Zellen aber keine *realen*. Aus diesem Grund kommen wir nicht umhin, *alle* Gene, Transkripte und Proteine zu untersuchen. Dies erfordert den massiven Einsatz von parallelen Untersuchungsverfahren, d.h. letztendlich Automatisierung und

Robotik – die Alternative wäre langweilige, vermutlich jahrzehntelange Routinearbeit.

Gene oder Proteine auf Eigenschaften zu untersuchen, auf die man schon Dutzende oder Hunderte andere Gene oder Proteine untersucht hat, mag nicht besonders originell sein – vor allem weil die Wahrscheinlichkeit dramatisch abnimmt, neue „Prinzipien“ und damit aufregende neue Mechanismen zu finden. Aber diese Art von Forschung ist trotzdem von entscheidender Bedeutung in der Biologie, die nun einmal von *wenigen* Prinzipien und endlos *vielen* Variationen über diese Themen geprägt ist – erst dadurch kommt ja auch die unglaubliche Vielfalt an Lebewesen zustande – obwohl es nur ein paar Dutzend „Hauptbaupläne“ gibt. Die Herausforderung wird deshalb darin bestehen, biologische Phänomene aus einer Unzahl von „Faktoiden“ zu erklären und gleichzeitig aus der Masse an Daten wiederum „emergente“ Prinzipien herauszufiltern, die eher Systemeigenschaften denn Eigenschaften einzelner Bauteile sind. Das wird die Hauptaufgabe der „Systembiologie“ im 21. Jahrhundert sein, der wir uns stellen müssen.

---

---

## Literatur

- [1] B. Lüscher,  
*Function and regulation of the transcription factors of the Myc/Max/Mad network.*  
*Gene* 277 (2001) 1
- [2] E.M. Blackwood, R.N. Eisenman,  
*Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc.*  
*Science* 251 (1991) 1211
- [3] P. Uetz, et al.  
(2000) *A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae.*  
*Nature* 403 (2000) 623
- [4] A.-M. Gavin, et al.  
*Functional organization of the yeast proteome by systematic analysis of protein complexes.*  
*Nature* 415 (2002) 141
- [5] C.L. Tucker, et al.  
*Towards an understanding of complex protein interaction maps.*  
*Trends in Cell Biology* 11 (2001) 102
- [6] A. Kumar, M. Snyder,  
*Protein complexes take the bait.*  
*Nature* 415 (2002) 123
- [7] B. Schwikowski, et al.  
*A network of interacting proteins in yeast.* *Nature Biotechnology* 18 (2000) 1257
- [8] <http://pronet.doubletwist.com/>  
Datenbank wurde Anfang Februar geschlossen
- [9] <http://itgmv1.fzk.de/www/itg/uetz/>