

Computergestützte Auswertung von Protein-Protein-Interaktionen

Jorge Silva¹, Rainer Stotzka¹,
Nicole Rüter¹, Peter Uetz²

¹Institut für Prozessdatenverarbeitung und Elektronik,

²Institut für Toxikologie und Genetik,

Forschungszentrum Karlsruhe, 76344 Eggenstein

Email: stotzka@ipe.fzk.de

Zusammenfassung. Bei der Untersuchung von "Two-Hybrid"-Systemen handelt es sich um eine visuelle Methode für die Detektion von Protein-Protein-Interaktionen. Dabei müssen Matrizen mit 384 Feldern auf wachsende Hefekolonien mit untersucht und die Ergebnisse in eine Datenbank eintragen werden. Die Auswertung ist subjektiv und zeitaufwändig. Die vorliegende Arbeit präsentiert ein erstes computergestütztes System, um digitale Bilder von solchen "Two-Hybrid"-Systemen automatisch auszuwerten.

Satz zu den Ergebnissen

1 Protein-Protein-Interaktionen

Proteine sind die aktive Bestandteile aller lebenden Zellen. Sie erfüllen sehr vielfältige Aufgaben, z. B. katalysieren sie als Enzyme fast alle Reaktionen eines Organismus und als Strukturproteine prägen sie dessen Gestalt von der Zelle bis zum kompletten Lebewesen. Dazu kommen regulatorische Proteine, welche die Dynamik lebender Systeme steuern und einige andere [1]. Traditionell werden Proteine einzeln untersucht. Man hat aber schon lange festgestellt, dass die meisten Proteine ihre Aufgabe in Zusammenarbeit mit anderen Proteinen erfüllen und nicht alleine. Daher ist eine Beschreibung der Protein-Interaktionen in einer Zelle für das Verständnis der Zellstruktur und der dynamischen Prozesse in der Zelle notwendig [2].

In der heutigen Forschung ist die Hefe der Modellorganismus schlechthin, weil sie der erste höhere Organismus ist, dessen Gene sequenziert und bei dem systematische Studien an allen Proteinen durchgeführt wurden. Nachdem das Hefegenom vollständig sequenziert war, kannte man zwar die darin kodierten 6000 Proteine, aber nicht deren Funktion und Anordnung in der Zelle. Aus diesem Grund wurde gleich nach Abschluss der Sequenzierung des Hefegenoms im Jahre 1996 begonnen, Protein-Interaktionen in der Hefe systematisch zu untersuchen [1].

Ein „Two-Hybrid“-System ist eine genetische Methode, Protein-Protein-Interaktionen zu detektieren. Es basiert auf einem genetischen Trick, bei dem

die Zelle veranlasst wird, nur dann zu wachsen, wenn zwei bestimmte Proteine interagieren. Damit kann man eine mögliche Protein-Interaktion an einer simplen Hefekolonien ablesen. Da alle Proteine in der Hefe bekannt sind, können alle Proteine systematisch paarweise auf solche Interaktionen getestet werden. Bei 6000 verschiedenen Proteinen der Hefe existieren auf 36 Millionen Kombinationsmöglichkeiten. Auf Grund der Komplexität von biologischen Prozessen ist es notwendig, die Gesamtzahl dieser Kombinationen zu untersuchen, um eine vollständige Beschreibung der Aktivitäten in diesem Organismus zu erzielen.

Zur Automatisierung der Versuche werden die Hefekolonien in Array-Form, sogenannte Two-Hybrid-Screens, unter Verwendung eines Roboters angeordnet. In jeder Kolonie wird ein bestimmtes Paar Proteine exprimiert. Diese Arrays ermöglichen die systematische Untersuchung aller möglichen Protein-Paare auf Interaktion. Kolonien mit positiver Interaktion sind als weiße Spots auf den Two-Hybrid-Screens zu erkennen. Durch das Array-Format werden die Einzelversuche reproduzierbar und vergleichbar und die Feststellung von einzelnen Fehldetektionen [3] wird vereinfacht. Zur Zeit erfolgt die Auswertung der Two-Hybrid-Screens manuell durch einen Experten, der die Ergebnisse sieht und in eine Datenbank einträgt.

Der Nachteil solcher manuellen Auswertungen liegen in den subjektiven Interpretationen, die nicht immer zu reproduzierbaren Ergebnissen führen, dem Zeitaufwand der Experten und den damit verbundenen Kosten. In der vorliegenden Arbeit wird ein erstes computergestütztes System für die automatische Auswertung von Two-Hybrid-Screens vorgestellt, die die Auswertezeit von mehreren Minuten auf wenige Sekunden pro Screen reduziert.

2 Computergestützte Auswertung

Für die automatische Auswertung der Two-Hybrid-Screens wurde eine System-Architektur [4] entwickelt, die mehrere Clients gleichzeitig bedienen kann (siehe Abb. 1). Die Datenaufnahme erfolgt durch eine handelsübliche Digitalkamera. Auf einem Client werden lokal die aufgenommenen Bilder gespeichert. Über ein Web-Interface werden die Bilder zu einem Auswerteserver gesendet. Auf dem Server wird die Bildauswertung durchgeführt. Die Ergebnisse werden in einer XML-Seite zusammengefasst und zurück zum Client geschickt. Dort überprüft ein Experte die Ergebnisse und kopiert sie in eine Datenbank.

Die Bildauswertung erfolgt auf dem Auswerteserver in fünf Schritten (siehe Abb. 2): Vorverarbeitung, Registrierung, Segmentierung, Quantifizierung und Klassifikation.

Durch die Bildaufnahme mittels einer Digitalkamera kann die Bildqualität aufgrund der Beleuchtung und Rauschen leicht variieren.

Bei der **Vorverarbeitung** handelt es sich um Kontrastverbesserung und Filterung von Störungen. Zuerst wird das Farbbild, bei dem ein Weißabgleich von der Kamera automatisch durchgeführt wurde, in ein 8-bit-Grauwertbild umgewandelt. Durch eine Grauwertspreizung auf den gesamten Grauwertbereich (0

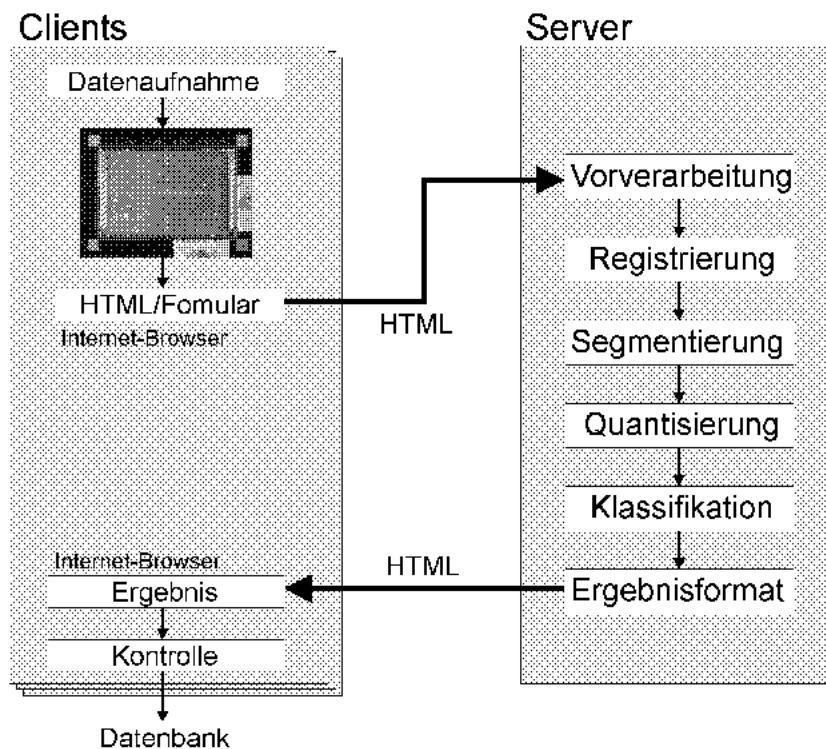


Abb. 1. Software-Struktur. Die Datenaufnahme erfolgt durch eine Digitalkamera. Über ein Web-Interface kommen die Bilder zu einem Auswerteserver. Auf dem Server werden die einzelnen Schritte der Bildauswertung durchgeführt und das Ergebnis entsprechend formatiert zum Client gesendet. Ein Experte kontrolliert die Ergebnisse und überträgt sie in einer Datenbank.

bis 255) wird der Kontrast verstärkt. Anschließend wird durch mehrfache Anwendung eines Median-Filters mit einer 3x3-Maske das Pixelrauschen reduziert.

Das Hefewachstum an einer bestimmten Stelle soll einem der 384 Felder im Two-Hybrid-Screen zugeordnet werden. Die Lage der Hefekolonien kann auf den Bildern verschoben, rotiert und in leicht unterschiedlicher Skalierung vorliegen. Durch eine rigide **Registrierung** werden die Bilder mit einem Gitter-Template in Übereinstimmung gebracht. Die Quadrate an den Ecken des Screens werden als Marker für die Ausrichtung verwendet. Nach der Anwendung eines geeigneten Schwellwertes auf das zu registrierende Bild wird nach den Ecken der Quadrate gesucht, die nach der Schwellwertoperation weiß erscheinen. Aus deren Position und der Position der Quadrate im Gitter-Template werden die Matrizen für die affine Transformation des Bildes berechnet und das Bild wird transformiert. Auf diese Art und Weise kann jedes Pixel des auszuwertenden Bildes eindeutig einer Gitterposition zugeordnet werden.

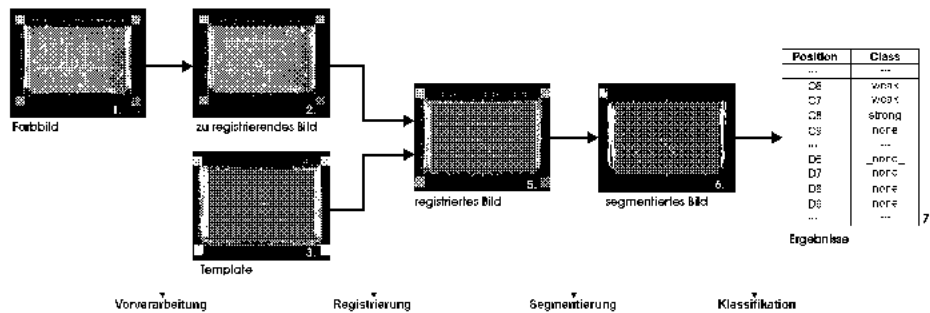


Abb. 2. Bilddauswertung. eqwuiof

Für die **Segmentierung** der Spots wird das Grauwertbild mit einer Schwellwertoperation in ein Binärbild umgewandelt. Der Schwellwert wird so gewählt, dass weiße Punkte innerhalb der Gitter den Spots mit Hefewachstum entsprechen.

In den Arbeitsschritten **Quantifizierung** und **Klassifikation** wird die Größe der Spots gemessen und einer Klasse zugeordnet. Mit der Anzahl der weißen Pixel in einem Gitterfeld wird das Hefewachstum in eine der Klassen „none“ (kein Wachstum), „weak“ (schwaches Wachstum) und „strong“ (starkes Wachstum) eingeordnet. Die Ergebnisse werden in XML formatiert und zum Client gesendet.

Die Server-Software wurde als Web-Service in JAVA unter Verwendung von „KHOROS“, „ITK“ und der im Forschungszentrum Karlsruhe entwickelten Komponentensoftware „ICE“ auf einem Linux-Rechner implementiert und getestet.

3 Ergebnisse

Der Erfolg bei der Erkennung von Kolonien mit positiver Interaktion wurde mittels 58 Bilder aus sechs verschiedenen Proben evaluiert. Die sechs Proben wurden a-priori von einem Experten ausgewertet, wobei jedes Feld im Screen in eine der drei Klassen („none“, „weak“ oder „strong“) eingeordnet wurde. Diesen qualitativen Aussagen wurden quantifiziert, um diese mit den quantitativen Ergebnissen der Mustererkennung vergleichen zu können. Die Software gibt zurück, wieviel Prozent von den Pixeln in einem Feld sind vom weißen Spot belegt. Es wurde folgende Einteilung verwendet:

- „none“ Weniger als 10 %,
- „weak“ zwischen 10 % und 24 % und
- „strong“ größer als 24 %.

Von jeder Probe wurden bei unterschiedlicher Position, Skalierung, Rotation und jeweils zehn Bilder aufgenommen, um mögliche Variationen bei der Bildaufnahme abzudecken. Dabei wurden auch extreme Situationen geschaffen, die in der Routine eigentlich nicht auftreten sollten.

Die computergestützte Auswertung wurde mit der menschlichen Auswertung verglichen. In 97 Prozent der Fälle bei 22272 untersuchten Feldern stimmen die Ergebnisse der Software mit der Auswertung des Experten überein.

Wie sehen Die ERGEBNISSE AUS, IN DENEN ES NICHT GEKLAPPT HAT???

4 Diskussion

Das vorgestellte System zur computergestützten Auswertung von Protein-Protein-Interaktionen kann die manuelle Auswertung der Hefekolonien ablösen. Die Ergebnisse können direkt in verwendete Datenbanken kopiert werden. Damit ist nicht nur eine erhebliche Zeitersparnis, sondern auch eine standardisierte Auswertung mit reproduzierbaren Ergebnissen verbunden. Die Softwarearchitektur erlaubt eine einfache Handhabbarkeit und Erweiterbarkeit des Systems. Man kann davon ausgehen, dass auf jedem modernen Computer ein WWW-Browser zur Verfügung steht, so dass auf den Labor-PCs der Biologen keine weitere Software installiert und gewartet werden muss. Neue Versionen und Erweiterungen der Auswertesoftware lassen sich im Betrieb einarbeiten und testen. Theoretisch kann der Auswerteserver beliebig viele Labor-PCs gleichzeitig bedienen; eine praktische Begrenzung liegt in der Rechen- und Speicherkapazität.

Die Bildauswertung ist modular aufgebaut und kann sukzessive an neue Anforderungen angepasst werden. Die verwendeten Bildverarbeitungsalgorithmen sind größtenteils sehr einfach. Insbesondere kann die Schwellwert-Segmentierung durch andere Verfahren ersetzt werden müssen, bei denen mehrere Felder überlappende Spots identifiziert werden. So können die 3 Prozent Fehler reduziert werden.

Seit Januar 2004 ist eine erste Version im Betrieb und wird von den Biologen getestet.

Literaturverzeichnis

1. Uetz Peter: Proteomik: Proteine im (Hefe-)Kontext. NACHRICHTEN - Forschungszentrum Karlsruhe 34(1):55-60, 2002.
2. Uetz Peter: Protein-Protein-Interaktionen im Modell Hefe. BIOforum 25(1-2):2-4, 2002.
3. Uetz Peter: Two-Hybrid Arrays. Curr Opin Chem Biol 6:57-62, 2001.
4. Stotzka R, Silva J, Uetz P, et al.: Computer-aided analysis of protein-protein-interactions In German Conference of Bioinformatics GCB'03:14-15, 2003.