

Proteomanalyse durch automatisierte Datenverarbeitung

Andreas Wattenberg, Martin Blüggel, Protagen AG, Dortmund

Die Automatisierung von Proteomanalysen ist von großem Interesse, allerdings hat die Vielfalt der Analysetechniken sowie die Komplexität der Analyseschritte eine allgemein verwendbare Automatisierung bislang erschwert. Für Proteomstudien mit Hilfe von 2D-Gelen und massenspektrometrischer Analyse gibt es inzwischen einen weitgehend standardisierten, automatisierbaren Arbeitsablauf. Mehrere Firmen bieten komplette Geräte-Lösungen (Proteomics-Strassen) oder Teillösungen an. Die elektronische Datenverarbeitung soll dabei die Aufgaben eines ‚Labor Informations Management Systems‘ (LIMS) und der bioinformatischen Datenauswertung erfüllen. Die softwaretechnische Umsetzung muß hohe Anforderungen an die Flexibilität der Anbindung weiterer Techniken, des Durchsatzes, der Prozeß-Stabilität, der Transparenz sowie eine Integration in die bioinformatische Auswertung erfüllen. Die Umsetzung des vielschichtigen Anforderungsprofils für die Datenverarbeitung einer Proteomanalysestraße wird am Beispiel der PROTEINEER™ Line (Bruker Daltonik) dargestellt.

Bei einer großen Zahl an automatisch analysierbaren Proben ist die Kontrolle der Daten- und Probenflüsse im Labor und eine parallelisierte Datenauswertung wichtig für eine erfolgreiche Proteomanalyse. Zum einen müssen Daten für die Steuerung einzelner Geräte und von Geräten erzeugte Daten weitergegeben und zentral verwaltet werden (LIMS). Zum anderen müssen vom Massenspektrometer (MS) erzeugte Daten weiterverarbeitet, in Primärstrukturinformationen umgewandelt und mit den biochemischen Daten der Probe unter Berücksichtigung der biochemischen Fragestellung der Proteomstudie korreliert werden (Bioinformatik). Eine Trennung beider Ebenen in ein LIMS (Sample Tracking) und eine Proteom-Datenbank (Knowledge Library) ist sowohl aus Stabilitäts- und Kompatibilitäts- als aus Transparenzgründen vorteilhaft.

Proteomanalysestraße

Kommerziell erhältliche Roboterplattformen und Implementierungen von Laborprotokollen sind bei unterschiedlichen Proteomanalysestraßen sehr verschiedenen. Generell läßt sich aber die Analyse in diskrete Schritte unterteilen.

Zunächst wird das Proteom mittels 2D-Gel-elektrophorese in die einzelnen Proteine aufgetrennt. Diese Technik erlaubt auch ohne Automatisierung eine hohe Trennleistung und Parallelisierung für mehrere 10.000 Proteine pro Woche. Anschließend werden die angefärbten Proteine mit einem Roboter aus dem 2D-Gel ausgestochen und einzeln in eine Mikrotiterplatte abgelegt. Im nächsten Schritt werden die Gelstücke gewaschen, die Proteine enzymatisch verdaut und für die Massenspektrometrie vor-

bereitet (z.B. MALDI-Target-Präparation). Im letzten Schritt erfolgt die Analyse der Proben im Massenspektrometer. Als MS-Verfahren finden die LC-ESI-MS oder die MALDI-MS Einsatz, wobei das letztere Verfahren für eine Automatisierung bevorzugt verwendet wird. Der Ablauf der Analyse und die entsprechenden Datenflüsse sind in Abbildung 1 dargestellt.

Die erste Komponente in der Analysestraße ist ein Roboter (Spot-Picker), der die angefärbten Proteine aus dem 2D-Gel aussticht. Dabei müssen die Koordinaten der einzelnen Proteinspots im Gel möglichst exakt bestimmt werden, um ein präzises Ausschneiden zu gewährleisten. Ein hochaufgelöstes Bild des 2D-Gels wird deshalb direkt auf dem Roboter mit Hilfe eines speziell entwickelten Durchlichtscanners erzeugt. Neben der Kommunikation mit einer externen Imageanalysesoftware ist eine Positionsbestimmung der Spots durch eine auf dem Roboter integrierte Spoterkennung möglich. Bei der Auswahl der zu analysierenden Proteinspots können auch Analyseprotokolle für die folgenden Schritte festgelegt und zum LIMS exportiert werden.

Die zweite Komponente bildet ein Verdau-Roboter, der die ausgestochenen Gelstücke in einem mehrstufigen Prozeß wäscht, enzymatisch verdaut und vollautomatisch auf einen MALDI-Probenträger präpariert. Für diesen Roboter sind die verwendeten Protokolle entscheidend, um auch kleinste Proteinmengen erfolgreich zu analysieren. Da dieser Prozeß für die Nachweisempfindlichkeit der gesamten Analyse limitierend ist, wurden bei der Entwicklung des Roboters die Protokolle durch ein definiertes Testsystem optimiert. Parallel dazu sind definierte Chemikalienkits, die alle für den Roboter nötigen Reagenzien in verschlossenen Behältern enthalten, für eine optimale Empfindlichkeit und Zuverlässigkeit notwendig.

Die dritte Komponente der Analysestraße ist ein MALDI-MS, in dem die Proben aus dem Verdauroboter analysiert werden. Die MS-Daten bilden die Basis für eine bioinformatische Analyse der Primärstruktur der Proteine. Die Besonderheit der neuen MALDI-MS-Generation ist, daß neben MALDI-Peptid-Massen-Fingerprint- (PMF) Spektren auch Fragmentspektren einzelner Peptide direkt aus dem komplexen Gemisch heraus in hoher Qualität erzeugbar sind (Peptid Fragmentierungs Fingerprints, PFF). So ist es möglich, die Aminosäuresequenz eines Peptids zu bestimmen und eine höhere Signifikanz der Proteinidentifikation und der Analyse von posttranslationalen Modifizierungen zu erhalten.

LIMS

Das LIMS ist für die Speicherung der Prozeßdaten und die Probeninformationen elementar. Die Stati einzelner Proteinspots („detec-

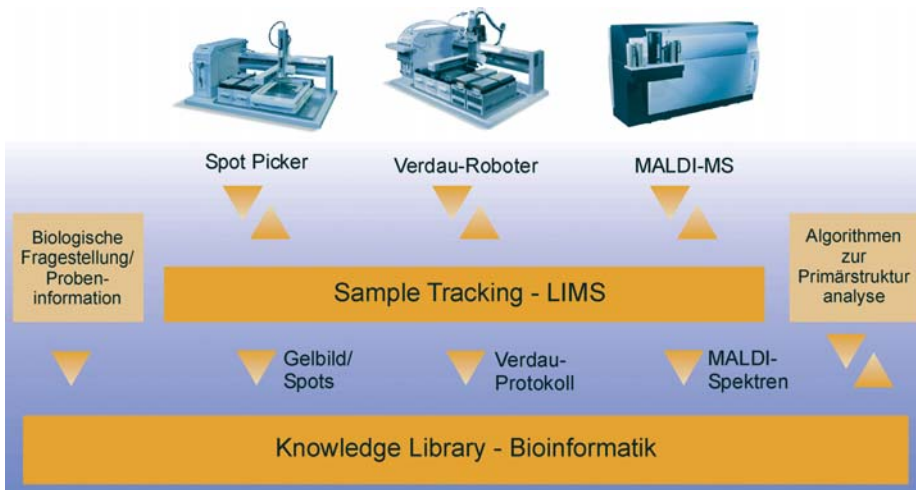


Abb. 1: Informationsfluß in der Proteomanalysestraße. Prozeß-spezifische Informationen werden von den einzelnen Robotern in das LIMS geschrieben und abgerufen. Die erzeugten Daten werden in der Knowledge Library mit weiteren Daten kombiniert und ausgewertet.

ted“, „picked“, „digested“, „acquired“, „identified“) werden zentral erfasst und den Robotern zur Verfügung gestellt. Damit kann die Software der Robotersteuerung vom LIMS die benötigten Informationen erhalten und die Informationen nach Prozessbeendigung zurückschreiben. Die Protokolle, die an den einzelnen Stellen der Straße angewendet werden sollen, sind im voraus definierbar und erhöhen so den Automatisierungsgrad. Voraussetzung dafür ist, daß die Nachverfolgung der Probe (des Proteinspots) an allen Stellen der Straße möglich und eindeutig ist. Dazu werden die drei Probenräger – das Passepartout des 2D-Gels, der Halter der Mikrotiterplatte und das MALDI-Target – mit einem Transponder versehen. Auf dem Transponder wird für die Probe eine Zahlenkombination zur eindeutigen Identifizierung gespeichert. Der Vorteil eines Transponders gegenüber einem Barcode ist, daß er neben dem Auslesen auch beschrieben werden kann. Der Status kann damit nach Beendigung des Protokolls inkrementiert werden. Jede Probenposition der Proteomanalysestraße ist mit einer Lese- und Schreibstationen ausgestattet und stellt so eine verwechslungsfreie Nachverfolgung der Proben sicher. Durch eine definierte Schnittstelle ist eine Einbindung weiterer Roboter gleichen Typs oder auch von Systemen verschiedener Hersteller möglich.

Bioinformatik

Für die bioinformatische Analyse der Daten ist eine komplexe Datenbanksoftware wie die Proteomdatenbank-Software Proteinscape (Bruker Daltonik) nötig. Zuerst werden Daten zur biochemischen Fragestellung der Proteomstudie und zu den Proben vor der Proteinauftrennung erfasst. Die Datenbank speichert die Daten aus den einzelnen Schritten der Proteomanalysestraße in einer hierarchischen Struktur. So können die MS-Spektren (PMF und PFF) vollautomatisch Gelbildern und Spotkoordinaten des Spotpickers zugeordnet werden (Abb. 2). Der Umwandlung von MS-Daten in die Primärstrukturinformation der Proteine kommt eine zentrale Rolle zu.

Der Ablauf einer typischen MALDI-Datenanalyse ist in Abbildung 3 dargestellt. Die Massenspektren werden aufgrund immer wieder auftretender Ionen (z.B. tryptischen Eigenpeptiden, Ionen des Farbstoffs, Keratin-Hintergrund) nach der Peak-Erkennung erneut intern kalibriert. Dann werden die Molekulargewichte der Kalibranten eliminiert, so daß nur unbekannte Massen der Peptide an die Suchmaschinen geschickt werden. Für eine optimale Proteinidentifizierung muß diese Kalibrantenliste automatisch und spezifisch für jedes Labor und Projekt erzeugt werden. Für die Auswertung der Massenspektren werden oft mehrere Suchalgorithmen mit verschiedenen Suchstrategien verwendet. Die Suchergebnisse müssen für eine übersichtliche Ergebnisdarstellung zusam-

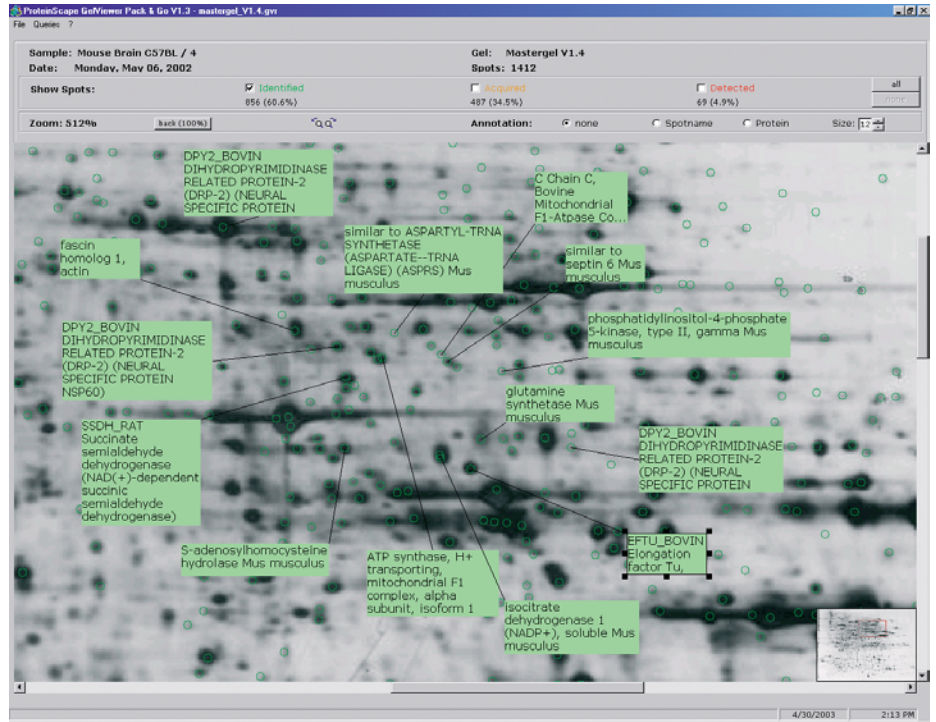


Abb. 2: Annotiertes 2D-Gel. Die Stati der Prozessierung der einzelnen Proteinspots können nachverfolgt werden. Die identifizierten Proteine werden direkt im 2D-Gel annotiert.

mengefaßt werden. Für eine automatisierte Auswertung ist es dabei wichtig, aus den Ergebnissen aller Suchmaschinen einen Metascore zu errechnen, der für die Beurteilung der Signifikanz unabhängig von verwendeten Suchmaschinen und Strategien ist. Die Datenauswertung muß eine intelligente und parallele Prozessierung vieler Spektren ermöglichen, da große Datenmengen manuell nicht mehr einzeln prozessierbar sind. So

dauert eine automatisierte Datenanalyse für einen Proteinspot weniger als zwei Minuten. Durch eine Aufgabenverteilung auf mehrere Computer ist ein Durchsatz von mehreren tausend Proteinproben pro Tag möglich. Die identifizierten Proteine können durch komplexe Abfragemöglichkeiten in Beziehung zu der Projektfragestellung gebracht werden. Dazu werden Proteinlisten geclustert, verglichen sowie die Ontologie der Proteine verglichen und bewertet.

Ausblick

Die Einführung von Proteomanalysestraßen ermöglicht eine Analyse ganzer 2D-Gele mit mehreren tausend Proteinspots. Eine enge Verzahnung der Datenverarbeitung mit den einzelnen Komponenten der Analysestraße bietet einen hohen Automatisierungsgrad und eine parallele Datenverarbeitung in vielen Einzelschritten. So können erstmals komplexe Proteome auf molekularer Ebene analysiert und auf Ebene der identifizierten Proteine verglichen werden. Dieses neue Proteomanalyse-Werkzeug wird neue Ansätze der Proteomforschung ermöglichen.

Korrespondenzadresse

Martin Blüggel
 Protagen AG
 Emil-Figge-Str. 76A, D-44227 Dortmund
 Tel./Fax: +49-(0)231-9742-890 / -891
 eMail: martin.blueggel@protagen.de
 www.protagen.de

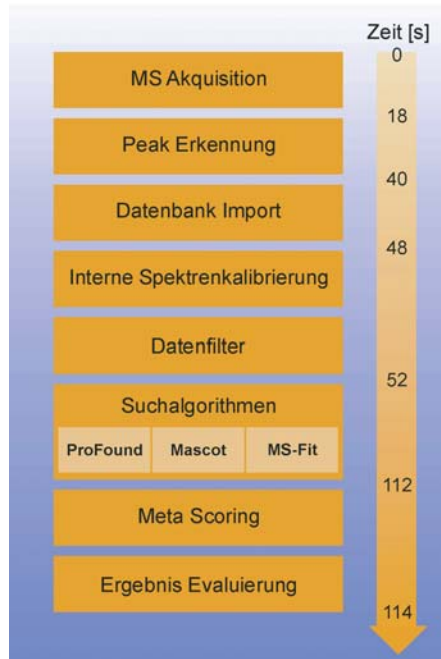


Abb. 3: Analyse von massenspektrometrischen Daten der Proteom-Analysestraße zur Primärstruktur-Identifizierung von Proteinen