

Aus dem Max Planck Institut für Molekulare Genetik

**Entwicklung neuer Methoden zur  
massenspektrometrischen Charakterisierung von  
Biomolekülen**

Inaugural-Dissertation  
zur Erlangung des Grades eines  
Doktors der Naturwissenschaften  
bei dem Fachbereich Biologie - Chemie - Pharmazie  
der Freien Universität Berlin

vorgelegt von  
**Volker Egelhofer**  
Diplom-Biochemiker, Diplom-Ingenieur aus Kaiserslautern

Berlin, 2001

Gutachter:

1. Prof. Dr. Hans Lehrach

2. Prof. Dr. Wolfram Saenger

Ort und Tag der Disputation:

Berlin, 20.06.2002

Diese Arbeit wurde im Juni 1999 am Max-Planck-Institut für Molekulare Genetik in der Abteilung von Prof. Dr. Hans Lehrach begonnen und im August 2001 abgeschlossen.

Mein Dank gilt Prof. Dr. Hans Lehrach, der mir die Durchführung dieser Arbeit ermöglichte. Prof. Dr. Wolfram Saenger sei ebenfalls für seine Unterstützung und Begutachtung gedankt.

Besonders möchte ich mich bei Dr. Eckhard Nordhoff für die allzeit gewährte Unterstützung, Beratung und Förderung bedanken.

Weiterhin bedanke ich mich bei:

Dr. Johan Gobom für engagierte Diskussionen und konstruktive Kritiken.  
Patrick Giavalisco und Prof. Dr. Dr. Joachim Klose für die Zurverfügungstellung von 2D-Gelen und Dr. Konrad Büssow für die Bereitstellung der gentechnisch hergestellten rekombinanten Proteinproben.

Dorothea Theiss, Christine Lübbert, Dr. Klaus-Dieter Klöppel und Dr. Gabriele Thiele für ihre umfangreiche Unterstützung und das angenehme Arbeitsklima.

## INHALTSVERZEICHNIS

<b>1</b>	<b>EINLEITUNG</b>	<b>1</b>
<b>1.1</b>	<b>2D-Gelelektrophorese</b>	<b>3</b>
<b>1.2</b>	<b>Massenspektrometrie</b>	<b>4</b>
1.2.1	<b>ELEKTROSPRAY- IONISATIONS- MASSENSPEKTROMETRIE (ESI-MS )</b>	<b>6</b>
1.2.2	MATRIX-ASSISTIERTE LASERDESORPTIONS/IONISATIONS- MASSEN-SPEKTROMETRIE (MALDI-MS)	7
1.2.3	MALDI-MS VERSUS ESI-MS	14
<b>1.3</b>	<b>„PEPTIDE MASS FINGERPRINTING“ UND PROTEINIDENTIFIKATION</b>	<b>14</b>
<b>1.4</b>	<b>PROTEOMANALYSE UND AUTOMATISIERUNG</b>	<b>16</b>
1.4.1	PROTEINSPOTENTNAHME	16
1.4.2	PROBENPRÄPARATION	18
<b>1.5</b>	<b>BIOINFORMATIK</b>	<b>20</b>
1.5.1	ANALYSE MASSENSPEKTROMETRISCHER DATEN	20
1.5.2	DATENORGANISATION	24
1.5.3	DATENINTERPRETATION	24
<b>2</b>	<b>ZIELSETZUNG DER ARBEIT</b>	<b>28</b>
<b>3</b>	<b>MATERIALIEN UND CHEMIKALIEN</b>	<b>29</b>
<b>3.1</b>	<b>VERWENDETE CHEMIKALIEN UND DEREN BEZUGSQUELLEN</b>	<b>29</b>
<b>3.2</b>	<b>METHODEN</b>	<b>30</b>
3.2.1	MASSENSPEKTROMETRIE	30
3.2.2	PROBENPRÄPARATION	30
3.2.3	PRÄPARATION DER KALIBRANTEN	30
3.2.4	2D-GELELEKTROPHORESE	31
3.2.5	BIOINFORMATIK	32

<b>4</b>	<b>ERGEBNISSE UND DISKUSSION</b>	<b>33</b>
<b>4.1</b>	<b>ALGORITHMEN ZUR VERBESSERUNG DER MASSENRICHTIGKEIT</b>	<b>33</b>
<b>4.2</b>	<b>„SCORING“-ALGORITHMUS ZUR SICHEREREN IDENTIFIKATION VON PROTEINEN IN GROßEN SEQUENZDATENBANKEN</b>	<b>47</b>
4.2.1	BEISPIEL FÜR DIE EFFIZIENZ DES „SCORING“-ALGORITHMUS BEI HOHER TREFFERZAHL	50
4.2.2	BEISPIEL FÜR DIE EFFIZIENZ DES „SCORING“-ALGORITHMUS BEI NIEDRIGER TREFFERZAHL	55
4.2.3	BEISPIEL FÜR DIE EFFIZIENZ DES „SCORING“-ALGORITHMUS BEI DER SUCHE IN GROßEN DATENBANKEN	58
<b>4.3</b>	<b>SOFTWARE PAKET MS-PROTEOMICS</b>	<b>66</b>
4.3.1	GRUNDLEGENDE FUNKTION UND EIGENSCHAFTEN	66
4.3.2	ERWEITERTE FUNKTION UND EIGENSCHAFTEN	73
<b>4.4</b>	<b>ANALYSE AUSGEWÄHLTER 2D-GELSPOTS</b>	<b>73</b>
<b>5</b>	<b>ZUSAMMENFASSUNG</b>	<b>86</b>
<b>6</b>	<b>SUMMARY</b>	<b>88</b>
<b>7</b>	<b>LITERATURVERZEICHNIS</b>	<b>90</b>
	<b>ABKÜRZUNGEN</b>	<b>97</b>
	<b>AMINOSÄUREN</b>	<b>98</b>
	<b>PUBLIKATIONEN</b>	<b>99</b>
	<b>POSTER PRÄSENTATION</b>	<b>99</b>
	<b>DIPLOM-ARBEITEN</b>	<b>100</b>
	<b>SOFTWARE ENTWICKLUNGEN</b>	<b>100</b>

## 5 Zusammenfassung

In dieser Arbeit werden von mir neu entwickelte Strategien und Algorithmen vorgestellt, welche die Proteinidentifikation mittels MALDI-TOF-MS „*peptide mass fingerprinting*“ unter Verwendung einer externen Kalibrierung so weit verbessern, dass auf die Anwendung einer internen Kalibrierung gänzlich verzichtet werden kann. Die Algorithmen basieren auf der Beobachtung, dass die Variation in den bestimmten Flugzeiten der sich an unterschiedlichen Positionen auf dem MALDI-Probenträger befindlichen Peptide auf zwei systematische Fehler zurückführen lässt. Zum einen wird bei Wechsel der Position der Nullpunkt des Massenspektrums verschoben, das heißt alle Massen weichen danach um einen bestimmten konstanten Betrag von den vorherigen Werten ab. Zum anderen können die Massenspektren nach Wechsel der Position noch zusätzlich linear gestreckt oder gestaucht sein. Im ersten Fall wird jede einzelne gemessene Masse mit einem bestimmten für alle Massen gleichen Fehlerbetrag behaftet. Die Differenzen zwischen den einzelnen gemessenen Massen bleiben davon jedoch unberührt. Im zweiten Fall werden sowohl die Absolutwerte, als auch die Massendifferenzen verändert. Die Algorithmen erkennen diese systematischen Fehler und ermöglichen, auch wenn die Massenrichtigkeit der generierten Daten sehr gering ist, eine korrekte Identifizierung der analysierten Proteine.

Um Proteine eindeutig in großen Proteinsequenzdatenbanken zu identifizieren, wurde von mir ein Algorithmus entwickelt, der mit Hilfe der Parameter: Standardabweichung, Trefferanzahl und prozentualer Sequenzabdeckung für jedes Protein einen „*Scoring*“-Faktor  $Z$  berechnet. Mit diesem neu entwickelten „*Scoring*“-Verfahren konnten z.B. 52 von 96 gentechnisch hergestellten Proteinen ohne Eingreifen des Menschen

eindeutig identifiziert werden. In keinem Fall wurde ein falsch-positives Ergebnis geliefert.

Weiterhin wurde von mir das Softwarepaket „MS-Proteomics“ entwickelt, das in kurzer Zeit vollautomatisch eine große Anzahl von massenspektrometrischen Datensätzen einliest, die Proteinidentifikation durch Abgleich mit einer in wenige Sekunden aus einer ausgewählten Proteindatenbank berechneten Peptidsequenzdatenbank durch Anwendung der oben erwähnten Algorithmen vornimmt und die Ergebnisse übersichtlich darstellt. Die Software liest darüber hinaus 2D-Gelbilder ein und weist den detektierten Proteinspots die Ergebnisse der Datenbanksuche automatisch zu.

Alle relevanten Ergebnisse wurden publiziert oder eingereicht zur Veröffentlichung in den anerkannten wissenschaftlichen Fachzeitschriften "Analytical Chemistry" und "Electrophoresis" (die entsprechenden Literaturhinweise finden sich Anhang). Eine Internetversion von MSA 2.0 wird nach Veröffentlichung dieser Arbeit unter <http://www.scienion.de/msa> der wissenschaftlichen Gemeinde zur Verfügung gestellt.

## 6 Summary

Within this work I present a new protein identification strategy that overcomes the need for performing internal or close external calibration in MALDI-TOF-MS peptide mass fingerprinting. The strategy is based on the observation that the variation of peptide flight times, when measured on different positions on the sample support, are systematic and affect mainly the linear components (offset and slope) of the correlation between  $m/z$  and the square of the flight time. Consequently, the mass errors obtained when using a single set of calibration constants, determined at one position of the sample support, to calibrate all other time-of-flight spectra recorded from that support, are also systematic. The developed search algorithm recognizes these systemic trends in the mass errors, thereby allowing protein identification even with a low mass accuracy of the input data.

For the retrieval of the correct protein in a database search, I have developed a new scoring algorithm, which uses the parameters: standard deviation, number of matching peptide masses and the sequence coverage of the protein to calculate the score for each protein. Using this algorithm it was possible to correctly identify 52 out of 96 recombinant proteins of known identity, without any false identification. Moreover, I implemented the above identification strategy and scoring algorithms in a software package designated "MS-Proteomics", which automatically reads many peptide mass maps in a short time and performs all calculations for protein identification. For protein identification from 2D-gel electrophoresis, MS-Proteomics also comprises a 2D-gel viewer that links the search results to its corresponding spots on the gel image. All relevant results have been published or submitted for publication in the peer-reviewed scientific journals "Analytical Chemistry"



and "Electrophoresis" (references are part of the appendix).

In addition, some of the results have been presented at the 48th ASMS Conference on Mass Spectrometry and Allied Topics, LA, California, USA, June 11-15, 2000. A web-based version of the program MSA 2.0 will be made available to the scientific community at <http://www.sciencion.de/msa>, following publication of this thesis.

## Publikationen

- V. Egelhofer, K. Büssow, C. Luebbert, H. Lehrach, and E. Nordhoff,  
*"Improvements in protein identification by MALDI-TOF-MS peptide mapping"*,  
Anal. Chem. **2000**, 72, 2741-2750.
- V. Egelhofer, J. Gobom, H. Seitz, P. Giavalisco, H. Lehrach, and E. Nordhoff  
*"Protein Identification by MALDI-TOF-MS peptide mapping: A New Strategy"*,  
Anal. Chem. **2002**, 74/8, 1760-1771 (Accelerated Article).
- V. Egelhofer *"Taking some of the "guesswork" out of protein identification "*,  
Interview in Anal.Chem. **2002**, 74/9, 248A-249A.
- E. Nordhoff, H. Eickhoff, M. Horn, T. Przewieslik, V. Egelhofer, P. Giavalisco,  
D. Theiss, H. Lehrach, and J. Gobom *"Large-Gel 2-DE - MALDI-TOF-MS –  
An Analytical Challenge for Studying Complex Protein Mixtures"*,  
Electrophoresis **2001**, 22, 2844-2855.
- F. Schmidt., A. Lueking, E. Nordhoff, J. Gobom, J. Klose, H. Seitz, V.  
Egelhofer, H. Eickhoff, H. Lehrach and D.Cahill *"Generation of minimal  
protein identifiers of proteins from 2D gels and recombinant proteins"*,  
Electrophoresis **2002**, 23, 621-625.
- J. Gobom, M. Mueller, V. Egelhofer, H. Lehrach, and E. Nordhoff *„Improved  
mass accuracy in MALDI-TOF-MS peptide analysis“*, Proceedings of the 49th  
ASMS Conference on Mass Spectrometry and Allied Topics **2001**, Chicago,  
California, USA, May 27 – 31.
- P. Giavalisco, E. Nordhoff, V. Egelhofer, J. Klose, H. Lehrach and J. Gobom  
*„Proteome analysis of Arabidopsis thaliana“*, Proceedings of the 49th ASMS  
Conference on Mass Spectrometry and Allied Topics **2001**, Chicago,  
California, USA, May 27 - 31.

## **Poster Präsentation**

V. Egelhofer, K. Büssow, C. Luebbert, H. Lehrach, and E. Nordhoff  
*„Improvements in protein identification by MALDI-TOF-MS peptide mapping“*,  
Proceedings of the 48th ASMS Conference on Mass Spectrometry and Allied  
Topics **2000**, LA, California, USA, June 11 - 15.

## **Diplom-Arbeiten**

V. Egelhofer *„Biologische Untersuchungen zur Charakterisierung von Steroid-Dehydrogenasen“*, Diplom-Arbeit, Technische Fachhochschule Berlin **1995**.

V. Egelhofer *„Investigations for fast identification of proteins by means of mass-spectroscopically peptide mapping“*, Diplom-Arbeit, Freie Universität Berlin **1999**.

## **Software Entwicklungen**

Datenbank Search engine MSA Version 0.99, frei zugänglich via  
*„[http://www.molgen.mpg.de/~mass-spec/ASMS-Conference / msaProgram.htm](http://www.molgen.mpg.de/~mass-spec/ASMS-Conference/msaProgram.htm)“*.

Datenbank Search engine MSA Version 2.0, frei zugänglich via  
*„<http://www.scienion.de/msa>“*.

Softwarepaket MS-Proteomics Version 1.0, komplette Client-Server  
Anwendung inklusive vollautoautomatischen Batchbetrieb, Protein-  
identifikation und interaktiven Ergebnisdarstellung.

## ***Curriculum vitae***

**Name** Volker Egelhofer

**E-mail** [egelhofer@scienion.de](mailto:egelhofer@scienion.de), [v\\_egel@gmx.de](mailto:v_egel@gmx.de),

### **Ausbildung**

**1990 – 1995** TFH Berlin, Studiengang Biotechnologie  
**08.06.1995** **Diplom in Biotechnologie, Titel der Arbeit:**  
*„Biochemische Charakterisierung der Enzyme D1-Dehydrogenase und 17 $\beta$ -Dehydrogenase aus Nocardia restrictata 26EB“*

**1995 – 1997** Freie Universität Berlin, Studiengang Sinologie  
**1996** **Abschlusstest Sprachpropädeutikum**

**1996 – 1998** Freie Universität Berlin, Studiengang Informatik

**1995 – 1999** Freie Universität Berlin, Studiengang Biochemie  
**27.05.1999** **Diplom in Biochemie, Titel der Arbeit:**  
*„Investigations on fast identification of proteins by means of mass spectrometric peptide mapping“*

**1999 – 2001** **Doktorand am Max-Planck-Institut für Molekulare Genetik, Abteilung Prof. Dr. H. Lehrach, Arbeitsgruppe Dr. E. Nordhoff**

**08.2001** **Abgabe der Dissertation, Titel der Arbeit:**  
*„Entwicklung neuer Methoden zur massenspektrometrischen Charakterisierung von Biomolekülen“*

**20.06.2002** **Disputation**

**present address:**

**Scienion AG**

**Volmerstrasse 7a**

**12489 Berlin (Germany)**

**Phone: [+49] 30-6392-1708**

**Fax: [+49] 30-6392-1701**

**e-mail: [egelhofer@scienion.de](mailto:egelhofer@scienion.de)**