Reiner Westermeier, Tom Naven, and
Hans-Rudolf Höpker

Proteomics in Practice

A Guide to Successful Experimental Design

Second, Completely Revised Edition

WILEY-VCH Verlag GmbH & Co. KGaA

Reiner Westermeier, Tom Naven,
and Hans-Rudolf Höpker
Proteomics in Practice

Related title from Wiley-VCH

Hamacher, M., Marcus, K., Stühler, K., van Hall, A., Warscheid, B., Meyer, H. E. (eds.)

Proteomics in drug Research

2006
ISBN 978-3-527-31226-9

Liebler, D. C., Petricoin, E. F., Liotta, L. A. (eds.)

Proteomics in Cancer Research

2005
ISBN 9780-471-44476-3

Hamdan, M. H., Righetti, P. g., Desiderio, D. M., Nibbering, N. M.

Proteomics Today

Protein Assessment and Biomarkers Using Mass Spectrometry, 2D Electrophoresis, and Microarray Technology

2005
ISBN 978-0-471-64817-8

Sanchez, J.-C., Corthals, G. L., Hochstrasser, D. F. (eds.)

Biomedical Applications of Proteomics

2004
ISBN 978-3-527-30807-1

Reiner Westermeier, Tom Naven, and
Hans-Rudolf Höpker

Proteomics in Practice

A Guide to Successful Experimental Design

Second, Completely Revised Edition

WILEY-VCH Verlag GmbH & Co. KGaA

The Authors

Dr. Reiner Westermeier
GE Healthcare Life Sciences
Oskar-Schlemmer-Str. 11
80807 München
Germany

Dr. Tom Naven
GE Healthcare UK Limited
Pollards Wood
Nightingales Lane
Chalfont St Giles
Bucks HP8 4SP
United Kingdom

Hans-Rudolf Höpker
Weberweg 6
79183 Waldkirch
Germany

Library of Congress Card No.:
applied for

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Bibliographic information published by the Deutsche Nationalbibliothek
Die Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.d-nb.de>.

Composition Kühn & Weyh, Software GmbH, Freiburg
Printing Betz-Druck GmbH, Darmstadt
Bookbinding Litges & Dopf GmbH, Heppenheim

Printed in the Federal Republic of Germany
Printed on acid-free paper.

ISBN 978-3-527-31941-1

Contents

Preface

More than a decade after the beginning of the "Proteomics Rush" the methodological and systematic approaches for the analysis of proteomes have evolved from being holistic and non-hypothesis driven to phenomenon-based, dedicated protein studies. Importantly and correctly so, it has been widely reported that it is very difficult to obtain all necessary information from one single workflow, e.g. 2-D gel-mass spectrometry; and that different workflows deliver complementary information rather than similar, overlapping results. Therefore it is necessary to make additions to the manuscript for the second edition of Proteomics in Practice and include a comprehensive description of chromatography methods, written mainly by Hans-Rudolf Höpker.

The objective of the second edition of Proteomics in Practice is to provide the reader with a comprehensive reference and practical guide for the successful analysis of proteins by 2-D electrophoresis, chromatography and mass spectrometry. The book includes a theoretical introduction into the most-applied methodologies, a practical section complete with worked examples, a unique troubleshooting section and a thorough reference list to guide the interested reader to further details.

The theoretical section introduces the fundamentals behind the techniques applied in proteomics and describes how the techniques are used for proteome analysis. However, the practical aspects of the book focus on 2D-DIGE technology and mass spectrometry. 2-D DIGE is increasingly cited for studying differential protein expression and, as such, a considerable section of the text is dedicated to this technique. The core components of 2D-DIGE, sample preparation and labeling, 2-D electrophoresis and image analysis are addressed in considerable detail. Further, the importance of mass spectrometry, sequence databases and search engines for successful protein identification are discussed.

The practical section of the book is, in principle, a course manual, which has been optimized over a number of years. The experimental

section describes how to achieve consistent, reliable and reproducible results using a single instrumental setup, instead of presenting a wide choice of techniques and instruments.

Fundamentally, the book celebrates the attention to detail that is necessary to perform proteome analysis routinely, with confidence.

As the technical developments in this field are proceeding quickly, the contents of the book will need to be updated every few months. The reader can have access to a web-site at http://www.wiley-vch.de/books/info/3-527-31941-7/index.html/, which will contain the updated chapters and recipes.

The authors would like to thank Professor Richard Simpson, LICR in Melbourne, for writing the foreword, and Dr. Axel Parbel, GE Healthcare in Munich, for critical reading of the manuscript and delivering valuable contributions to the LC MS sections.

Reiner Westermeier
Tom Naven
Hans-Rudolf Höpker January 2008

Further thanks to:
Philippe Bogard
Josef Bülles
Matrixscience.com
Staffan Renlund
Günter Thesseling
Jenny Samskoog

Foreword

Biological macromolecules are the main actors in the makeup of life. To understand biology and medicine at a molecular level, we need to visualize the activity and interplay of large macromolecules such as proteins. To study protein molecules, the principles of their separation, quantitation, and determination of their individual characteristics had to be developed. One of the most important separation techniques used today for the characterization and analysis of proteins is electrophoresis: a separation technique involving the movement of charged species through a matrix under the influence of an applied electric field. In 1948, the Nobel Prize in Chemistry was awarded to Arne Tiselius "for his research on electrophoresis and adsorption analysis, especially for his discoveries concerning the complex nature of the serum proteins". This acknowledgement followed his seminal work in 1937, which led to the development of an apparatus purposely designed for the separation of serum proteins – the Tiselius moving-boundary apparatus. Explosive developments in electrophoresis occurred in the 1940s and 1950s when, in addition to zone electrophoresis, two other electrophoretic techniques emerged: isolectric focusing and isotachophoresis. Concomitant with these discoveries was the development of the matrices employed for these techniques (e.g., paper, polymer gels, such as agar or starch, and in 1959, polyacrylamide gels), each yielding distinct advantages for different samples. Of these, acrylamide gel support media emerged as the most widely used in the separation of proteins, in particular SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and two-dimensional gel electrophoresis, independently discovered in 1975 by Joachim Klose and Patrick O'Farrell. Today, electrophoresis still remains the seminal technique in the armory of methods that biologists apply to protein separation and characterization problems.

More and more, as students and experienced researchers from different disciplines delve into intricate biological questions that require protein chemistry input, they are confronted with the pressing need to learn fundamental protein separation methods and techniques.

Often, finding suitable resources to accomplish this task may present as big a challenge as mastering the subject field itself. In 2002, Reiner Westermeier and Tom Naven accomplished this formidable task by condensing background information, electrophoretic theory, didactic protocols, complete source lists for the tested materials, practical tips, and information resources into a single volume: Proteomics in Practice – A Laboratory Manual of Proteome Analysis. Of immense value are the sections that cover sample preparation (considered the "Achilles' heel" of proteomics) and the development of purification strategies. Given the ever-broadening landscape of proteomic technique development, Reiner Westermeier and his coauthors Tom Naven and Hans-Rudolf Höpker have now completely rewritten most parts of the First Edition according to the new developments which have happened since 2002.

Proteomics in Practice – A Guide to Successful Experimental Design (Second, completely revised edition) by Reiner Westermeier, Tom Naven and Hans-Rudolf Höpker is an invaluable information resource both for the experienced protein chemist venturing into cutting-edge electrophoretic separation methodologies tailored for a mass spectrometric protein identification endpoint and for researchers from diverse biological fields who are novices to analytical protein chemistry. This volume represents an essential tool for every laboratory involved in contemporary proteomics research.

Richard Simpson
Member, Ludwig Institute for Cancer Research, Melbourne
Professor, University of Melbourne

Abbreviations, Symbols, Units

1D electrophoresis	One-dimensional electrophoresis
2D electrophoresis	Two-dimensional electrophoresis
1D-LC	One-Dimensional liquid chromatography
2D-LC	Two-Dimensional liquid chromatography
^{12}C	Monoisotopic peak in a peptide isotopic envelope
A	Ampere
AC	Affinity chromatography
A,C,G,T	Adenine, cytosine, guanine, thymine
AEBSF	Aminoethyl benzylsulfonyl fluoride
AIEX	Anion exchange
Å	Ångström
ANOVA	Analysis of variance
API	Atmospheric pressure ionization
APS	Ammonium persulfate
Asn-Xxx-Ser/Thr	N-linked glycosylation sequon, where Asn = Asparagine, Ser = Serine, and Thr = Threonine
AU	Absorbance units
16-BAC	Benzyldimethyl-*n*-hexadecylammonium chloride
BAC	Bisacryloylcystamine
Bis	N, N′-methylenebisacrylamide
BLAST	Basic local alignment search tool
BN Page	Blue native polyacrylamide gel electrophoresis
bp	Base pair
BPB	Bromophenol blue
BSA	Bovine serum albumin
C	Crosslinking factor (%)
CAF	Chemically assisted fragmentation

CAM	Co-analytical modification
CAPS	3-(cyclohexylamino)-propanesulfonic acid
CBB	Coomassie brilliant blue
CCD	Charge-coupled device
CF	Chromatofocusing
CHAPS	3-(3-cholamidopropyl)dimethylammonio-1-propane sulfonate
CIEX	Cation exchange
CE	Capillary electrophoresis
CID	Collision induced dissociation
conc	Concentrated
CM	Carboxylmethyl
CMOS	Complementary metal oxide semiconductor
CMW	Collagen molecular weight marker
const.	Constant
CSF	Cerebrospinal fluid
CTAB	Cetyltrimethylammonium bromide
CV	Column volume
Da	Dalton
DALPC	Direct analysis of protein complexes
DB	Database
DBM	Diazobenzyloxymethyl
DDRT	Differential display reverse transcription
DEA	Diethanolamine
DEAE	Diethylaminoethyl
DGGE	Denaturing gradient gel electrophoresis
2,5-DHB	2,5-dihydroxybenzoic acid
DIGE	Difference gel electrophoresis
Disc	Discontinuous
DMF	Dimethyl formamide
DMSO	Dimethylsulfoxide
DNA	Desoxyribonucleic acid
dpi	dots per inch
DTE	Dithioerythreitol
DTT	Dithiothreitol
E	Field strength in V/cm
ECD	Electron capture dissociation
EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray ionization
EST	Expressed sequence tag
FAB	Fast atom bombardment
FDR	False discovery rate

FT-ICR	Fourier transform – Ion cyclotron resonance
GF	Gel filtration
GLP	Good laboratory practice
GMP	Good manufacturing practice
h	Hour
H_3PO_4	Phosphoric acid
HCCA	α-cyano-4-hydroxycinnamic acid
HED	Hydroxyethyldisulfide
HeNe	Helium neon
HEPES	N-2-hydroxyethylpiperazine-N′-2-ethanane-sulfonic acid
HFBA	Heptafluorobutyric acid
HMW	High Molecular Weight
HPLC	High Performance Liquid Chromatography
HUPO	Human Proteome Organization
I	Current (A, mA)
ICAT	Isotope coded affinity tags
i.d.	Internal diameter
ID	Identification
IEF	Isoelectric focusing
IEP	Isoelectric point
IEX	Ion exchange
IgG	Immunoglobulin G
IMAC	Immobilized metal affinity chromatography
IP	Immunoprecipitation
IPAS	Intact protein analysis system
IPG	Immobilized pH gradients
IRMPD	Infra red multiphoton dissociation
ITP	Isotachophoresis
kB	Kilobase
kDa	Kilodalton
L	Liter
LC	Liquid chromatography
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LMW	Low molecular weight
LOD	Limit of detection
LWS	Laboratory workflow system

M	mass
mA	Milliampere
MALDI	Matrix assisted laser desorption ionization
MDLC	Multidimensional liquid chromatography
min	Minute
mol/L	Molecular mass per liter
m_r	Relative electrophoretic mobility
mRNA	messenger RNA
MS	Mass spectrometry
MS^n	Multistage tandem mass spectrometry where *n* is greater than 2
MS/MS	Tandem mass spectrometry
M_r	Relative molecular mass
MudPIT	Multi dimensional protein identification technology
m/z	mass/charge ratio (*x*-axis in a mass spectrum)
Nonidet	Non-ionic detergent
NEPHGE	Non equilibrium pH gradient electrophoresis
NHS	*N*-hydroxy succinimide
NPS	Non-porous silica
NR	Non-redundant
NSE	Neuron-specific enolase
NTA	Nitrilotriacetic acid
O.D.	Optical density
P	Power (W)
PAG	Polyacrylamide gel
PAGE	Polyacrylamide gel electrophoresis
PAGIEF	Polyacrylamide gel isoelectric focusing
PBS	Phosphate-buffered saline
PC	Peak capacity
PEG	Polyethylene glycol
PEEK	Polyetherether ketone
PFPA	Pentafluoroproprionic acid
pI	Isoelectric point
p*K* value	Dissociation constant
PMF	Peptide mass fingerprint
PMSF	Phenylmethyl-sulfonyl fluoride
PPA	Piperidinopropionamide
PPF	Protein pre-fractionation
ppm	Parts per million (measure of mass accuracy)
PSA	Prostate-specific antigen
PSD	Post-source decay

PTM	Post-translational modification
PVC	Polyvinylchloride
PVDF	Polyvinylidene difluoride
QTOF	quadrupole time-of-flight
r	Molecular radius
Rf value	Relative distance of migration
R_m	Relative electrophoretic mobility
RNA	Ribonucleic acid
RP	Reversed Phase
RPC	Reversed phase chromatography
rpm	Revolutions per minute
RuBPS	Ruthenium II tris (bathophenanthroline disulfonate)
s	Second
SAX	Strong anion exchange
SCX	Strong cation exchange
SDS	Sodium dodecyl sulfate
SILAC	Stable isotope labeling of amino acids in cell culture
S/N	Signal/noise ratio
SOP	Standard operation procedure
SP	Sulfopropyl
T	Total acrylamide concentration (%)
t	Time (h, min, s)
TAP	Tandem affinity purification
TBP	Tributylphosphine
TBS	Tris-buffered saline
TCA	Trichloroacetic acid
TCEP	Tris(2-carboxyethyl) phosphine
TEMED	N,N,N′,N′-tetramethylethylenediamine
TFA	Trifluoroacetic acid
THPP	Tris(hydroxypropyl)phosphine
TiO_2	Titanium dioxide
TNF	Tumor necrose factor
TOF	Time of flight
Tricine	N,tris(hydroxymethyl) methyl glycine
Tris	Tris(hydroxymethyl) aminoethane
U	Volt
UV	Ultraviolet

V	Volume (L)
v	Speed of migation (m/s)
v/v	Volume per volume
W	Watt
w/v	Weight per volume (mass concentration)
ZiO_2	Zirconium dioxide

Introduction

In a living cell, most activities are performed by proteins. Therefore proteins are the subject of intense research in life science. "Proteomics" is the study of quantitative changes of protein expression levels and their application to drug discovery, diagnostics and therapy. Thereby it is important to apply the correct strategy to discover induced biological changes against the background of inherent biological variations of the sample sources.

Proteomics research has many different application areas: Pharmaceutical companies search for faster identification of new drug targets in transformed cell lines or diseased tissues. Also the validation of the detected targets, *in vitro* and *in vivo* toxicology studies, and checks for side effects can be performed with this approach. Clinical researchers want to compare normal versus disease samples, diseased versus treated samples, find molecular markers in body fluids for diagnosis and prognosis, monitor diseases and their treatments, determine and characterize post-translational modifications. In clinical chemistry it would be interesting to subtype individuals to predict response to therapy. Biologists study basic cell functions and molecular organizations. Another big field is microbiology for various research areas. Proteomics is also applied for plant research for many different purposes, for instance for breeding plants of higher bacterial, heat, cold, drought, and other resistances, increasing the yield of crop and many more.

1 History

The original definition of the "Proteome" analysis means "The analysis of the entire PROTEin complement expressed by a genOME, or by a cell or tissue type" (Wasinger *et al.* 1995). Originally the technologies behind proteome analysis were two-dimensional electrophoresis and identification of proteins by subsequent MALDI mass spectrome-

Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL, Humphery-Smith I. Electrophoresis 16 (1995) 1090–1094.

try with peptide mass fingerprinting. Therefore the proteins spots of interest were picked from the gel and digested with trypsin. In case of failure of identification the peptide mixtures were submitted to sequencing by tandem mass spectrometry. Although the concept of Proteome analysis is older than the phrase, it only began to become widely employed, because several prerequisites came real at the same time:

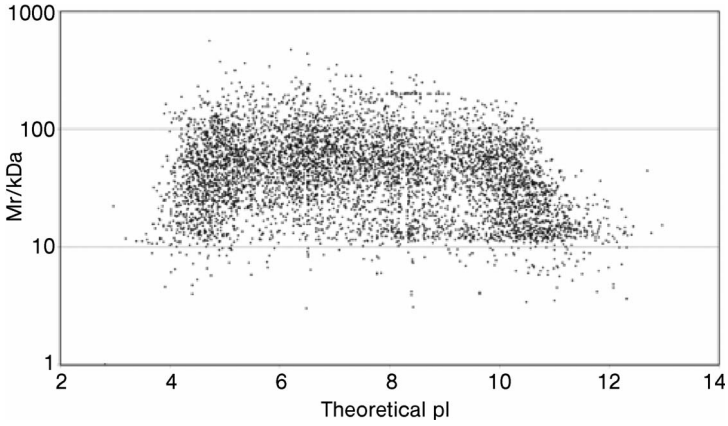
There are more definitions to find. Often they are linked to the application area.

- Availability of genomic sequence information
- Development of novel techniques of mass spectrometry.
- Availability of computing power, memory, and database accessibility.
- Improvement of separation technologies.

Furthermore it became obvious that the genomic sequence and protein function cannot be directly correlated: Co- and post-translational protein modifications cannot be predicted from the genome sequence. And it is known, they play a very important role in causing diseases. However, the DNA sequence can be "in silico" translated into the protein sequence, and therefore genome databases can be used for identification.

The theoretical 2-D maps of other organisms look in principle similar; they differ mainly in the complexity.

As an example a plot of the molecular masses versus the isoelectric points of the theoretically expressed proteins of the yeast genome is shown in Figure 1. There are many reasons, why this picture looks very different from the result of a 2-D electrophoresis of a yeast cell extract (see Figure 2):



Wildgruber R, Harder A, Obermaier C, Boguth G, Weiss W, Fey SJ, Larsen PM, Görg A. Electrophoresis 21 (2000)

Fig. 1: Theoretical two-dimensional map of masses and isoelectric points calculated from the protein sequences which have been "in silico" translated from the open reading frames of the yeast genome (from Wildgruber *et al.* 2000).

- A proteome reflects the actual metabolic state of a cell. It is a highly dynamic object and strongly dependent on many parameters.
- The plot cannot reflect the protein expression levels.
- Not all possible proteins are expressed.
- Many proteins are expressed in low copy numbers, often they are below the detection limit. Particularly proteins in the basic area, like regulatory proteins, transcription factors, and other DNA-binding proteins are mostly missed.
- A number of proteins have become modified in different ways during or after translation.
- A number of proteins are outside the working range of 2-D electrophoresis.

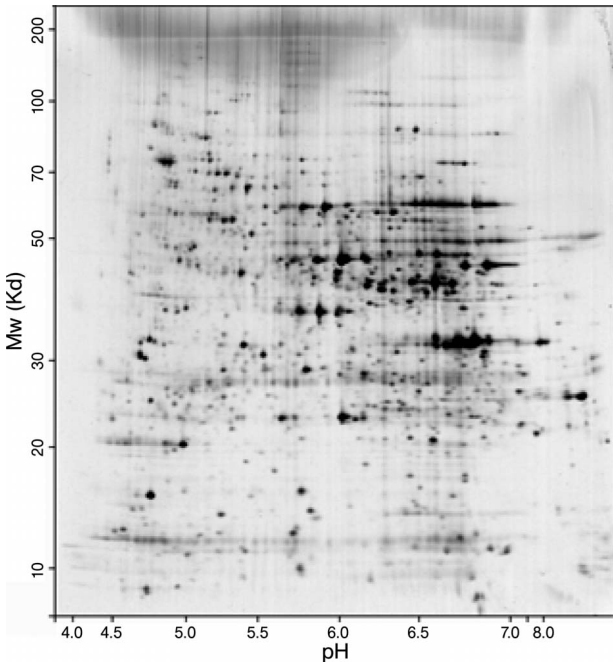


Fig. 2: Two-dimensional electrophoresis of yeast proteins as shown on the SWISS-2DPAGE database on the free accessible Expasy server (from Sanchez *et al.* 1996).

Sanchez JC, Golaz O, Frutiger S, Schaller D, Appel RD, Bairoch A, Hughes GJ, Hochstrasser DF. Electrophoresis 17 (1996) 556–565.

A large 2-D electrophoresis gel of 20 × 20 cm has a theoretical separation space of about 10,000 proteins.

A view on the working range of 2-D electrophoresis – as displayed in Figure 3 – can explain, why 2-D electrophoresis had been selected as the first choice of separation methods for the analysis of proteomes. Still the separation according to two completely independent physicochemical parameters of proteins, isoelectric point and size, offers the highest resolution. Several thousands of proteins can be separated, displayed and stored in one gel. Proteins in the size range from 10 kDa to 200 kDa and with isoelectric points between 3 and 11 can be analyzed. Because the separation takes place under completely denaturing conditions, also quite hydrophobic proteins are included in the work range. It seems like two-dimensional electrophoresis will remain the major separation technique, because its resolution and the advantage of storing the isolated proteins in the gel matrix until further analysis is unrivalled by any of the alternative techniques.

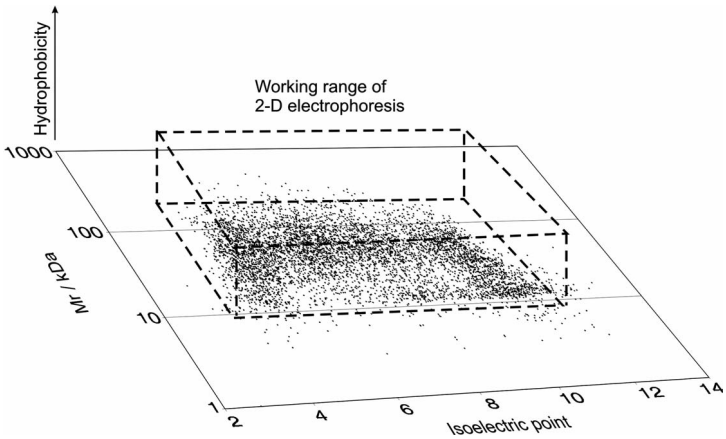


Fig. 3: Estimated working range of 2-D electrophoresis for separating highly complex protein mixtures.

However, there also some shortcomings of 2-D electrophoresis:

- Small, very large, very basic, and very hydrophobic proteins are widely excluded.
- 2-D electrophoresis is rather complex, not automated, labor-intensive, and therefore dependent on the skills of the operator.
- Even optimal separations show gel-to-gel variations. This results in difficult image analysis procedures.
- The peptide yield after in-gel digestion of proteins is considerably lower than in liquid phase. This leads to limited sensitivity in the subsequent mass spectrometry analysis.

Therefore proteomics researchers started to look for alternatives to either replace or – at least – to complement the results acquired with the 2D gel-based workflow. The most successful approach employs tryptic digestion of the entire protein mixture and analysis of the peptides with the combination of nanoscale liquid chromatography and electrospray mass spectrometry. This procedure was either called DALPC (Direct analysis of protein complexes, see Link *et al.* 1999) or MudPIT (Multidimensional protein identification technology, see Washburn *et al.* 2001). The major advantages of the LC-based workflows are the superior sensitivity and the possibility of automation by an LC-ESI MS via on-line connection. Several orthogonal separation techniques are combined to MDLC (Multi Dimensional Liquid Chromatography).

Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR III. Nature Biotech 17 (1999) 676–682.

Washburn MP, Wolters D, Yates JR III. Nature Biotech 19 (2001) 242–247.

At the present time, most multi-dimensional LC applications in proteomics deal with the separation of tryptic peptides. A variety of semi-automated off-line and fully automated on-line, as well as high-throughput configurations are available as commercial systems or can be customized according to the individual needs and preferences of the operators. Although this type of advanced tryptic peptide separation is often referred as multi-dimensional, actually it only utilizes two dimensions, namely ion exchange chromatography – cation exchange chromatography preferred – in combination with reversed phase chromatography.

Still in its infancy, multi-dimensional chromatography is enjoying more and more acceptance as a sample preparation tool for the pre-fractionation of intact proteins further upstream the proteomics workflow. The techniques and methods applied in protein pre-fractionation have been derived and adapted from protein purification, which are in use since decades with great success and reliability.

Finally, the orthogonal, high resolution separation at both protein and peptide level would deserve the term multidimensional liquid chromatography (MDLC).

Practice has shown that these different workflows develop different subsets of the same proteome with surprisingly little overlaps. A typical example can be found in the paper by Vanrobaeys *et al.* (2005). Thus none of them can be replaced by the other one. But it has been recognized that several complementary workflows need to be employed in order to keep the number of missed proteins as low as possible.

Vanrobaeys F, Van Coster R, Dhondt G, Devreese B, Van Beeumen J. J Proteome Res 4 (2005) 2283 – 2293. There are even differences within the same workflows, caused by different design of equipment.

Furthermore, another important aspect is stated in a paper by Chamrad and Meyer (2005): Today … "there are no basic rules on how to perform a proteomic study and manuscripts can frequently be found that publish results from single … experiments without any repetition, which can become problematic for further independent validation steps. Thus, search strategies and data evaluation methods

Chamrad D, Meyer HE. Nat Methods 2 (2005) 647–648.

Elias JE, Haas W, Faherty BK, Gygi SP. Nat Methods 2 (2005) 667–675.

in … proteome studies must be improved, and the manuscript by Gygi and colleagues gives some very useful directions …".

Other combinations than 2-D gel-MS and LC-MS have been introduced, which deliver highly satisfying results for special samples and experiments. For instance, very frequently one-dimensional SDS PAGE followed by tryptic digestion of proteins with subsequent LC-MS is employed. Also for separations on the peptide level electrophoretic alternatives have been developed to complement liquid chromatography, at least in the first stage. Furthermore, it became obvious that pre-fractionation of the highly complex protein mixtures leads to more successful protein identifications than direct analysis of crude samples. Figure 4 shows an overview of analysis modules applied in proteomics, which can be assembled to various workflows.

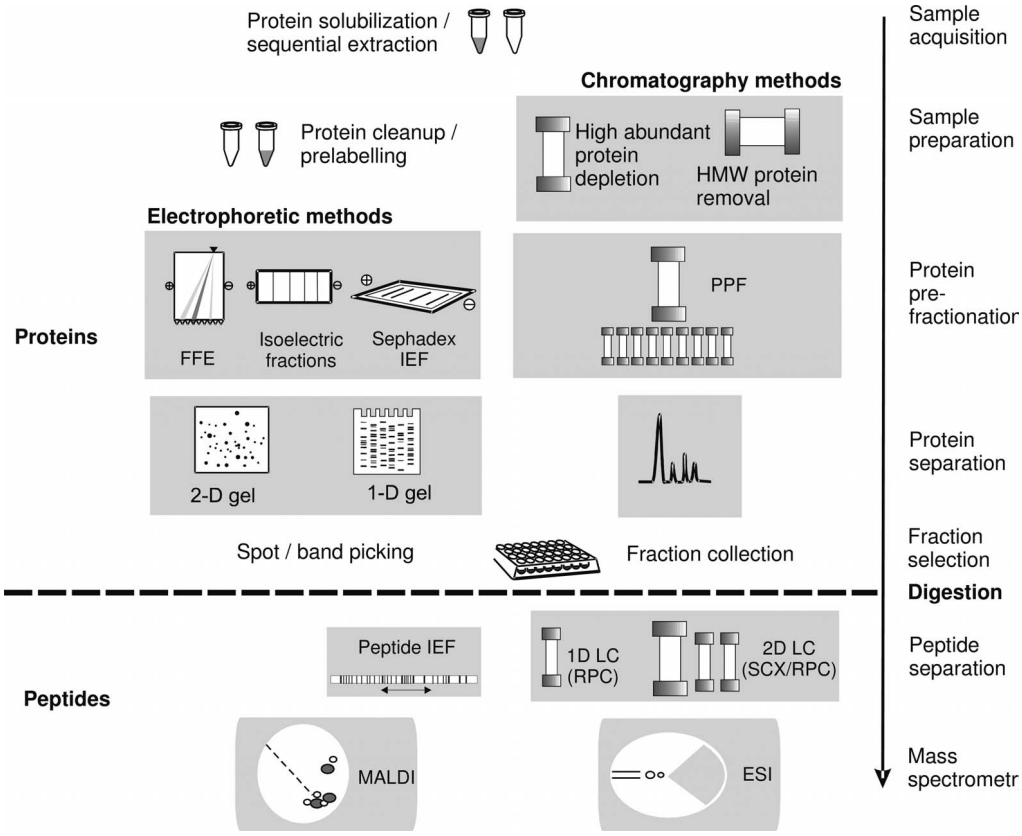


Fig. 4: Toolbox for proteome analysis. The modules can be combined to various workflows in different ways. The functions and features of the techniques displayed here will be described in more detail in the following book chapters. On the right hand side the chronological order of the analysis is indicated. Note the important division between protein and peptide level.

These technologies and their combinations will be described in the first part of the book.

Since the start of the "Proteomics Era" huge progress has been made in the instrumental development for improved nanoscale liquid chromatography, higher resolution and more sensitive mass spectrometers, evaluation software, and peripherical technologies.

A great step forward is the concept of difference gel electrophoresis (DIGE). With this method, introduced by Ünlü *et al.* (1997), protein samples are pre-labeled with modified cyanine dyes (CyeDye™), mixed, and separated together in the same gel. The co-migrated protein spots of the different samples are detected by scanning at different wavelengths; their abundance ratios are determined with dedicated software, which employs a spot co-detection algorithm. This approach makes it now possible to use an internal standard in 2-D gel electrophoresis (Alban *et al.* 2003). In this way gel-to-gel variations are compensated, which leads to highly confident quantitative and qualitative results. The technique has been applied on almost all different sample types, and during the last couple of years the number of papers on the DIGE method has increased exponentially (see Figure 5).

Ünlü M, Morgan EM, Minden JS. Electrophoresis 19 (1997) 2071–2077.

Alban A, David S, Bjorkesten L, Andersson C, Sloge E, Lewis S, Currie I. Proteomics 3 (2003) 36–44.

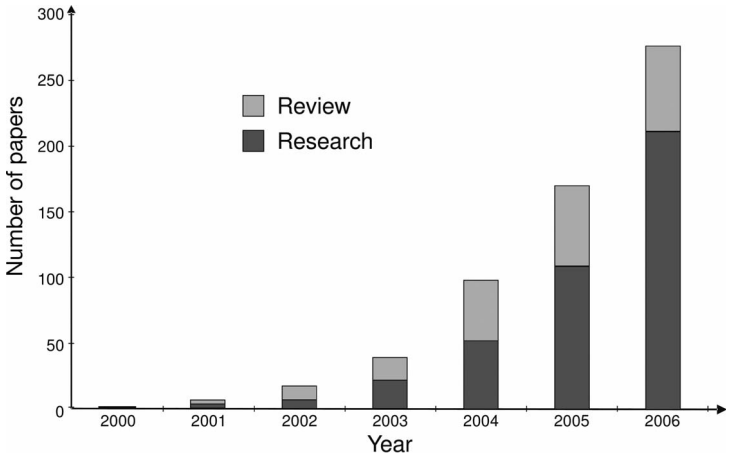


Fig. 5: Graphical representation of the number of published papers in DIGE until end of the year 2006.

At present, some major projects and developments are pursued, which raise high expectations for proteomics. Here are a few examples:

- The systematic exploration of the human proteome with Affinity (Antibody) Proteomics to generate quality assured antibodies to all non-redundant human proteins (Uhlén and Ponten, 2005).

Uhlén M, Ponten F. Mol Cell Proteomics 4 (2005) 384–393.

Misek DE, Kuick R, Wang H, Galchev V, Deng B, Zhao R, Tra J, Pisano MR, Amunugama R, Allen D, Walker AK, Strahler JR, Andrews P, Omenn GS, Hanash SM. Proteomics 5 (2005) 3343–3352.

- The combination of DIGE labeling, liquid chromatography of proteins, SDS PAGE and LC-MS for finding biomarkers in samples with very wide dynamic ranges of protein expression levels (Misek *et al.* 2005).

The generation of small subsets of intact proteins is still a challenge.

- The further developments for the top-down approach with FT-ICR mass spectrometry.
- The development of protein arrays.

Many of these critical points will be described in the following sub-chapter.

During the first few years of the proteomics era holistic approaches, mostly not hypothesis driven, were preferred in order to study complete proteomes at once by high-throughput methods. It was assumed that a proteome could be analyzed in a similar way like a genome, just with a higher effort. Unfortunately it turned out that these protein samples have more challenges in store than expected. Thus it can be observed that Proteomics is now evolving from a high-throughput industrial-scale concept ("shotgun proteomics") to carefully planned experiments and hypothesis driven analyses in order to answer certain biological questions.

2 Critical Points

2.1 Challenges of the Protein Samples

As many steps as necessary, but as little as possible!

Example: the human genome contains about 22,000 genes. With PTMs a few hundred thousand human proteins can be expected.

Usually the complexity of the protein and/or peptide mixture lies beyond the theoretical separation space of any separation method. This issue can only be solved by intelligent pre-fractionation of the sample and analyzing smaller protein subsets. But it should be noted that the more separation steps are involved, the more proteins can get lost due to technical reasons. Furthermore, the analysis of one complex sample can take quite a long time.

Five steps with 80% recovery each – which is not too bad – gives less than 40% overall recovery (see Figure 6). It becomes obvious, if not choosing a proper strategy, that there is a high risk of losing the entire sample.