New and Emerging Proteomic Techniques

Edited by

Dobrin Nedelkov
Randall W. Nelson
New and Emerging Proteomic Techniques
Preface

Ever since its first definition, proteomics has been referred to as the study of the full set of proteins encoded by a genome, and qualitative and quantitative comparisons of proteomes under different conditions to further unravel biological processes. Initially, proteomics was mainly a technology-driven field, centered on two-dimensional gels/mass spectrometry and LC-MS/MS approaches, which offered the ability to detect hundreds of proteins in a single analysis. Proteomics has since evolved into a wide-ranging discipline that includes a plethora of technologies and approaches whose subject is the study of proteins. However, as proteomics is entering into the realms of clinical and diagnostic applications, the enabling techniques are coming under an increased level of scrutiny. Significant weaknesses in some technological aspects have thus become apparent. As a result, new and improved proteomics techniques are being developed and propagated.

It is the intent of *New and Emerging Proteomic Techniques* to present some of the newer and still developing proteomics tools and techniques that enable enhanced protein analyses. The techniques span the entire spectrum of top-down and bottom-up approaches, and in their sum offer a clear example of how proteomics has embraced essentially all techniques that contend with protein analysis. From microarrays and gels, to chromatography and affinity separations, the proteomics techniques described in this book are addressing every aspect of the human proteome, both quantitative and qualitative. The methods of protein detection utilized are also very diverse, ranging from fluorescence and resonance light scattering, to surface plasmon resonance and mass spectrometry. Furthermore, several chapters describe a combination of two or more distinct techniques, resulting in enabling approaches for proteome analysis. There are also three chapters that describe advanced bioinformatics approaches, as they are becoming increasingly important in the analysis of the complex proteomics data.

*New and Emerging Proteomic Techniques* is aimed at both beginners and more experienced practitioners in the field of proteomics. Beginners will find it very useful to have such a diverse set of techniques assembled in a single book, serving as a valuable reference when choosing a technique that can address a specific proteomics question. For experienced proteomics researchers, the book offers protocols and know-how from the pioneers and the expert users of each of these techniques, with details that are usually not found in a typical research publication. We are well aware that there are other proteomics
approaches that are not represented in this book. However, we feel that the
fifteen chapters included describe some of the most promising new and emerg-
ing proteomics techniques, and hope that at least some of them will become
proteomics mainstays in the years to come.

Finally, we would like to thank all of our colleagues who kindly contributed
their time and expertise for the assembly of New and Emerging Proteomic
Techniques.

Dobrin Nedelkov
Randall W. Nelson
Contents

Preface .................................................................................................................................................. v
Contributors .......................................................................................................................................... xi

1 On-Chip Protein Synthesis for Making Microarrays
Niroshan Ramachandran, Eugenie Hainsworth, Gokhan Demirkan, and Joshua LaBaer .......................................................... 1

2 RCA-Enhanced Protein Detection Arrays
Brian B. Haab and Paul M. Lizardi ................................................................. 15

3 Antibody Microarrays Using Resonance Light-Scattering Particles for Detection
Bernhard H. Geierstanger, Petri Saviranta, and Achim Brinker ........ 31

4 Chemical Proteomics Profiling of Proteasome Activity
Martijn Verdoes, Celia R. Berkers, Bogdan I. Florea, Paul F. van Swieten, Herman S. Overkleeft, and Huib Ovaa ....... 51

5 Two-Dimensional Difference Gel Electrophoresis
Terence L. Wu ................................................................................................. 71

6 Oligomeric States of Proteins Determined by Size-Exclusion Chromatography Coupled With Light Scattering, Absorbance, and Refractive Index Detectors
Ewa Folta-Stogniew .......................................................................................... 97

7 Surface Plasmon Resonance Imaging Measurements of Protein Interactions With Biopolymer Microarrays
Terry T. Goodrich, Alastair W. Wark, Robert M. Corn, and Hye Jin Lee ............................................................................................. 113

8 Surface Plasmon Resonance Mass Spectrometry for Protein Analysis
Dobrin Nedelkov and Randall W. Nelson ...................................................... 131

9 High-Throughput Affinity Mass Spectrometry
Urban A. Kiernan, Dobrin Nedelkov, Eric E. Niederkofler, Kemmons A. Tubbs, and Randall W. Nelson ............................................. 141

10 Isotope-Coded Affinity Tags for Protein Quantification
Christopher M. Colangelo and Kenneth R. Williams................................. 151
11 Proteomic Analysis by Multidimensional Protein Identification Technology  
Laurence Florens and Michael P. Washburn ........................................... 159

12 Isolation of Glycoproteins and Identification of Their N-Linked Glycosylation Sites  
Hui Zhang and Ruedi Aebersold ................................................................... 177

13 N-Glycosylation Analysis Using the StrOligo Algorithm  
Martín Ethier, Daniel Figeys, and Hélène Perreault ................................. 187

14 MALDI-MS Data Analysis for Disease Biomarker Discovery  
Weichuan Yu, Baolin Wu, Junfeng Liu, Xiaoye Li, Kathy Stone,  
Kenneth R. Williams, and Hongyu Zhao .................................................. 199

15 Using the Global Proteome Machine for Protein Identification  
Ronald C. Beavis ...................................................................................... 217

Index ........................................................................................................... 229
Contributors

RUEDI AEBERSOLD • Institute for Systems Biology, Seattle, WA, and Swiss Federal Institute of Technology (ETH) and Faculty of Natural Sciences, University of Zurich, Zurich, Switzerland

RONALD C. BEAVIS • Beavis Informatics Ltd., Winnipeg, Canada

CELIA R. BERKERS • Division of Cellular Biochemistry, Netherlands Cancer Institute, Amsterdam, the Netherlands

ACHIM BRINKER • Genomics Institute of the Novartis Research Foundation, San Diego, CA

CHRISTOPHER M. COLANGELO • W. M. Keck Foundation Biotechnology Resource Laboratory, Yale University, New Haven, CT

ROBERT M. CORN • Department of Chemistry, University of California, Irvine, CA

GOKHAN DEMIRKAN • Harvard Institute of Proteomics, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Cambridge, MA

MARTIN ETHIER • Department of Chemistry, University of Manitoba, Winnipeg, Canada, and Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, Canada

DANIEL FIGEYS • Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, Canada

BOGDAN I. FLOREA • Leiden Institute of Chemistry, Leiden University, Leiden, the Netherlands

LAURENCE FLORENS • Stowers Institute for Medical Research, Kansas City, MO

EWA FOLTA-STOGNIEW • W. M. Keck Foundation Biotechnology Resource Laboratory, Yale University, New Haven, CT

BERNHARD H. GEIERSTANGER • Genomics Institute of the Novartis Research Foundation, San Diego, CA

TERRY T. GOODRICH • Department of Chemistry, University of California, Irvine, CA

BRIAN B. HAAB • The Van Andel Research Institute, Grand Rapids, MI

EUGENIE HAINSWORTH • Harvard Institute of Proteomics and Technology and Engineering Center, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Cambridge, MA

URBAN A. KIERNAN • Intrinsic Bioprobes Inc., Tempe, AZ
JOSHUA LABAER • Department of Biological Chemistry and Molecular Pharmacology, Harvard Institute of Proteomics, Harvard Medical School, Cambridge, MA
HYE JIN LEE • Department of Chemistry, University of California, Irvine, CA
XIAOYE LI • Department of Applied Mathematics, Yale University, New Haven, CT
JUNFENG LIU • Department of Statistics, West Virginia University, Morgantown, WV
PAUL M. LIZARDI • Department of Pathology, School of Medicine, Yale University, New Haven, CT
DOBRIN NEDELKOV • Intrinsic Bioprobes Inc., Tempe, AZ
RANDALL W. NELSON • Intrinsic Bioprobes Inc., Tempe, AZ
ERIC E. NIEDERKOFLER • Intrinsic Bioprobes Inc., Tempe, AZ
HUIB OVAAN • Division of Cellular Biochemistry, Netherlands Cancer Institute, Amsterdam, the Netherlands
HERMAN S. OVERKLEEFT • Leiden Institute of Chemistry, Leiden University, Leiden, the Netherlands
HÉLÈNE PERREAULT • Department of Chemistry, University of Manitoba, Winnipeg, Canada
NIROSHAN RAMACHANDRAN • Department of Biological Chemistry and Molecular Pharmacology, Harvard Institute of Proteomics, Harvard Medical School, Cambridge, MA
PETRI SAVIRANTA • Genomics Institute of the Novartis Research Foundation, San Diego, CA
KATHY STONE • W. M. Keck Foundation Biotechnology Resource Laboratory, Yale University, New Haven, CT
KEMMONS A. TUBBS • Intrinsic Bioprobes Inc., Tempe, AZ
PAUL F. VAN SWIETEN • Leiden Institute of Chemistry, Leiden University, Leiden, the Netherlands
MARTIJN VERDOES • Leiden Institute of Chemistry, Leiden University, Leiden, the Netherlands
ALASTAIR W. WARK • Department of Chemistry, University of California, Irvine, CA
MICHAEL P. WASHBURN • Stowers Institute for Medical Research, Kansas City, MO
KENNETH R. WILLIAMS • W. M. Keck Foundation Biotechnology Resource Laboratory, Yale University, New Haven, CT
BAOLIN WU • Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN
TERENCE L. WU • W. M. Keck Foundation Biotechnology Resource Laboratory, Yale University, New Haven, CT
Contributors

WEICHUAN YU • Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT
HUI ZHANG • Institute for Systems Biology, Seattle, WA
HONGYU ZHAO • Department of Epidemiology and Public Health, Yale University, New Haven, CT
On-Chip Protein Synthesis for Making Microarrays

Niroshan Ramachandran, Eugenie Hainsworth, Gokhan Demirkan, and Joshua LaBaer

Summary

Protein microarrays are a miniaturized format for displaying in close spatial density hundreds or thousands of purified proteins that provide a powerful platform for the high-throughput assay of protein function. The traditional method of producing them requires the high-throughput production and printing of proteins, a laborious method that raises concerns about the stability of the proteins and the shelf life of the arrays. A novel method of producing protein microarrays, called nucleic acid programmable protein array (NAPPA), overcomes these limitations by synthesizing proteins in situ. NAPPA entails spotting plasmid DNA encoding the relevant proteins, which are then simultaneously transcribed and translated by a cell-free system. The expressed proteins are captured and oriented at the site of expression by a capture reagent that targets a fusion protein on either the N- or C-terminus of the protein. Using a mammalian extract, NAPPA expresses and captures 1000-fold more protein per feature than conventional protein-printing arrays. Moreover, this approach minimizes concerns about protein stability and integrity, because proteins are produced just in time for assaying. NAPPA has already proven to be a robust tool for protein functional assays.

Key Words: Protein microarrays; functional proteomics; protein expression; protein purification; microarray surface chemistry.

1. Introduction

The recent development of functional protein microarrays has stirred excitement in the proteomics community (1–7). The power of this approach is that, by spotting many proteins on a single array surface, many biochemical activities can be studied simultaneously. These activities include identifying interacting proteins, examining the selectivity of drug binding, finding substrates for active enzymes, and looking for unintended drug interactions. Typically, the array is probed with a labeled query molecule to identify interactions with proteins on
the array (Fig. 1). For example, a labeled candidate kinase inhibitor might be used to screen an array of kinases to determine which kinase(s) the inhibitor binds directly. In order to build protein microarrays, one needs the content to spot on the array and an appropriate binding chemistry to capture the protein. These components must be optimized to produce and present proteins of good integrity and stability. The goal is to preserve the functionality of the protein in order to minimize false-negatives. Here we will address the issues pertaining to building functional protein microarrays.

1.1. Issues for Protein Array Production

1. **Availability of array content.** Assembling the proteins for printing on the array remains a major challenge for most researchers, because recombinant expression of proteins in the numbers anticipated for protein microarrays relies on the availability of large collections of cDNAs in protein expression-ready formats. It also requires methods to produce and purify the proteins. Although several collections of cDNAs are available, the methods and robotic equipment required for the high-throughput (HT) expression and purification of thousands of proteins remain outside the realm of most laboratories (8–10).

2. **Protein integrity.** Ensuring that properly folded proteins can be produced and captured remains a challenge. Proteins are more likely to fold naturally if heterologous
systems can be avoided and if proteins are synthesized in a milieu as close to their natural setting as possible. For example, mammalian proteins are more likely to fold naturally in mammalian (or at least animal) cells. Yet, many expression systems are too cumbersome and expensive to allow thousands of proteins to be easily processed. Bacterial cells, which are readily adapted to HT protein expression, can be counted on to produce only 50% of mammalian proteins (10).

3. **Protein stability.** Proteins are notoriously fragile, raising concerns about the stability of the isolated proteins before and after they have been arrayed on a glass slide. We expect some proteins to remain relatively stable, with good shelf life, whereas others display greater lability and are unable to withstand prolonged array conditions. Moreover, it is difficult to determine which of the proteins remain active at the time of the assay. In general, to ensure proper assay conditions and minimize false-negatives, it is best to use the array soon after the protein is synthesized.

4. **Microarray surface chemistry.** Several surface chemistries have been developed and validated for microarray platforms, particularly for DNA microarrays. The chemistry for binding DNA is simple compared with the chemical demands necessary to immobilize functional protein. DNA molecules, which are all negatively charged, bind to surfaces based on charge alone, enabling positively charged arrays such as a polylysine-coated slides to bind all DNA. In contrast, proteins display a staggering range of hydrophobicity and charge, making it a challenge to find a single method that provides good binding for most, let alone all, proteins. Factors to consider include:
   a. Generality of binding: ability to bind all proteins that will be spotted on the array.
   b. Binding capacity: maximum amount of protein captured per feature.
   c. Efficiency of capture: fraction of spotted protein that is captured on the array.
   d. Orientation: specific vs random orientation. Proteins can be immobilized either in an orientation-specific manner (e.g., by binding via either an N-terminus or a C-terminus tag) or in random orientations (e.g., by chemical attachment). Random rather than specific orientations may allow many areas of the protein to be exposed, increasing accessibility to the protein. Although this may increase the likelihood of an interaction, there have been no significant differences observed between these approaches (11); it may be necessary to evaluate this on an experiment by experiment basis.
   e. Distance from surface: some attachment methods allow for a spacer (e.g., a large polypeptide tag) that separates the protein from the array surface; other methods (e.g., chemical attachment) bring the proteins in direct contact with the array surface. Increasing the distance between the protein and the array surface might alleviate some of the steric hindrance caused by the surface and potentially increase accessibility to the protein.
   f. Native or denatured protein: surface chemistry can be formulated to contain hydrophobic or hydrophilic residues. Given that many proteins have a hydrophilic exterior and a hydrophobic interior, the choice of the surface chemistry could support the binding of non-denatured or denatured protein (12).
Early demonstration of the feasibility of printing proteins on a microscopic surface has been promising. To demonstrate that spotted proteins maintain their functional integrity upon immobilization, well characterized and specific interactions among proteins, lipids, and small molecules, as well as enzyme-substrate screens, were recapitulated with proteins on the arrays (1,3,6,7,11,13–20). Even in light of these achievements, the widespread use of this technology has remained limited, largely as a result of the labor-intensive protein production, the quality of proteins expressed in heterologous systems, and the stability of the proteins during storage. To address these persistent concerns, we developed a self-assembling protein microarray method.

1.2. Nucleic Acid Programmable Protein Array

To circumvent the need to express, purify, and spot the protein, this approach prints the plasmids bearing the genes on the array and the proteins are synthesized in situ. The genes are configured such that each expressed protein contains a polypeptide tag used to capture the protein to the array surface. The proteins are expressed using a cell-free transcription/translation extract, which can be selected to match the source of the genes (e.g., rabbit reticulocyte lysate for mammalian genes), thus enabling the proteins to be expressed in a more native milieu. The use of appropriate cell-free extracts helps to encourage natural folding and, at least in the case of reticulocyte lysate, is highly successful at expressing most proteins. In addition, some natural posttranslational modifications occur in these extracts and/or can be induced by using supplemented lysates (21,22).

Arranging the genes so that each has an appropriate capture tag is facilitated by using vectors with recombinational cloning sites. Coding regions inserted in recombinational cloning systems, such as the Invitrogen Gateway system or Clontech Creator system, can be readily moved into expression vectors that append the appropriate tag(s) to the coding regions. The transfer reactions themselves are simple, highly efficient, error-free, and automatable. The assembly of large collections of genes in these systems is currently in progress (10,23–27).

A significant advantage of the nucleic acid programmable protein array (NAPPA) approach is that it eliminates concerns about protein stability. Proteins on the array are not produced until the array is ready for use in experiments; that is, they are made just in time. Prior to activation with the cell-free transcription/translation extract, the arrays are stable and can be stored dry on the bench for months.

Using this approach in a recent study, 30 human DNA replication proteins were expressed and captured on NAPPA microarrays (28). The yield of captured protein was 400–2700 pg/feature, which was 1000-fold more than con-
On-Chip Protein Synthesis of Microarrays

Conventional protein-spotting arrays, 10–950 fg/feature (11). Arrays were used to determine protein-protein interactions (recapitulating 85% of the previously known interactions), to map protein interaction domains by using partial-length proteins, and to assemble multiprotein complexes.

2. Materials

2.1. Equipment

1. Arrayer with solid pins, humidity control.
2. Microarray scanner.
3. Programmable chilling incubator.
4. SpeedVac.
6. Ultraviolet (UV) light, UVP UVLMS-38, set at 365 nm.

2.2. Preparation of the Slides

2. Solution of 2% aminosilane (Pierce 80370) in acetone. Make up 300 mL just before use.
3. Stainless steel 30-slide rack (Wheaton), handle removed.
5. Lock & Lock 1.5 cup boxes (Heritage Mint Ltd., ZHPL810).
6. Prepare a 50 mM dimethyl suberimidate·2 HCl (DMS) stock solution: 1 g of DMS linker (Pierce 20700) in 40 mL dimethylsulfoxide (DMSO). Store at –20°C.
7. To coat slides with linker only (used if NAPPA strategy is to spot avidin/streptavidin along with plasmid DNA and anti-glutathione S-transferase [GST] antibody): 2 mM DMS in phosphate-buffered saline (PBS), pH 9.5 (see Note 1).
8. To coat slides with avidin/streptavidin (used if NAPPA strategy is to spot only plasmid DNA and anti-GST antibody): 2 mM DMS, plus avidin (Cortex CE0101) at 1 mg/mL or streptavidin (Cortex CE0301) at 3.5 mg/mL, in PBS, pH 9.5. For material in either step 7 or 8, make fresh at the time of coating, otherwise the DMS linker may hydrolyze over time (see Note 1).
10. Bioassay dishes with dividers (Genetix x6027).

2.3. DNA Preparation

1. The plasmid DNA is prepared in 300-mL cultures usually grown in Terrific Broth media. The DNA preparation is derived from Sambrook et al. (29) and is summarized below.
2. Solution 1 (GTE): 50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM ethylenediamine tetraacetic acid (EDTA), pH 8.0, and 0.1 mg/mL RNAs. Store at 4°C.
3. Solution 2: 0.2 N NaOH with 1% sodium dodecyl sulfate (SDS).
4. Solution 3: 3 M KOAC; add glacial acetic acid until pH is 5.5.
5. 250-mL conical Corning centrifuge bottle.
6. Glass fiber 0.7-µm filter plate, long drip (Innovative Microplate F20060).
7. 96-well deep-well block (Marsh AB-0661).

2.4. Preparation of Samples and Arraying

1. Plasmid DNA (prepared in Subheading 2.3.).
2. Microcon YM-100 (100 kDa) tube (Millipore), or DNA binding plate: 100 kDa 96-well filter plate (Millipore plasmid plate).
3. BrightStar Psoralen-biotin kit (Ambion 1480). Just before use, prepare psoralen-biotin: dissolve the contents (4.17 ng) of the kit in 50 µL DMF (also in kit).

Or

4. EZ-Link Psoralen-PEO-Biotin (Pierce 29986). Prepare stock solution of 5 mg/mL in water and store at –20°C.
5. UV-transparent 96-well plate (Corning 3635).
6. Sephadex G50 (Sigma-Aldrich).
7. 1.2-µm glass fiber filter plate, long drip (Innovative Microplate F20021).
8. Collection plate, round bottom (Corning 3795).
9. 384-well plate for arraying (Genetix x7020).
11. Purified GST protein (Sigma G5663). Prepare stock solution of 0.03 mg/mL in PBS.
12. Whole mouse immunoglobulin (Ig)G antibody (Pierce 31204). Prepare stock solution of 0.5 mg/mL in PBS.
13. Bis(sulfo)succinimidyl) suberate (BS3) linker (Pierce 21580).
14. Bioassay dish dividers to be used as slide racks (Genetix x6027) and deeper bioassay dishes (e.g., Corning 431111 or 431272; do not use “low profile” dishes).

2.5. Expression of Proteins

1. HybriWell gaskets (Grace HBW75).
2. Cell-free expression system (Rabbit reticulocyte lysate) (Promega L4610).
3. RNaseOUT (Invitrogen 10777–019).
4. SuperBlock blocking solution in TBS (Pierce 37535).
5. Milk blocking solution: 5% milk in PBS with 0.2% Tween-20 (Sigma).

2.6. Detection and Analysis

1. Primary AB solution: mouse anti-GST (Cell Signaling 2624) 1:200 in SuperBlock (Pierce 37535). Store at 4°C.
2. Primary AB solution: mouse anti-HA (Cocalico) 1:1000 in SuperBlock. Store at 4°C.
4. Tyramide Signal Amplification (TSA) stock solution: use TSA reagent (PerkinElmer SAT704B001EA). Prepare per kit directions. Keep this solution at 4°C.
5. Milk blocking solution: 5% milk in PBS with 0.2% Tween-20 (Sigma).
6. Cover slips (VWR 48393–081).
7. PicoGreen (Molecular Probes P11495) stock solution: to the 100 µL/vial that comes in kit, add 200 µL TE buffer. Before use, do a 1:600 dilution in SuperBlock.

3. Methods

NAPPA chemistry relies on efficient immobilization of plasmid DNA onto a solid surface without compromise to integrity, and on rapid capture of the expressed target proteins. In order to immobilize the plasmid, we use a psoralen-biotin bis-functional linker that derivatizes the plasmid DNA (Fig. 2). Under long-wave UV (365 nm), psoralen intercalates into the DNA, creating a biotinylated plasmid. The reaction is fairly robust over a wide range of pH and salt concentrations. The biotinylated plasmid is tethered to the array surface by high-affinity binding to either avidin or streptavidin. In addition to the plasmids, target protein capture molecules are also immobilized on the slide. Currently, our plasmids are programmed to express target proteins with a C-terminal GST fusion protein; therefore, a polyclonal anti-GST antibody is bound to the array as the capture molecule to immobilize the expressed target proteins (Fig. 3). The presence of the C-terminal fusion tag can later be confirmed by incubating the slides with an antibody that recognizes a different epitope on the tag than the antibody used for capture. The presence of the C-terminal tag indicates that the full-length protein was expressed.

In order to make this chemistry robust and reproducible, we have used high-affinity capture reagents that are well characterized and stable throughout arraying and storage. Moreover, the schemes outlined previously can be altered by the user to accommodate different immobilization chemistries for the plasmid DNA and/or target proteins.

3.1. Preparation of the Slides

1. Prepare 300 mL of aminosilane coating solution (2% aminosilane reagent in acetone).
2. Put slides in metal rack (30-slide Wheaton rack).
3. Treat glass slides in the aminosilane coating solution, approx 1–15 min in glass staining box on shaker. Rinse with acetone in rack using wash bottle. Briefly rinse with Milli-Q water. Spin dry in SpeedVac or use 0.2-µm filtered air cans or use house air with 2 × 0.25 µm filters (see Note 2). It is important to use clean air to dry slides in order to prevent contaminating debris from binding to the surface.
4. Store at room temperature in metal rack in Lock & Lock box.
5. Just before use, prepare linker solution according to Subheading 2.2. step 7 or step 8, depending on the array strategy.
6. Set slides on divider in bioassay dish, with water in the bottom of the tray. Treat each slide with 150–200 µL linker solution and cover slip (see Note 3). Incubate for 2–4 h at room temperature or overnight in cold-room.