

## Laboratory Exercises

### Teaching Proteomics

Received for publication, November 6, 2002

Leigh Plesniak‡ and Ellis Bell§¶

From the ‡Chemistry Department, University of San Diego, San Diego, California 92110 and the §Department of Chemistry, University of Richmond, Richmond, Virginia 23173

The sequencing of the human genome opened doors to a vast array of information and the explosion of bioinformatics, an essential part of any undergraduate curriculum in biochemistry or molecular biology. While the genomic revolution has given us much information about when and where proteins are expressed in a cell and bioinformatic approaches are increasingly able to predict function and potential interactions between types of proteins, no information is obtained about actual activities and interactions of proteins without actually examining and studying the proteins themselves: hence proteomics, or the study of the expressed protein in a given cell. Questions that are not answered (although clues may be given) by genomics or bioinformatics include issues of the actual covalent structure of the expressed protein (is the protein post translationally modified?), the interactions of the protein with other cellular components, and the overall activity of the protein. These questions require a wide variety of old and new techniques to answer, and while studies of the isolated protein can play a role in providing some of this information it is clear that the future will require techniques that increasingly look inside the cell at activities *in situ*. To answer the question “what is proteomics?” is complex. Proteomics is a rapidly evolving field, and there are many different answers to this question. At present the answer is most likely that proteomics is the documenting of the products of the expressed genome and the nature of interactions between gene products in any given cell. As work toward these ends progresses, proteomics will evolve to include an understanding of the functions of the expressed proteins in a given cell type. With ~50% of all gene products having unknown structure or function at present, the focus of most proteomic research is currently in the area of “structural” proteomics. As we understand more of the answers that are being provided by current work, the focus will inevitably change to the more complex issues of “functional” proteomics. The answers to issues of functional proteomics requires understanding the interactions and effects of the internal milieu on the activity of the protein and its various complexes and interactions.

Currently much research in proteomics is focused on the question of identifying the structures of unknown proteins and defining what cellular components interact with each

other. With massive “structural genomic” projects underway to identify currently unknown structures, the structural focus of proteomics is on the identification of post-translationally modified proteins and defining protein-protein interactions. These areas of research were featured in the “Teaching Proteomics” workshop at the annual Protein Society meeting in San Diego. The workshop, sponsored by the recently formed Education committee of the Protein Society, involved presentations on current research in proteomics by Andrew Link, Vanderbilt University and Bernhard Geierstanger from the San Diego-based Genomics Institute of Novartis Research Foundation (GNF).<sup>1</sup> These presentations brought both an academic and an industry perspective to the ensuing discussion of how students should be educated to prepare them for both the current challenges of proteomics and the needs of the future of this rapidly evolving area of the life sciences. Link, one of the founding faculty of the new Cold Spring Harbor workshop on Proteomics, a 2-week intensive training course in current proteomic methodology, first described the structure of that course with its emphasis on high throughput expression cloning, two-dimensional gel electrophoresis, protein purification techniques for isolating protein complexes, and the various mass spectrometry techniques used to characterize the components of such isolated protein complexes. These include MALDI-TOF, LC-MS-MS, and multidimensional LC-MS-MS.

The basic concept behind the isolation and characterization of multiprotein complexes is that if you know the structure and function of one or more of the components of such a complex you may be able to infer the function of other components in the complex. You may also be able to infer multiple, previously unrecognized, functions for proteins, particularly those that may be involved in regulatory activity since these proteins may well be subjected to post-translational covalent modification. Link’s own research focuses on establishing networks of protein interactions and, put simply, involves the isolation of interacting proteins using tandem affinity purification techniques, the identification of protein components in such complexes by trypsin digestion followed by peptide identifica-

<sup>1</sup> The abbreviations used are: GNF, Genomics Institute of Novartis Research Foundation; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; LC-MS-MS, liquid chromatography-tandem mass spectrometry; RLS, resonance light scattering.

¶ To whom correspondence should be addressed. E-mail: jbell2@richmond.edu.

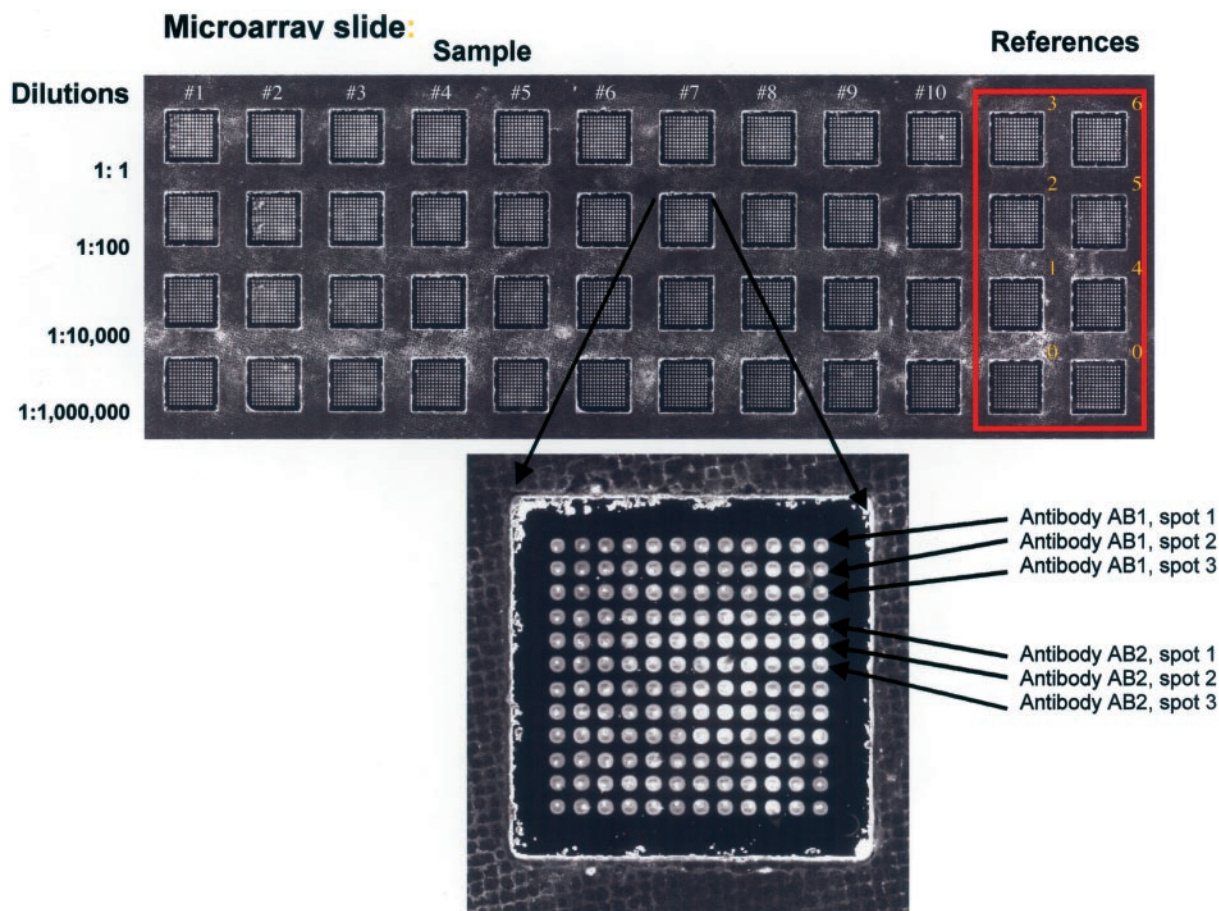


FIG. 1. **Layout of a microarray slide for Micro Sandwich assay.** The entire array is the size of a microscope slide containing 6912 spots for reference and assay points.

tion using mass spectrometry, the establishment of post-translational modification, and most importantly the validation, using reverse genetics, of the interaction networks uncovered.

Unlike Bernhard Geierstanger's work (described below), Link's work uses mainly experimental techniques that are often, or could be (with the exception of mass spectrometry), part of everyday undergraduate biochemistry courses. The problem is that they are usually presented at the undergraduate level as stand alone techniques, not as potential players in proteomic research. Most courses talk about affinity purification of proteins even if only in the context of histidine-tagged proteins. Think how easy it would be to introduce the idea of tandem affinity purification and the purposes of tandem affinity purification in a course for undergraduates. It would simply require introducing some of the big picture concepts and problems facing modern biochemistry and molecular biology and then setting a variety of experiments in context. Most courses talk about proteolytic digestion by trypsin, often in the old fashioned context of fragmenting a protein to sequence by Edman degradation (still in virtually every biochemistry textbook). How easy it would be to introduce the idea of using tryptic cleavage of a multiprotein complex, separation and sequencing peptides by LC-MS-MS, and then identifying the components of the complex by searching sequence databases with the obtained peptide sequences.

At the workshop, Dr. Bernhard Geierstanger, group leader of protein array technology at GNF, spoke about the work being carried out in his laboratory to design high throughput screening with protein arrays. At GNF, chemistry, genomics, proteomics, structural biology, computational biology, and engineering are all integrated in the development of new methods to identify and control biological processes that are relevant to human disease. As one portion of the GNF proteomic approach, they are developing parallel assays using protein arrays.

Protein arrays have been developed for the purpose of screening protein-protein interactions, protein-ligand interactions, protein abundance, or protein function in the proteome. For example, the Micro "Sandwich" assay at GNF is a flexible platform for screening protein abundance and disease markers in mouse serum using antibody arrays. The goal of this platform is to simultaneously assay 96 analytes in the serum of 200 mice per week. Because viability of the mice is essential to the study, the assay has an upper sample volume limit of serum of 20  $\mu$ l. The dynamic range of analyte concentration is from 1 pg/ml up to 1 mg/ml. The scaffold for the protein array is a microscope slide containing 48 wells on which 5  $\mu$ l of analyte-containing sample is placed. Each well contains 48 "capture" antibodies and 144 spots (see Fig. 1). Different capture antibodies are stereotactically placed on different spots in the wells robotically. A detection antibody that is conjugated to biotin binds to the analyte. Detection of the

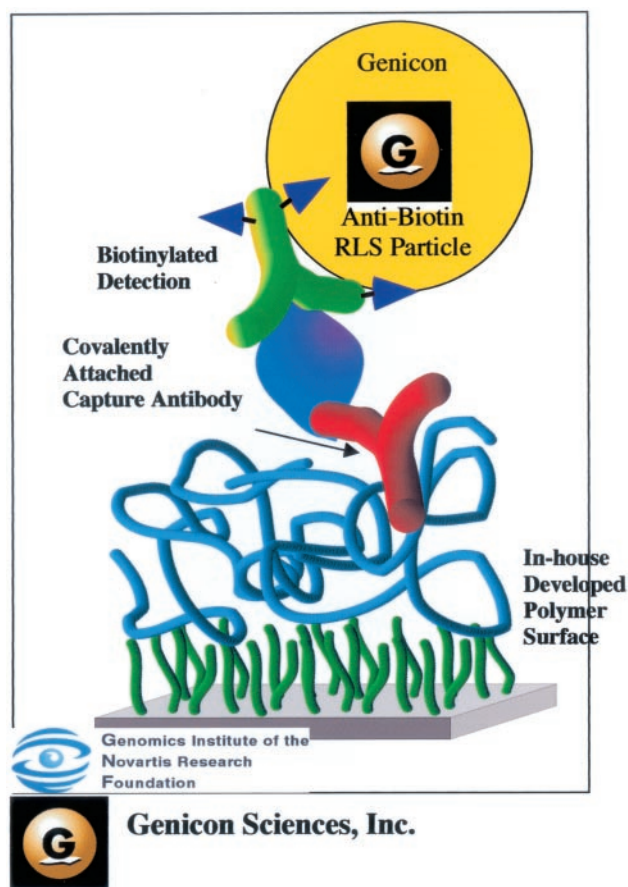


FIG. 2. The antibody array is built upon a polymer developed at GNF. The capture antibody is specific for an analyte. The analyte is then attached to another biotinylated antibody that is in turn conjugated to an Anti-Biotin RLS Particle for detection.

analyte is achieved with Anti-Biotin Resonance Light Scattering (RLS) Gold Particles from Genicon Sciences, Inc. (Fig. 2). With the RLS Gold Particles, the detection limits of the assay are as low as 1 pg/ml. The dynamic range issue is handled by applying several dilutions to different wells on a single microscope slide (Fig. 1). Specific reproducible conditions for sample deposition are identified for uniform application on the microarray plates.

With the involvement of robotics, large instrumentation, and automated sample placement on the arrays, one may wonder where biochemistry fits in or what skills will enable students to be competitive in the emerging field of proteomics. Scientists working in the protein array group at GNF have diverse backgrounds including biochemistry, molecular biology, biotechnology, surface science, biology, experimental pathology, and biophysics. Collaboration and teamwork among scientists of different backgrounds is essential. Dr. Geierstanger highlighted the need for skills in protein chemistry and analytical chemistry and in data handling. Group members need to be prepared for quantification of results and not be afraid of working with robotics. His work has relied heavily upon the Excel macroprogramming language, database management, and statistics. He also cites the importance of having some coursework in cell biology for knowing which questions to ask.

While the educational issues and challenges are more

complex with Geierstanger's experimental approaches, there are still many areas where the concepts of the approach can be introduced (array technology) and experimental issues discussed in an undergraduate setting. In both presentations there were remarkable agreement and overlap in areas of basic undergraduate education that need addressing. These focused on the needs of better understanding of analytical techniques and of quantitative approaches. Both speakers emphasized the need for students to hear about the "big picture" issues facing the molecular life sciences.

Let us examine the techniques that can contribute to such knowledge. Many of them are traditional protein chemistry techniques that for many years have been regarded as "old fashioned" such as two-dimensional gel electrophoresis and HPLC, often used in combination with newer techniques such as mass spectrometry. Adaptations such as gel shift assays to identify interacting molecules could easily be introduced into an undergraduate curriculum. Some techniques utilize chemical probes to give information about protein interactions such as chemical cross-linking. Others involve the development of new adaptations of genetic techniques such as two-hybrid systems. Still others have utilized entirely new developments in microscopy using atomic force microscopy and electron probe x-ray microanalysis. The techniques of current proteomics often emphasize high throughput approaches, miniaturization, and automation. While such approaches are often beyond undergraduate laboratories the issues can easily be introduced, perhaps in the context of plate readers, often the centerpiece of related technologies such as combinatorial chemistry approaches to determine protease specificity. How many of these techniques appear in the average undergraduate biochemistry curriculum and at what cost in terms of more traditional material? How many topics in current curricula could be spun toward proteomics to good effect?

The discussion so far has dealt with current proteomics, but proteomics is a rapidly evolving field. Where is it going? Paraphrasing the words of Mark Lesney, the future of proteomics is in the ultimate "omics," physiomics, a logical extension of proteomics. Physiomics will be the system wide informatic science capable of modeling and understanding the biology of complex temporal and environmental states involved in the life of individual organisms, not just cells [1].

What do we need to teach students to prepare them for even the logical next step of proteomics, understanding the roles of interactions between biomolecules within a given cell type, diseased or normal? Enzyme kinetics and ligand binding are good examples: How many courses give a brief overview of Michaelis-Menten kinetics and reversible inhibitors and set problems about calculating  $K_m$  and  $V_{max}$  or  $K_i$ ? Most, we suspect. How many go into any detail about kinetic mechanisms and allosteric regulation so that a student has a better chance of understanding the true impact of intracellular conditions on the activity of an enzyme? Not many, we would be prepared to wager. How many address issues of multienzyme complexes and the kinetic and regulatory implications? With substrate channeling, etc.? How many emphasize that *in vitro* enzyme

kinetics and the properties of enzymes determined by standard enzyme kinetics may have little relationship to *in vivo* situations? *In vivo* enzymes are often stoichiometric with the substrate and product concentrations and hence are participants in reactions, not simple catalysts, which may change apparent equilibrium constants for reactions. The conformation of the protein may further complicate this situation. These are the issues that the study of enzyme kinetics must prepare students to understand and investigate in the future. Likewise, how much coverage of the physical chemistry of ligand interactions is given in the average undergraduate course? How many students understand the quantitative aspects of protein-ligand interactions and the concepts of linked equilibria? These are necessary prerequisites for both current and future proteomics. As both Link and Geierstanger emphasized, in both these areas the level of understanding necessary to take on the challenges of proteomics is in the quantitative aspects, and detailed understanding is required rather than the memorization of a glossary of terms.

Below are listed several websites, reviews, articles, and books that may be valuable in the development of a course that includes aspects of proteomics.

#### *Websites of Interest*

Protein Society Education: [www.haverford.edu/biology/PSEducation/](http://www.haverford.edu/biology/PSEducation/)

Two-dimensional Gel Databases:

Swiss-Prot 2D Gel PAGE: [www.expasy.ch/ch2d/](http://www.expasy.ch/ch2d/)

Danish Center for Human Genome Research's 2-D PAGE Databases: [proteomics.cancer.dk](http://proteomics.cancer.dk)

Mass Spectrometric Protein Identification:

MASCOT: [www.matrixscience.com/](http://www.matrixscience.com/)  
 Protein Prospector: [prospector.ucsf.edu/](http://prospector.ucsf.edu/)  
 Chip Technology and Molecular Diagnostics:  
 K. K. Jain: [www.devicelink.com/ivdt](http://www.devicelink.com/ivdt)

#### *Reviews and Journals*

*A Trends Guide to Proteomics*, Vol. 19, No. 10 (suppl.), October 2001 (a supplement to *Trends in Biotechnology*: [www.trends.com](http://www.trends.com)), American Genomic/Proteomic Technology, ISC Publications, Shelton, CT.

*Computational Proteomics* (a supplement to *Biotechniques*, March 2002, [www.BioTechniques.com/supplements.html](http://www.BioTechniques.com/supplements.html)).

"Making the Most of Microarrays" by Christopher Roberts in *Drug Discovery and Development*, Vol. 5(4), pp. 20–26, 2002. The same issue contains an excellent article about high throughput screening of kinases including a discussion of mobility shift assays.

*Molecular and Cellular Proteomics*, a new journal published by the American Society for Biochemistry and Molecular Biology: keep up with latest research activities at the forefront of proteomics.

#### *Book Recommendations by Andrew Link*

Books Covering Fundamentals of Proteomics: *Introduction to Proteomics: Tools for the New Biology* by Daniel C. Liebler, Humana Press, Totowa, NJ, 1st edition, December 2001, ISBN: 0896039927.

#### REFERENCES

- [1] M. S. Lesney (2001) Pathways to the proteome: from 2DE to HPLC, *Mod. Drug Discov.* **October**, 33–39.