Perspective

Proteomics: An Exciting New Science, but Where are the Chemical Engineers?

Kelvin H. Lee
School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853
Kenneth F. Reardon
Dept. of Chemical Engineering, Colorado State University, Fort Collins, CO 80523

Introduction

At the time of writing this article, there is an ongoing debate about the future of AIChE as an organization. This discussion has also come at a time when many chemical engineers and academic chemical engineering departments are thinking about their role relative to modern life science research and education. One of the benefits of this community-wide evaluation of the future of the profession is the identification of fruitful areas for chemical engineering research and application. Although there may well be changes in our profession, we argue that chemical engineering science is already well poised to tackle important problems in a new area of life science research—proteomics.

What is Proteomics?

Proteomics refers to the system-wide analysis of proteins where the system could be the set of proteins expressed in a cell, tissue, microbial community, or other biological system of interest. The emergence of the field of proteomics is a direct outgrowth of the rapid evolution of life science research and the growth in the field, as measured by publications cited in the National Library of Medicine (PubMed, Figure 1), has paralleled the growth in genomics and functional genomics. The Human Genome Initiative, with the initial goal of determining the complete DNA sequence of humans and other species, has had a tremendous impact on life science research.

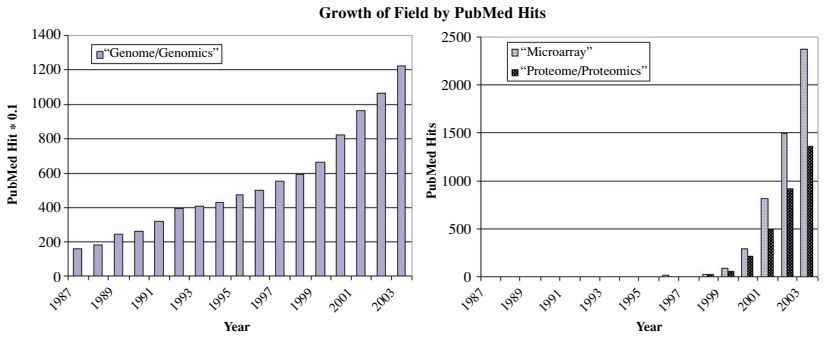

Figure 1. Growth in the number of publications available through the National Library of Medicine (PubMed or Medline).
Papers identified by "genome" and "genomics" since 1987 are shown (numbers divided by 10). Papers identified by "microarray" which mostly refers to DNA microarrays for functional genomic analysis are shown, as well as papers identified by "proteome" and "proteomics." There is a clearly increasing number of papers in these fields as the field continues to grow, with some of the growth fueled by advances in technology. Values for 2003 extrapolated from partial year data.

Key outcomes of this effort are the investment in the development of new technologies for analyzing biomolecules and an accompanying paradigm shift about how biological problems are studied. For example, the earliest phases of the Human Genome Initiative focused on the development of high-throughput technologies for the analysis of DNA sequence information. These efforts produced capillary electrophoresis based DNA sequencers that rely on sophisticated technology platforms to automate the acquisition of DNA sequence information and new software algorithms for the analysis of data from "shotgun" sequencing efforts. A second example of novel high-throughput technologies is the DNA microarray, which permits the simultaneous and quantitative measurement of the expression levels of tens of thousands of messenger RNA molecules present inside cells and tissues. The successful development of these technologies has helped to shift the manner in which biological systems are studied (discussed in Hatzimanikatis, 2000). The change from studying single molecules and genes to studies that include system-wide measurements demands access to even more technology, as well as the use of statistical and mathematical frameworks for the analysis of data. These needs are especially important in the field of proteomics.

Although proteomics involves the system-wide or genome-wide analysis of proteins, a definitive description of this new field has still not been articulated (in our opinion). Perhaps, paralleling the complexity of proteins as a class of molecules, the complexity of defining proteomics as a field suggests that this will remain an active area of discussion for some time. For simplicity, in this article we consider proteomics to be a system-wide, quantitative analysis of changes in protein expression among samples of

interest. The goal of any such study is to know which proteins may be upregulated or downregulated in response to a stimulus, in response to a disease state or to serve as a molecular marker. Such studies often rely on a process workflow that involves several key steps (Figure 2).

In the first step, samples of interest (perhaps proteins from a bioreactor or proteins from an engineered tissue) are collected and the protein content is separated using a gel- or liquid-based strategy. Generally speaking, cells express many thousands of proteins at a single time and a successful proteomics approach will employ high-resolution separation strategies to resolve this complex mixture of molecules into individual components. The two most commonly used methods are:

(1) Two-dimensional protein electrophoresis (2-DE), a gel-based technique in which proteins are separated based on isoelectric point and on size (see Figure 3 for an example).

(2) Multidimensional liquid chromatography (MDLC), in which the separation is based on multiple characteristics such as charge, hydrophobicity, size, and/or affinity.

Although neither of these techniques can resolve all of the proteins present in a typical cell lysate (there is tremendous diversity in physical and chemical properties of proteins), these two approaches provide the foundation for virtually all proteomics studies, although other strategies, such as protein array technologies, show promise.

While these separations strategies allow one to separate and purify a protein of interest, the great utility of the approach was realized only recently with the advent of protein and peptide mass spectrometry. The separations-only approach to proteomics described above provides an investigator with the hint that there are changes in protein expression among samples that may be of interest. However, these separations-only experiments do not normally provide any useful clues about the identity of the proteins or genes themselves, and, thus, do not lead readily to mechanistic understanding of the problem. The combination of separations strategies with soft ionization techniques such as matrix assisted laser desorption/ionization and electrospray ionization, coupled with mass spectrometry, permit the investigator to use the separations step as both an analytical tool, as well as a preparative tool. The coupling of separations with mass spectrometry has revolutionized the field of proteomics by permitting one to connect changes observed in protein expression patterns with underlying genes.

The successful integration of a proteomics workflow into a biochemical engineering laboratory thus relies not only on access to relevant technologies, but also on interfaces with both biochemistry and analytical chemistry. Although there are several examples of the successful application of this approach to important problems in biochemical engineering, there remain many important unmet needs in proteomics—needs that could be addressed by the chemical engineering community because of our understanding of chemical kinetics, transport, separations, and biological molecules and systems. Before we discuss these opportunities, it is pertinent to review the work that chemical engineers have already done in this field.

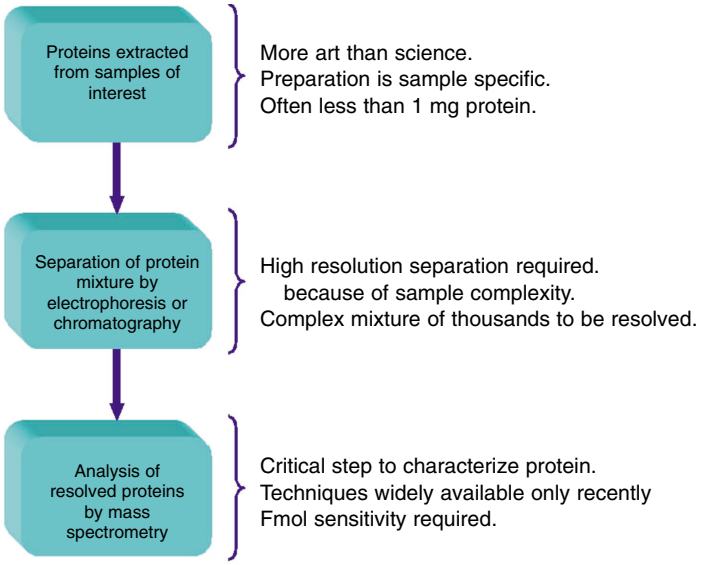

Figure 2. Process workflow for a typical proteomics experiment.
The advent of mass spectrometry for characterization of separated proteins has revolutionized proteomics. Nonetheless, technical challenges remain which need to be addressed.

What have Chemical Engineers done to Advance and Use the Tools of Proteomics?

Although the techniques of proteomics date back to the 1970s, chemical engineers did not become involved with the use or development of these methods for about 15 years and, even today, only a few work in this area. One of the first applications of proteomic tools by a chemical engineering group was published by Birnbaum and Bailey (1991), who used 2-DE to investigate phenomena associated with the production of recombinant proteins by bacteria. Although it was known that bacteria carrying many copies of the cloned gene on plasmids grew more slowly than did wild-type cells, the nature and scope of the impact of the plasmid burden was unclear. The proteomic approach of Birnbaum and Bailey allowed them to determine that plasmid presence in recombinant bacteria affected the levels of many host cell proteins, as well as the ribosome content. A wide range of proteins were impacted, illustrating the ability of proteomics to reveal effects that were completely unpredicted and would have been overlooked with a more specific approach.

Around the same time, Dykstra and Wang (1990, 1992) used 2-DE to understand product-induced stress response in an antibiotic-producing *Streptomyces* fermentation. Although they were unable to identify differentially expressed proteins, four proteins were found to correlate with successful process improvement strategies. These researchers also had the foresight to suggest using 2-DE for process monitoring and even proposed the use of microfluidics for this purpose. However, the next application of a proteomic method in the context of bioprocess monitoring did not come until a decade later, when Champion and coworkers used 2-DE and mass spectrometry for the purpose of validating an immunoassay protocol used in quality control (Champion et al., 2003).

Another type of application of proteomics has been for bioprocess analysis, in which the vast amount of information provided by examination of protein expression patterns is used to better

understand and ultimately to improve a bioprocess. For example, biochemical engineers often add or alter genes in bacteria and other cells to produce a large amount of a single protein product or, in the case of metabolic engineering, to produce new products. Frequently, these genetic changes have unexpected consequences that limit the productivity of the altered cells. Since the impacts on the cells are likely to be global in nature, the study of a single gene or protein is ineffective, but proteomics has been shown to provide the necessary information (Birnbaum and Bailey, 1991; Kosinski et al., 1992; Pferdeort et al., 2003). Other examples of proteomics as a tool in bioprocess analysis are for the analysis of protein inclusion bodies (precipitated proteins) from fermentations of recombinant bacteria (Hart et al. 1990), determination of the effects of an inhibitory metabolic byproduct (Lasko et al., 1997), the effects of high cell concentration in cultivations (Yoon et al., 2003), and the influence of temperature on product formation by recombinant mammalian cells (Kaufmann et al., 1999). In an environmental context, proteome analysis provided critical insights into the biodegradation kinetics of a pollutant mixture and led to formulation of an alternative model (Reardon and Kim, 2002).

As illustrated by the above examples, the primary use by chemical engineers of proteomics has been in the *application* of the tools to solve problems. While these efforts signal the importance of proteomic approaches to address problems related to biological systems, there remain many technological challenges in proteomics. These challenges relate to the sensitivity of detection (because there are no methods to amplify the copy number of low abundance proteins) and the inability to resolve all proteins in a single experiment (because of the diverse physicochemical properties). There are but a few examples of the application of chemical engineering science, in its broadest sense, to aid in the development of new proteomics technologies.

One such example of technology development relates to the use of polymeric materials for the electrophoretic separation of mixtures of proteins. Gel-based methods rely on polyacrylamide, which is not well suited for the analysis of hydrophobic proteins. The use of alternative polymeric materials to separate hydrophobic proteins was done in the 1990s (Harrington et al., 1994) and relied on the use of sintered porous materials for 2-D separations. Greenlee and Ivory (1998) developed a theoretical and experimental approach to protein focusing based on conductivity gradient methods in which a constant convective flow of buffer is opposed by an electric field gradient resulting in protein focusing. Beyond these efforts, most of the relevant work done by the chemical engineering community has involved theoretical modeling of electrophoresis such as the pioneering efforts by Dudley Saville (Bier et al., 1983) in the development of computational simulations and mathematical modeling of electrophoresis.

Opportunities: Proteomics Needs Chemical Engineers

The field of proteomics will continue to grow and evolve (recall the lack of an effective definition at this time for what proteomics includes). Although the intersection between chemical engineering and proteomics has not (yet) been considerable, it is apparent that chemical engineers could have a substantial positive impact on the development of this field in multiple ways. These opportunities include at least four levels of input. First, chemical engineers could make a significant contribution to the development of high-throughput technologies for protein separations and analysis. Second, chemical engineers can continue to apply existing proteomics methods to important problems encountered in the biotechnology industry. Third, academic chemical engineering programs can educate students to be skilled in technology and application and to communicate effectively with biochemists and analytical chemists. Fourth, chemical engineers have an appropriate foundation to build computational frameworks to integrate proteomic data with data from other types of experiments such as DNA sequencing efforts and DNA microarray studies (discussed in Hatzimanikatis, 2000).

Consider the basic proteomics strategies depicted in Figure 2. Such a process workflow relies on effective protein separations technologies. While the current methods of 2-DE and MDLC are providing important answers to questions in the life sciences, there is a need for improved resolution and enhanced capabilities. Such technology development must be based upon fundamental studies

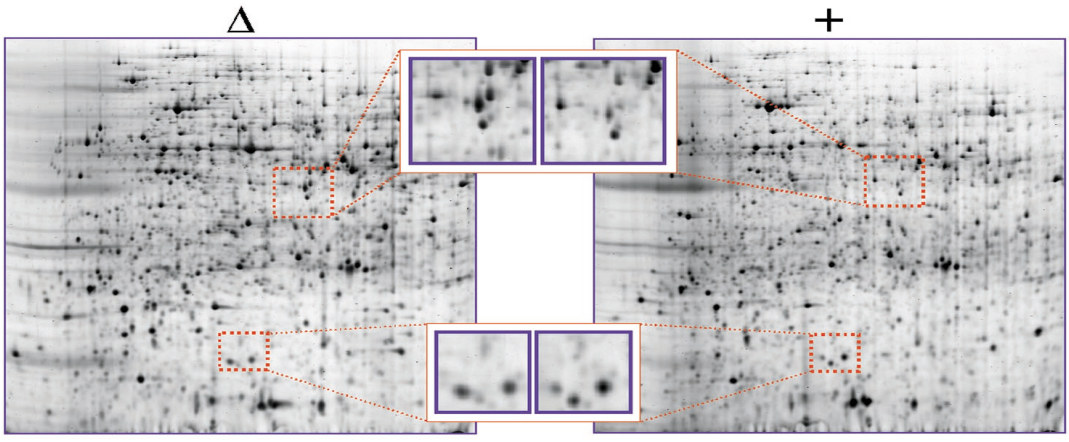

Figure 3. A typical 2-DE experiment involves the separation of proteins from two samples of interest. Proteins are separated in the horizontal direction by charge and in the vertical direction by size. Comparisons (typically done with image analysis software) reveal changes in protein expression specific to samples.

of protein electrophoresis (Figure 4). While significant efforts have been devoted to understanding electrophoresis of DNA in polymers, relatively little effort has been devoted to understanding electrophoresis of protein mixtures in polymers. If an appropriate framework that describes the transport of these molecules in polymers could be developed, then the field could be guided by rational changes to the existing separations schemes rather than the *ad hoc* developments which have characterized advances in proteomics to date. A second area in which chemical engineering science could play an important role in proteomics is in the modeling and understanding of the chemical and enzymatic digestion of purified proteins prior to mass spectrometry. In practice, femtomolar quantities of proteins isolated in polyacrylamide gel plugs are digested by specific proteases. The selection of reaction conditions is typically done without regard to any rigorous testing or analysis of optimal conditions, and these reactions sometimes are performed under arbitrarily-defined conditions. The optimization of the reaction conditions for such experiments could yield much improved sensitivity for the characterization of proteins.

A particularly promising area of proteomics technology development that has already engaged chemical engineers involves the use of microfluidic devices (Chow, 2002; Stone and Kim, 2001) and/or protein-based microarrays (Figure 4). The ability to create microfluidic devices for protein separations has been demonstrated by a number of chemical engineering laboratories (Throckmorton et al., 2002; Mohan and Lee, 2002) and will have a significant impact on workflows based on MDLC. Although these technologies have not yet achieved the resolution required to be an effective alternate to larger-scale separations, the promise of the ability to analyze small quantities of fluids with minimal carry over from sample to sample is exciting. The development of protein activity-based microarrays using printed chemical libraries (Gosalia and Diamond, 2003) offers an alternate route to monitor protein activities. Unlike the more traditional gel and liquid-based methods that measure protein expression, these arrays offer insight into changes in protein and enzymatic activities in a high-throughput manner.

Conclusions

Much of the technology development that has helped life science research evolve in the past decade has been developed by biologists, chemists, and physicists. These advances have come about because of a desire to answer specific biological questions for which the current technology had been inadequate. The skills available within the chemical engineering community are ideal to solve problems, apply the technologies, and generally make a large impact in the area of proteomics. While some of the technological developments that may arise from chemical engineers in the area of proteomics may be motivated by a desire to answer specific biological questions of interest to the bioprocess and biotechnology communities, other developments may arise from a natural evolution in research activities within the broader chemical engineering community. It would not be the first time that the core skills of thermodynamics, kinetics, and transport phenomena would be applied to technology development. Historically, these skills, which involve both theoretical and empirical studies, played a significant role in the development of technologies associated with typical unit operations in chemical processing plants. Although the length scales and molecules are different in proteomics, the challenges are analogous.

Given the opportunities and needs in proteomics and the current interest in extending the reach of chemical engineering into the life

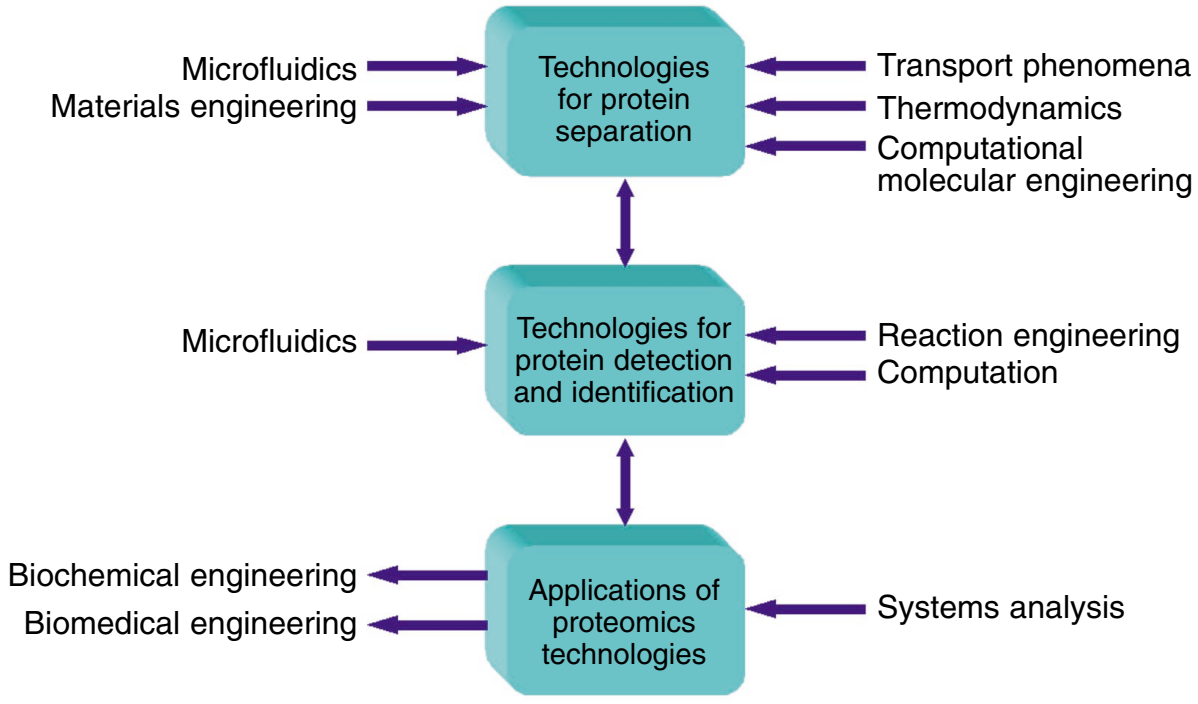

Figure 4. Relationships between chemical engineering fields and major areas of proteomics technology and application where contributions could be made.
Other disciplines, including analytical chemistry and biological sciences, also play major roles in proteomics.

sciences, it seems to be an ideal time for the chemical engineering community to apply our knowledge to the development of technologies for the analysis of complex mixtures of proteins. The *fundamental* expertise in transport, thermodynamics, computer simulations, and materials offers tremendous opportunities to impact protein expression studies. Further, the diverse set of *new* research areas studied by the chemical engineering community—microfluidics and electronic materials processing, among others—intersect needs of the proteomics community. At this time, the impact of proteomics on human health and in the biotechnology arena is still not fully realized, and the impact of chemical engineering on proteomics is even more nascent. One first step should be in education, to produce a pool of undergraduate and graduate students comfortable with working on problems at the interface of proteomics and chemical engineering science. One of the strengths of chemical engineering science has always been the ingenious application of its principles to solve problems in new areas, especially those outside the traditional boundaries. The field of proteomics provides a wonderful set of challenges and opportunities for exactly this pursuit.

Acknowledgment

KHL acknowledges support from NSF (BES0120315) and the USDA Agricultural Research Service (SCA #58-1907-1-146). KFR acknowledges support from NSF (BES-0004248) and National Institute of Environmental Health Sciences, NIH (5 P42 ES05949-05).

Literature Cited

Bier, M., O. A. Palusinski, R. A. Mosher, and D. A. Saville, "Electrophoresis: Mathematical Modeling and Computer Simulation," *Science*, **219**, 1281 (1983).

Birnbaum, S., and J. E. Bailey, "Plasmid Presence Changes the Relative Levels of Many Host Cell Proteins and Ribosome Components in Recombinant *Escherichia coli*," *Biotechnol. Bioeng.*, **37**, 736 (1991).

Champion, K. M., J. C. Nishibara, I. S. Aldor, G. T. Moreno, D. Andersen, K. L. Stults, and M. Vanderlaan, "Comparison of the *Escherichia coli* Proteomes for Recombinant Human Growth Hormone Producing and Nonproducing Fermentations," *Proteomics*, **3**(7), 1365 (2003).

Chow, A., "Lab-on-a-Chip: Opportunities for Chemical Engineering," *AIChE J.*, **48**, 1590 (2002).

Dykstra, K. H., and H. Y. Wang, "Feedback Regulation and the Intracellular Protein Profile of *Streptomyces griseus* in a Cycloheximide Fermentation," *Appl. Microbiol. Biotechnol.*, **34**(2), 191 (1990).

Dykstra, K. H., and H. Y. Wang, "Monitoring Intracellular Protein Profiles with Two-Dimensional Gel Electrophoresis," *Biotechnol. Prog.*, **8**(4), 369 (1992).

Gosalia, D. N., and S. L. Diamond, "Printing Chemical Libraries on Microarrays for Fluid Phase Nanoliter Reactions," *Proc. Natl. Acad. Sci. USA*, **100**, 8721 (2003).

Greenlee, R. D., and C. F. Ivory, "Protein Focusing in a Conductivity Gradient," *Biotechnol. Prog.*, **14**, 300 (1998).

Hatzimanikatis, V., "Bioinformatics and Functional Genomics: Challenges and Opportunities," *AIChE J.*, **46**(2), 2340 (2000).

Harrington, M. G., K. H. Lee, J. E. Bailey, and L. E. Hood, "Sponge-like Electrophoresis Media: Mechanically Strong Materials Compatible with Organic Solvents, Polymer Solutions and Two-Dimensional Electrophoresis," *Electrophoresis*, **15**(2), 187 (1994).

Hart, R. A., U. Rinas, and J. E. Bailey, "Protein Composition of *Vitreoscilla* Hemoglobin Inclusion Bodies Produced in *Escherichia coli*," *J. Biol. Chem.*, **265**(21), 12728 (1990).

Kaufmann, H., X. Mazur, M. Fussenegger, and J. E. Bailey, "Influence of Low Temperature on Productivity, Proteome and Protein Phosphorylation of CHO Cells," *Biotechnol. Bioeng.*, **63**, 573 (1999).

Kosinski, M. J., U. Rinas, and J. E. Bailey, "Proteolytic Response to the Expression of an Abnormal beta-galactosidase in *Escherichia coli*," *Appl. Microbiol. Biotechnol.*, **37**(3), 335 (1992).

Lasko, D. R., C. Schwerdel, J. E. Bailey, and U. Sauer, "Acetate-Specific Stress Response in Acetate-Resistant Bacteria: An Analysis of Protein Patterns," *Biotechnol. Prog.*, **13**(5), 519 (1997).

Mohan, D., and C. S. Lee, "On-line Coupling of Capillary Isoelectric Focusing with Transient Isotachophoresis-Zone Electrophoresis: A Two-Dimensional Separation System for Proteomics," *Electrophoresis*, **23**, 3160 (2002).

Pferdeort, V. A., T. K. Wood, and K. F. Reardon, "Proteomic Changes in *Escherichia coli* TG1 after Metabolic Engineering for Enhanced Trichloroethene Biodegradation," *Proteomics*, **3**(6), 1066 (2003).

Reardon, K. F., and K. H. Kim, "Two-Dimensional Electrophoresis Analysis of Protein Production during Growth of *Pseudomonas putida* F1 on Toluene, Phenol, and their Mixture," *Electrophoresis*, **23**(14), 2233 (2002).

Stone, H. A., and S. Kim, "Microfluidics: Basic Issues, Applications, and Challenges," *AIChE J.*, **47**, 1250 (2001).

Throckmorton, D. J., T. J. Shepodd, and A. K. Singh, "Electrochromatography in Microchips: Reversed-Phase Separation of Peptides and Amino Acids Using Photopatterned Rigid Polymer Monoliths," *Anal. Chem.*, **74**, 784 (2002).

Yoon, S. H., M. J. Han, S. Y. Lee, K. J. Jeong, and J. S. Yoo, "Combined Transcriptome and Proteome Analysis of *Escherichia coli* during High Cell Density Culture," *Biotechnol. Bioeng.*, **81**(7), 753 (2003).