

# Proteomics – advances, applications and the challenges that remain

Ruedi Aebersold and Benjamin F. Cravatt

With complete genome sequences now available for several prokaryotic and eukaryotic organisms, biological researchers are faced with the unprecedented scientific challenges of assigning molecular and cellular functions to thousands of newly predicted gene products and explaining how these products cooperate in complex physiological processes. To address this problem, the field of proteomics has emerged with the goals of developing and applying methods for the global analysis of protein expression and function. It is anticipated that the creation of effective methodologies for the rapid and parallel analysis of proteins will accelerate the ‘functionalization’ of these biomolecules and, by doing so, discover, among other things, new biomarkers and therapeutic targets for the diagnosis and treatment of diseases of humans and generally increase our mechanistic understanding of biological processes. Nonetheless, for proteomic researchers to achieve these lofty goals, they must first confront and eventually overcome several technical challenges that currently limit efforts to characterize systematically proteins from highly complex samples.

Indeed, although it might be conceptually attractive to view proteomics as simply the ‘protein equivalent’ of genomics, such a view falls far short of depicting the myriad of methodological problems that are uniquely associated with the scientific investigation of proteins. For example, in contrast to RNA or DNA, proteins cannot be amplified by methods analogous to PCR, that is, the amount of protein present in a given sample is the amount of protein that must be analyzed. Thus, proteomic researchers are confronted with the difficult task of detecting, identifying and characterizing numerous low abundance proteins at their natural cellular concentrations, even when the levels of these biomolecules are 6–8 orders of magnitude lower than those of high-abundance proteins present in the same sample (as is often the case). Similarly, unlike RNA or DNA, proteins do not inherently possess well defined high-affinity and/or high-selectivity binding partners. Thus, whereas the field of gene microarrays has been able to capitalize on the special interactions that oligonucleotides share with their antisense partners, protein microarray researchers must devise a specific capture reagent for each protein of interest, a

process that promises to be both expensive and laborious, but nonetheless of crucial importance. Finally, proteins exhibit a range of biochemical properties that far exceeds the relatively homogenous behavior of oligonucleotides and is crucially dependent on the precise 3D structure of the folded polypeptides. The diversity of traits exhibited by proteins, ranging from extreme pI values to membrane association to post-translational modification, means that methods to handle and process these molecules are rarely as generally applicable as we would like.

Considering all of the challenges that face the field of proteomics, it might be worth asking – isn’t genomics good enough? Clearly, however, the answer is becoming an increasingly emphatic ‘no’. Accumulating examples in the literature of the limited correlation that exists between steady-state mRNA and protein abundance, as well as the changes in mRNA and protein abundance induced by perturbations, serve as a constant reminder of the complex manner in which cells regulate protein expression in a post-transcriptional manner. Additionally, proteins are modified by an ever-increasing number of post-translational events, many of which are dynamic in nature, highlighting the need to not only detect, but also characterize proteins from native proteomes. Finally, it is important to recognize that the physiological and pathological events that form the basis for health and disease are, at their core, protein-driven processes. To understand these events in molecular detail, the proteome, in all of its spatially and temporally regulated forms, must be analyzed.

In this *TRENDS Guide to Proteomics*, we have solicited contributions from leading authors in a diverse number of fields that collectively form the foundation for contemporary proteome research. We believe that these articles should interest both the general reader and dedicated proteomic researcher alike, because they highlight not only the significant advances made to date in proteome analysis, but also the numerous technical challenges that still persist.

In the first paper, Smith describes the considerable advances that continue to be made in mass spectrometry, a technology of central importance to proteomics, especially because it relates to increasing the throughput and coverage of complex proteome analysis. Link provides a broad overview of current strategies being employed to

## Ruedi Aebersold

Institute for Systems Biology, 1441 North 34th Street, Seattle, WA 98103-8904, USA.  
e-mail: raebersold@systemsbiology.org

## Benjamin F. Cravatt

The Skaggs Institute for Chemical Biology and Departments of Cell Biology and Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla CA 92037, USA.  
e-mail: cravatt@scripps.edu

fractionate the proteome by multidimensional separation methods so that its low abundance constituents can be visualized. Focusing on the emerging field of protein chips, Lee and Mrksich offer a thoughtful review of the potential that these arrays have to accelerate greatly the functional characterization of proteins, whilst simultaneously reminding us of the significant technical problems that must be overcome to realize this goal. In a complementary article, Elia and colleagues make a strong case that the future value of protein chips will be inextricably tied to the quality of the capture reagents contained on these arrays, as well as the methods used to detect bound proteins. Flory and colleagues review the current state of quantitative proteome analysis by stable isotope tagging, a method that has rapidly moved from proof-of-concept to general applicability for the comparative analysis of complex protein samples and therefore for the detection of dynamic changes in the proteome.

Shifting the focus from basic research to clinical application, Petricoin and Liotta detail the power of combining mass spectrometry with artificial intelligence-based informatics to generate serum proteomic patterns that can be used to diagnose disease. Also highlighting the importance of bioinformatics, Fenyő and Beavis discuss the fundamental challenges that face the systematic processing, analysis and presentation of proteomic data, emphasizing the need to evaluate this information with statistics so that

estimates of significance can be achieved. Not only must proteomic data be properly analyzed and presented, but it must also be stored, made accessible to the general public and integrated in a dynamic way with other molecular, cellular and organismal information. These topics are the subject of an intriguing review by Hancock and colleagues. Finally, Ryan and Patterson discuss both the promises and problems associated with the industrialization of proteomics for the discovery of new disease markers and therapeutic targets.

Given the limited space available in this *TRENDS Guide to Proteomics*, we recognize that several worthy topics in proteome research are necessarily absent. Nonetheless, we are confident that all of the subjects covered here represent matters of crucial importance to the youthful, but rapidly expanding field of proteomics. We hope that the reader will agree that our unabashed enthusiasm for the future of proteomics has been tempered throughout by a critical tone highlighting the many challenges that still face this field. Regardless, it should be clear to all who read these reviews that proteomics researchers are deeply committed to developing and applying rigorous methods for complex protein analysis and that very rapid progress is being achieved. Proteomics, therefore, is making an increasingly powerful impact on our understanding of the physiological and pathological processes that underlie health and disease.



Ruedi Aebersold is a co-founder and full faculty member of the Institute for Systems Biology (Seattle, WA, USA). He leads a research effort that is focused on the development of new quantitative proteomics technologies and their application to increase the understanding of the structure, function and control of complex biological systems.

Aebersold's research and teaching have been recognized by a long-term fellowship from the European Molecular Biology Organization, a scholarship from the Swiss National Science Foundation, the Killam Research Prize and the Pehr Edman Award. In 2002, he was the recipient of the Widmer Award, the Biemann Medal and the World Technology Network Award for Biotechnology.

Aebersold is senior editor of *Physiological Genomics*, an associate editor of *Molecular Cellular Proteomics* and on the Editorial Advisory Boards of *Proteomics* and the *Journal of Proteome Research*.



Benjamin Cravatt is an Associate Professor at The Scripps Research Institute (La Jolla, CA, USA) in the Departments of Cell Biology and Chemistry. He is also a member of The Skaggs Institute for Chemical Biology (La Jolla, CA, USA). Cravatt's research group is interested in understanding the roles that enzymes play in complex physiological and pathological settings and uses a combination of chemical, biochemical and genetic methods to investigate this subject. Cravatt received his BSc in Biological Sciences and BA in History from Stanford University (Stanford, CA, USA) in 1992. He received his PhD from The Scripps Research Institute in 1996.