

Proteomics from the clinical perspective: many hopes and much debate

Laura Beretta

Mass spectrometry has been rapidly maturing as the core technology at the heart of proteomics. The application of these powerful methods to the study of human diseases and their translation to the clinic, however, has been beset with unique challenges.

The escalation of interest in the application of proteomics to the study of human diseases and translation of this technology to the clinic is reflected by the growth of a new field defined as 'clinical proteomics'. Clinical proteomics aims at providing clinicians with tools to accurately diagnose and treat patients in an individualized manner. For optimal health management, the clinician needs the means to identify an individual at risk of developing a specific disease, to diagnose this disease accurately at an early stage and to monitor disease progression. Subsequently the clinician needs the means to select a specific therapeutic strategy best suited to an individual patient and a disease phenotype, and to assess treatment efficacy and susceptibility to adverse events.

Clinical proteomics has focused on the discovery of diagnostic and prognostic disease biomarkers as well as of novel drug targets, but the most intense interest has been in applying proteomics to develop new serological biomarkers for early diagnosis of chronic diseases such as cancer and cardiovascular diseases^{1–7}. Biomarkers can have a tremendous impact in clinical oncology, in identifying individuals at risk for developing cancer, in preclinical detection of cancer and in real-time monitoring of drug response.

The explosion of proteomics in general, and of clinical proteomics in particular, is often associated with the completion of the human genome project. Although the

availability of genome-sequence databases undoubtedly contributed to this explosion, we should not minimize the major impact of mass spectrometry technology. Over the past several years, mass spectrometry has improved greatly in sensitivity and accuracy, and mass spectrometry-based proteomics has become the method of choice for the analysis of complex protein samples. Within a decade, the field evolved from having great difficulty in identifying a single protein to being able to assign peaks with high confidence, to a large number (sometimes hundreds) of proteins within a single mass spectrometry run. New horizons have been opened; new hopes for medicine have been raised.

The inherent analytical advantages of mass spectrometry, including sensitivity, resolution, speed and throughput, combined with advanced bioinformatics for data interpretation, allow for the rapid and systematic analysis of thousands of proteins. Beyond identification, mass spectrometry also facilitates quantitation—an important requirement in clinical proteomics for comparing two or more samples for discriminating features. The application of mass spectrometry to the analysis of tissue and plasma or serum specimens from patients at different stages of disease or with different diseases has the potential to provide unique information about disease-associated alterations at the protein level.

Although mass spectrometry-based proteomics has become an extremely powerful laboratory tool, pre-analytical, analytical and post-analytical challenges associated with the methodology (see reviews^{8–11})

have hindered the application of proteomics in the clinic. It is widely perceived that the methodology is not yet ready to meet the constraints necessary for routine use in the clinical setting. The process for mass spectrometry-based biomarker or drug target discovery follows a long and complicated pipeline from the collection of clinical samples and associated clinical data, processing and handling of specimens, sample selection, sample fractionation and preparation, mass spectrometry analysis, data interpretation and statistical analysis. The challenge is that, for each of these steps, many factors may affect reproducibility, and biases can be introduced affecting the final results (for example, collection, processing and storage of samples, classification of patients, patients' medication(s), analytical incompleteness, limitations of bioinformatics tools and databases for protein identification). These challenges are increased by the inherent instability of proteins and the difficulty of detecting proteins that are present in small amounts in complex proteomes.

One way to circumvent the limitations of the dynamic range of the mass spectrometry instrumentation itself is to simplify complex mixtures of proteins into less complex components. This is particularly true for serum or plasma. Low-abundance proteins can be successfully detected by enhancing the pre-analytical step using multidimensional fractionation strategies or by enrichment of a subproteome. It is imperative, however, to balance benefits against additional sources of variability that might be introduced (for excellent reviews, see refs. 12–15). A collective effort to define protocols for sample

Laura Beretta is at the Molecular Diagnostics Program, Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. e-mail: lberetta@fhcrc.org

collection, to optimize the pre-analytical techniques and to further develop and improve bioinformatics tools and databases is underway^{16–18}. With rigorous and extensive multiplication of runs, mass spectrometry is now the most robust component in the pipeline.

To move the field forward, fundamental biological questions remain to be addressed, and experimental designs need to be improved. Most study designs so far have been based on intuitive concepts and assumptions. For example, one assumption is that a subset of proteins present in the blood reflects, reproducibly and specifically, a single disease at a particular stage. We also assume that we can identify disease-specific signals over noise in the blood. There is now no information on the biological variations of the blood proteome. Unless we have knowledge of the composition and the dynamics of the blood proteome in healthy individuals, critical elements of how to design biomarker studies will evade us.

Another big debate in the proteomics community is which type of sample—blood or diseased tissue—should be used to discover serologic biomarkers. Proteomic studies using blood as a source of proteins still have not resolved the question of whether plasma or serum can serve as a window into the state of a patient's disease, and at present there is no scientific evidence that a linear relationship exists between protein changes in diseased tissue and in blood. The basic assumption that proteins derived from a variety of tissues are indeed detectable in plasma using presently available proteomics technologies has been challenged only recently¹⁹. Additionally, the 'one-size-fits-all' rule seems to have been applied to studies targeting different organs and different diseases.

A particularly difficult question is how to design a study aimed at discovering biomarkers for early detection of small, localized, pre-invasive cancer lesions that cannot be diagnosed with presently available tools. Most studies so far have used samples (tissue or plasma or serum) from patients with advanced disease, with the hope that some of the identified markers will have utility in early diagnosis. This is based on the oversimplified assumption that the molecular

composition of very early tumors and more advanced tumors are very similar. Recent genomics studies indicate, however, that this is not the case.

Finally, one of the biggest hurdles in biomarker discovery studies is the lack of means to gauge the significance of the generated results and to measure success. Without independent, large validation studies in clinical trials, it is very difficult to determine whether the expression of the markers that are found to discriminate between groups will have clinical utility. Similarly, it is very difficult to recognize those proteins identified in the discovery phase that may turn out to be the best diagnostic or therapeutic biomarkers. It is vital to precisely define a clinical problem and to focus the experimental design around appropriate study populations and samples. Translation from the laboratory to the clinic is a long process, and it will take time to move a new target or biomarker from discovery to regulatory approval. Unreasonable expectations for the possible outcomes of proteomics-based clinical studies are detrimental to the field.

We are now in a position to critically assess the lessons learned to define future strategies to develop diagnostic biomarkers, predictors of disease progression or new therapeutic targets for individualized patient management. Efforts to integrate proteomics and other technologies might greatly accelerate the translation of basic discoveries into daily clinical practice. The use of relevant animal models such as mouse models in clinical proteomics has the potential to accelerate discovery, as well as the initial validation processes, in part by reducing heterogeneity and sample preparation variability. Diagnostic imaging techniques can be combined with proteomics methods to improve sensitivity for detecting disease and have the unique potential to provide whole-body assessment. Clinical proteomics studies have focused mainly on early detection of cancer, one of the most challenging of clinical problems. Other clinical problems have only rarely been tackled using proteomics, for example, viral pathogenesis. Knowledge of the protein composition of the infectious viral particle would provide important insights into the mechanisms of viral release, the receptor(s) for cell entry

and the strategies used to evade the host's immune system. This information could be translated into novel antiviral drugs or vaccines.

As the initial enthusiasm for clinical proteomics has faded, it is important to recognize that mass spectrometry has matured into a powerful tool for defining proteomes. With its ability to meet the two most important needs in clinical proteomics (qualitative and quantitative proteome-wide coverage), mass spectrometry is ideally suited for disease-driven studies. Novel mass spectrometry methodologies are emerging and show potential for several applications in clinical proteomics. But disease-driven studies, by their nature, are extremely difficult, and will require a complex multidisciplinary effort with experts from clinical, basic as well as statistical or computational research teams. Undoubtedly, mass spectrometry will remain at the core of the expanding effort to find a solid basis for individualized patient management.

1. Petricoin, E., Wulfkuhle, J., Espina, V. & Liotta, L.A. *J. Proteome Res.* **3**, 209–217 (2004).
2. Conrads, T.P., Hood, B.L., Petricoin, E.F., III, Liotta, L.A. & Veenstra, T.D. *Expert Rev. Proteomics* **2**, 693–703 (2005).
3. Hanash, S. *Nature* **422**, 226–232 (2003).
4. Massion, P.P. & Caprioli, R.M. *J. Thorac. Oncol.* **1**, 1027–1039 (2006).
5. Marko-Varga, G. *et al. J. Proteome Res.* **6**, 2925–2935 (2007).
6. Mayr, M. *et al. Mol. Cell. Proteomics* **5**, 1853–1864 (2006).
7. Fu, Q. & Van Eyk, J.E. *Expert Rev. Proteomics* **3**, 237–249 (2006).
8. Diamandis, E.P. *J. Natl. Cancer Inst.* **96**, 353–356 (2004).
9. Karsan, A. *et al. Clin. Chem.* **51**, 1525–1528 (2005).
10. Solassol, J. *et al. Expert Rev. Proteomics* **3**, 311–320 (2006).
11. van der Merwe, D.E., Oikonomopoulou, K., Marshall, J. & Diamandis, E.P. *Adv. Cancer Res.* **96**, 23–50 (2007).
12. deVera, I.E., Katz, J.E. & Agus, D.B. *Clin. Adv. Hematol. Oncol.* **4**, 541–549 (2006).
13. Veenstra, T.D. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **847**, 3–11 (2007).
14. Mischak, H. *et al. Proteomics Clin. Appl.* **1**, 148–156 (2007).
15. Hernandez, P., Muller, M. & Appel, R.D. *Mass Spectrom. Rev.* **25**, 235–254 (2006).
16. Barker, P.E. *et al. Clin. Chem.* **52**, 1669–1674 (2006).
17. Prakash, A. *et al. Mol. Cell. Proteomics*, published online 27 July 2007 (doi:10.1074/mcp.M600470-MCP200).
18. Taylor, C.F. *et al. OMICS* **10**, 145–151 (2006).
19. Zhang, H. *et al. Mol. Cell. Proteomics* **6**, 64–71 (2007).