Clinical Proteomics

ISSN 1542-6416/06/02:01–04/$30.00 (Online)

Editorial

Clinical Proteomics

My View of the Current Issues

Daniel W. Chan, PhD, DABCC, FACB

Editor-in-chief

Clinical Proteomics

Personalized medicine is expected to revolutionize health care delivery in the next decade. I believe that targeted proteomic diagnostics and therapeutics will be the basis of personalized molecular medicine.

Clinical proteomics is the application of proteomic techniques to the field of medicine. The application of clinical proteomic research is growing rapidly in the field of biomarker discovery, especially in the area of cancer diagnostics. Because most diseases like cancer are complex, it is unlikely that a single biomarker could be used successfully for the early detection of cancer. It is likely that a multiple marker panel would be required for cancer diagnostics. Recent advances in proteomics and bioinformatics technologies, such as mass spectrometry (MS) and affinity-based methods, such as protein microarrays, have been used successfully to detect disease-associated biomarkers in complex biological specimens, for example, cell lysates, serum, plasma, and other body fluids. Extensive validation will be needed to transform these biomarkers into targeted clinical use.

Clinical proteomics holds the potential of "taking a snapshot" of the total protein complement of a cell, or body fluid, and identifying proteins as potential biomarkers for the differentiation of disease and health. For successful biomarker discovery, validation, and translation of these biomarkers into clinical practices, many issues need to be addressed including study design, bioinformatics, specimen acquisition, handling, quality control, standard operating procedures, and good laboratory practices ***(1)***.

The majority of the biomarker discovery is focused on the proteomic exploration of body fluids, as these specimens are information rich and more accessible. Unlike other tissues, blood (plasma or serum) is the most complex human-derived proteome. In addition to containing proteins specific to blood, it also contains proteins released, either through leakage, injury or other factors, from other tissues in the

body. The extreme dynamic range in plasma protein concentrations of various proteins poses a challenge to proteomic analysis (2).

A good understanding of how preanalytical variables affect test results is important in biomarker research. Specimen handling, whether sample collection, pipetting, or diluting, all contribute to preanalytical error. Examples are the different blood collection tubes (3), coagulation times, storage conditions, or sample type (whole blood, plasma, or serum). The Human Proteome Organization (HUPO) established a Specimen Committee in 2002 to study and address many of these issues. A report from this committee was published recently (4). Another source of variation is biological variation (i.e., between-subject variation and within-subject variation). For any individual, many analytes fluctuate based on the time of day, fasting state, or age. Although these fluctuations may not be clinically relevant, they do add an additional level of complexity in elucidating disease-induced protein changes from changes caused by biorhythmic fluctuations.

Serum/plasma contains roughly 60–80 mg of protein per milliliter. It is estimated that there are more than 10,000 proteins commonly found in serum—most of them are present in very low concentrations. Roughly 22 high-abundant proteins—albumin, immunoglobulin, haptoglobin, and transferrin, to name a few—comprise approx 97% of the protein content of serum. The remaining 3% of proteins are present in low concentrations and are referred to as low-abundant proteins. Most attempts to separate/remove high-abundant proteins from serum have focused on albumin, the most abundant serum protein, which comprises 45–55% of all serum proteins, and immunoglobulin, which comprises approx 15% of all serum proteins. Eliminating or reducing these two high-abundant proteins will remove approx 60% of the total serum proteins and make it easier to analyze lower-abundant proteins. Recently, devices such as protein depletion columns have been developed with antibodies against the most abundant 6, 12, or 22 proteins. One should be aware of the potential of removing other proteins of interest along with these high-abundant proteins. Furthermore, one should assess not only the efficiency of the depletion method, but also the reproducibility of this procedure. In addition to selecting an effective fractionation method that will remove high-abundant proteins, an equally important issue is selecting the methodology for biomarker discovery. One common approach to the application of MS to biomarker discovery is the use of protein profiling from matrix-assisted laser desorption/ionization (MALDI)-time-of-flight (TOF) or surface-enhanced laser desorption/ionization (SELDI)-TOF MS as a "protein fingerprint" to identify a diseased state, followed by the identification of specific proteins as potential diagnostic biomarkers. The other approach is to digest proteins into peptides for liquid chromatography MS or MALDI-TOF MS for analysis. Once peptides of interest are identified, they could be traced back to the proteins. With either one of these two approaches, sophisticated bioinformatics tools are needed.

The enormous amount of data produced during proteomic analysis, coupled with the inherent variation in the instrumentation used, require the utilization of sound experimental design, proper calibration of instruments, and appropriate bioinformatics methods in order to generate good quality data from which valid conclusions can be drawn (5). In biomarker research, samples are generally collected from multiple sites and randomly divided into a discovery (training) data set and a validation (testing) data set. Differences in collection practices, sample handling, or storage conditions will differ between institutions and, as such, may influence the proteins present in a given sample. For results to be meaningful, a sufficiently large number of

collection sites must be employed and the sample population must be diverse so as to best represent the target population of interest. Because both data sets are derived from the same-pooled samples, it will naturally follow that the discovery set will represent the validation set. This might result in artificially high performance (i.e., sensitivity and specificity). An alternative approach that is employed in our laboratories is to use each sites data sets separately ***(6)***. For example, potential biomarkers discovered from one site would then be cross-compared and validated using other sites data sets. This type of model mimics the multicenter validation process that all clinically useful biomarkers must conform with prior to clinical use. An example of the success of this approach can be found in the work of Zhang et al. who created two different discovery and validation cohorts using specimens from four different test sites ***(6)***. Candidate biomarkers were cross-validated between discovery and validation cohort sets, as well as against a currently used cancer biomarker CA125. Using such an approach, the authors successfully identified three candidate proteins that have the potential to distinguish ovarian cancer from noncancer controls.

Finally, overwhelmed by the large number of reports and at the same time the sharp criticisms, "does it work?" has become a frequently asked question. Specific experiments and studies should be planned to answer this question. Such studies will undoubtedly provide invaluable information. For any emerging technology, whether it works or not at the beginning is not as important as understanding whether the physical principles behind the new technology are sound. If the answer is yes, the imperfections will eventually be resolved. Right now we should try to answer the more relevant question of "will it work?" If the evidence supports it, we then should let scientific discovery and technological innovation have the opportunity to proceed ***(7)***.

Clinical proteomics is the most exciting clinical laboratory science of the 21st century. In the past year, we have witnessed significant commitments from private and public sectors in the research and development of clinical proteomics. Recently, the National Cancer Institute announced a $104 million program for clinical proteomics technology assessment for cancer. In the private sector, the pharmaceutical industry has been the leader in using proteomic approach for biomarkers and drug discovery. Diagnostic industries have begun to explore opportunities in clinical proteomics. In the end, clinical laboratories and, ultimately, patients will benefit from these advances in proteomics.

Clinical Proteomics is in the unique position to provide a scholarly forum for novel scientific research in the field of translational proteomics, with special emphasis on the application of proteomic technology to all aspects of clinical investigations including academic, clinical laboratory, pharmaceutical, and diagnostic industries. We have a strong editorial team of leading scientists in the field of clinical proteomics with four associate editors, Robert J. Cotter, PhD, Eleftherios P. Diamandis, MD, PhD, Samir Hanash, MD, PhD, and Lance Liotta, MD, PhD, and the entire editorial board members. We expect that *Clinical Proteomics* will be the authoritative forum for this field.

References

1 Colantonio, D. A. and Chan, D. W. (2005) The clinical application of proteomics. *Clin. Chim. Acta* **35,** 151–158.

2 Anderson, N. L. and Anderson, N. G. (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell Proteomics* **1,** 845–867.

3 Drake, S. K., Bowen, R. A., Remaley, A. T., and Hortin, G. L. (2004) Potential interferences from blood collection tubes in mass spectrometric analyses of serum polypeptides. *Clin. Chem.* **50,** 2398–2401.

4 Rai, A. J., Gelfand, C. A., Haywood, B. C., et al. (2005) Human Proteome Organization-plasma proteome project specimen collection and handling: Towards the standardization of parameters for plasma proteome samples. *Proteomics* **5,** 3262–3277.

5 White, N. C., Chan, D. W., and Zhang, Z. (2004) Bioinformatics strategies for proteomic profiling. *Clin. Biochem*. **37,** 636–641.

6 Zhang, Z., Bast, R. C., Jr., Yu, Y., et al. (2004) Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res.* **64,** 5882–5890.

7 Zhang, Z. and Chan, D. W. (2005) Cancer proteomics: in pursuit of "true" biomarker discovery. *Cancer Epidemiol. Biomarker Prev.* **14,** 2283–2286.