Population Proteomics – Enabled by Mass Spectrometry, Driven by Clinical Outcomes

a report by

Dobrin Nedelkov

Director of Research and Technology Development, Intrinsic Bioprobes, Inc., and Founding President, Institute for Population Proteomics

Population proteomics is the study of protein diversity in human inhabitations. It can be defined as the investigation of human proteins across and within populations to delineate and understand protein differences, and to facilitate the discovery and validation of disease-specific protein modulations.[1–3] In broader terms, population proteomics can be compared with population genomics, where individuals catalogue common genetic variants and determine how they are distributed among people around the world. Although population proteomics cannot (yet) claim such outreach and goals, it has the potential to become an important proteomics subdiscipline. As the tools and approaches that enable population proteomics are increasingly being applied, it stands to provide answers about the extent of protein modifications in humans and their association with disease.

Several attributes define population proteomics. The first is the use of targeted proteomics approaches. Population proteomics does not engage in the study of entire proteomes because it is likely that for a specific cell or tissue proteome there is no definitive set and number of proteins common to all within a group or a larger population. Instead, population proteomics focuses on the interrogation of a selected number of proteins from a large number of individuals to delineate the distribution of specific protein modifications within these subpopulations.

Another important attribute of population proteomics is the use of mass spectrometry (MS)-based approaches. MS is the only detection method that can provide information about specific protein structural modifications without *a priori* knowledge of the modification. MS interrogates the protein mass, which is an intrinsic property of each fully expressed and functional protein. The mass contains information about the gene that encodes the protein and the post-expression processing that the protein undergoes. Any changes in the gene sequence and/or post-expression protein processing will be reflected in the mass of the whole protein.

Mass Spectrometric Immunoassay

The MS methods utilised in population proteomics must be capable of analysing hundreds, if not thousands, of samples per day with high reproducibility and sensitivity. Hence, top-down MS approaches utilising affinity ligands are the most likely methods of choice for population proteomics. Surface-immobilised ligands can be utilised to affinity-retrieve a protein of interest from a biological sample, after which the protein (with or without the affinity ligand) is introduced in a mass spectrometer. One of the first affinity MS methods developed was the MS immunoassay (MSIA).[4,5] This approach combines targeted protein affinity extraction with rigorous characterisation using matrix-assisted laser desorption/ionisation-time of flight (MALDI-TOF) MS (see *Figure 1*). Protein(s) are extracted from a biological sample with the help of affinity pipettes derived by utilising polyclonal antibodies. The proteins are eluted from the affinity pipettes with a MALDI matrix and are MS-analysed. Enzymatic digestion, if needed, is performed on the MALDI target itself.

Specificity and sensitivity, as in traditional immunoassays, are dictated by the affinity-capture reagents, i.e. the antibodies. However, a second measure of specificity is incorporated into the resulting mass spectra, wherein each protein registers at a specific mass-to-charge ratio. During data analysis, the major signal in the mass spectrum that corresponds to the targeted protein is initially evaluated. It should be within a reasonable range (e.g. error of measurement $<0.05\%$) from the value of the empirically calculated mass obtained from the sequence of the protein deposited in the Swiss-Prot databank. Once this mass value is confirmed (or observed to be shifted), the presence of protein modifications is noted by the appearance of other signals in the spectra of the mass (usually in the vicinity of the native protein peaks) or by mass shifts of the major protein signal. Modifications can be tentatively assigned by accurate measurement of the observed mass shifts (from the wild-type protein signals and/or *in silico* calculated mass) and knowledge of the protein sequence and possible modifications. The identity of the modifications is then verified using proteolytic digestion and mass mapping approaches in combination with high-performance MS.

Determining Normal Biological Variations

In the case of healthy populations, the first candidates for population proteomics studies are well-studied (and generally higher-concentration) plasma proteins, because of availability of well-characterised affinity reagents (antibodies). Then, proteins at lower concentrations in human plasma can be progressively addressed. It is important to emphasise that there are virtually no data on the distribution of specific post-translational modifications across the general populace, even for the most abundant proteins. Deglycosylation, sequence truncations, side-chain residue modifications (phosphorylation, sulfonation, oxidation, etc.) have been reported for myriad proteins, yet, to date, a concerted effort has not been undertaken to assess the incidence of these structural modifications in individual proteins in the general population. Hence, the first aim of population proteomics is to catalogue protein modifications and establish their frequency in the general population. Those modifications that are

Dobrin Nedelkov is a Director of Research and Technology Development at Intrinsic Bioprobes, Inc., and a Founding President of the Institute for Population Proteomics, both in Tempe, Arizona. He introduced the concept of population proteomics, which is based on the application of affinity-based mass spectrometry approaches for high-throughput protein analysis. Prior to joining Intrinsic Bioprobes in 1999, Dr Nedelkov was a Post-doctoral Associate at Yale University in the Department of Molecular Biophysics and Biochemistry. He received his PhD in Chemistry and Biochemistry from Arizona State University in 1997.

© TOUCH BRIEFINGS 2007

found to occur at a high frequency can be declared wild-type and, unless they undergo a quantitative modulation in response to a specific disease, they stand to bear little significance to future biomarker discovery efforts. It is the low-frequency protein modifications that are most likely to be of greater significance as potential biomarkers of disease. To detect them, thousands of individuals will need to be analysed.

Samples

Both plasma and urine are good media for such a wide ranging proteome examination as they contain almost every protein (at some point in time, and in some shape and form) that the human body and cells produce (including cancer cells). While the wide dynamic ranges of protein concentrations and their molecular masses in plasma remain big challenges for the wide specificity proteomics approaches, these obstacles are minimised with targeted proteomics methods that are employed in population proteomics. Certainly, some protein-specific pre-treatment of those samples (e.g. addition of buffers, stabilisers, etc.) might be implemented, but major fractionation steps should be avoided so that the overall process and method remains simple and high in throughput.

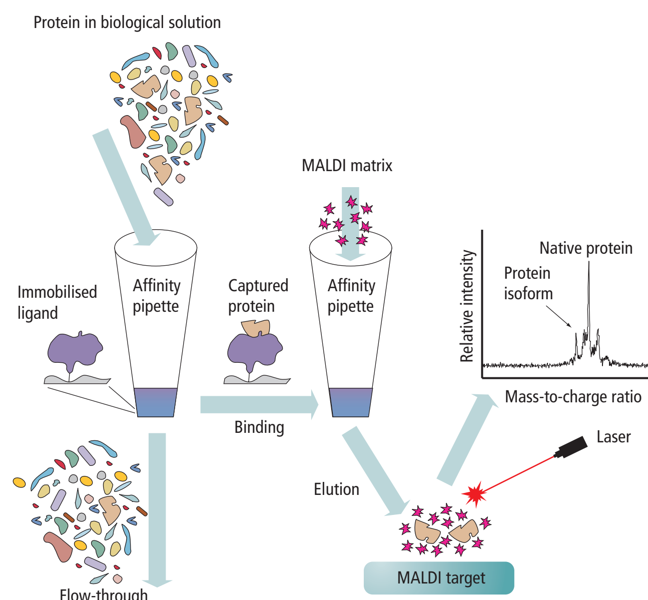
Discovery

It is expected that such large-scale population proteomics studies will result in the discovery of a plethora of novel protein structural modifications. For example, recent applications of the MSIA approach to a relatively small number of human plasma samples resulted in the discovery of new post-translational modifications for several plasma proteins, including apolipoprotein A-I,[6] apolipoprotein A-II,[6] C-reactive protein,[7] insulin-like growth factor II,[8] retinol-binding protein,[9,10] serum amyloid A[11] and serum amyloid P.[12] Consequently, the Swiss-Prot entries for these proteins have been annotated with the new modifications. Expanding the protein knowledge database via discovery of novel modifications is a key element of population proteomics studies.

Validation

Population proteomics can also be implemented for the validation of putative protein biomarkers discovered through clinical proteomics efforts, which is the most challenging aspect of this programme. Such validation efforts usually involve targeted analysis of the biomarkers (assayed either individually or in groups for better predictive values) via traditional immunoassay methods such as enzyme-linked immunosorbent assay (ELISA) and/or protein microarrays. However, these approaches cannot detect structurally modified proteins that have been indicated as putative cancer biomarkers.[13–15] Standard immunoassay approaches utilise detection labels that cannot discriminate between structural protein modifications. This is due to the

Figure 1: Mass Spectrometric Immunoassay Approach



MALDI = matrix-assisted laser desorption/ionisation.

fact that the resulting quantitative signal is the sum of signals from all isoforms for a given protein captured by the primary affinity ligand. It is only logical to validate protein modification implicated as potential biomarkers via immunoassays, which incorporate the same method of detection that was used in their discovery, i.e. MS. Otherwise, one is faced with the daunting task of generating highly specific monoclonal antibodies that will be able to discriminate between sequences that differ in as little as a single amino acid or a small side-chain modification.

Summary

In the last 10 years, proteomics has evolved from a technology-driven field into an application-driven discipline, the subject being the study of proteins in humans. Population proteomics, as described here, is an applied proteomics subdiscipline engaging in the long-term study of the human protein diversity. Assessing human proteins variations among and within populations is a paramount undertaking that can facilitate the effort of clinical proteomics in the discovery and validation of protein features that can be used as markers for early disease diagnosis, monitoring of disease progression and assessment of therapy. ■

1. Nedelkov D, Population proteomics: addressing protein diversity in humans, *Expert Rev Proteomics*, 2005;2(3):315–24.
2. Nedelkov D, Kiernan UA, Niederkofler EE, et al., Investigating human plasma proteins diversity, *Proc Natl Acad Sci U S A*, 2005;102(31):10852–7.
3. Nedelkov D, Kiernan UA, Niederkofler EE, et al., Population proteomics: The concept, attributes, and potential for cancer biomarkers research, *Mol Cell Proteomics*, 2006;5(10):1811–18.
4. Nelson RW, Krone JR, Bieber AL, Williams P, Mass-spectrometric immunoassay, *Anal Chem*, 1995;67(7):1153–8.
5. Krone JR, Nelson RW, Williams P, Mass spectrometric immunoassay, *Proc Soc Photo Opt Instrum Eng*, 1996;26:415–21.
6. Niederkofler EE, Tubbs KA, Kiernan UA, et al., Novel mass spectrometric immunoassays for the rapid structural characterization of plasma apolipoproteins, *J Lipid Res*, 2003;44(3):630–39.
7. Kiernan UA, Nedelkov D, Tubbs KA, et al., Selected expression profiling of full-length proteins and their variants in human plasma, *J Clin Proteomics*, 2004;1(1):7–16.
8. Nelson RW, Nedelkov D, Tubbs KA, Kiernan UA, Quantitative mass spectrometric immunoasay of insulin like growth factor 1, *J Proteome Res*, 2004;3(4):851–5.
9. Kiernan UA, Tubbs KA, Nedelkov D, et al., Comparative urine protein phenotyping using mass spectrometric immunoassay, *J Proteome Res*, 2003;2(2):191–7.
10. Kiernan UA, Tubbs KA, Nedelkov D, et al., Comparative phenotypic analyses of human plasma and urinary retinol binding protein using mass spectrometric immunoassay, *Biochem Biophys Res Commun*, 2002;297(2):401–5.
11. Kiernan UA, Tubbs KA, Nedelkov D, et al., Detection of novel truncated forms of human serum amyloid A protein in human plasma, *FEBS Lett*, 2003;537(1–3):166–70.
12. Kiernan UA, Nedelkov D, Tubbs KA, et al., Proteomic characterization of novel serum amyloid P component variants from human plasma and urine, *Proteomics*, 2004;4(6):1825–9.
13. Zhang Z, Bast RC Jr, Yu Y, et al., Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer, *Cancer Res*, 2004;64(16):5882–90.
14. Malik G, Ward MD, Gupta SK, et al., Serum levels of an isoform of apolipoprotein A-II as a potential marker for prostate cancer, *Clin Cancer Res*, 2005;11(3):1073–85.
15. Li J, Orlandi R, White CN, et al., Independent validation of candidate breast cancer serum biomarkers identified by mass spectrometry, *Clin Chem*, 2005;51(12):2229–35.

THE WHOLE WORLD IS IN THE MARKET FOR YOUR COMPANY'S NEW DRUG.
DO YOU REALLY WANT TO KEEP 6.5 BILLION CUSTOMERS WAITING?

Getting a new drug to market is a challenge for laboratories. Making these quality products affordable for everyone who needs them is a daunting task. With Waters' breakthrough technologies, powerful analytical software and unparalleled service and support, laboratory scientists can improve the efficiency of the drug development process. Eliminate unnecessary costs. Ensure regulatory compliance. And help get life-saving products to market faster. With laboratory science, affordable healthcare is possible. And a healthier world can become a reality.

To learn more, visit www.waters.com/b2.

©2007 Waters Corporation. Waters and The Science of What's Possible
are trademarks of Waters Corporation.