A new era for proteomics

John F Kellie^{*1}, Timothy W Sikorski¹, Bo An¹, Zhuo Chen¹, Ahmed H Moghieb¹, Maria G Busz¹, Matthew E Szapacs¹ & Thomas E Angel¹

¹Bioanalysis, Immunogenicity & Biomarkers, 1250 S Collegeville Rd, Collegeville, PA 19426, USA

^{*}Author for correspondence: john.x.kellie@gsk.com

“With technological advancements in both MS- and array-based proteomics, areas such as biomarker research and PD are poised for emerging applications by drug metabolism and PK groups, who have not traditionally engaged in such research.”

First draft submitted: 26 July 2019; Accepted for publication: 15 August 2019; Published online: 16 October 2019

Keywords: biomarkers • drug discovery • mass spectrometry • proteomics

Proteomics has seen various applications in research from the pharmaceutical industry. Recently, technological advancements have brought increased use in support of PD and biomarker investigations. Affinity proteomics platforms and MS-based approaches are being increasingly implemented. MS-based approaches, in particular, offer promising advantages such as protein-turnover measurement, proteoform discovery and monitoring. In addition to an overview of the current landscape, considerations for MS sample handling and preparation are discussed. Emerging technologies and future directions for implementing commercialized proteomics workflows are also highlighted. For laboratories where proteomics-based approaches and data processing practices are not well established, researchers should proceed with care in order to design optimal studies and deliver the highest of quality and most informative data possible.

Background

Identification and quantification of the proteins and proteoforms present in a biological sample provide a necessary foundation to generate insights into the complexities and emergent properties of biological systems to support drug discovery and development activities. To fully understand these complex systems, researchers have turned to MS-based proteomic technologies as well as highly multiplexed affinity-based platforms to quantitate hundreds to thousands of proteins in a single biological sample. In the pharmaceutical industry, proteomics has long been utilized as a drug-discovery tool to help understand changes in protein profiles for disease states or protein expression in relation to genomic studies for target discovery or identification [1]. With technological advancements in both MS- and array-based proteomics, areas such as biomarker research and PD are poised for emerging applications by drug metabolism and PK groups, who have not traditionally engaged in such research. Considerations of the pros and cons of proteomics methods, sample preparation and future implementations can guide researchers into this new era of proteomics.

Biomarker discovery, PD & on/off-target binding

Proteomics analysis for cellular systems undergoing experimental drug therapeutic intervention may yield identification of biomarkers useful for future assay development complementing already existing biomarkers. High-throughput sample workflows enable testing of certain subsets/subtypes of preclinical animal models or patient populations to understand global changes in protein concentrations with respect to disease state in plasma, urine, sputum or tissues. Proteins with shared pathway functions can help elucidate drug mechanism of action (e.g., up/downregulation) and play roles in drug discovery, biomarker discovery and PD [2]. It should be highlighted that additional testing is necessary to validate screening results, and researchers interested in using proteomics for biomarker discovery applications should plan to have such biomarker confirmation/validation tools in place to follow-up on changes observed from the discovery studies [3,4].

newlands
press

PD, a modulation due to compound exposure, plays a key role in drug development. Relationships between drug levels and protein expression, activity or turnover can provide a critical end point for candidate progression. A confident, statistically backed confirmation of target engagement or pathway modulation to confirm mechanism of action, when added to existing milestone evaluations, may ultimately lead to better drug candidates in stage II and III clinical trials, as was a key indication from the pipeline review at Pfizer [5]. However, the use of proteomics and the data considered for establishing such end points should be thoroughly considered and evaluated in advance, for example, establishing which proteins or protein families will be considered and measured to indicate favorable PD response.

The importance of understanding drug-binding target has also been identified as a critical step in drug development [5]. Chemical proteomics platforms have emerged as viable approaches for evaluating on-target and off-target drug binding or interactions, and monitoring drug-target engagement as proteome-wide assessments of different treatments can be examined. In the cellular thermal shift assay (CETSA) method, wherein cell or tissue samples are heated, and molecule binding effects examined, proteome dynamics are assessed with the aim of determining effects of on-target binding [6,7]. In addition to on-target binding, effects of off-target binding on proteins can also be determined by the assay. Others have also examined off-target binding for assessment of toxicity [8]. Typically results from cells and model systems may require additional experiments, but proteomics can serve an important role in the process of assessing on-target or off-target drug-binding effects.

Identification and quantitation of proteins based on absolute or relative abundance carries importance but simple information regarding identity and amount may not enable a deep understanding of complex relationships within biological systems. Measuring synthesis and degradation rates of biomolecules reveals the protein dynamics which are intrinsic to many biological processes. Stable isotope labeling combined MS enables the analysis of biomolecular flux (synthesis and degradation) and has successfully been applied to analysis of lipids, metabolites, proteins and cells [9]. For proteins, mass isotopomer distribution measurement from enzymatically digested peptides can provide individual protein fractional synthesis rate measurement [10]. For example, stable isotope labeling can be used to understand dynamics, synthesis and/or degradation of proteins can provide insight in disease or disease modulation to lead to future biomarkers [11].

Current proteomics platforms

Affinity-based platforms to perform quantitation of protein biomarkers to support clinical studies utilize either antibodies (e.g., O-link, Myriad, Kiloplex1000) or aptamers (e.g., Somascan) to capture the proteins of interest with quantitation varying from traditional immunoassay to array-based quantitation. Affinity platforms provide high sensitivity quantitation for up to thousands of biomarker proteins in different matrices (i.e., plasma, serum, urine, sputum or CSF), are highly multiplexed (up to 5000-plex) and can be automated for high-throughput sample analysis. Typically, reagents for affinity platforms are developed to capture human proteins, therefore cross reactivity is limited in preclinical species, which ultimately limits utility of these platforms in nonclinical studies. Affinity platforms are designed to quantitate total concentration of the protein biomarker and cannot typically detect post-translational modifications. Affinity-based assays have emerged for various formation/degradation biomarkers and neo-epitopes to try and close this gap, for example, neo-epitope monitoring to measure specific protein formation [12]. However, specific affinity assays for protein variants may ultimately hinder throughput of affinity-based proteomics platforms.

Advances in MS-based technologies (ion sources, optics and detectors) along with progress in sample preparation approaches such as isolation, purification and fractionation have led to modern studies which can identify, quantify or characterize hundreds or even thousands of proteins from a single biological sample source [3,4]. While affinity-based protein arrays can provide high-throughput quantification with high sensitivity, specificity is often uncertain and can be a limitation in understanding the resulting data. In contrast, MS provides high specificity, is not limited to the subset of proteins and proteoforms with available affinity reagents, and as such enables broad protein/proteoform and post-translational modification characterization and quantification. Commercialized MS platforms now enable analyses that provide critical information in milestone and decision-making checkpoints for drug discovery and development. In contrast to affinity approaches, it can be noted that LC-MS offers limited throughput and is more difficult to automate. Focusing more toward application (as opposed to development), a need exists for experts in biological systems partnered with trained analytical scientists as a two-way mentorship of sorts, to maximize potential gains when performing global sample analysis.

Sample preparation & analytical considerations

Measurement of protein abundance can provide expression or turnover information for PD analyses, yet simple, automated protein extractions are difficult for large studies [13]. Generally, protein extraction, pretreatment and cleanup are critical components of tissue/cell sample preparation to achieve reliable and robust quantitative proteomics [13]. Conditions including high salt buffers, surfactants, acidic/basic pH, organic solvents and sonication can also be used as a harsher extraction conditions, especially for membrane proteins.

For MS-based proteomic sample analysis, the digestion step is also critical for proper detection and quantitation. Filter-aided sample preparation and in-gel digestion protocols may be used if in-solution digestion is not ideal for tissue sample preparation (due to low peptide recovery and increased variance). Recently, a more-straightforward surfactant-aided on-pellet digestion protocol was reported for both target protein quantification and quantitative proteomics analysis, delivering superior digestion efficiency and reproducibility for various cells, bacterial tissues and biofluids [4]. Often, multi-dimensional separation and fractionation are implemented for simplifying protein or peptide mixtures at the point of analysis (e.g., at the mass spectrometer). For example, gel and gel-free separations of proteins can simplify complex matrices by resolving proteins based on molecular weight and isoelectric point. Alternatively, affinity immuno-depletion can remove from matrices abundant, unwanted proteins such as albumin or immunoglobins. While effective for simplifying the sample matrix and increasing the effective concentration of the analyte of interest, depletion of abundant proteins has high cost per sample, often prohibitive for the analysis of large sample numbers. Fractionation at the peptide level (after enzymatic digestion) has become the most common approach to decrease sample complexity for proteomic experiments. 2D LC separation, with low-pH reverse phase as the second dimension, is a common strategy to fractionate a peptide mixture prior to tandem MS analysis. Strong cation exchange chromatography has been implemented in the first dimension of 2D LC workflows because the mechanism of separation is considerably different from that of reverse phase, offering good orthogonal separation [14]. Strong cation exchange has drawbacks, such as extra time required to desalt samples before and after fractionation with poor sample recovery. More recently, high-pH reverse-phase chromatography has become a popular alternative for peptide fractionation prior to traditional low-pH LC-MS [15]. It is important to remember that any fractionation strategy significantly increases sample analysis time, as in most cases multiple fractions from a single sample need to be analyzed in the 2D. Intelligent fraction selection strategies have been developed to minimize the number of fractions that are chosen for LC-MS/MS analysis, leading to significant improvements in sample analysis throughput without sacrificing the gains in selectivity and sensitivity that come from peptide fractionation [16].

Future technologies

For the affinity-based proteomic platforms, researchers are continuing to increase the diversity of proteins that can be quantitated and as with all reagent-based platforms, the number of targets that can be quantitated is only limited by availability of a purified or recombinant version of the protein of interest and the ability to develop an affinity reagent(s) against that protein. In addition, one of the limitations of this technology is the inability to distinguish proteoforms or post-translationally modified proteins but as an understanding of these modifications becomes key to understanding disease, affinity-based reagents will become available and can be added as an end point to the various platforms available.

For MS-based proteomics platforms, a solution to address LC-MS separations with low separation peak capacity is the use of gas-phase separation by ion mobility spectrometry. The high-field asymmetric waveform ion mobility (FAIMS) system has demonstrated gas-phase separation of peptide ions, separation of co-eluting isobaric interferences and has shown promise coupling with existing GC- and LC-MS methods for potential use in clinical proteomics analyses [17]. CE-coupled MS has also been applied in clinical proteomics studies due to low sample consumption, low carryover, reproducible retention times and applicability to hydrophilic peptides. CE and HPLC can yield unique and complementary peptide identifications due to differences in separation mechanism [18]. Microscale multi-dimensional fractionation such as dynamic pH-junction-based CE have delivered practicality and performance of 2D-LC-CE-MS/MS and similar approaches may have utility for intact protein analyses [19,20]. However, wide-spread adaptation of CE-MS instrumentation remains a hurdle for routine application of CE-MS in clinical proteomics due to the limited adaptation of CE-MS compared with LC-MS, in addition to cost implications for new CE equipment.

Directions for implementation of commercialized proteomic technologies

The aim of many reagent and sample preparation companies, instrument vendors and software developers has been largely to automate and simplify the preparation, acquisition and data processing steps involved in MS. The marketing goals of a streamlined, integrated path to a product of proteomic data is appealing to new customer bases, but the same customers should be cognizant of the risks inherent to experiment planning; essentially, if a researcher cannot ask the right (experimental) question, the researcher might get a misguided answer. The disconnect between fresh, unprepared sample and protein search output can be a chasm. Education (ideally vendor agnostic), analytical training (or talent on-boarding) and diligent data review (with connection to the raw data) can narrow the chasm to a manageable and less-risky state. If a research laboratory already has LC and HRMS instrumentation, the only new investments may be in data processing, experimental planning expertise and oversight necessary to ensure high quality, compliant data. It is also important to note that the generation of a proteomics dataset is not the endgame of a study. Once proteins of interest for further interrogation are identified, targeted follow-up experiments (e.g., immunoassay or western blot-type analyses) or further targeted MS investigations should be planned.

A few hurdles will need to be cleared for instances of new implementation of proteomics workflows in laboratories seeking to demonstrate proteomic applications in emerging areas. Education of both analysts and management will be essential. Topics of proteomic databases, identification and reporting statistical significance cut-offs, post-translational modification reporting, sample processing options, mass spectrometer operation or proteomic experimental design may be foreign or under-appreciated in emerging proteomics laboratories. Training in areas of data acquisition and review may be necessary. Bench scientists will need to evaluate instrument performance over time, troubleshoot and identify sample and data processing solutions. Data handling, particularly for CROs may need to be addressed on a case-by-case basis. For all laboratories, management will need awareness of data acquisition, access, retention, compliance, quality assurance and implication in filings.

Acknowledgments

The authors gratefully acknowledge C Evans and E Yang for support.

Financial & competing interest disclosure

All authors are employees of GSK and may be eligible for GSK stock options or have stock ownership. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

1. Parker CG, Galmozzi A, Wang Y *et al.* Ligand and target discovery by fragment-based screening in human cells. *Cell* 168(3), 527–541 (2017).
2. Schenone M, Dančik V, Wagner BK, Clemons PA. Target identification and mechanism of action in chemical biology and drug discovery. *Nat. Chem. Biol.* 9(4), 232 (2013).
3. Keshishian H, Burgess MW, Specht H *et al.* Quantitative, multiplexed workflow for deep analysis of human blood plasma and biomarker discovery by mass spectrometry. *Nat. Protocols* 12, 1683 (2017).
4. Cominetti O, Núñez Galindo A, Corthésy J *et al.* Proteomic biomarker discovery in 1000 human plasma samples with mass spectrometry. *J. Prot.Res.* 15(2), 389399 (2015).
5. Morgan P, Van Der Graaf PH, Arrowsmith J *et al.* Can the flow of medicines be improved? Fundamental pharmacokinetic and pharmacological principles toward improving Phase II survival. *Drug Discov. Today* 17(9–10), 419–424 (2012).
6. Savitski MM, Reinhard FB, Franken H *et al.* Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* 346(6205), 1255784 (2014).
7. Molina DM, Jafari R, Ignatushchenko M *et al.* Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* 341(6141), 84–87 (2013).
8. Hoffmann BR, El-Mansy MF, Sem DS, Greene AS. Chemical proteomics-based analysis of off-target binding profiles for rosiglitazone and pioglitazone: clues for assessing potential for cardiotoxicity. *J. Med. Chem.* 55(19), 8260–8271 (2012).
9. Decaris ML, Li KW, Emson CL *et al.* Identifying nonalcoholic fatty liver disease patients with active fibrosis by measuring extracellular matrix remodeling rates in tissue and blood. *Hepatology* 65(1), 78–88 (2017).
10. Shankaran M, King CL, Angel TE *et al.* Circulating protein synthesis rates reveal skeletal muscle proteome dynamics. *J. Clin. Invest.* 126(1), 288–302 (2016).

11. Anderson NL. Dynamics of clinically important proteins: measuring turnover of drug targets and biomarkers. *Clin. Chem.* 64(2), 247–248 (2018).
12. Nielsen MJ, Nedergaard AF, Sun S *et al.* The neo-epitope specific PRO-C3 ELISA measures true formation of type III collagen associated with liver and muscle parameters. *Am.J. Transl. Res.* 5(3), 303 (2013).
13. Feist P, Hummon AB. Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *Int. J. Mol. Sci.* 16(2), 3537–3563 (2015).
14. Slebos RJ, Brock JW, Winters NF *et al.* Evaluation of strong cation exchange versus isoelectric focusing of peptides for multidimensional liquid chromatography–tandem mass spectrometry. *J. Prot. Res.* 7(12), 5286–5294 (2008).
15. Yang F, Shen Y, Camp DG, Smith RD. High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. *Expert Rev. Prot.* 9(2), 129–134 (2012).
16. Nie S, Shi T, Fillmore TL *et al.* Deep-dive targeted quantification for ultrasensitive analysis of proteins in nondepleted human blood plasma/serum and tissues. *Anal. Chem.* 89(17), 9139–9146 (2017).
17. Pfammatter S, Bonneil E, McManus FP *et al.* A novel differential ion mobility device expands the depth of proteome coverage and the sensitivity of multiplex proteomic measurements. *Mol. Cell. Prot.* 17(10), 2051–2067 (2018).
18. Pontillo C, Filip S, Borràs DM, Mullen W, Vlahou A, Mischak H. CE-MS-based proteomics in biomarker discovery and clinical application. *Proteomics – Clinical Applications* 9(3-4), 322–334 (2015).
19. Yang Z, Shen X, Chen D, Sun L. Microscale reversed-phase liquid chromatography/capillary zone electrophoresis–tandem mass spectrometry for deep and highly sensitive bottom-up proteomics: identification of 7500 proteins with five micrograms of an MCF7 proteome digest. *Anal. Chem.* 90(17), 10479–10486 (2018).
20. Li Y, Compton PD, Tran JC, Ntai I, Kelleher NL. Optimizing capillary electrophoresis for top-down proteomics of 30–80 kDa proteins. *Proteomics* 14(10), 1158–1164 (2014).