Qoronfleh, J Proteomics Bioinform 2017, 10:10
DOI: 10.4172/jpb.1000448

Journal of
Proteomics & Bioinformatics

Short Communication Open Access

Protein Biomarker Validation: A Mass Spectrometry Approach

M Walid Qoronfleh*

Department of Healthcare Research & Policy, WISH Center, Qatar Foundation, Doha, Qatar

Abstract

This concept communication describes a mass spectrometry workflow specific to biomarker assay development and validation. The primary objective of this workflow is to significantly decrease the cost and timeline for assay development and biomarker validation. The biomarker development workflow can be separated into five development stages that take less than 6 months to complete.

Keywords: Proteomics; Protein biomarkers; Biomarker validation; Assay development; Mass spectrometry; Peptide multiple reaction monitoring (pMRM); LC-MS/MS; GeLC/MS

Introduction

In the post-genome era, the use of molecular biomarkers is becoming de rigueur for programs in research and in drug development. The biomarker testing market has a proven record of revenue generation ($612 MM in 2007) and is estimated to have an annual growth rate of 23.5% based on currently available biomarker assays. Proteomic technologies have been used successfully for biomarker discovery projects producing lists of many candidate protein biomarkers [1]; when integrated with genomic data and literature mining there can be hundreds of candidates for a given study. However, further verification work is typically limited by the small number of proteins for which there are commercially available assays (~500 human proteins). If researchers opt to develop assays for these candidates using traditional antibody-based approaches, the cost would likely be over $25K/protein and the timeline would be a couple of years or more. Thus, for protein biomarkers, assay development is a current bottleneck.

In this communication article we explain a Liquid Chromatography/Mass Spectrometry (LC/MS) based approach for the validation of protein biomarkers. Quantitation is achieved using surrogate peptides generated from an enzymatic digest of the native protein in a biological sample [1,2]. The application LC-Multiple Reaction Monitoring mass spectrometry (LC-MRM/MS) technology enables the quantitation of the surrogate peptide in the digested biological sample [2]. The peptide multiple reaction monitoring (pMRM) assay is very specific for targeted proteins. The stoichiometric relationship between the peptide and the native protein can be used to confer the protein level in a given sample. Ultimately the use of an isotope labeled internal standard peptide yields absolute quantitation data. The primary objective of this workflow is to significantly decrease the cost and timeline for assay development and biomarker validation. The biomarker development workflow can be separated into five components (Figure 1) described below.

Workflow Concept

Stage 1: Identification of candidate protein biomarkers

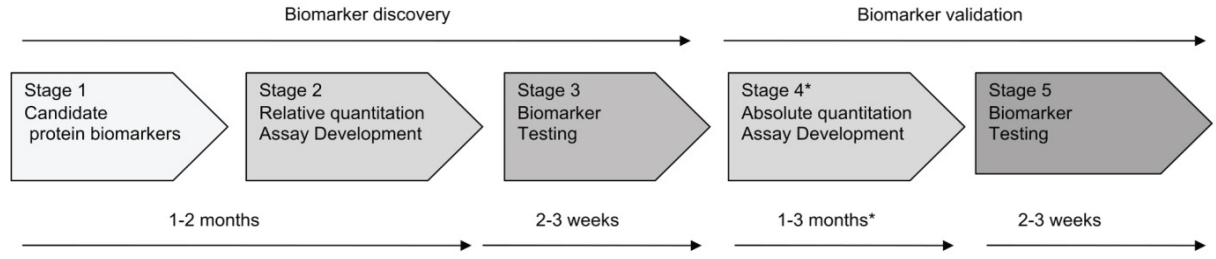
Candidate protein biomarkers may come from a number of sources including proteomic discovery experiments, a discovery experiment in another discipline (e.g., transcript profiling, pharmacogenomics) or be based on literature sources. In the biomarker development workflow described here, mass spectrometry data for candidate protein biomarkers enables the rapid design of protein assays. Mass spectrometry data that can be used in protein biomarker assay development include peptide fragmentation spectra, protein sequence coverage maps, peptide abundance or ion intensity and observed charge state information. Other useful analytical information that can be garnered from mass spectrometry based experiments include chromatographic behavior and protein isoform information. All the aforementioned information can all be collected from a one-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis followed by liquid chromatography-tandem mass spectrometry (GeLC/MS) experiment [3] outlined in Figure 2. In a GeLC/MS experiment sample characterization is achieved through a combination of SDS-PAGE for protein fractionation and liquid chromatography-tandem mass spectrometry (LC-MS/MS) for detection and quantitation of proteolytically derived peptides. Sample preparation is matrix-dependent and is no more complex than removal of abundant proteins. One-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) is used as a protein fractionation strategy to reduce the complexity of the sample prior to LC-MS/MS analysis and to ensure the deepest proteome coverage possible within an economical time frame. Post electrophoretic separation the polyacrylamide gel is excised into 40 equal segments. Each segment is enzymatically digested (usually with trypsin although alternative enzymes may be employed to target specific regions of proteins) and each peptide sample is then analyzed by data-dependent LC-MS/MS using nano-scale chromatography and nano-spray ionization coupled with a hybrid linear ion trap LTQ–Orbitrap mass spectrometer. The data are searched using a search engine against available protein databases and compiled into non-redundant lists using commercial software tools.

GeLC/MS is a highly sensitive approach that yields a large number of identifications and matches a high number of peptides per protein (median value typically 6-10 unique peptides); having multiple peptide choices is a key feature for subsequent assay development. Libraries of protein/peptide mass spectrometry data can be generated for different biological samples and then quickly referenced with software tools. The private protein/peptide library information produced by NextGen Sciences, Inc. (Ann Arbor, USA) termed BiomarkerLibrary™ or other public resources containing similar information such as PeptideAtlas

***Corresponding author:** M Walid Qoronfleh, Director, Department of Healthcare Research & Policy, WISH Center, Qatar Foundation, PO Box: 5825, Doha, Qatar, Tel: +974-4454-6475; +974-6690-5093; E-mail: wqoronfleh@qf.org.qa

Received October 06, 2017; **Accepted** October 23, 2017; **Published** October 27, 2017

Citation: Qoronfleh MW (2017) Protein Biomarker Validation: A Mass Spectrometry Approach. J Proteomics Bioinform 10: 246-251. doi: 10.4172/jpb.1000448



Stage 1: Compile candidate protein biomarkers list. Sources include Differential GeLC/MS, transcriptomics, literature, etc.

Stage 2: Develop relative quantitation assay for proteins identified in Stage 1.

Stage 3: Apply relative quantitation assay from Stage 2 to a cohort of samples representative of normal and stressed state.

Stage 4: Refinement of assay using data from Stage 3. Elevation to absolute quantitation accompanied with fit for purpose validation .

Stage 5: Relative or Absolute quantitation of target proteins across hundreds of samples.

** Optional*

** Validation timelines vary with the number of proteins/peptides included in the assay and the extent of assay validation.*

Figure 1: Components of the proposed protein biomarker assay development pipeline.

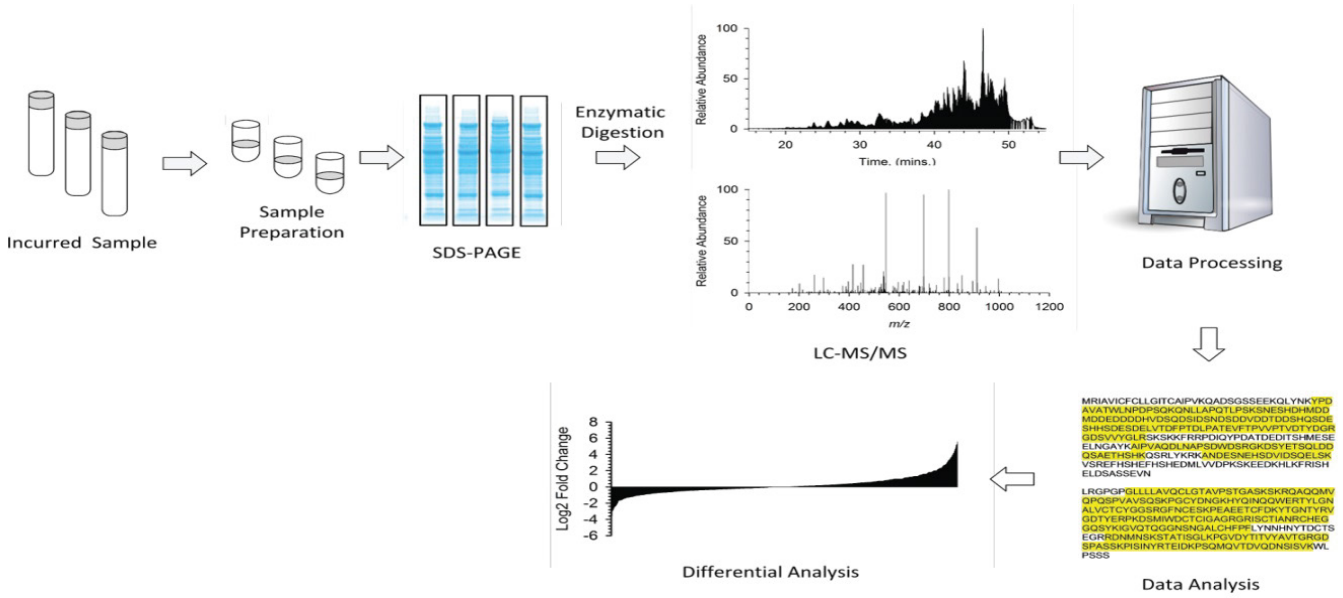


Figure 2: The GeLC/MS - based biomarker discovery workflow. The general processes involved in characterizing the proteome of a biological sample is outlined. Sample preparation can vary depending on the nature of the sample but is typically no more complex than removing the most abundant proteins in a sample. A sample is fractionated by 1D SDS-PAGE before enzymatic digestion and subsequent LC-MS/MS analysis on a hybrid linear ion trap – Orbitrap mass spectrometer. Peptide fragmentation data are processed using bioinformatics tools to produce lists of identified proteins. Spectral count data is then used to perform measurements of protein abundance across a number of samples. These data can be used to calculate p-values and determine fold change on a global proteome scale.

and BiblioSpec, available at www.peptideatlas.org and http://proteome.gs.washington.edu/software/bibliospec, respectively, are examples of such databases. Differential GeLC/MS experiments can be used as an integrated part of the biomarker development workflow described in this article. In this case samples or pools of samples representing different states of a disease or toxic insult can be analyzed by GeLC/MS and the data compared directly. Spectral counting is used to provide a metric of protein/peptide abundance. The data from differential GeLC/MS experiments can be subjected to statistical analysis and the relative changes in protein levels assigned p-values. These p-values can then be used in combination with biological oversight and reasoning to generate a list of candidate protein biomarkers to be moved forward into assay development.

Stage 2: Relative quantitation assay development

The objective of this stage of the biomarker development workflow is to develop the assay that will be used to confirm candidate biomarkers identified in Stage 1. The timelines for this stage of the biomarker development workflow is governed by the number of proteins included in the assay; an assay for a panel of twenty proteins is typically complete in less than a month. For quantitative assays peptide detection is accomplished using liquid chromatography with Multiple Reaction Monitoring (MRM) mass spectrometry (LC-MRM/MS), this technique is widely accepted as the most selective and sensitive mass spectrometry coupling for quantitative bioanalysis [2,4]. The surrogate peptide workflow

requires that biological samples must be subjected to enzymatic digestion prior to analysis. Candidate surrogate peptides that have previously been identified in a biological matrix can be identified for instance by referencing the biomarkerlibrary™, or a similar repository of mass spectrometry data. During assay development, many peptides for each protein are initially selected and then through an iterative process of assay development the peptides with the best detection characteristics are ultimately selected to move forward as part of the assay. The attrition process initially involves the consideration of certain fundamental aspects of the peptide chemistry these include the presence of oxidation sites, alkylation sites, glycoslyation motifs, basic residues and missed cleavage sites. These factors can impact the analytical, technical and biological variability of a peptide and are an undesired source of bias in an assay. Peptides containing these functionalities are excluded from the surrogate candidate list. The presence of protein isoforms or biological processing products must also be considered, the data in Figure 3a, demonstrate the detection of multiple forms of a protein in cerebral spinal fluid. The GeLC/MS characterization platform allows visualization of the different forms of a protein in a sample and peptides can be selected accordingly. The presence of multiple protein forms is important consideration in assay development.

Typical data acquired during method development for three candidate surrogate peptides are presented in Figure 3b. One observation that can be made from the chromatographic data is the poor sensitivity of peptide 2 relative to peptides 1 and 3. This peptide was previously identified in a GeLC/MS experiment but it does not perform well on the LC-MRM/MS platform. This failure to transfer from one platform to another can be attributed to many factors and is part of the natural attrition process in peptide selection for protein assays of this nature. Peptides 1 and 3 have good selectivity with the best sensitivity was achieved for peptide 1. The bar chart in panel 3b shows some other characteristics of peptides that must be considered in the surrogate peptide selection process. Peptide 1 (the blue bars) has poor analytical reproducibility with a broad range of CV's in excess of 20% and poor technical reproducibility. Peptide 2 (the red bars) is the most sensitive peptide with good analytical and technical reproducibility. Peptide 3 (the green hatched bars) shows good analytical reproducibility but poor technical reproducibility. Peptides 1 and 2 display all the characteristics necessary to have utility in an LC/MRM-MS assay.

Multiple methods for candidate surrogate peptides must be tested against samples representative of the different states of the incurred samples. We have found that this is best accomplished with pooled samples representing extreme states of the control and stressed

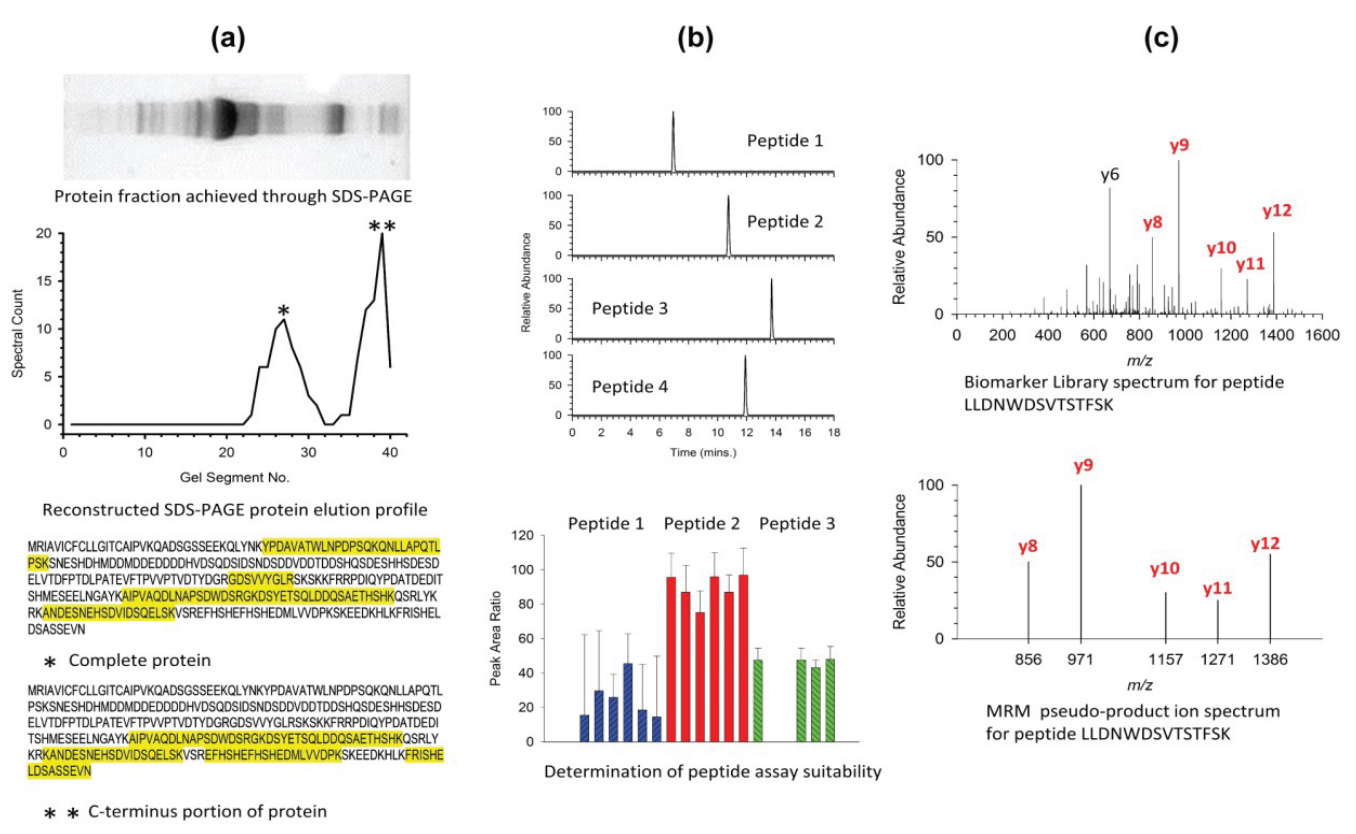


Figure 3: Highlight of some of the key points in biomarker assay development. Figure 3a shows two forms of a protein are clearly resolved using 1D SDS-PAGE. Upon examination these are revealed as the intact protein and a biological artifact consisting of the c-terminus of the protein. Figure 3b, an example of the chromatographic output during assay development. Figure 3c, the usefulness of peptides as surrogates for the intact protein is determined during assay development. Here three peptides are tested and their relative merits and demerits factored in to the selection process. Here three peptides for a single protein have been tested in six technical replicate samples. Only peptide two has the properties desired to move forward in the assay. Figure 3c is an example of the cross platform portability of the product ion spectra of the peptide LLDNWDSVTSTFSK. The peak ratios in the less complex MRM spectrum closely match those of the more complex ion-trap spectrum. On a routine basis the correlation between the two spectra are determined using bioinformatics tools. Classification of differently expressed proteins with numbers of respective proteins according to the cellular localizations Gene Ontology (GO). The analysis was performed using the String 10.0 database with a minimum false discovery rate P_{FDR} <0.05. The numbers refer to the number of proteins corresponding to the respective GO terminology.

samples. The inclusion of multiple product ions for a candidate peptide for a protein biomarker further increases the specificity of the LC-MRM/MS platform. Furthermore the data generated using multiple transitions can be used as a corroborative tool to confirm the identification of peptides detected on the LC/MRM-MS platform. The rank order of the y ion series from the pseudo-product ion spectrum reconstructed from MRM data can be compared to the series observed in the biomarkerlibrary™. An example of this comparison is presented in Figure 3c for the peptide LLDNWDSVTSTFSK. The rank order of the y ions is y9, y12, y8, y10 and y11 this pattern is in agreement between the quantitative and qualitative data.

If the assay is to be applied to a small sample set and is being used to provide data as part of exploratory or demonstrative studies then a relative quantitation assay may be sufficient to answer the questions asked. For relative quantitation assays the extent of analytical method validation can be limited to be fit for the purpose of the application. However, if the assay is intended for such applications as patient stratification or dose selection then it is likely data need to be reported as protein concentrations. Thus, an absolute quantitation assay must be developed and more stringent validation criteria must be adhered to (see Stage 4). Other considerations during the development of the assay are the number of samples that are likely to be analyzed and the manual labor required to prepare the samples. For large sample sets automation must be considered to minimize errors and for optimum productivity and efficiency.

Stage 3: Biomarker testing

Stage 3 experiments are typically performed using 10's of samples representing clearly delineated populations [5]. The timelines for this stage of the biomarker development workflow are typically 2-3 weeks depending on the number of samples submitted to the assay. Figure 4 is an overview of the general scheme for sample testing. Incurred samples are typically biological fluids such as plasma, cerebral spinal fluid, and urine. Sample preparation is typically specific to the biological matrix and can be minimal such as centrifugation or more involved such as solid phase extraction or immuno-affinity depletion and/or purification [5]. Often the abundance of a protein can inferred from pre-existing data and this will influence the sample preparation required for its detection [5]. Data collected during Stage 3 can be readily visualized using comparative plots and early decisions made whether or not to move a protein biomarker forward to be tested within a larger more variable cohort.

A key feature of the MRM platform is that many peptides, and therefore proteins, can be multiplexed in a single chromatographic run. In Figure 4 an example of data from a 25-plex assay, where each protein is represented by two peptides, are presented as a three dimensional plot. The chromatographic runtime was seventeen minutes with a total injection to injection cycle time of twenty minutes or 1.25 mins/protein. The chromatographic data are integrated using quantitation software provided by the instrument vendor. In this instance the data are reported in peak area ratio's calculated using an internal standard protein added to the incurred samples.

Stage 4: Absolute quantitation assay development

For biomarkers that may be used in critical decision-making processes, such as patient stratification or dose selection, protein concentrations are generally considered appropriate [2,4,6]. Part of the biomarker validation process is establishing a protein level that is considered normal and out of the normal range. In this stage of the biomarker development workflow the relative quantitation assay (Shown in Stage 2) is further developed to provide concentration values for proteins in the assay. For absolute quantitation an internal standard is employed that is a stable isotope labeled peptide (typically containing 13C and/or 15N atoms) with an amino acid sequence identical to the native form [4]. This peptide will have analytical performance metrics identical to the native peptide but be shifted in mass. Figure 5a is an example of data from the analysis of four peptides and their equivalent labeled internal standards. Calibrations curves for the four peptides were developed in a proxy matrix; an example of a typical calibration curve is presented in Figure 5b. In this way absolute concentrations for the peptides in the assay can be calculated. The resulting absolute quantitation MRM assay is validated and the extent of validation typically fits the purpose of the assay as illustrated in Figure 5c. The timelines for this stage of the biomarker development workflow are 1-3 months; this broad range is a result of variations in the number of proteins/peptides in the assay and the extent of the validation required for the desired application. This stage of biomarker development is also application specific and it may be the case that relative quantitation is sufficient thus further decreasing timelines in the development of a biomarker.

Stage 5 of Biomarker testing

Following data review from stage 3, the panel of candidate protein biomarkers may already have been reduced in size or some proteins may have been replaced with others. In this regard the development of a multiplex protein biomarker assay may be considered as an iterative process, where the throughput and flexibility of mass spectrometry is recognized as an enabling tool. Stages 2 and 3 can be repeated several times before a panel is ready for scaling up to absolute quantitation or application to precious and limited clinical samples [5,7].

Stage 5 in the biomarker development workflow involves the application of a well characterized assay to hundreds of samples.

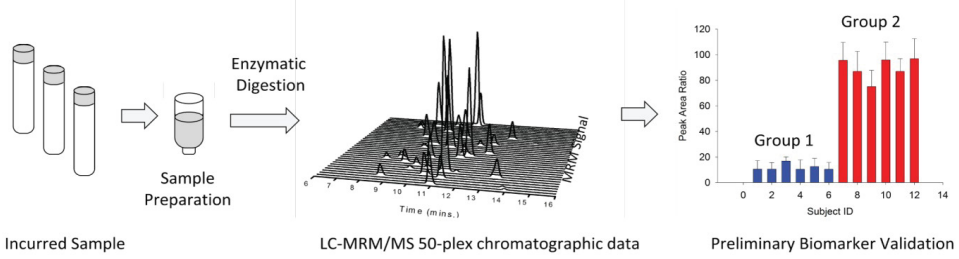


Figure 4: Relative quantitation assay scheme. Incurred biological samples are typically biofluids such as plasma, urine, and cerebral spinal fluid. Sample preparation is typically specific to the biological matrix and can be minimal such as centrifugation or more involved such as solid phase extraction or immuno-affinity depletion and/or purification. Samples are analyzed by LC-MRM/MS and the data processed using commercial bio analytical software tools.

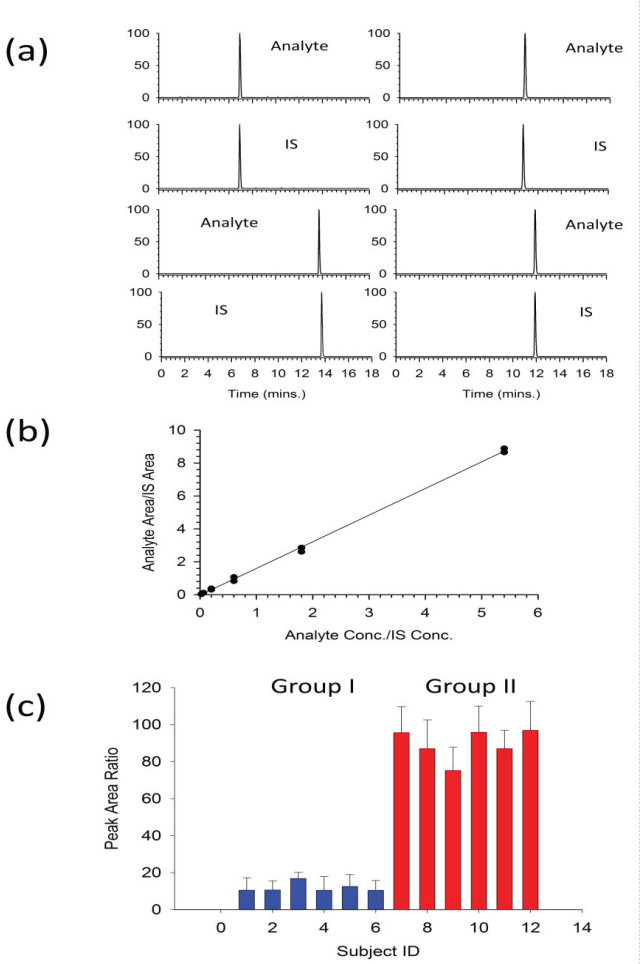


Figure 5: Absolute quantitation assay scheme. Panel A) Typical chromatographic data for native peptides and their respective isotope labeled partners. Panel B) Calibration data for a six point curve with analytical duplicate measurements. Panel C) Absolute quantitation data for a protein.

Fidelity of all peptide signals across many samples is assessed in detail, and discriminatory power of individual signals taken alone or in combination is evaluated. NextGen Sciences, Inc. has used the workflow described in this article with a number of groups (academic, biotech and large pharma) to develop protein biomarkers. Over the last two years they have demonstrated that the MRM platform can be used to develop multiplex assays and perhaps more importantly, that we can build assays and test samples in the timelines presented in this workflow. As with Stage 3, the timelines in biomarker testing timelines are typically defined by the number of samples submitted to the assay.

Conclusion

A mass spectrometry workflow specific to biomarker assay development and validation is presented. The biomarker development workflow can be separated into five development stages that take less than 6 months to complete. The mass spec-based assay platform approach reduces cost and development time. The multiplexing capability of the pMRM assay easily processes 30–50 proteins simultaneously and can go as high as 100 proteins. The workflow has been shown to accelerate biomarker assay development and validation. NextGen Sciences, Inc. in the USA has used this concept effectively to launch its first protein biomarker discovery panel, plasmadiscover41, an MRM-MS-based panel comprising 41 plasma proteins thought to be potential biomarkers for breast, prostate, and lung cancer [5,6]. It also released a number of additional protein marker panels in 2011 and 2012, including an expanded human plasma panel and cerebrospinal fluid (CSF) panels for human and rat. The assay simultaneously measures 43 human CSF proteins that are thought to have potential as biomarkers in CNS diseases, particularly Alzheimer's, Parkinson's, multiple sclerosis (MS) and amyotrophic lateral sclerosis (ALS) diseases.

Acknowledgements

The author wants to acknowledge the technical contribution of Dr. Richard Jones, Dr. Michael Ford, Dr. Ravi Amunugama and David Allen of MS Bioworks LLC (USA). The author also wish to thank NextGen Sciences, Inc. (USA) and its scientific team for utilizing the biomarker concept to develop human assay disease panels.

References

1. Biomarkers Definitions Working Group (2001) Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 69: 89-95.

2. Gerber SA, Rush J, Stemman O, Kirscher MW, Gygi SP (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci 100: 6940-6945.

3. Dzieciatkowska M, Hill R, Hansen KC (2014) GeLC-MS/MS analysis of complex protein mixtures. Methods Mol Biol 1156: 53-66.

4. Anderson NL, Anderson NG, Haines LR, Hardie DB, Olafson RW, et al. (2004) Mass spectrometric quantification of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). J Proteome Res 3: 235-244.

5. Cummings J, Ward TH, Greystoke A, Ranson M, Dive C (2008) Biomarker method validation in anticancer drug development. British J Pharmacol 153: 646-656.

6. Anderson L, Hunter CL (2006) Quantitative mass spectrometry MRM assays for major plasma proteins. Mol Cell Proteomics 5: 573-588.

7. Usdin S (2008) Getting real about biomarkers. SciBx 1: 1-4.