BIOTECH SUPPORT GROUP

APPLICATION REPORT
APRIL 16, 2015

On-Bead Digestion
Tackling the challenges of serum proteomics (LC-MS)

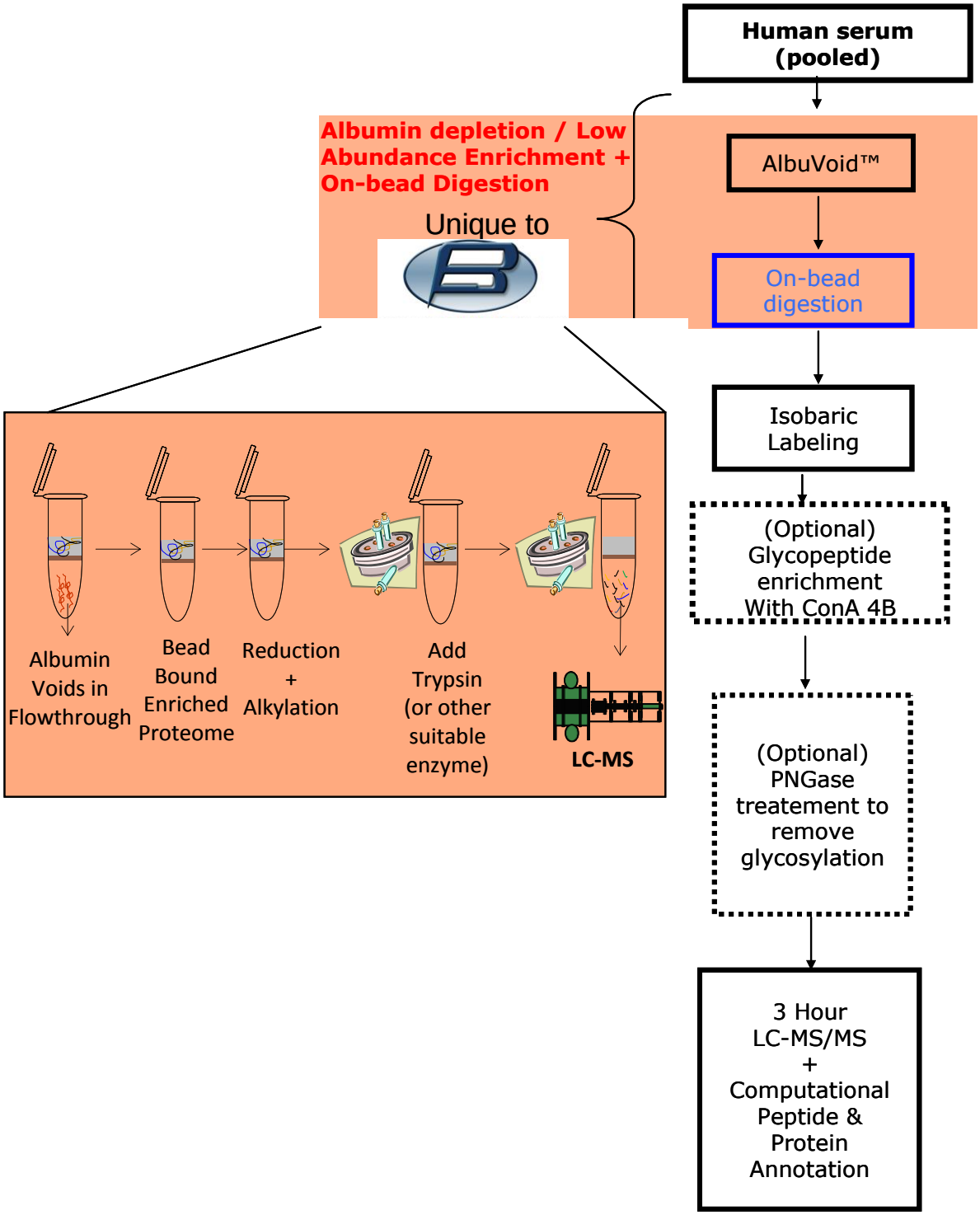
Introduction

Serum and plasma proteomics can be challenging for two reasons: 1) Albumin accounts for about 50% of the total protein mass, and 2) Serum and plasma as a whole presents a challenging proteolytic sample type; approximately 40% of plasma peptides in the public data repository – PeptideAtlas, correspond to partly tryptic sequence[1]. For the first challenge, the AlbuVoid™ bead and associated method can efficiently deplete albumin by a voidance strategy, enriching the remaining low abundance proteins on the bead. The second challenge - proteolytic resistance, is due in part by the substantial presence of glycoproteins, a particularly resistant class of proteins. As Trypsin requires access to a protein's interior, it has been suspected that the large size of glycan groups can block such access, resulting in missed cleavages not suitable to computational data-mining[2].

As a result, both missed and non-specific cleavages can greatly impact the final data profile, impacting all proteomic applications, be it for discovery or quantification. Therefore short, efficient and consistent digestions are of paramount importance in all proteomic workflows. For these reasons, we continue to evaluate and report these efficiencies in our methods development for on-bead digestion using AlbuVoid™ beads. In the data to follow, we consider the efficiencies gained by an optimized workflow of albumin depletion, low abundance enrichment, and on-bead digestions at pH 7, to identify peptides and proteins from serum including glycoproteins.

We previously reported for AlbuVoid™, on-bead Trypsin digestion was equivalent or better than in-solution, with LC-MS/MS derived peptide and protein identifications comparing favorably with immuno-depletion[3,4]. These results were based on the optimal in-solution condition at pH 8. We subsequently have observed that for on-bead digestions with AlbuVoid™, pH 7 is optimal (patent pending). We suspect at pH 7, much of the protein remains bound to the AlbuVoid™ beads, presenting better proteolytic access than would otherwise be available in solution or when marginally bound to the AlbuVoid™ beads, both of which occur at higher pH.

Serum Proteomic Workflow Using AlbuVoid™ Beads



Material & Methods

AlbuVoid™ Protocol for Albumin Depletion and Low Abundance Serum Protein Enrichment

Processes 50-100 µl serum per prep. In bold are the **AlbuVoid™ kit** components.

1. Weigh out 25 mg of **AlbuVoid™** bead in a spin-tube (0.45µ SpinX centrifuge tube filter from Corning).

2.Add 125 µl of **Binding Buffer AVBB.** Vortex for 5 minutes at room temperature followed by centrifugation at 3000 rpm. Discard the supernatant.

3.Repeat step-2

4.Condition by adding 100 µl of **AVBB** and 50-100 µl of the **Serum.** Centrifuge for 5 minutes at 10,000 rpm. Add clarified sample to the AlbuVoid™ bead in step 3. Vortex for 10 minutes and then centrifuge for 4 min. at 10,000 rpm.

5. Remove the albumin enriched supernatant (Flow-Through) **FT**.

6. To the beads, add 250 µl of **Wash Buffer AVWB.** Vortex for 5 min and centrifuge for 4 minutes at 10,000 rpm. Discard the **Wash.**

7. Repeat Step-6.

The AlbuVoid™ bead is now enriched with albumin depleted low abundance proteins. For LC-MS sample preparation, an on-bead digestion protocol can be applied.

8. After the final wash steps from Step 7 from the enrichment, add 10 µL 100mM DTT + 90 µL 10mM HEPES, pH 7, vortex 10 min, incubate ½ hr at 60 °C.

9.After cooling, add 20µl 200mM Iodoacetamide, and 80 µL 10mM HEPES, pH 7, incubate in dark for 45 min at room temp.

10.Centrifuge at 10,000 rpm (microfuge max setting) for 5 minutes, and discard supernatant.

11.Add 40 µL Sequencing-grade trypsin (0.4µg/µl, in 50mM acetic acid) + 60 µL 10mM HEPES, pH 7 to the beads. Digest overnight at 37°C or other optimized time period.

12. Centrifuge at 10,000 rpm (microfuge max setting) for 5 minutes, and retain peptide filtrate.

13.To further extract remaining peptides, add 150 µL 10% formic acid, vortex 10 min, centrifuge at 10,000 rpm (microfuge max setting) for 5 mins., and add this volume to the first volume.

14.Total is about 250µl. Prepare to desired final concentration. Store at -80 °C until LC-MS/MS.

LC-MS/MS Analysis, Peptide & Protein Identifications

The following results were obtained by LC-MSMS at the Mass Spectrometry facility at Rutgers Proteomics Center. A nanoLC-MS/MS coupled a RSLC system (Dionex, Sunnyvale CA) interfaced with a LTQ Orbitrap Velos (ThermoFisher, San Jose, CA). Samples were loaded onto a self-packed 100µm x 2cm trap packed with Magic C18AQ, 5µm 200 A (Michrom Bioresources Inc, Aubum, CA) and washed with Buffer A(0.2% formic acid) for 5 min with flow rate of 10ul/min. The trap was brought in-line with the homemade analytical column (Magic C18AQ, 3µm 200 A, 75 µm x 50cm) and peptides fractionated at 300 nL/min with a multi-stepped gradient (4 to 15% Buffer B (0.16% formic acid 80% acetonitrile) in 35 min and 15-25%B in 65 min and 25-50%B in 50 min). Mass spectrometry data was acquired using a data-dependent acquisition procedure with a cyclic series of a full scan acquired in Orbitrap with resolution of 60,000 followed by MSMS scans (CID 35% of collision energy) of 20 most intense ions with a repeat count of two and the dynamic exclusion duration of 60 sec. The LC-MS/MS data was searched against the Human & Rat respective Ensembl databases using X!tandem with carbamidomethylation on cysteine as fixed modification and oxidation of methionine and deamidation on Asparagine as variable modifications using a 10 ppm precursor ion tolerance and a 0.4 Da fragment ion tolerance. Glyco-peptides were determined as deamidation at the asparagine where NXS/T motif. We included only the peptides with loge<=-1.5 and charge state <=4. For spectra with multiple top hit peptides, only one was selected.

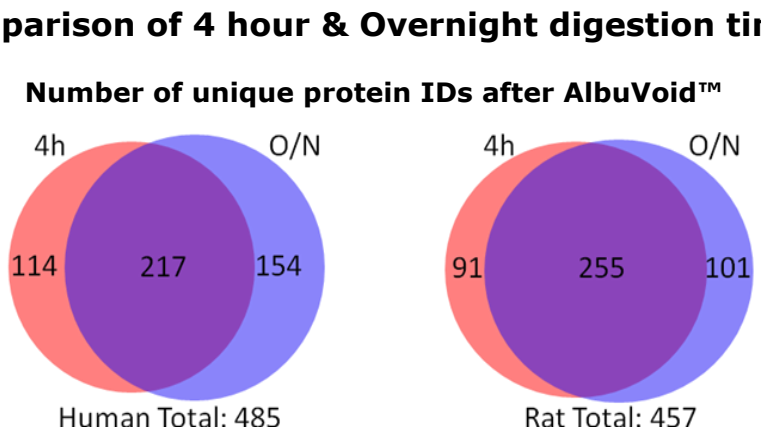


Figure 1 – The total protein identifications from AlbuVoid™ are compared for human and rat sera, each at 2 digestion times, 4 hours and overnight. 217 human serum proteins were overlapping from each digestion condition, that is, they were identified regardless of the digestion time. Similarly, 255 rat serum proteins were overlapping from each digestion time.

Results and Discussion

The efficiency and consistency of proteolysis is often taken for granted in proteomic workflows. Although perfect specificity and complete digestion of proteomes is often assumed and certainly desirable, unfortunately it is not realistic as some tryptic peptide sites are slow to hydrolyze and difficult to digest[5]. Furthermore, all commercial sequencing grade Trypsins have some chymotryptic side-activity presumably from autoproteolysis, resulting in further digestion of the tryptic peptides at non-canonical sites, or over-digestion[6]. Thus digestion efficiency can impact the total number of peptides and proteins identifiable through the duration of any practical or economic limits to LC-MS/MS instrument time – 3 hours in this case.

So while preliminary digestion data can establish proteolytic efficiency metrics for missed cleavages, the peptides contributed from non-specific cleavages (over-digestion) can often lead to misassignments or missed identifications. This is because peptide features are either permanently lost or masked, and are especially hard to recover computationally[1,6]. One can however evaluate indirectly, the number of proteins identifiable under different digestion conditions. Such an indirect measure is reported in Figure 1; the number of total proteins identified at two different digestions times. In like manner, the total number of tryptic identifiable peptides are reported in Figure 2.

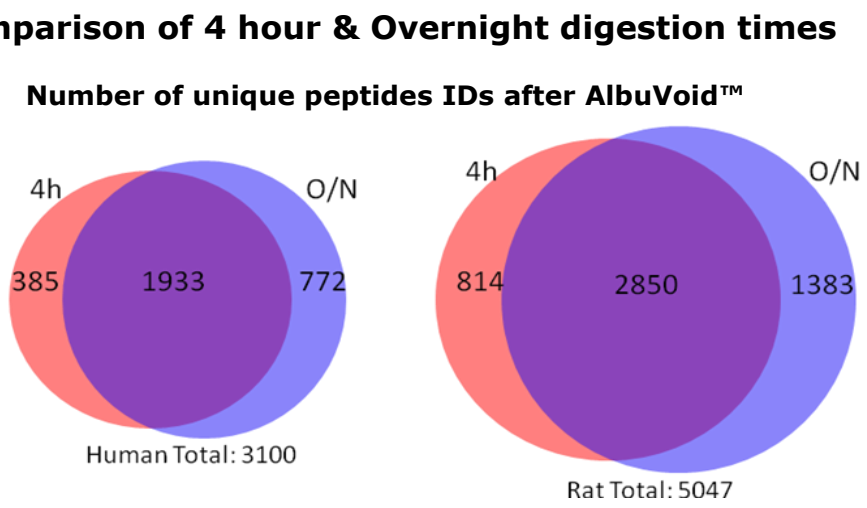


Figure 2 – The total peptide identifications from AlbuVoid™ are compared for human and rat sera, each at 2 digestion times, 4 hours and overnight. 1933 human serum peptides were overlapping from each digestion condition, that is, they were identified regardless of the digestion time. Similarly, 2850 rat serum peptides were overlapping from each digestion time.

For quantification applications, requisite peptide data quality for quantification limits the number of quantifiable proteins to a fraction of the total identified. Yet, quantification especially relies on the consistency and quality of limit peptides, defined as those peptides to which peptide bonds that can be cleaved, have been cleaved and cannot be subject to further hydrolysis by the protease in use[7]. For labeled peptide quantification, complete proteolysis ensures reliable peptide quantification, while for label-free, incomplete digestion diminishes the ion signal attributable to limit peptides. Indeed several reports show that quantitative precision is strongly influenced by variations in enzymatic digestion efficiency[8-10].

Thus proteolytic efficiency affects all proteomic applications as computational methods rely on the corresponding predicted tryptic digestion in which every theoretical tryptic site is cleaved *in silico* and compared to the spectral profile from the LC-MS/MS instrument for sequence identification. Nevertheless, even with the best of the commercial sequencing grade Trypsins, about 10-20% of the peptides observed are either semi-tryptic or non-specifically cleaved[6]. Notwithstanding this, many proteomic identifications are tolerant to a small number of mis-cleavages[7]. Such tolerance for one mis-cleavage was even allowable in a quantitative label-free analysis of tuberculosis vs control sera[11].

Missed Cleavage Comparison of 4 hour & Overnight digestion times

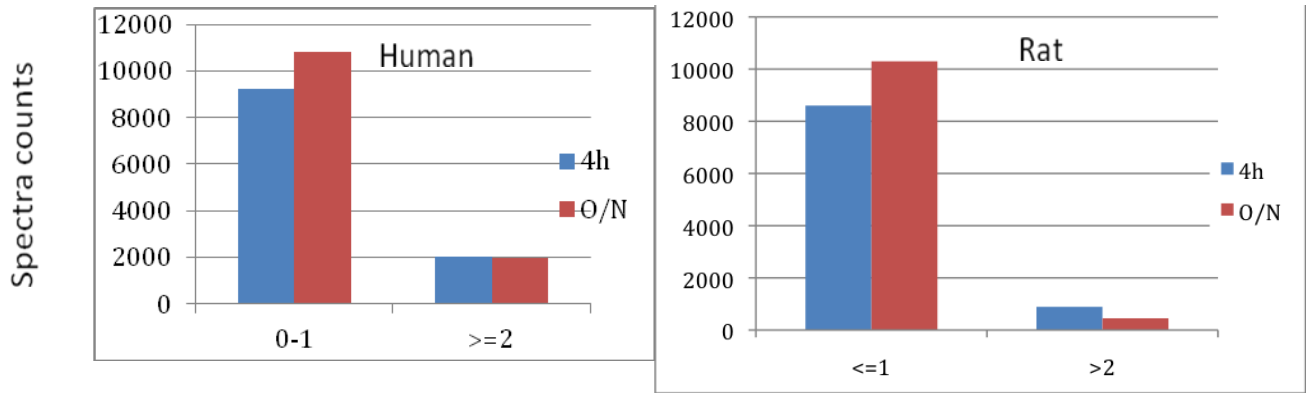


Figure 3. Complete digest means the peptide satisfies all of the following criteria:
1: No miss-cut (no K or R in the middle of the sequence unless followed by P)
2. the peptide is either protein N-terminal peptide or the residue immediately before the 1st aa is K or R
3. The peptide is either protein C-terminal peptide or is ended with K or R.

Our results demonstrate high efficiency digestions with minimal mis-cleavages, even at short digestion times, Figure 3. So while longer digestion times are required for some proteins, such time may contribute to a higher non-specific proteolytic background from the more abundant proteins. Such proteolytic background can mask the signal from the tryptic peptides produced from the lower abundance proteins[1]. So in addition to supporting higher throughput workflows with short digestion times, discovery applications can benefit by protein identifications from more than one digestion time, in our case, from short (4 hours) and overnight digestions. As can be seen from the Venn Diagrams, many protein identifications overlap, while others remain in one or the other digest time population. This suggests that one consider digested peptides as being generated from three populations of observable proteins:

Those proteins that are digested efficiently at short digestion times, but to which some peptides may be prone to non-specific digestion at long digestion times, or otherwise be masked by high proteolytic background.

Those proteins that are digested efficiently even at short digestion times, but whose observation is not negatively affected by long digestions.

Those proteins that require overnight digestion for proper identification.

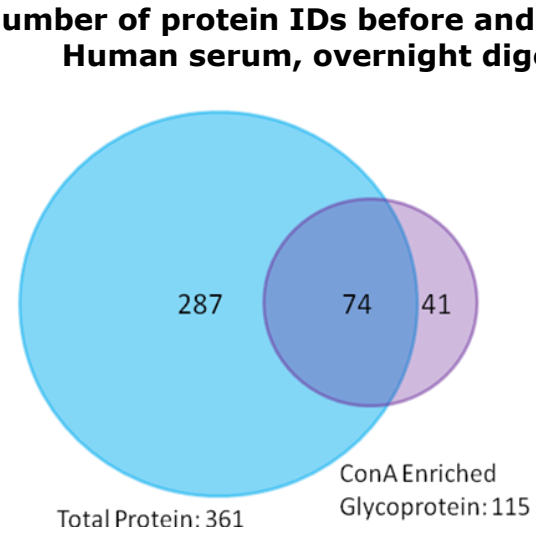


Figure 4. After proteolytic Trypsin digestion, the resultant peptides were separated by immobilized ConA lectin. The bound fraction was de-glycosylated with PNGase, and then analyzed by LC-MS/MS. Only the overnight digest from Human serum was analyzed in this manner.

Lets consider how these results impact the proteomic application. For discovery, more proteins will be identified when a variety of on-bead digest conditions are prepared. In like manner, for quantitative proteomics, exploratory studies can establish optimal conditions for on-bead digestion assessing the level of specific limit peptides of interest. Such limit peptides can represent the parent protein to be quantified. Brownridge and Beynon have reported such a similar strategy, to evaluate a time course of proteolysis and evaluate its influence on peptide generation and subsequent protein quantifications[7].

To consider the efficiency of proteolysis for the observation of glycoproteins, as a secondary step we chose immobilized ConA to enrich for glycopeptides, followed by enzymatic cleavage of the glyco-bond, resulting in peptides specific to the glycoprotein fraction. With such enrichment we observed 36% more glycoproteins then were identified in the initial total protein analyses, Figure 4. This suggests that there are many glycopeptides generated during on-bead proteolysis, but are nevertheless either poorly represented by computational assignments, or are not fully resolved throughout the LC-MS cycle, when the carbohydrates remain attached to the peptide. This is an area for future investigation.

Conclusions

AlbuVoid™ selectively voids albumin and enriches the low abundance proteome providing a suitable alternative to other depletion strategies; in a previous report, 20-30% more peptide identifications were obtained when compared to immuno-affinity depletion[4]. Our combined method of low abundance enrichment integrated with on-bead digestion offers many workflow advantages:

- Unique chemically derived beads, it is a consumable, one-time use product with no potential for cross-contamination or reduced performance upon regeneration.
- Species agnostic; human, rat, mouse, goat, sheep, porcine and bovine sera have been tested
- No in-gel digests, no solution digests, no C18 desalting, more consistent, reproducible results
- Compatibility with quantitative label (i.e., iTRAQ) and label-free LC-MS methods

The data reported here supports the importance of evaluating different digestion conditions as such conditions can impact both missed cleavages, and non-specific cleavages. When non-specific cleavages occur, higher proteolytic background can obscure sequence-rich features leaving many proteins unaccounted for. For quantitative applications, the proper selection of limit peptides that are not subject to inconsistent or inefficient proteolysis will greatly minimize artifactual errors. Therefore, the speed and efficiency of AlbuVoid™ on-bead digestions can improve workflows and minimize many of the inconsistencies of proteolytic hydrolysis during the generation of serum or plasma peptides.

Nevertheless, as with all digestion procedures, there is not one ideal method, and care must be taken to choose the right digestion time and condition for the application, whether it be for total discovery identifications, and/or for label or label-free quantification of select proteins.

Acknowledgements
We thank Dr. Haiyan Zheng and her colleagues at the Rutgers Proteomics Center for help with experimental design and LC-MS data.

References

1. Picotti, P., Aebersold, R., Domon, B. The Implications of Proteolytic Background for Shotgun Proteomics. Mol & Cell Proteomics 6.9 (2007) 1589-1598. doi:10.1074/mcp.M700029-MCP200.

2. Chen R. et al. Glycoproteomics Analysis of Human Liver Tissue by Combination of Multiple Enzyme Digestion and Hydrazide Chemistry.J Proteome Res., 2009, 8 (2), pp 651–661.

3. Roy, S,, Kuruc, M., Patel, K., Ackloo, S., Nahnsen, S. New On-Bead Digestion Protocols Improve LC-MS Workflows Of Albumin Depleted Samples. Presented as a poster US HUPO Conference March 11,12, 2013. http://biotechsupportgroup.com/sites/default/files/New%20On-Bead%20Digestion%20Protocols%20Improve%20LC-MS%20Workflows%20Of%20Albumin%20Depleted%20Samples_0.pdf

4. Roy, S., Zheng, H., Zhao, C., Arpa, A., Kuruc, M. Improved proteomic enrichment and workflow strategies. Presented As A Poster At US HUPO Conference, 04/06/2014-04/09/2014 http://biotechsupportgroup.com/sites/default/files/Biotech%20Support%20Group%20Improved%20Proteomic%20Enrichment%20and%20Workflow%20Strategies%20US%20HUPO%202014.pdf

5. Siepen, J.A, E.J. Keevil, D. Knight, S.J. Hubbard, J. Proteome Res. 6 (2007) 399-408.

6. Burkhart, J.M., Schumbrutzki, C., Wortelkamp, S., Sickmann, A., Zahedi, R.P. Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of typsin quality on MS-based proteomics. J of Proteomics 75 (2012) 1454-1462. doi:10.1016/j.jprot.2011.11.016

7. Brownridge, P., Beynon, R.J. The importance of the digest: Proteolysis and absolute quantification in proteomics. Methods 54 (2011) 351-361.

8. Addona, T.A., et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. Nat. Biotechnol. (2009) 27, 633-685.

9. Kuzyk, M.A., Smith, D., Yang, J.C., Cross T.J., Jackson, A.M., Hardie, D.B., Anderson, N.L., and Borchers, C.H. (2009) Multiple reaction monitoring-based, multiplexed, absolute quantification of 45 proteins in human plasma. Mol.Cell. Proteomics 8, 1860-1877.

10. Shuford, C.M., Sederoff, R.R., Chiang, V.L., and Muddiman, D.C. (2012) Peptide production and decay rates affect the quantitative accuracy of protein cleavage isotope dilution mass spectrometry (PC-IDMS). Mol. Cell. Proteomics 11, 814-823.

11. Song, S.H., et al. Proteomic Profiling of Serum from Patients with Tuberculosis. Ann Lab Med. 2014 Sep;34(5):345-353. http://dx.doi.org/10.3343/alm.2014.34.5.345