

AlbuVoid™ Coupled to On-Bead Digestion - Tackling the Challenges of Serum Proteomics

Haiyan Zheng¹, Caifeng Zhao¹, Meiqian Qian¹, Swapan Roy², Absari Arpa², Amenah Soherwardy² and Matthew Kuruc^{2*}

¹Rutgers Center for Integrative Proteomics, Piscataway, NJ, USA

²Biotech Support Group LLC, Monmouth Junction, NJ, USA

Abstract

For cancer research, serum and plasma are especially attractive sample types as collection of blood is common, simple and only minimally invasive. Yet serum samples can offer unique challenges in LC-MS proteomic analyses. The two biggest challenges being: 1) the high abundance of Albumin accounting for about 50% of the total protein mass and, 2) proteolytic resistance, in large part due to substantial amount of glycoprotein, a modification that manifests proteolytic resistance. In this short report, we describe new methods using a surface/bead based product, AlbuVoid™, which depletes Albumin through a negative selection or avoidance strategy, retaining the vast amount of the remaining serum proteome on the bead. We then combine this novel enrichment, with a direct and seamless integration with Trypsin digestion, a method conventionally referred to as on-bead digestion. We evaluated the digestion time as a parameter to identify whether different sub-populations of peptides and proteins can be observed by LC-MS analyses. Using 2 different allotted digestion times - 4 hours, and overnight, each with a singular 3 hour gradient LC-MS run, between 400-500 total proteins were observed for both human and rat sera, with overlapping and distinct sub-populations observable at each digest time. These results support that the described methods gain efficiencies over other high abundance depletion and in-solution digestion workflows. We solicit that such workflows will minimize many of the inconsistencies of proteolytic hydrolysis for both discovery and quantitative serum proteomic applications.

Keywords: Serum biomarker; Quantitative On-Bead Digestion; Proteomics Isobaric labels; Label-free

Introduction

The discovery of cancer biomarkers that can personalize a treatment process has become an important research area in the proteomics field. For this, many proteomics approaches including two dimensional polyacrylamide gel electrophoresis (2DPAGE & 2D-DIGE), surface enhanced laser desorption/ionization time of flight (SELDI-ToF), protein arrays, and multidimensional protein identification technology (MudPIT) are being implemented in cancer research [1].

So for cancer research, serum and plasma are especially attractive sample types as collection of blood is common, simple and only minimally invasive. Yet these samples can offer unique challenges in differential proteomic analyses. The two biggest challenges being: 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 [2].

These challenges are being met by developments surrounding the AlbuVoid™ depletion product, coupled to on-bead digestion of the remaining serum proteome. For the first challenge, the AlbuVoid™ bead and method can efficiently deplete Albumin and Transferrin by a avoidance strategy, enriching the remaining low abundance proteins on the bead. The second challenge - proteolytic resistance, is in large part due to substantial amounts of glycoprotein, such modification manifesting proteolytic resistance. It is suspected that the large size of glycan groups can block Trypsin access, resulting in missed cleavages not suitable to computational data-mining [3]. In this report we demonstrate that combining AlbuVoid™ enrichment followed by subsequent on-bead digestion can effectively identify low abundance peptides, proteins and glycoproteins from serum. This combined

product and method has now been commercialized under the product name **AlbuVoid™ LC-MS On- Bead**. Furthermore, for quantitative proteomics, we solicit that workflows that can greatly reduce one or more highly abundant proteins, along with short, efficient and consistent digestions will be highly desirable.

In the initial characterization of on-bead digestion coupled to AlbuVoid™ enrichment, we observed that on-bead Trypsin digestion was equivalent or better than the same proteins first eluted and then subject to in-solution digestion [4,5]. These results while promising, were nevertheless 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. This is because during early stage proteolysis, it is suspected that much of the protein remains bound to the bead, rather than becoming unbound which starts to occur at higher pH conditions. The destabilized higher order protein structure of bound protein is likely to improve Trypsin's endoprotease activity with access to the interior regions of the polypeptide chains. With proteins virtually coating the bead surface, Trypsin remains soluble, maintaining its optimal reactivity. 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.

***Corresponding author:** Matthew Kuruc, Biotech Support Group LLC, Monmouth Junction, NJ, USA, Tel: 732-230-3470; E-mail: mkuruc@biotechsupportgroup.com

Received July 14, 2015; **Accepted** September 26, 2015; **Published** September 30, 2015

Citation: Zheng H, Zhao C, Qian M, Roy S, Arpa A, et al. (2015) AlbuVoid™ Coupled to On-Bead Digestion - Tackling the Challenges of Serum Proteomics. J Proteomics Bioinform 8: 225-230. doi:10.4172/0974-276X.1000373

Copyright: © 2015 Zheng H, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Material and Methods

AlbuVoid™ Protocol for albumin depletion and low abundance serum protein enrichment

For these tests, 50 μ l of serum per prep was used. In bold are the AlbuVoid™ kit components.

1. Weigh out 25 mg of **AlbuVoid™** matrix in a spin-tube (0.45 μ SpinX centrifuge tube filter from Corning).
2. Add 125 μ l 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 μ l of **Binding Buffer AVBB**, and 50 μ l of the **Serum**. Centrifuge for 4 min. at 10,000 rpm to clarify serum. Add the clarified conditioned serum to the beads from step 2. Vortex for 10 min. and then centrifuge for 4 min. at 10,000 rpm.
5. Discard the albumin enriched filtrate (Flow-Through) fraction. The beads now contain the remaining serum proteome.
6. To the beads, add 250 μ l of **Wash Buffer AVWB, pH 7**. Vortex for 5 min and centrifuge for 4 minutes at 10,000 rpm. Discard the **Wash**.
7. Repeat Step-6. Reproducibility of the separation method is reported in Figure 1.

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

8. After the final wash steps from Step 7 from the enrichment, add 10 μ l 100mM DTT + 90 μ l 10mM HEPES, pH 7, vortex 10 min, incubate ½ hr at 60°C.
9. After cooling, add 20 μ l 200mM Iodoacetamide, and 80 μ l 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 μ l Sequencing-grade trypsin (0.4 μ g/ μ l, in 50mM acetic acid) + 60 μ l 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 μ l 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 250 μ l. Prepare to desired final concentration. Store at -80°C until LC-MS/MS.

Glyco-peptide enrichment with immobilized Con A

Briefly, we modified a protocol reported by Udea et al. [6]

1. The overnight digested human sample was solubilized in 80 μ l of 10% acetonitrile, 50mM NH_4HCO_3 (50 μ l serum) at RT for 2h
2. Take 80 μ l of 50% ConA 4B (GE Healthcare, Piscataway NJ) slurry
3. Wash with 160 μ l of water, spin, discard the supernatant
4. Repeat step 3 for two more times
5. Wash with 160 μ l of binding buffer (20mM Tris-HCl, 0.5M NaCl, 1mM MnCl_2 , 1mM CaCl_2 , pH7.4)
6. Repeat step 5 for four more times
7. Add sample from step 1 to the beads, incubate on a vortex at RT for 1h
8. Centrifuge at 25,000 g for 30sec, save the supernatant as flow-through
9. Wash with 160 μ l of washing buffer (10% acetonitrile, 50mM NH_4HCO_3), centrifuge at 25,000g for 30sec, save the supernatant to the same tube of flow-through
10. Repeat step 9 for four more times
11. Add 40 μ l of 0.5M mannoside to the beads, incubate on a vortex at RT for 30min. Centrifuge at 25,000g for 30sec, save supernatant to a new tube labeled as the "sample name" and "elution"

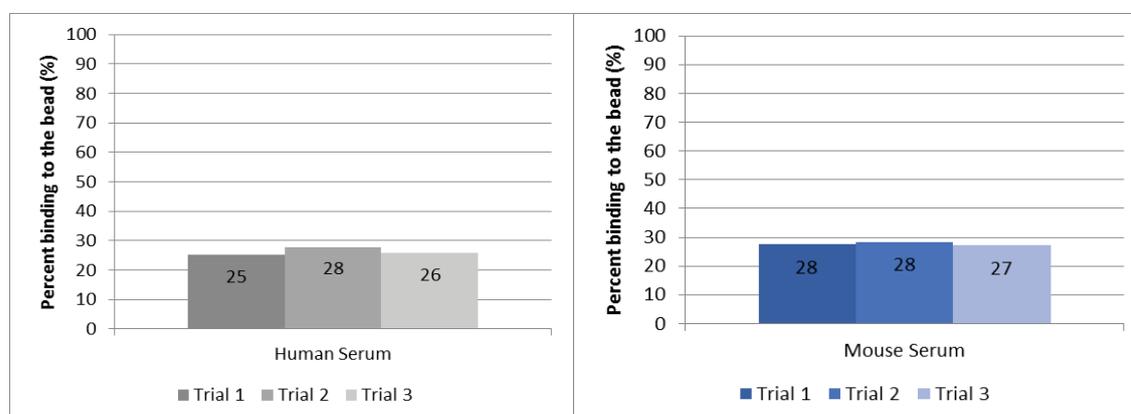


Figure 1: Reproducibility of the AlbuVoid™ protein separations and enrichment. Left: Human Serum was separated in triplicate parallel runs, total protein amount was measured in the untreated sample containing the binding buffer, and compared to the total protein amount in the combined flow-through (the unbound) and wash fractions. The percent protein bound to the bead is reported for all three trials. Right: Mouse Serum was used and reported in like fashion, demonstrating that the protein fractionation is not dependent on species.

12. Repeat step 11 one more time, and save supernatant to the same elution tube

13. Add 5 μ l of 500mM NH_4HCO_3 , 10 μ l of PNGaseF to the elution, incubate at 37 degree for 2h

14. Add 5 μ l of PNGaseF and incubate at 37 degree for overnight

15. Acidify with 27.5 μ l of 1% TFA and desalt with spec18.

a. Wet Varian tip A57203 (lot 0000091101) with 100 μ l of 0.1% TFA in acetonitrile, 50 g, 20 sec; b. Wash the tip with 100 μ l with 0.1% TFA in water, 50g, 20sec; c. Repeat step b for two more times; d. Add sample to the tip, 23g, 1min, save the filtrate as FT; e. Wash the tip with 100 μ l with 0.1% TFA in water, 50g, 20sec; f. Repeat step e for two more times; g. Elute with 100 μ l of 15% acetonitrile/0.1% TFA, 50g, 20sec; h. Elute with 100 μ l of 30% acetonitrile/0.1% TFA, 50g, 20sec; i. Elute with 100 μ l of 50% acetonitrile/0.1% TFA, 50g, 20sec; j. Elute with 100 μ l of 80% acetonitrile/0.1% TFA, 50g, 20sec; k. Pool the eluate fractions, dry in the vacuum.

16. Solubilized in 5% acetonitrile, 0.1% TFA and load 2.5 μ l serum equivalent enriched glycol- peptide to LC-MS/MS

LC-MS/MS analysis, peptide and 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 100 μ m \times 2cm trap packed with Magic C18AQ, 5 μ m 200 A (Michrom Bioresources Inc, Auburn, CA) and washed with Buffer A (0.2% formic acid) for 5 min with flow rate of 10 μ l/min. The trap was brought in-line with the homemade analytical column (Magic C18AQ, 3 μ m 200 A, 75 μ m \times 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 (thegpm.org) 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.

The searches were done using an in-House version of X! Tandem with protein filters set based on FPR supplied by the software. Human: valid $\log(e) < -0.4$, $\rho = 87$, FPR = 0.72% Rat: valid $\log(e) < -0.4$, $\rho = 87$, FPR = 0.72%. All proteins were counted regardless of number of peptides identified. The entire list of protein IDs and spectral counts can viewed in supplemental materials.

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 [7]. 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 [8]. 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 [2,8]. One can however evaluate indirectly, the number of proteins identifiable under different digestion conditions. Such an indirect measure is reported in Figure 2; 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 3.

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

Number of unique protein IDs after AlbuVoid™ Comparison of 4 hour & Overnight digestion times

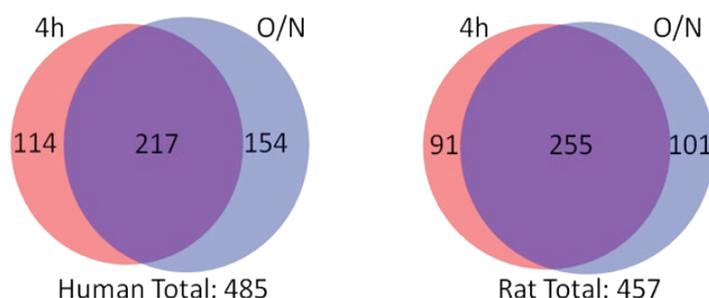


Figure 2: The total protein identifications from AlbuVoid™ are compared for human and rat sera, each at 2 digestion times, 4 hours and overnight. 217 human serum proteins were overlapping from each digestion condition, that is, they were identified regardless of the digestion time. Similarly, 255 rat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

be subject to further hydrolysis by the protease in use [9]. 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 [10-12].

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 [8]. Notwithstanding this, many proteomic identifications are tolerant to a small number of mis-cleavages [8]. For example, such tolerance for one mis-cleavage was allowable in a quantitative label-free analysis of tuberculosis vs control sera [13].

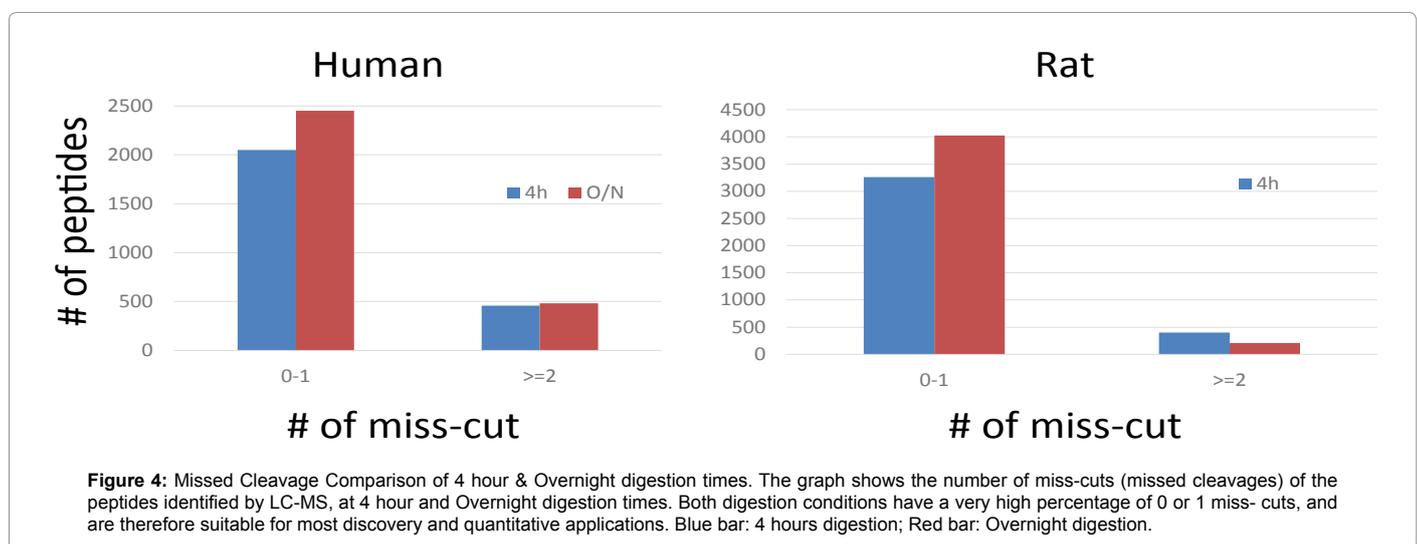
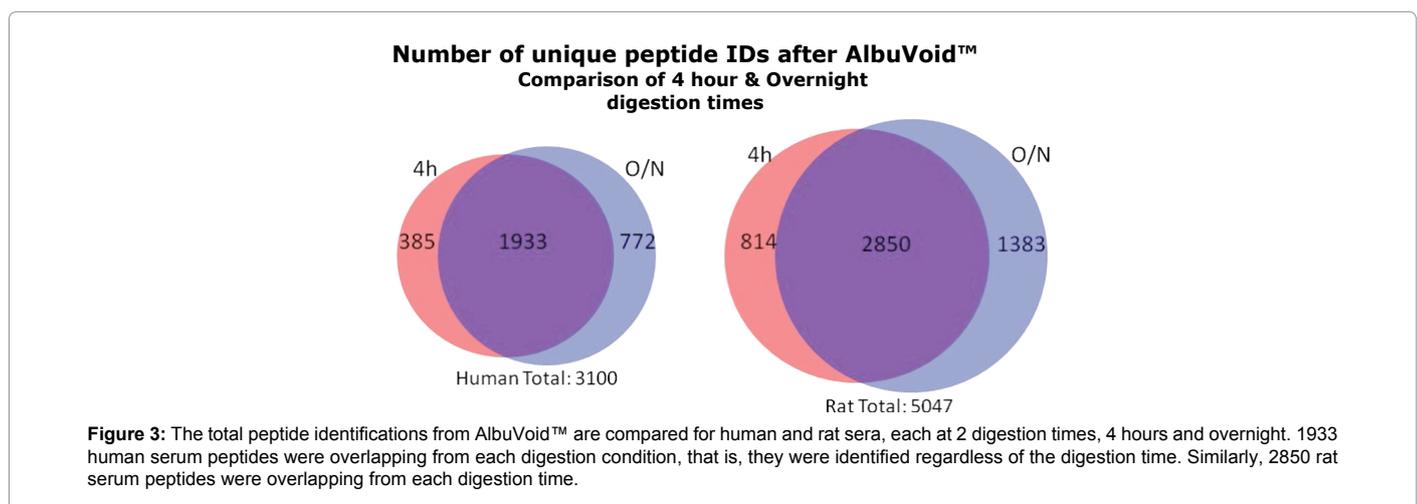
Our results demonstrate high efficiency digestions with minimal mis-cleavages, even at short digestion times, Figure 4. So while longer digestion times are required for some proteins, such time may contribute to a higher non-specific proteolytic background from the

more abundant proteins. Such proteolytic background can mask the signal from the tryptic peptides produced from the lower abundance proteins [2]. So in addition to supporting higher throughput workflows with short digestion times, discovery applications can benefit by protein identifications from both pools, those that come from short (4 hours in this example) 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.



We consider how these results impact the proteomic application. For discovery, more proteins will be identified when more than one on-bead digest times 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 [9].

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 than were identified in the initial total protein analyses, Figure 5. This suggests that there are many glycopeptides generated during on-bead proteolysis, but are nevertheless either poorly represented by computational assignments, do not ionize to the same extent, 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

As has been discussed by others in the field, there is always a tradeoff between effort, throughput and proteome coverage [14]. However it is accounted for, the utility cost of LC-MS instrument time is by far the largest singular expense within the overall workflow. Therefore, LC-MS productivity in terms of total proteins identified and quantified per allocated instrument service time, is an important metric to consider in the optimization of workflows. Because of the complications of integrating quantitative intensities from multiple peptide fractions, along with the associated costs and only marginal gains in proteome coverage, so far we have elected not to consider peptide level fractionation in our investigations using **AlbuVoid™ LC-MS On-Bead**. We recommend users of this product to consider peptide level fractionation and its relative benefits as appropriate or not for their particular investigation. The same is true for gains in productivity using more than one on-bead digestion time. Prospectively, this may be more productive than peptide level fractionation, and is an area for future research. This allows the **AlbuVoid™ LC-MS On-Bead** user

great latitude in designing a workflow that meets the needs of any particular investigation.

Nevertheless, we conclude that the **AlbuVoid™ LC-MS On-Bead** product and protocol supports efficient workflows as it greatly reduces the abundance of Albumin and Transferrin peptides, and enriches the low abundance proteome. It provides a suitable alternative or addition to other depletion strategies, and enrichment performance can be optimized for the application and goals of the investigation. For example, the product protocol specifies a starting volume range of 50-100 μ l. Within this range, the influences of volume per prep can then be optimized for performance, with consideration for total protein amounts, low abundance identifications or quantitative precision. Thus, this new 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
- Versatile, cost-effective workflows

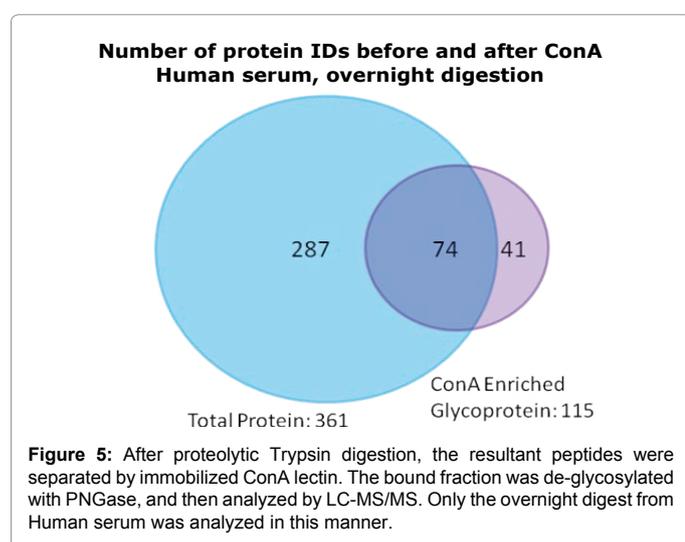
We suggest that it is practical and productive to evaluate different digestion times as this can impact both missed cleavages, and non-specific cleavages. When non-specific cleavages occur, higher proteolytic background can obscure sequence-rich features leaving many proteins unaccounted for. For quantitative applications, the speed and efficiency of **AlbuVoid™** coupled to on-bead digestions can minimize many of the inconsistencies of proteolytic hydrolysis during the generation of serum or plasma peptides. This will prove advantageous for both quantitative discovery (shotgun) as well as targeted SRM/MRM applications for biomarker diagnostics and drug development.

Acknowledgements

We thank Dr. Haiyan Zheng and her colleagues at the Rutgers Proteomics Center for help with experimental design and LC-MS data. We thank Dr. Xing Wang at Array Bridge Inc. (St. Louis, MO) for providing the serum samples.

References

1. Maurya P, Meleady P, Dowling P, Clynes M (2007) Review: Proteomic Approaches for Serum Biomarker Discovery in Cancer. *Anticancer Res* 27: 1247-1256.
2. Picotti P, Aebersold R, Domon B (2007) The Implications of Proteolytic Background for Shotgun Proteomics. *Mol Cell Proteomics* 6: 1589-1598.
3. Chen R, Jiang X, Sun D, Han G, Wang F, et al. (2009) Glycoproteomics Analysis of Human Liver Tissue by Combination of Multiple Enzyme Digestion and Hydrazide Chemistry. *J Proteome Res* 8: 651-661.
4. Roy S, Kuruc M, Patel K, Ackloo S, Nahnsen S (2013) New On-Bead Digestion Protocols Improve LC-MS Workflows Of Albumin Depleted Samples. US HUPO Conference March.
5. Roy S, Zheng H, Zhao C, Arpa A, Kuruc M (2014) Improved proteomic enrichment and workflow strategies. US HUPO Conference.
6. Ueda K, Takami S, Saichi N, Daigo Y, Ishikawa N, et al. (2010) Development of serum glycoproteomic profiling technique; simultaneous identification of glycosylation sites and site-specific quantification of glycan structure changes. *Mol Cell Proteomics* 9: 1819-1828.



7. Siepen JA, Keevil EJ, Knight D, Hubbard SJ (2007) Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics. *J Proteome Res* 6: 399-408.
8. Burkhart JM, Schumbrutzki C, Wortelkamp S, Sickmann A, Zahedi RP (2012) Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS- based proteomics. *J Proteomics* 75: 1454-1462.
9. Brownridge P, Beynon RJ (2011) The importance of the digest: Proteolysis and absolute quantification in proteomics. *Methods* 54: 351-361.
10. Addona TA, Abbatiello SE, Schilling B, Skates SJ, Mani DR, et al. (2009) Multi-site assessment of the precision and reproducibility of multiple reaction monitoring based measurements of proteins in plasma. *Nat Biotechnol* 27: 633-685.
11. Kuzyk MA, Smith D, Yang JC, Cross TJ, Jackson AM, et al. (2009) Multiple reaction monitoring-based, multiplexed, absolute quantification of 45 proteins in human plasma. *Mol Cell Proteomics* 8: 1860-1877.
12. Shuford CM, Sederoff RR, Chiang VL, Muddiman DC (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.
13. Song SH, Han M, Choi YS, Dan KS, Yang MG, et. al. (2014) Proteomic Profiling of Serum from Patients with Tuberculosis. *Ann Lab Med* 34: 345-354.
14. Pan J, Chen HQ, Sun YH, Zhang JH, Luo XY (2008) Comparative Proteomic Analysis of Non-small-cell Lung Cancer and Normal Controls Using Serum Label-Free Quantitative Shotgun Technology. *Lung* 186: 255-261.