

A Novel Multidimensional Protein Identification Technology Approach Combining Protein Size Exclusion Prefractionation, Peptide Zwitterion–Ion Hydrophilic Interaction Chromatography, and Nano-Ultraperformance RP Chromatography/nESI-MS² for the in-Depth Analysis of the Serum Proteome and Phosphoproteome: Application to Clinical Sera Derived from Humans with Benign Prostate Hyperplasia

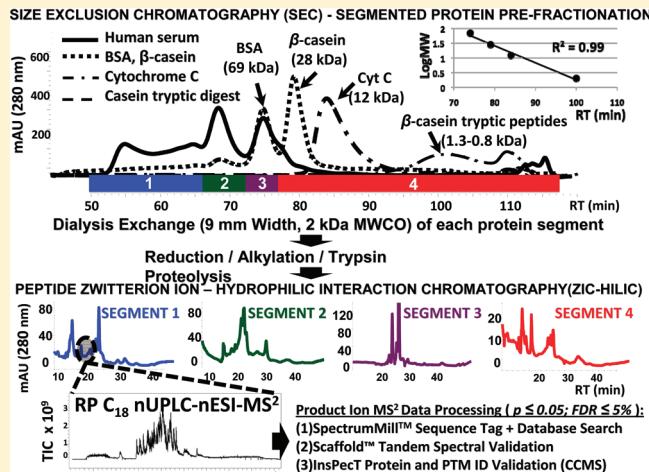
Spiros D. Garbis,^{*†} Theodoros I. Roumeliotis,[†] Stavros I. Tyritzis,[‡] Kostas M. Zorbas,[†] Kitty Pavlakis,[‡] and Constantinos A. Constantinides[‡]

[†]Center for Basic Research, Division of Biotechnology, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

[‡]Department of Urology, Athens University Medical School, “LAIKO” Hospital, Athens, Greece

 Supporting Information

ABSTRACT: The current proof-of-principle study was aimed toward development of a novel multidimensional protein identification technology (MudPIT) approach for the in-depth proteome analysis of human serum derived from patients with benign prostate hyperplasia (BPH) using rational chromatographic design principles. This study constituted an extension of our published work relating to the identification and relative quantification of potential clinical biomarkers in BPH and prostate cancer (PCa) tissue specimens. The proposed MudPIT approach encompassed the use of three distinct yet complementary liquid chromatographic chemistries. High-pressure size-exclusion chromatography (SEC) was used for the prefractionation of serum proteins followed by their dialysis exchange and solution phase trypsin proteolysis. The tryptic peptides were then subjected to offline zwitterion–ion hydrophilic interaction chromatography (ZIC-HILIC) fractionation followed by their online analysis with reversed-phase nano-ultraperformance chromatography (RP-nUPLC) hyphenated to nanoelectrospray ionization-tandem mass spectrometry using an ion trap mass analyzer. For the spectral processing, the sequential use of the SpectrumMill, Scaffold, and InsPecT software tools was applied for the tryptic peptide product ion MS² spectral processing, false discovery rate (FDR) assessment, validation, and protein identification. This milestone serum analysis study allowed the confident identification of over 1955 proteins ($p \leq 0.05$; FDR $\leq 5\%$) with a broad spectrum of biological and physicochemical properties including secreted, tissue-specific proteins spanning approximately 12 orders of magnitude as they occur in their native abundance levels in the serum matrix. Also encompassed in this proteome was the confident identification of 375 phosphoproteins ($p \leq 0.05$; FDR $\leq 5\%$) with potential importance to cancer biology. To demonstrate the performance characteristics of this novel MudPIT approach, a comparison was made with the proteomes resulting from the immunodepletion of the high abundant albumin and IgG proteins with offline first dimensional tryptic peptide separation with both ZIC-HILIC and strong cation exchange (SCX) chromatography and their subsequent online RP-nUPLC–nESI-MS² analysis.



Mass spectrometry based proteomics has the potential to uncover novel candidate protein markers and to give insights to poorly understood mechanisms of disease, via the systematic investigation of expressed and tissue relevant proteins in less invasive clinical matrices such as plasma or serum.^{1–6} One of the several challenges of the serum and plasma proteomic

methods has involved the removal of highly abundant proteins (i.e., albumin, IgGs, etc.) for the in-depth analysis of the less

Received: April 10, 2010

Accepted: December 3, 2010

Published: December 21, 2010

abundant proteins, where potential biomarkers can be revealed.^{7–10} However, their removal may have also resulted in the coremoval of a significant percentage of the less abundant proteins due to their propensity to mutually interact on the basis of an ensemble of individually weak, but collectively strong, noncovalent electrostatic and hydrogen-bond forces.^{11–14} By the same token, the coanalysis of both high- and low-abundance proteins and their endogenously occurring cleavage products (serum degradome) may confer greater insight on serum biochemistry and cancer biology.^{13,15,16} Building on the success of the multidimensional protein identification technology (MudPIT) in analyzing complex biological samples including serum,^{17–19} the study aimed in the design and application of a novel MudPIT approach for the bottom-up proteome analysis of a wide spectrum of both low- and high-abundance proteins in clinical sera derived from patients with benign prostate hyperplasia (BPH). This study constituted a continuation of the BPH and prostate cancer (PCa) tissue proteomic study reported by our group.²⁰ Briefly, this proof-of-principle study involved the use of a novel MudPIT approach to the analysis of a pooled BPH serum sample ($N = 25$) incorporating high-pressure size-exclusion chromatography (SEC) for the prefractionation of serum proteins followed by their dialysis exchange and solution phase trypsin proteolysis. The tryptic peptides were then subjected to offline zwitterion–ion hydrophilic interaction chromatography (ZIC-HILIC) fractionation followed by their online analysis with reversed-phase nano-ultraprecision chromatography (RP-nUPLC) hyphenated to nanoelectrospray ionization-tandem mass spectrometry using an ion trap mass analyzer. For the spectral processing, the sequential use of the SpectrumMill, Scaffold, and InsPecT software tools was applied for the tryptic peptide product ion MS² spectral processing, false discovery rate (FDR) assessment, validation, and protein identification.

Tissue proteomics has been pursued as a logical first step for the novel discovery of cancer markers given that tumor-derived proteins exist in higher concentrations due to their more direct proximity to cancer cells.⁶ However, it is not well understood how protein expression in tissues reflect measurable levels in the serum to monitor their pathophysiological status.^{5,6} As such, this single executed pilot analysis study of a pooled BPH clinical serum specimen, serving as a proof-positive specimen set, provided the forum to examine the utility of our proposed MudPIT approach in capturing proteins expressed in their respective BPH tissue proteome reported by the authors.²⁰ To further demonstrate the performance characteristics of our proposed MudPIT approach, a systematic comparison was made with other approaches involving the immunodepletion of the albumin and IgG proteins, given their highest abundance and predisposition to interact with other serum proteins, followed by the trypsin proteolysis of the depleted protein extract and subsequent offline first dimensional tryptic peptide separation with both ZIC-HILIC and strong cation exchange (SCX) chromatography and their subsequent online RP-nUPLC–nESI-MS² analysis.

MATERIALS AND METHODS

Serum Sample Processing. A total of 25 fasting level serum specimens were obtained from patients with a history of BPH. Bioethics, patient characteristics, clinical inclusion and exclusion criteria, and specimen procurement procedures have been previously described.²⁰ Mean age \pm SD of the patients was 70 ± 6 years (range 49–80) and mean PSA \pm SD was 6.0 ± 3.9 ng/mL (range 1.8–17.7).

The collection and handling of the serum from the clinical blood specimens was applied in accordance to the recommendations of the Standard Operating Procedure Integration Working Group (SOPIWG).²¹ The serum specimens were thawed and vortex mixed for about 2 min. Individual 200 μ L aliquots from each of the 25 specimens were pooled together as the serum stock solution with an estimated protein concentration of $37 \pm 5 \mu\text{g}/\mu\text{L}$, as determined with the Bradford assay.

MudPIT Approach Definitions. Three different bottom-up MudPIT approaches were used for this study. The MudPIT approach A involved size-exclusion chromatography (SEC) for the prefractionation of proteins as segments and their dialysis exchange and trypsin proteolysis, followed by the offline first dimensional fractionation of the tryptic peptides with ZIC-HILIC separation (Figure 1A). The MudPIT approach B incorporated the immunodepletion of the albumin/IgG proteins and the trypsin proteolysis of the affinity purified protein extract, followed by the offline first dimensional fractionation of the tryptic peptides with ZIC-HILIC separation (Figure 2A). The MudPIT approach C incorporated the immunodepletion of the albumin/IgG proteins and the trypsin proteolysis of affinity purified protein extracts, followed by the offline first dimensional fractionation of the tryptic peptides with SCX separation (Figure 2A). All three MudPIT approaches utilized the online RP C₁₈ nUPLC-nESI ion trap tandem MS² system for the separation and analysis of the tryptic peptides (Figures 1A and 2A). The data processing was common for all three components. The experimental details of each component are described below. All offline chromatographic experiments were performed on the model P680 HPLC pump equipped with PDA-100 photodiode array detector and thermostated column compartment (Dionex, Germering, Germany).

SEC Protein Prefractionation. A 20 μ L volume aliquoted from the serum stock solution (equivalent of 1.48 mg total serum protein) was mixed with 180 μ L size SEC mobile phase to the final 200 μ L injection volume and passed through a 0.22 μ m PVDF filter (Millipore Corp., Billerica, MA). The SEC mobile phase composition of 10% methanol (Sigma, 99.9%), 50 mM KH₂PO₄ (Fluka, 99.5%), 10 mM Tris-HCl (Bio-Rad), 50 mM ammonium acetate (Fluka, 99%), 0.3 M NaCl (Carlo Erba, 99.9%), and 6 M guanidine HCl (Gibco BRL) at pH 5.3 ensured the thorough protein denaturation and enzyme inhibition during the sample handling process. The SEC prefractionation was performed isocratically with two serially connected 8.0 mm i.d. \times 300 mm L Shodex KW-804 SEC columns at a flow rate of 0.2 mL/min at 15 °C and monitored at 280 nm retrofitted with suitable guard column. The SEC experimental conditions such as mobile phase composition and flow rate, column temperature, and total protein amount injected were iteratively selected to achieve optimum separation conditions for a set of protein standards, namely, bovine serum albumin (69 kDa), β -casein (28 kDa), cytochrome *c* (12 kDa), and β -casein tryptic peptides (1.3–0.8 kDa), resulting in the linear chromatographic response between the logMW and retention time for these species as shown in Figure 1A. The SEC separation was reproducibly applied five times for the prefractionation of a total of 7.4 mg of serum protein.

A total of four distinct SEC segments were selected for serial fraction collection on the basis of the UV-signal response at 280 nm so as to reduce chromatographic peak undersampling (Figure 1). The serially pooled SEC fractions were then lyophilized to dryness and reconstituted with 15 mL of aqueous 100 mM ammonium bicarbonate. Buffer exchange to the resulting samples was performed

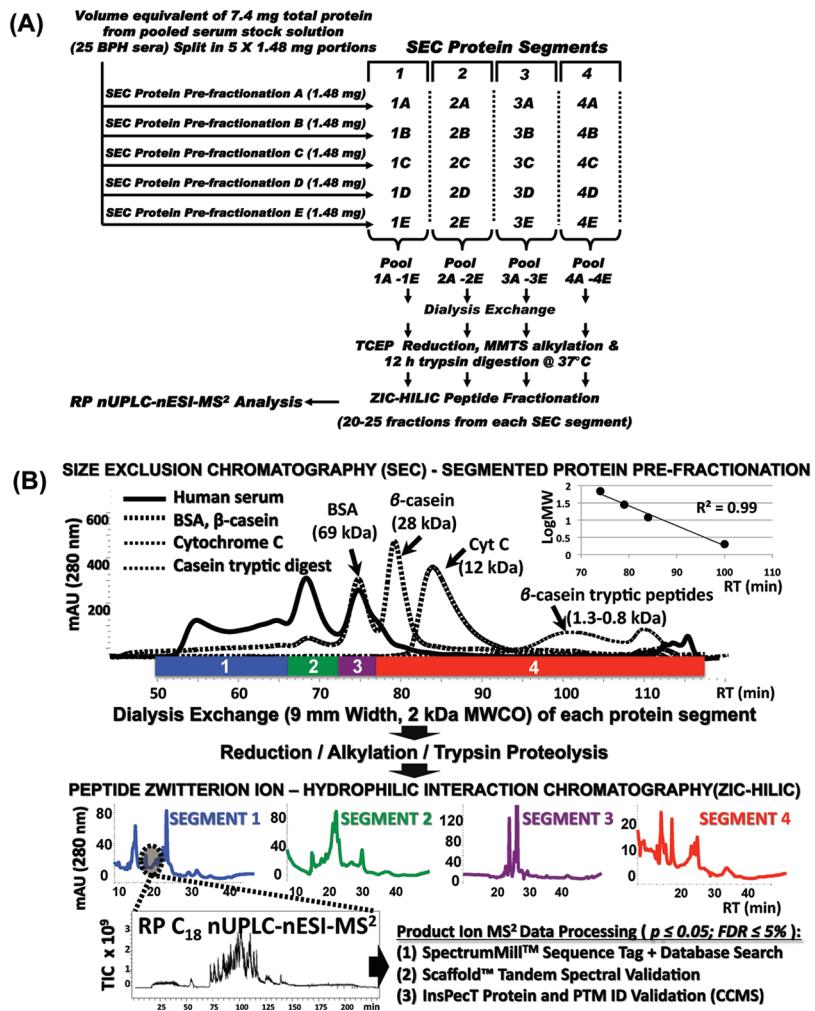


Figure 1. (A) Depiction of the workflow used for the MudPIT approach A in the analysis of the BPH sera samples. (B) Top HPLC trace: A representative size exclusion chromatography (SEC) trace of a pooled serum sample. Calibrant SEC traces are also shown along with their log MW vs RT (min) linear response curve. Middle HPLC traces: Post-SEC sample treatment and ZIC HILIC tryptic peptide traces in concordance with the SEC protein segment. The ZIC-HILIC peptide fractionation was performed in a peak-dependent manner. Bottom HPLC trace: Each lyophilized peptide fraction was reconstituted in MP and individually analyzed with RP C₁₈ nUPLC-nESI-MS² analysis. The resulting product ion MS² peptide spectra were processed with Mascot, with Scaffold validation and SpectrumMill software programs.

using 30 cm dialysis exchange tubing (9 mm width, 2 kDa MWCO, Sigma). The membranes were immersed into a beaker containing 3 L aqueous 100 mM ammonium bicarbonate and were kept overnight at 4 °C with constant agitating for a total of three volume exchanges. The desalted protein solutions were collected into 50 mL tubes at 4 °C. The desalted fractions were concentrated to a final volume of 500 μL.

Immunodepletion of Albumin and IgG Proteins. Immunodepletion of 400 μL (equivalent of about 14.8 mg total protein) sampled from the serum stock solution was performed with the ProteoSeek antibody-based albumin/IgG spin column removal kit (Thermo Scientific Pierce, Deerfield, IL) in strict accordance to manufacturer instructions. The resulting affinity processed protein extract was dissolved in a solution containing 0.5 M triethylammonium bicarbonate and 0.1% SDS with 30 min of heating at 60 °C and tip sonication. The total protein amount after albumin/IgG depletion was measured at 2.4 mg with the Bradford assay and was split in two equal amounts for their subsequent ZIC-HILIC and SCX peptide fractionation, respectively (Figure 2A).

In Solution Trypsin Digestion of Serum Proteins. The protein extracts from the SEC and immunodepleted steps were

enzymatically digested overnight at 37 °C using sequencing grade trypsin (Roche Diagnostics, GmbH, Mannheim, Germany) at a ratio of 1:30 (w/w, trypsin to protein) as instructed by the manufacturer using tris(2-carboxyethyl)phosphine (TCEP) as the reducing agent and, methyl methanethiosulfonate (MMTS) as the alkylation agent.

ZIC-HILIC Peptide Fractionation. For the ZIC-HILIC peptide fractionation, the SeQuant ZIC-pHILIC column (150 × 4.6 mm, 5 μm) was used with mobile phase (A) 100% ACN, 0.1% formic acid, 15 mM ammonium formate and mobile phase (B) 100% H₂O, 0.1% formic acid, 15 mM ammonium formate, pH = 3. The peptide tryptic digests from the SEC and immunodepletion steps were SpeedVac concentrated to 50 μL, diluted with 500 μL of mobile phases A and B at a 9:1 ratio, and filtered using 0.22 μm Millipore PVDF filters (Billerica, MA). The separation method was as follows: 5 min isocratic 10% B, 40 min gradient up to 40% B, 20 min gradient up to 90% B, 10 min isocratic 90% B, and 10 min gradient up to 10% B. The injection volume was 500 μL (full-loop), flow rate was 0.4 mL/min, column temperature was at 30 °C, and the signal UV detection was monitored at 280 nm.

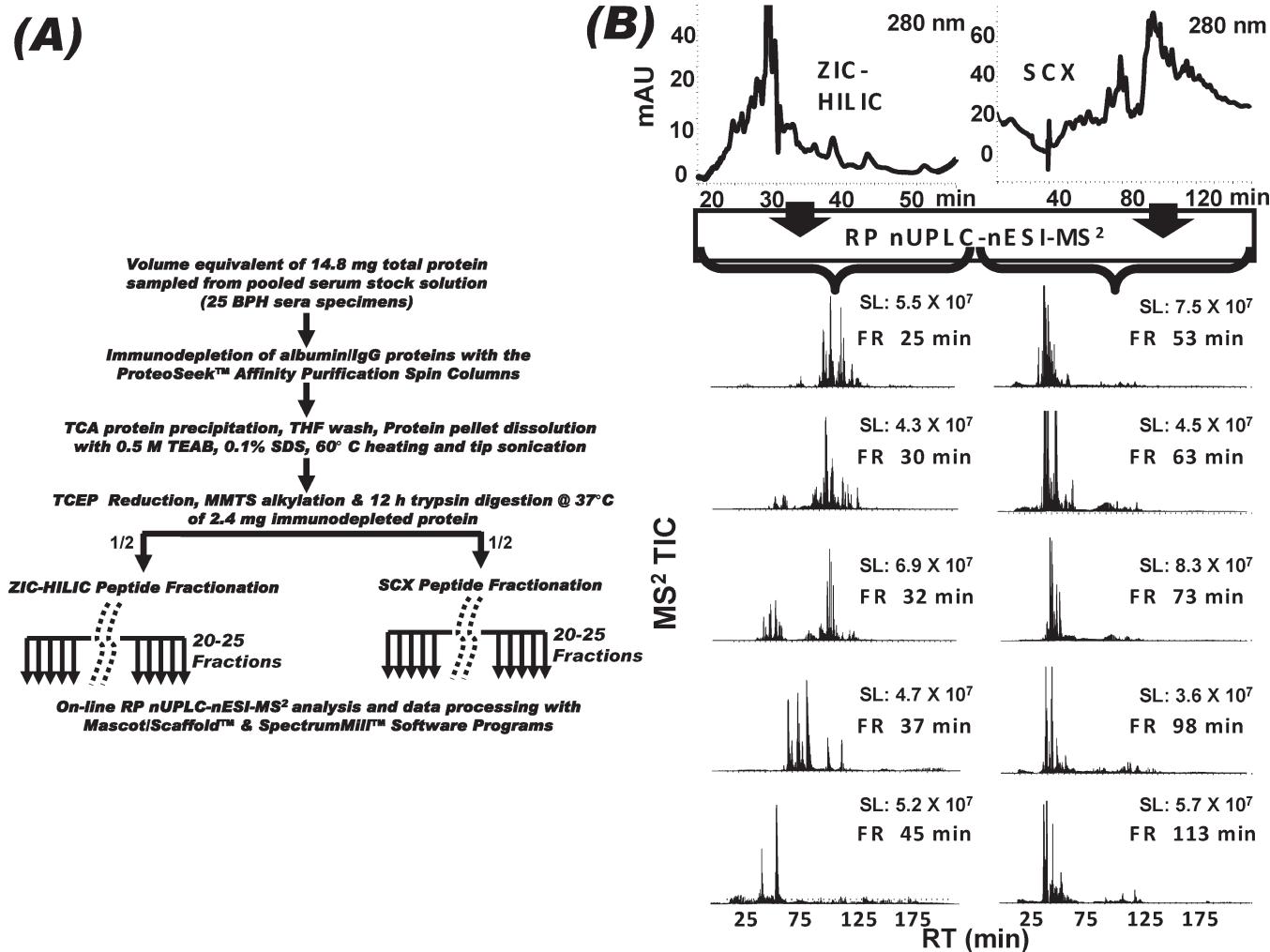


Figure 2. (A) Depiction of the workflow used for the MudPIT approaches B and C in the analysis of the BPH sera samples. (B) Left HPLC traces: The ZIC-HILIC peptide fractionation was performed in a peak-dependent manner, reconstituted in MP and analyzed with RP C₁₈ nUPLC-nESI-MS² analysis as shown by representative MS² TIC signal response vs RT of fraction. Right HPLC traces: The SCX peptide fractionation was performed in a peak-dependent manner, reconstituted in MP and analyzed with RP C₁₈ nUPLC-nESI-MS² analysis, as shown by representative MS² TIC signal response vs RT of the fraction. The resulting product ion MS² peptide spectra were processed in accordance to the procedure used for the MudPIT approach A.

Peptide fraction collection was conducted in a peak signal dependent mode. Approximately, 20–25 fractions were collected from each ZIC-HILIC experiment and lyophilized to dryness and immediately stored at –80 °C under a blanket of argon until LC–MS analysis.

SCX Peptide Fractionation. For the strong cation exchange (SCX) fractionation, a polysulfethyl A (1.0 mm × 200 mm; 5 μ particle; 300 Å pore, the Nest Group) column was used with mobile phases (A) aqueous 20 mM ammonium formate adjusted to pH 2.7 with formic acid and (B) 20% acetonitrile, 1.4 M ammonium formate, adjusted to pH 2.7 with formic acid. The peptide tryptic digests from the immunodepletion steps were SpeedVac concentrated to 50 μL, diluted with 500 μL of mobile phases A and B at a 9:1 ratio, and filtrated using 0.22 μm Millipore PVDF filters (Billerica, MA). The separation method was as follows: 55 min linear gradient from 0 to 30% B, 105 min linear gradient to 80% B, then 30 min isocratic at 90% B and 10 min gradient [back] down to 0% B. The injection volume was 200 μL (full-loop), flow rate was 0.2 mL/min, column temperature was at 30 °C, and the signal UV detection was monitored at 280 nm. Peptide fraction collection was conducted in a peak signal

dependent mode. Approximately, 20–25 fractions were collected from each SCX experiment and lyophilized to dryness and immediately stored at –80 °C under a blanket of argon until LC–MS analysis.

LC–MS² Analysis. All LC–MS² experiments were performed on a quadrupole ion trap mass analyzer (Agilent Technologies, model 6330) retrofitted to a 1200 nano-HPLC system equipped with a micro well plate autosampler and nanoelectrospray ionization source (Agilent Technologies, Karlsruhe, Germany). Individual lyophilized ZIC-HILIC fractions were freshly reconstituted in 30 μL mobile phase A (3% ACN, 0.1% formic acid). A 3 μL volume of the resulting sample solution was injected and then eluted at 150 nL/min onto a 0.075 × 150 mm reversed-phase capillary ultraperformance liquid chromatography column (RP-nUPLC, Zorbax C₁₈, 300 Å pore, 1.8 μm particle, Agilent Technologies, Karlsruhe, Germany) retrofitted onto the nanoelectrospray (nESI) source and connected to a 1P-4P coated, 8 μm tip × 360 μm o.d. × 75 μm i.d. PicoTip nESI emitter (New Objective, Woburn, MA). A 0.30 × 5 mm reversed-phase guard column (Zorbax C₁₈, 300 Å pore, 5.0 μm particle,

Agilent Technologies, Karlsruhe, Germany) was connected between the pump outlet tubing and the capillary column.

The RP-nUPLC separation started with an isocratic composition of 100% mobile phase A composed of 3% ACN, 0.1% formic acid for 40 min followed by gradient for 10 min up to 15% mobile phase B composed of 97% ACN, 0.1% formic acid, 130 min gradient up to 60%, 20 min isocratic at 60%, and for 30 min re-equilibration.

The ion trap analysis parameters were as follows: Selection of the four most abundant multiply charged precursors above absolute intensity threshold of 10^4 during the MS survey scan within a mass range of 300–2200 amu with 150 ms accumulation time and enhanced resolution. All MS² experiments were conducted with ultra-scan resolution with an on-flight adjustment of the collision energy using the smart fragmentation operation. Precursors were excluded after the acquisition of two spectra and released again after 0.2 min. The MS² window was set at 100–2200 amu. The analyses were applied for both sets of ZIC-HILIC fractions originating from the SEC and immunodepletion-based approaches.

Data Processing. All RP-nUPLC–nESI-MS² data batches were initially subjected to data extraction/peak picking/deisotoping and searching using the SpectrumMill software program (Agilent Technologies, Karlsruhe, Germany). Specifically, the data extraction and signal processing parameters were tailored for the quadrupole ion trap mass analyzer that included a signal-to-noise threshold for the precursor MS peak ≥ 25 and similarity merging for peaks with mass difference ± 1.4 Da eluted within a 15 s time window.

The SpectrumMill results were further processed with the Scaffold²² software tool (version 3.0, Proteome Software, Portland, OR) in order to extract all spectra in .mgf format. For the final tryptic peptide/protein result, the Scaffold processed .mgf data files were processed by the InsPecT²³ software tool (of the Center for Computational Mass Spectrometry at <http://proteomics.ucsd.edu/LiveSearch/>) that included the contaminant list option and oxidation and STY phosphorylation as the permitted PTMs. The specific data processing parameters chosen were as follows: instrument, ESI-ION-TRAP; protease, trypsin; parent mass tolerance, 2.5 Da; ion tolerance, 0.7 Da; maximum number of PTM allowed per peptide, 2; PTM, delta mass: +15.999 400 Da, affected: MW, type: opt; PTM, delta mass: +79.979 900 Da, affected: STY, type: opt; sequence, refs eq.human.2010.1001 sequence, common contaminants; results were filtered at a spectrum-level p-value of 0.05, measured by hits to the decoy database.

RESULTS AND DISCUSSION

As part of a cancer biomarker discovery program, the systematic proteomic study of the target organ system is a logical initial step given the high abundance levels of expressed proteins that can be traced to a particular pathological condition such as well-defined prostate cancer conditions.^{6,24–26} As a means to develop a noninvasive protocol to extend the utility of these potential biomarkers, the systematic proteomic study of the blood serum, to which target tissue proteins can be secreted or shedded becomes a potentially useful, albeit challenging, scientific undertaking. The factors that were accounted for in the rational and in-depth MudPIT analysis approach of the serum proteome were multi-fold and are discussed separately.

Table 1 lists the number of proteins and phosphoproteins identified with the sequential use of the SpectrumMill, Scaffold, and InsPecT software tools ($p < 0.05$; FDR ≤ 0.05) resulting

from the application of the three MudPIT proteomic approaches (detailed proteomes are available in the Supporting Information). Table 1 also lists secreted proteins, as classified by the Uniprot convention, not available in the Human Plasma Proteome Database (www.plasmaproteomedatabase.org) and, as such, may signify novel MS-based proteomic findings. Also listed in Table 1 are the low-abundance serum proteins identified with MudPIT approach A, namely, interleukin-5 (native concentration level at pg/mL per bioassay technique)¹ and cyclin-F (native concentration level at fg/mL per bioassay technique).²⁷ These concentration levels are approximately 12 orders of magnitude less than the highest abundant serum albumin protein occurring at the mg/mL concentration level.¹⁷ Such a wide analytical detection range is currently beyond the detection range possible with the MudPIT approaches utilizing the latest hybrid triple-quadrupole MRM-based, high-capacity trap (quadrupole/linear ion traps, Orbitrap) or QqTOF MS/MS platforms.^{1–6} Hyphenating such MS platforms with the orthogonal chromatographic strategy used in the MudPIT approach A imparts a more effective parsing, purification, and enrichment of the tryptic peptides, thus increasing their individual mass density (higher peptide signal intensity per chromatographic retention time window) at reduced coeluting peptide amounts arriving at the nanoelectrospray ionization source per unit time. This resulted in the generation of more information-rich tandem mass spectra at improved S/N ratios. Consequently, the collective attributes of the MudPIT approach A, rather than the linear dynamic range intrinsic to its constituent techniques (i.e., the ~ 3.5 orders linear dynamic range of the ion trap LC–MS technique), resulted in the identification of proteins differing by approximately 12 orders concentration range in terms of their native abundance levels in the naturally occurring serum matrix (as measured with bioassay technique such as ELISA). From this perspective and given the lack of standardized reference materials to collectively interrogate the MudPIT analysis approach of complex biological specimens as a whole, the pooled human BPH serum specimen served as a proof-positive test sample in detecting known serum proteins spanning this ~ 12 order native concentration range. Additionally, proteins highly expressed in prostate tissue captured by the MudPIT approach A, as discussed below, provide further testament to its analysis utility.

Figure 3 A shows the theoretical distribution of the analyzed proteins in terms of their theoretical MW distribution vs their pI values. As a means to visualize the proteins identified as a function of hydrophobic–hydrophilic character, a virtual map distribution of their molecular weight against their grand average of hydropathy^{28,29} (Gravy) index is shown in Figure 3B. A positive Gravy index value reflects a more hydrophobic character relative to the negative Gravy index values reflecting a more hydrophilic character.^{28,29} Although the Gravy index sorting of the proteins was entirely based on their surrogate tryptic peptide composition, it does provide a good estimate on how our study method exhibited selectivity on the basis of the intrinsic differences in hydropathy and polarity for the serum proteins. Parts C and D of Figure 3 are illustrative Venn diagrams of the number of proteins and phosphoproteins identified by the three approaches, respectively. The more substantial in-depth proteome coverage resulting from the MudPIT approach A may stem from the coanalysis of the higher abundance albumin and IgG proteins with their lower abundance serum proteome counterparts. The SEC mobile phase contained 6 M guanidinium cation (pH 5.1), which acted as an effective protein denaturant, minimizing the electrostatic

Table 1. Number of Proteins and Phosphoproteins Identified at $p \leq 0.05$; FDR $\leq 5\%$ with the Sequential Use of the SpectrumMill, Scaffold, and InsPecT Software Tools and Identified Secreted Proteins^a

accession no.	MudPIT approach	A	B	C
	number of proteins identified ($p \leq 0.05$; FDR $\leq 5\%$)	1955	563	499
	number of phosphoproteins identified ($p \leq 0.05$; FDR $\leq 5\%$)	375	143	148
Identified Secreted Proteins (Uniprot Convention) Not Included in the HUPO Human Plasma Proteome Database				
gi 110735441	ADAM metallopeptidase with thrombospondin type 1 motif, 16	✓	—	—
gi 40806187	ADAM metallopeptidase with thrombospondin type 1 motif, 18	✓	✓	—
gi 154275767	ADAMTS-like 1	✓	—	—
gi 133930786	egf-like module containing, mucin-like, hormone receptor-like 3	✓	—	—
gi 11128015	interferon, alpha 6	✓	—	—
gi 28144901	interleukin 28B (interferon, lambda 3)	✓	—	—
gi 4505185	macrophage migration inhibitory factor (glycosylation-inhibiting factor)	✓	—	—
gi 116006955	neuregulin 1	✓	—	—
gi 226342931	podocan-like 1	✓	—	—
gi 13325075	quiescin Q6 sulfhydryl oxidase 1	✓	—	—
gi 45935371	serglycin	✓	—	—
gi 14141195	stromal cell-derived factor 2	✓	—	—
gi 56682966	transforming growth factor, beta receptor III	✓	—	—
gi 16936520	wingless-type MMTV integration site family, member 10A	✓	—	—
gi 15082261	wingless-type MMTV integration site family, member 9A	✓	—	—
Low-Abundance Serum Proteins at the Low pg/mL–fg/mL Native Concentration Level				
gi 118572588	cyclin f	✓	—	—
gi 10834994	interleukin-11	—	✓	—
gi 4504671	interleukin-5	✓	—	—

^a Listed also are the secreted proteins identified by at least one of the MudPIT approaches applied, as classified by the Uniprot convention, and not available in the Human Plasma Proteome Database (www.plasmaproteomedatabase.org), signifying novel MS-based proteomic observations. The interleukin-5 (native concentration level at pg/mL) and cyclin-F (native concentration level at fg/mL) proteins were detected only with the MudPIT proteomic approach A. The ✓ denotes a protein identification and the dashed line (—) denotes a null observation.

interaction and protein folding potential of the larger and more abundant proteins with the smaller and less abundant proteins, leading to an enhancement of analytical efficiency.^{12,30–32} The four SEC protein segments fractionated were chosen on the basis of the UV-signal response at 280 nm so as to reduce chromatographic peak undersampling and to minimize the coelution of the higher abundance with the lower abundance tryptic peptides originating from the SEC protein segments. The bottom-up MudPIT analysis of the SEC segments identified high MW proteins (>200 kDa) in the low MW SEC regions that indicated the presence of endogenous cleavage products for these high MW proteins reported in the literature to occur in serum as potential surrogate markers of cancer biology.^{13,15,16} This fact made it practically impossible to identify distinct subproteomes on the basis of their relative molecular weight distribution corresponding to each SEC protein segment chosen for this study. Expanding on this notion, a more comprehensive SEC fractionation scheme based on the detection of absolute molecular weight differences using the principles of refractive index,³³ dynamic light scattering,³⁴ and multiangle light scattering³⁵ is warranted. Such SEC configurations, when combined with more targeted affinity capture protocols (i.e., lectin glycoprotein capture chemistries), would also be more sensitive to molecular weight differences as a result of *in vivo* protein modifications (i.e., protein glycosylation).^{36–38} However, these approaches would have prohibitively increased the overall analysis time and complexity. The SEC step as used in the MudPIT

proteomic approach A served as a crude prefractionation step to minimize the potential loss of the lower abundant proteins from the affinity depletion of the highly interacting albumin and IgG proteins observed in the case of the MudPIT approaches B and C.

The second dimensional ZIC-HILIC chemistry used for the separation of the tryptic peptides for the MudPIT approaches A and B was based on the bonded zwitterionic sulfobetaine group acting as the interactive aqueous layer. In principle, ZIC-HILIC separates the zwitterionic peptide ions by passing a hydrophobic organic mobile phase across the aqueous stationary phase formed, causing these peptides to elute in order of increasing hydrophilic character. Multiple factors can accentuate this hydrophilic character, including the ionization potential, polarity index, and diffusion kinetics in an aqueous medium that collectively depend on peptide amino acid composition and size.^{39,40} The separation attributes of HILIC have been successfully exploited in the analysis of polar metabolites endogenous in plasma.⁴¹ It is interesting to note that the ZIC-HILIC separation behavior for the tryptic peptides changed as a response to their SEC protein segment of origin, as seen in Figure 1B. The variability in the ZIC-HILIC separation response resulted from expected differences in the hydrophilic and hydrophobic tryptic peptide content originating from the four separate SEC protein segments. The orthogonal response of the ZIC-HILIC fractionation step was also evident in the variable retention time behavior of the tryptic peptides during their reversed-phase nUPLC separation for both MudPIT approaches

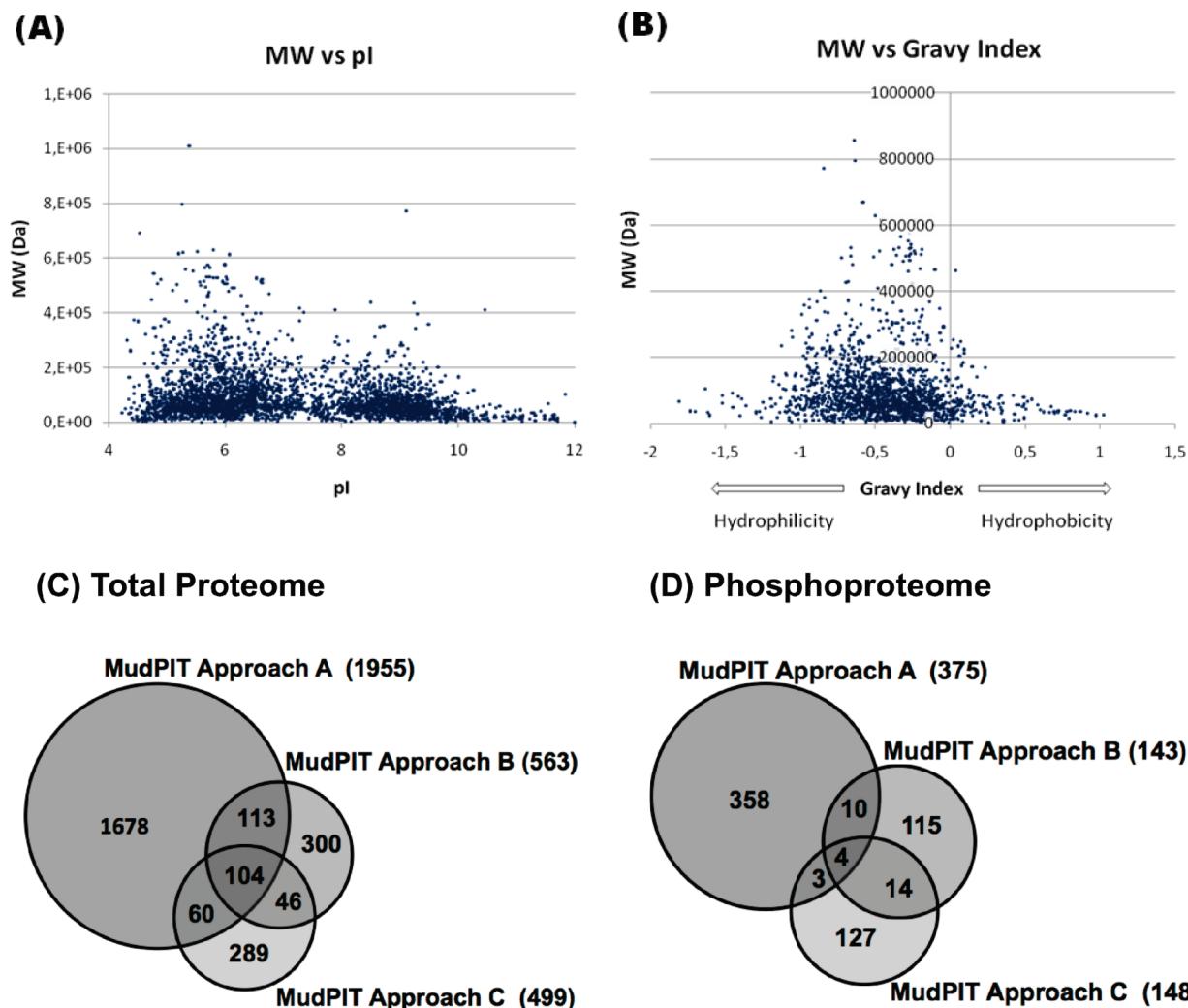


Figure 3. (A) Protein map distribution in accordance with their molecular weight (MW, kDa) vs isoelectric point (pI) and (B) their molecular weight (MW, kDa) vs Gravy index by the MudPIT approach A. (C) Venn diagram of the total number of proteins and (D) phosphoproteins as identified by the three MudPIT approaches. In all cases, the sequential use of the SpectrumMill \Rightarrow Scaffold \Rightarrow InsPecT software tools were used for the confident protein and phosphoprotein identifications ($p \leq 0.05$; FDR $\leq 5\%$).

A and B (Figure 2B). Such a trend allowed for a larger retention time window to be occupied, leading to reduced peptide coelution. Such a separation behavior was less pronounced when using SCX as the peptide fractionation step (Figure 2B) and was also observed in the prostate tissue study.²⁰ The nUPLC configuration (using 1.8 μm particles) for the reversed-phase separation was verified to be more effective than the nHPLC configuration (using 3.5 μm particles) with a standard six protein tryptic digest mix (Dionex) at a 10 fmol/ μL concentration, resulting in MS² product ion spectra with significantly improved ion scores under the same gradient conditions.

A significantly larger number of hydrophilic phosphopeptides and consequently more phosphoproteins were confidently identified with the MudPIT approach A relative to MudPIT approaches B and C (Table 1, Figure 3D, and Supporting Information). The ability to effectively detect phosphoproteins is of significant utility to cancer biology.^{6,14} To illustrate the sensitivity and selectivity of the MudPIT approach A, Figure 4 shows an example annotated tandem MS² spectrum of the +2 charged tryptic phosphopeptide T.ADSphosGEGDFLAEGGGV.R with precursor mass at 730.9 Da ($\Delta m = -0.8$ Da) traceable to the fibrinogen alpha chain

isoform alpha preproprotein (accession no.: gi|11761629|ref|NP_068657,1]). This phosphoprotein was identified at $p = 0.0075$; FDR < 1% as determined by the sequential use of the SpectrumMill, Scaffold and InsPecT software tools described above. The separate evaluation of the phosphate localization for the phosphopeptides detected was not performed in this proof-of-principle study. Perspective prostate cancer functional proteomic studies from our group, however, will incorporate a specialized phosphorylation site localization software tool such as that of PLS optionally available to the InsPecT software program (<http://proteomics.ucsd.edu>). The more targeted nature of these functional proteomic studies will allow the essential visual examination of all relevant data given the fact that guidance toward a reasonable PLS score for phosphate group localization is currently unavailable.

Although the ZIC-HILIC approach cannot be applied as part of an online two-dimensional LC platform, as in the case of SCX, it may be amenable to automated sample handling systems (such as the Quadra 4 platforms, Tomtec Corp., Hamden, CT, or the MassPrep Station, Waters Corp., Milford, MA) that allow the prior evaporation of the off-line fractions and subsequent

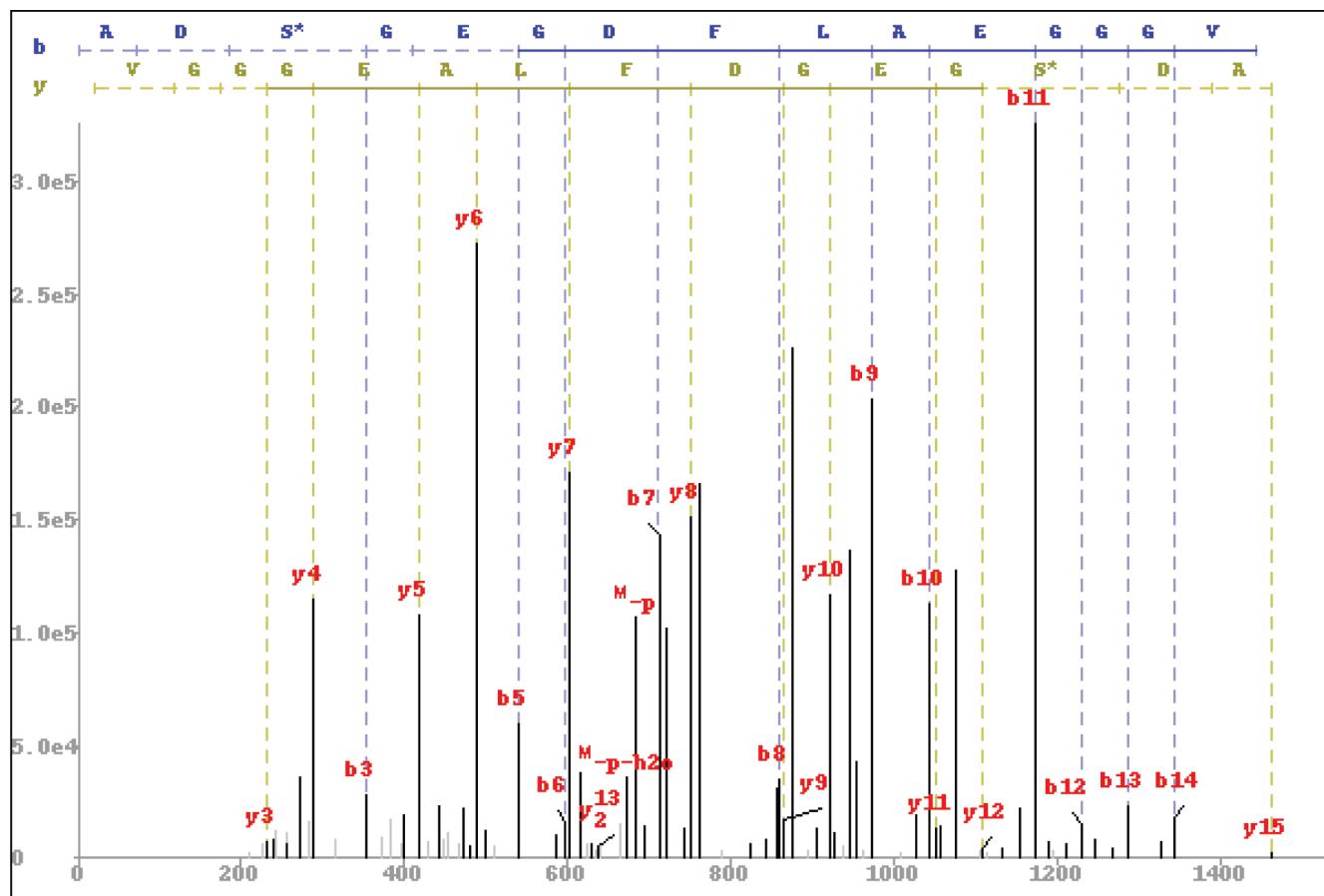


Figure 4. An example annotated ion-trap MS² product ion spectra for the +2 H tryptic phosphopeptide T.ADSphosGEGDFLAEGGGV.R with precursor mass at 730.9 Da ($\Delta m = -0.8$ Da) traceable to the fibrinogen alpha chain isoform alpha preproprotein (accession no.: gi|11761629|ref|NP_068657,1|). The phosphoprotein was identified at $p = 0.0075$; FDR < 1%, as determined by the sequential use of the SpectrumMill \Rightarrow Scaffold \Rightarrow InsPecT software tools. The spectral presentation format is in accordance with the InsPecT software tool convention.

reconstitution and mixing with reversed-phase compatible solvents followed by their online LC–MS analysis. Such platforms would maintain the higher sample loading capacities amenable to an off-line semipreparatory workflow more appropriate to in-depth proteome discovery studies.

Another major characteristic of the serum proteome as determined by the MudPIT approach A was its large protein heterogeneity in terms of the extensive range of biological ontologies and tissue expression space observed, as expected given the blood's significant role in human physiology (Figure 5). In particular, the protein functional annotation was performed with the DAVID software tool (<http://david.abcc.ncifcrf.gov>), which facilitated the categorization of the confidently identified proteome resulting from the MudPIT approach A in four distinct categories: biological process, cellular component, molecular function, and tissue expression, respectively.

Table 2 lists the differentially expressed proteins found in the BPH and PCa tissue study²⁰ also qualitatively found in the BPH sera samples with at least one of the MudPIT approaches used in this current study. A total of 12 out of the 15 tissue expressed proteins were also qualitatively detected in the BPH serum with the MudPIT approach A, the majority of which were up-regulated in the reported BPH prostate tissue category.²⁰ Also listed in Table 2 are other proteins designated by the Uniprot convention to be expressed in prostate tissue that were also identified in the BPH

serum by at least one of the MudPIT proteomic approaches. When reduced identification thresholds were applied to the combined use of SpectrumMill and Scaffold validation search tools, the marginal detection of the anticipated prostate specific antigen (PSA) was observed in the proteomes resulting from the MudPIT approaches A and C (see Supporting Information for a detailed listing of the tryptic peptides detected). These proteins, along with the BPH and PCa tissue expressed proteins listed in Table 2, could potentially be applied on a panel basis as predictive or predisposition markers in otherwise asymptomatic patients.⁴² Overall, this tissue–serum proteome correlation constituted a major milestone in establishing an essential link between the restricted expression at the target tissue level and its occurrence at measurable levels in the serum that correlated or reflected the BPH or PCa-tissue expressed proteins.

As a proof-of-concept study, the three MudPIT approaches were executed on a single pooled BPH serum sample ($N = 25$). The analytical techniques constituting the MudPIT approach A, however, exhibited reproducible behavior between replicate sample injections of either serum specimens or QC tryptic peptide standards. For example, the replicate serial injections for the SEC protein prefractionation step exhibited identical chromatographic profiles monitored with UV at 280 nm. Analogous behavior was observed for the commercially available chromatographic columns for the ZIC-HILIC separation (MudPIT approaches A and B), SCX separation (MudPIT approach C) monitored with

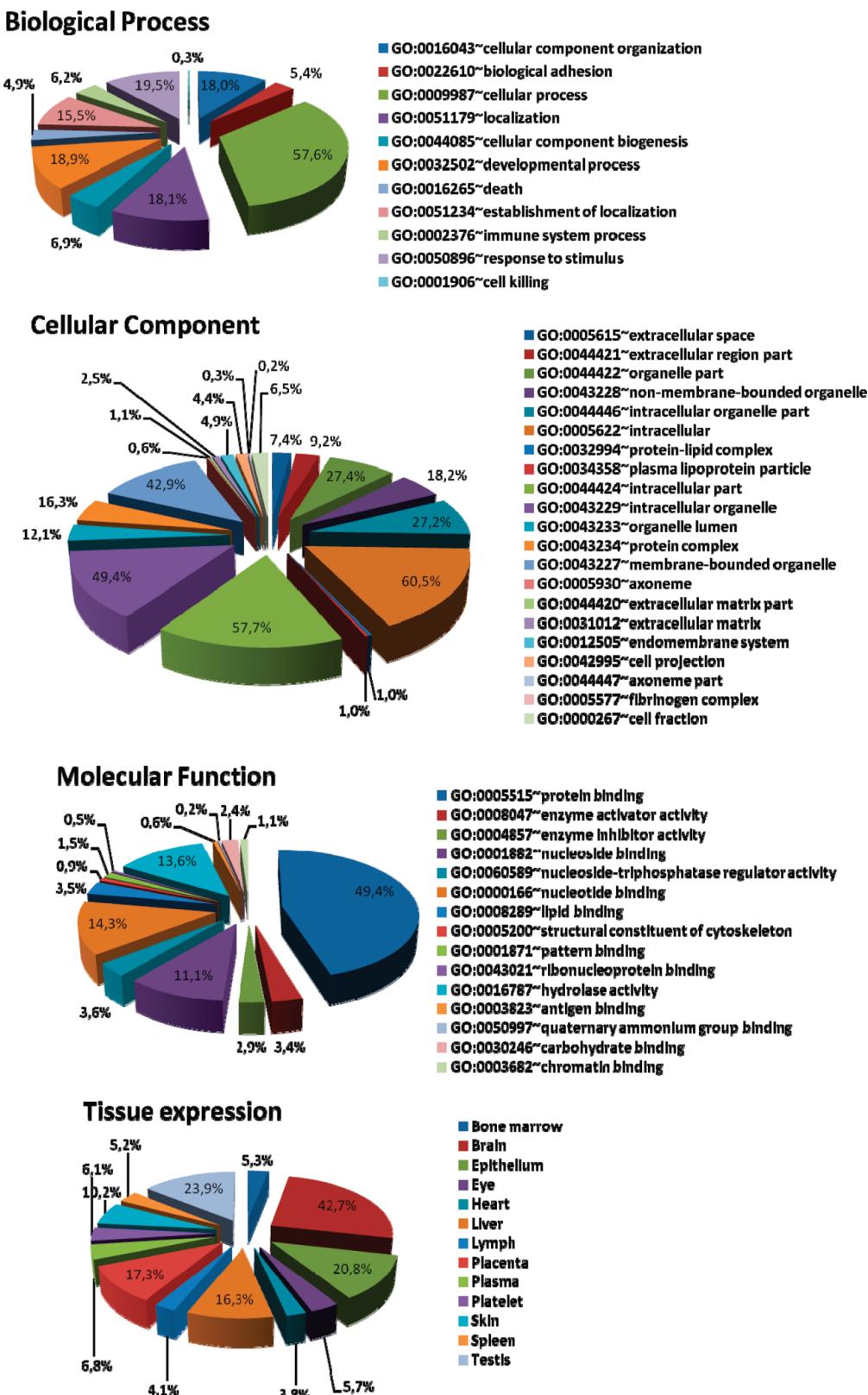


Figure 5. Functional annotation of the proteins identified by the MudPIT approach A in accordance with their biological process, cellular component, molecular function, and tissue expression gene ontologies, as determined by the DAVID software tool. Only proteins identified by the sequential use of the SpectrumMill \Rightarrow Scaffold \Rightarrow InsPecT software tools were functionally annotated ($p \leq 0.05$; FDR $\leq 5\%$).

UV at 280 nm and RP C₁₈ UPLC separation (MudPIT approaches A, B, and C) monitored with ESI ion trap MS of various tryptic

digest standards (i.e., BSA, casein, 6-protein mix, etc.) at specified concentration levels.

Table 2. List of Proteins Identified To Be Commonly Expressed in the BPH Serum Proteome Reported To Be Differentially Expressed in the BPH and PCa Tissue Proteome by at Least One of the MudPIT Approaches^a

accession no.	protein name	approach A	approach B	approach C
Identified Serum Proteins up-Regulated in BPH Tissue (ref 20)				
gi 4503273	angiotensin-converting enzyme isoform 1 precursor	✓	—	—
gi 71274107	cell surface glycoprotein MUC18	✓	—	—
gi 89276751	collagen alpha-1(V) chain preproprotein	—	✓	—
gi 87196339	collagen alpha-1(VI) chain precursor	✓	—	—
gi 23065544	glutathione S-transferase Mu 1 isoform 1	—	✓	—
gi 4504183	glutathione S-transferase P	✓	—	—
gi 21618357	kallikrein-11 isoform 2 precursor	✓	—	—
gi 21361091	ubiquitin carboxyl-terminal hydrolase isozyme L1	—	—	✓
gi 7669550	vinculin isoform meta-VCL	✓	✓	✓
gi 4508047	zyxin	✓	—	—
Identified Serum Proteins up-Regulated in PCa Tissue (ref 20)				
gi 50363217	alpha-1-antitrypsin precursor	✓	✓	✓
gi 161702986	ezrin	✓	—	—
gi 6006001	glutathione peroxidase 3 precursor	✓	—	—
gi 4758484	glutathione S-transferase omega-1 isoform 1	✓	—	—
gi 10835063	nucleophosmin isoform 1	✓	—	—
Identified Serum Proteins Expressed in the Prostate Tissue per Uniprot Annotation				
gi 46852388	cell division cycle and apoptosis regulator protein	✓	—	—
gi 11496281	kallikrein-13 precursor	✓	—	—
gi 4758650	kinesin heavy chain isoform 5C	✓	—	—
gi 45331215	leucine zipper putative tumor suppressor 2	✓	—	—
gi 34452734	peroxisomal membrane protein 4 isoform a	—	✓	—
gi 23238202	tumor necrosis factor receptor superfamily member 1	✓	—	—
gi 150378533	ubiquitin carboxyl-terminal hydrolase 7	—	✓	—

^a Additional proteins observed with bioassay measurement to be predominately expressed in prostate tissue also detected by at least one of the MudPIT approaches. The ✓ denotes a protein identification and the dashed line (—) denotes a null observation.

The putative detection of serum proteins in this study primarily by the MudPIT approach A was accomplished without the use of targeted antibody capture, as is commonly the case. This fact demonstrated the unbiased and in-depth analysis of the BPH proteome of our study, which is a major advantage over other studies. As part of the hallmark Medical Therapy of Prostatic Symptoms (MTOPS) clinical trial, the Prostatic Samples Analysis (MPSA) Consortium attempted to establish and validate potential biomarkers that could stratify the BPH patients according to their response to medical therapy, by using dual antibody sandwich immunoassay of ELISA, which is the most sensitive bioassay technique currently available.^{43,44} However, this was an *a priori* approach targeting specific biomarkers that bypassed the possibility of observing a plethora of low-abundance proteins that might play a significant role on the differential diagnosis between BPH and PCa. Along this theme, the proposed MudPIT approach A, used as a qualitative protein identification tool in this study, can be extended to include relative quantitative features made possible with the use of multiplex stable isotope labeling strategies at the protein or peptide level to further minimize analytical systematic error and to better stratify patients in accordance with prostate pathophysiology analogous to that of the BPH/PCa prostate tissue study reported by the authors.²⁰ Such an approach can serve as part of a more systematic serum biomarker discovery study.

■ CONCLUDING REMARKS AND FUTURE OUTLOOK

The proposed MudPIT approach, incorporating size-exclusion chromatographic (SEC) for the protein prefractionation as segments and their dialysis exchange and trypsin proteolysis, followed by the offline first dimensional fractionation of the tryptic peptides with ZIC-HILIC fractionation and subsequent online RP-nUPLC-nESI-MS² analysis, resulted in the identification of 1955 proteins ($p \leq 0.05$; FDR $\leq 5\%$) with a broad spectrum of biological and physicochemical properties including secreted, tissue-specific proteins spanning approximately 12 orders of magnitude in natural abundance levels, and encompassed in this proteome was the confident identification of 375 phosphoproteins ($p \leq 0.05$; FDR $\leq 5\%$) of potential importance to cancer biology. Its performance attributes were demonstrated against two other MudPIT approaches that were based on the immunodepletion of the highest abundant albumin and IgG proteins, followed by the offline first dimensional fractionation of the tryptic peptides with both SCX and ZIC-HILIC fractionation and subsequent online RP-nUPLC-nESI-MS² analysis. The significant proteome homology combined with the distinct proteomes obtained from the three MudPIT approaches, when used collectively, served as a large-scale proteome discovery and verification strategy. The versatility and adaptability of the constituent techniques constituting these MudPIT approaches can be extended to incorporate quantitative features using stable isotope labeling or

label free approaches or can be applied for the in-depth proteome analysis of protein extracts derived from cell culture, tissue, organ proximal fluid, plasma, or urinary specimens as part of a global and unbiased biomarker discovery program.

■ ASSOCIATED CONTENT

§ Supporting Information. Proteome results from the sequential use of the SpectrumMill \Rightarrow Scaffold \Rightarrow InsPecT software tools (including peptide annotation, surrogate protein accession, peptide charge state, MQScore, AA length, median PRM score, fraction nY, fraction nB, intensity, p-value, F-Score, delta score, precursor *m/z*, precursor *m/z* error, unique protein identification, and phophopeptide/phosphoprotein identifications); a table of prostate specific antigen data (containing the accession number, the score (Spectrum Mill), percent scored peak intensity, reverse score, delta forward reverse score, variable modifications, sequence, parent *m/z*, parent charge state, delta parent mass, delta ppm); and product ion (MS^2) spectra corresponding to this table are provided. This material is available free of charge via the Internet at <http://pubs.acs.org.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: 0030 210 6597129 (O) and 0030 2106597131 (L). Fax: 0030 210 6597545. E-mail: sgarbis@bioacademy.gr. Website: www.bioacademy.gr.

■ ACKNOWLEDGMENT

This study was supported by the E.LKE National and Kapodistrian University of Athens grant (No. 70/4/6573). We thank Mr. Harris Tsafaras for assisting in the SEC separation work.

■ REFERENCES

- (1) Anderson, N. L.; Anderson, N. G. *Mol. Cell. Proteomics* **2002**, *1*, 845–867.
- (2) Wu, S.-L.; Choudhary, G.; Ramström, M.; Bergquist, J.; Hancock, W. S. J. *Proteome Res.* **2003**, *2*, 383–393.
- (3) Omenn, G. S. *Proteomics* **2004**, *4*, 1235–1240.
- (4) Anderson, N. L.; Polanski, M.; Pieper, R.; Gatlin, T.; Tirumalai, R. S.; Conrads, T. P.; Veenstra, T. D.; Adkins, J. N.; Pounds, J. G.; Fagan, R.; Lobley, A. *Mol. Cell. Proteomics* **2004**, *3*, 311–326.
- (5) States, D. J.; Omenn, G. S.; Blackwell, T. W.; Fermin, D.; Eng, J.; Speicher, D. W.; Hanash, S. M. *Nat. Biotechnol.* **2006**, *3*, 333–338.
- (6) Hanash, S. M.; Pitteri, S. J.; Faca, V. M. *Nature* **2008**, *452*, 571–579.
- (7) Steel, L. F.; Trotter, M. G.; Nakajima, P. B.; Mattu, T. S.; Gonye, G.; Block, T. *Mol. Cell. Proteomics* **2003**, *2*, 262–270.
- (8) Liu, T.; Qian, W.-J.; Mottaz, H. M.; Gritsenko, M. A.; Norbeck, A. D.; Moore, R. J.; Purvine, S. O.; Camp, D. G., II; Smith, R. D. *Mol. Cell. Proteomics* **2006**, *5*, 2167–2174.
- (9) Seam, N.; Gonzales, D. A.; Kern, S. J.; Hortin, G. L.; Hoehn, G. T.; Suffredini, A. F. *Clin. Chem.* **2007**, *53*, 1915–1920.
- (10) Ringrose, J. H.; van Solinge, W. W.; Mohammed, S.; O'Flaherty, M. C.; van Wijk, R.; Heck, A. J. R.; Slijper, M. *J. Proteome Res.* **2008**, *7*, 3060–3063.
- (11) Zolotarjova, N.; Martosella, J.; Nicol, G.; Bailey, J.; Boyes, B. E.; Barrett, W. C. *Proteomics* **2005**, *5*, 3304–3313.
- (12) Granger, J.; Siddiqui, J.; Copeland, S.; Remick, D. *Proteomics* **2005**, *5*, 4713–4718.
- (13) Gundry, R. L.; Qin, F.; Jelinek, C. A.; Van Eyk, J. E.; Cotter, R. J. *Proteomics Clin. Appl.* **2007**, *1*, 73–88.
- (14) Zolla, L. *Drug Disc. Today* **2008**, *13*, 1042–1051.
- (15) Chignard, N.; Shang, S.; Wang, H.; Marrero, J.; Bréchot, C.; Hanash, S.; Beretta, L. *Gastroenterology* **2006**, *130*, 2010–2022.
- (16) van Winden, A. W.; van den Broek, I.; Gast, M. C.; Engwegen, J. Y.; Sparidans, R. W.; van Dulken, E. J.; Depla, A. C.; Cats, A.; Schellens, J. H.; Peeters, P. H.; Beijnen, J. H.; van Gils, C. H. *J. Proteome Res.* **2010**, *9*, 3781–3788.
- (17) Shen, Y.; Jacobs, J. M.; Camp, D. G., II; Fang, R.; Moore, R. J.; Smith, R. D. *Anal. Chem.* **2004**, *76*, 1134–1144.
- (18) Jin, W.-H.; Dai, J.; Li, S.-J.; Xia, Q.-C.; Zou, H.-F.; Zeng, R. *J. Proteome Res.* **2005**, *4*, 613–619.
- (19) Chen, E. I.; Hewel, J.; Felding-Habermann, B.; Yates, J. R. *3rd Mol. Cell. Proteomics* **2006**, *5*, 53–56.
- (20) Garbis, S. D.; Tyritzis, S. I.; Roumeliotis, T.; Zerefos, P.; Giannopoulou, E. G.; Vlahou, A.; Kossida, S.; Diaz, J.; Vourekas, S.; Tamvakopoulos, C.; Pavlakis, K.; Sanoudou, D.; Constantinides, C. A. *J. Proteome Res.* **2008**, *7*, 3146–3158.
- (21) Tuck, M. K.; Chan, D. W.; Chia, D.; Godwin, A. K.; Grizzle, W. E.; Krueger, K. E.; Rom, W.; Sanda, M.; Sorbara, L.; Stass, S.; Wang, W.; Brenner, D. E. *J. Proteome Res.* **2009**, *8*, 113–117.
- (22) Searle, B. C. *Proteomics* **2010**, *10*, 1265–1269.
- (23) Gupta, N.; Pevzner, P. A. *J. Proteome Res.* **2009**, *8*, 4173–4181.
- (24) Petricoin, E. F.; Zoon, K. C.; Kohn, E. C.; Barrett, J. C.; Liotta, L. A. *Nat. Rev. Drug Discovery* **2002**, *1*, 683–695.
- (25) Lescuyer, P.; Hochstrasser, D.; Rabilloud, T. *J. Proteome Res.* **2007**, *6*, 3371–3376.
- (26) Beretta, L. *Nat. Methods* **2007**, *4*, 785–786.
- (27) Herr, W.; Schneider, J.; Lohse, A. W.; Meyer zum Büschenfelde, K. H.; Wölfel, T. *J. Immunol. Methods* **1996**, *191*, 131–142.
- (28) Kyte, J.; Doolittle, R. F. *J. Mol. Biol.* **1982**, *157*, 105–132.
- (29) Kyte, J. *Biophys. Chem.* **2003**, *100*, 193–203.
- (30) Garbis, S.; Lubec, G.; Fountoulakis, M. *J. Chromatogr. A* **2005**, *1077*, 1–18.
- (31) Annesley, T. M. *Clin. Chem.* **2003**, *49*, 1041–1044.
- (32) Castellino, F. J.; Barker, R. *Biochemistry* **1968**, *7*, 4135–4138.
- (33) Folta-Stogniew, E.; Williams, K. R. *J. Biomol. Tech.* **1999**, *10*, 51–63.
- (34) Oliva, A.; Llabrés, M.; Fariña, J. B. *J. Pharm. Biomed. Anal.* **2001**, *25*, 833–841.
- (35) Ye, H. *Anal. Biochem.* **2006**, *356*, 76–85.
- (36) Zhang, H.; Yi, E. C.; Li, X.-J.; Mallick, P.; Kelly-Spratt, K. S.; Masselon, C. D.; Camp, D. G., II; Smith, R. D.; Kemp, C. J.; Aebersold, R. *Mol. Cell. Proteomics* **2005**, *4*, 144–155.
- (37) Zhang, H.; Liu, A. Y.; Loriaux, P.; Wollscheid, B.; Zhou, Y.; Watts, J. D.; Aebersold, R. *Mol. Cell. Proteomics* **2007**, *6*, 64–71.
- (38) Meany, D. L.; Zhang, Z.; Sokoll, L. J.; Zhang, H.; Chan, D. W. *J. Proteome Res.* **2009**, *42*, 221–224.
- (39) Alpert, A. *J. Chromatogr.* **1990**, *499*, 177–196.
- (40) Boersema, P. J.; Divecha, N.; Heck, A. J. R.; Mohammed, S. *J. Proteome Res.* **2007**, *6*, 937–946.
- (41) Garbis, S. D.; Melse-Boonstra, A.; West, C. E.; van Breemen, R. B. *Anal. Chem.* **2001**, *73*, 5358–5364.
- (42) Alcaraz, A.; Hammerer, P.; Tubaro, A.; Schröder, F. H.; Castro, R. *Eur. Urol.* **2009**, *55*, 864–873.
- (43) Mullins, C.; Lucia, M. S.; Hayward, S. W.; Lee, J. Y.; Levitt, J. M.; Lin, V. K.; Liu, B. C.; Chinnaianyan, A. M.; Rubin, M. A.; Slawin, K.; Star, R. A.; Getzenberg, R. H.; MPSA Consortium *J. Urol.* **2008**, *179*, 1243–1256.
- (44) Rissin, D. M.; Kan, C. W.; Campbell, T. G.; Howes, S. C.; Fournier, D. R.; Song, L.; Piech, T.; Patel, P. P.; Chang, L.; Rivnak, A. J.; Ferrell, E. P.; Randall, J. D.; Provuncher, G. K.; Walt, D. R.; Duffy, D. C. *Nat. Biotechnol.* **2010**, *6*, 595–600.