analytical chemistry

Article

pubs.acs.org/ac

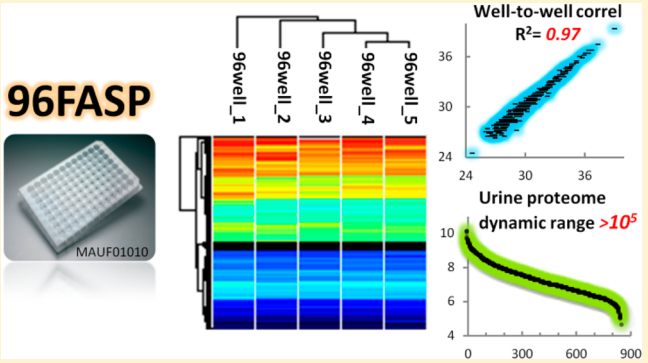
Urine Sample Preparation in 96-Well Filter Plates for Quantitative Clinical Proteomics

Yanbao Yu,* Moo-Jin Suh, Patricia Sikorski, Keehwan Kwon, Karen E. Nelson, and Rembert Pieper*

The J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, Maryland 20850, United States

Supporting Information

ABSTRACT: Urine is an important, noninvasively collected body fluid source for the diagnosis and prognosis of human diseases. Liquid chromatography mass spectrometry (LC-MS) based shotgun proteomics has evolved as a sensitive and informative technique to discover candidate disease biomarkers from urine specimens. Filter-aided sample preparation (FASP) generates peptide samples from protein mixtures of cell lysate or body fluid origin. Here, we describe a FASP method adapted to 96-well filter plates, named 96FASP. Soluble urine concentrates containing ~10 μg of total protein were processed by 96FASP and LC-MS resulting in 700−900 protein identifications at a 1% false discovery rate (FDR). The experimental repeatability, as assessed by label-free quantification and Pearson correlation analysis for shared proteins among replicates, was high ($R \geq 0.97$). Application to urinary pellet lysates which is of particular interest in the context of urinary tract infection analysis was also demonstrated. On average, 1700 proteins (±398) were identified in five experiments. In a pilot study using 96FASP for analysis of eight soluble urine samples, we demonstrated that protein profiles of technical replicates invariably clustered; the protein profiles for distinct urine donors were very different from each other. Robust, highly parallel methods to generate peptide mixtures from urine and other body fluids are critical to increase cost-effectiveness in clinical proteomics projects. This 96FASP method has potential to become a gold standard for high-throughput quantitative clinical proteomics.



Using liquid chromatography mass spectrometry (LC-MS) based technologies developed over the past decade, body fluid proteomes can be surveyed with protein abundances in a dynamic range of 5 orders of magnitude, allowing identification of thousands of proteins at a false discovery rate (FDR) of 1% without extensive fractionation prior to LC-MS.[1,2] Urine is a sample source of high importance for clinical proteomic studies because it is easily available and collected noninvasively, thus eliminating health risks for the donor. The identity and quantity of proteins excreted into urine may reflect pathological conditions that can be traced to different organs in the body, particularly the kidneys, prostate and urogenital tract.[3] The urinary proteome has been studied for more than a decade with LC-MS based technologies,[4−6] resulting in the identification of more than 1500 distinct proteins associated with at least 58 different gene ontology (GO) molecular function categories.[6] The functional diversity demonstrates the richness of urine as a source of identifying perturbations in biological pathways and organ malfunctions in the human body. An area of considerable interest in urinary proteomics is the identification of pathogens.[7−9] A survey for the year 2006 estimated that the occurrence of 1.7 million emergency room visits, 11 million physician visits, and half a million hospitalizations in the U.S. alone were due to urinary tract infections (UTI) resulting in 3.5 billion dollars of health care costs.[10] Shotgun proteomics succeeds in identifying microbial proteins from urine independent of the ability to detect microbes in a urine laboratory culture, the most commonly used method to determine the pathogen causing a UTI.[11] Urinary pellets derived from patients diagnosed with either asymptomatic bacteriuria or UTI were recently used to identify urinary tract-colonizing bacteria with metagenomic and metaproteomic methods.[9]

A typical shotgun proteomics workflow involves protein extraction from tissues, cells, or body fluids, enzymatic digestion and LC-based peptide fractionation in one or multiple dimensions followed by MS-based protein identification.[12] To extract, solubilize, and denature proteins, detergents, such as SDS and NP-40, and chaotropic reagents, such as urea and thiourea, are commonly used.[12] Solubilized protein mixtures are not directly applied to in-solution digestion because the presence of detergents and chaotropic reagents reduces the activity of the proteolytic enzyme (e.g., trypsin). Incomplete digestion decreases the number of available peptide analytes of LC-MS. For example, 0.1% SDS leads to an ~80% loss of trypsin activity.[13] Furthermore, trace amounts of SDS can cause significant signal suppression in LC-MS experiments.[14]

Received: February 21, 2014
Accepted: May 5, 2014
Published: May 5, 2014

   dx.doi.org/10.1021/ac5008317 | *Anal. Chem.* 2014, 86, 5470−5477

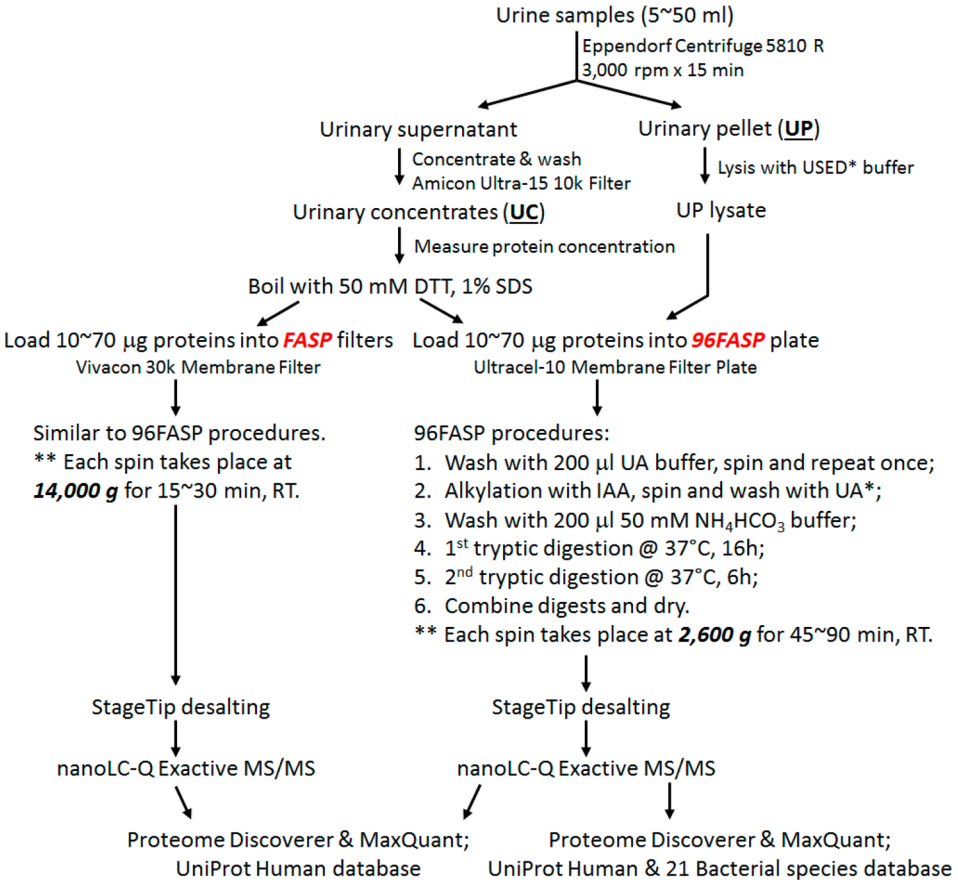


Figure 1. Overview of the experimental procedures performed in this study. Briefly, the urine sample is centrifuged to separate urinary pellet (UP) from urinary supernatant fraction, which is then followed by concentration using Amicon filter. The urinary concentrates (UC) and UP are then subjected to filter aided sample preparation processing using single filter (FASP) or multiwell format filter (96FASP). The peptides after on-filter digestion are desalted using StageTip, and analyzed by LC-Q Exactive MS/MS and computational database search. *USED buffer: 8 M urea, 1% SDS, 5 mM EDTA, and 50 mM DTT. *UA buffer: 8 M urea in 0.1 M Tris-HCl, pH 8.0. RT: room temperature.

Detergent removal by precipitation of proteins with organic solvents prior to digestion[15] or precipitation of peptides with KCl after digestion[16] have been employed to reduce the adverse effects of detergents and improve proteome coverage. Among the many detergent depletion techniques that have been explored, a filter aided sample preparation (FASP) method appears to be the most effective one to achieve high protein coverage in shotgun proteomic analyses.[17,18] A standard FASP ultrafiltration device, usually having a load volume of less than 500 μL and a 30 kDa molecular weight cutoff (MWCO), facilitates removal of detergents and chaotropic reagents and equilibration in buffers ideal for the reduction, alkylation and digestion steps. Enzymatic digestion of protein mixtures occurs directly on the filter membrane.[17,18] The digested peptides, most of which have MW values of less than 5 kDa, are smaller than the filter's MWCO and pass through the membrane upon centrifugation. These versatile features of the FASP method have resulted in its application in numerous proteomic projects.[19−23]

Recent method development efforts focused on pre- or post-FASP fractionation to reduce the sample complexity and to improve proteome coverage.[24−26] For instance, the separation of FASP-related protein digests into six fractions via strong anion exchange (SAX) allowed identification of 4,206 proteins from mouse hippocampus tissue.[25] Prefractionation of proteins using size exclusion chromatography followed by SAX and FASP allowed Nagaraj et al. to identify more than 10 000 HeLa cell proteins.[24] Peptide fractionation by high pH reversed-phase LC following FASP was described in a report on the BV-2 microglial cell proteome, which consisted of 5494 identified proteins.[26] All aforementioned studies focused on proteome coverage rather than parallel sample processing. The use of FASP was also reported for the comparative proteomic analysis of cell lines and tissues, processing up to 30 samples in parallel to examine disease-related and prognostic protein biomarkers.[23,27,28] In clinical proteomics, it is desirable to process large numbers of samples to achieve the statistical power necessary to identify promising biomarker candidates.[29] Therefore, development of a reliable, highly parallel FASP method using a 96-well filter plate is a worthwhile endeavor. The potential benefits are batch-mode operation of up to 96 samples, which could lower the requirements of experimental repetition, and overall cost-effectiveness.

To our knowledge, the first report on the use of FASP in 96-well plates was published recently. Switzar et al.[30] used a 96-well plate with built-in 10 kDa MWCO membrane filters and a stepwise process to generate cellular protein digests. This method was evaluated using different organic solvent wash steps and compared to a gel filtration method for purification of peptide mixtures. Up to 442 proteins with a FDR of less than 0.1% were identified from HEK293T cell lysates in a single LC-MS experiment, using the LTQ-Orbitrap Velos system for MS. In depth assessments of experimental repeatability of FASP in 96-well plates were not part of this study. Urine is a challenging

dx.doi.org/10.1021/ac5008317 | *Anal. Chem.* 2014, 86, 5470−5477

sample source due to high abundances of only a few proteins and the richness in functional protease inhibitors that interfere with tryptic digestion unless they are inactivated first.[6,31] Here, we show that 96-well based parallel FASP processing of human urine specimens including soluble and insoluble fractions, combined with LC-MS, is robust and repeatable and yields high proteome coverage.

EXPERIMENTAL SECTION

Urine Specimens and Urine Sample Preparation for 96FASP. One set of human urine specimens was from a study of juveniles diagnosed with Type 1 diabetes (T1D) including matched sibling controls. Participants were informed that urine specimens were to be used for research purposes. Midstream urine was collected in a doctor's office. A Human Subject Protocol was established and approved by JCVI's Internal Review Board. Human subject informed consent was obtained. The other set of human urine specimens was related to a study of urinary tract infections (UTI) where samples were obtained as medical waste from a urine diagnostics laboratory. The study was exempted from the requirement of a Human Subject Protocol by JCVI's Internal Review Board. For both studies, urine samples were received after specimen deidentification. They were stored at 4 °C for maximally 6 h prior to transfer to the JCVI laboratory and storage at −20 °C. The starting volumes of urine sample aliquots ranged from 20 to 50 mL. After centrifugation at 3000 × g for 15 min at 10 °C, urine supernatants were concentrated with an Amicon Ultra-15 centrifugal filter device (10 kDa MWCO, Millipore) at 3000 × g to a volume of ∼1.0 mL and are referred to as UC samples. Resulting urinary pellets from UTI specimens were recovered and are referred to as UP samples. The type of filter plate used here was equipped with cellulose membrane filters with a 10 kDa MWCO (MultiScreen Ultracel-10; catalogue number: MAUF01010; Millipore). From here on, we refer to this filter plate as the 96FASP plate. We are not aware of other 96-well filter plate products with a larger MWCO. UC samples were denatured with 1% SDS (w/v) and 50 mM DTT at 95 °C for 10 min; UP samples were lysed with a solution of 8 M urea, 1% SDS, 5 mM EDTA and 50 mM DTT prior to 96FASP plate analysis. A protocol described previously,[18] also used here for a comparative proteomic analysis with UC samples and the single-filter FASP device (Vivacon 30 kDa MWCO, Sartorius, Germany), was employed using 96FASP plates with minor modifications as shown in Figure 1. Experimental procedures are also described in detail in the Supporting Information.

NanoLC-MS/MS Method. The nanoLC-MS/MS analysis was performed on a Ultimate 3000 nano LC and Q Exactive mass spectrometer system coupled with a FLEX nano-electrospray ion source (all components were from Thermo Scientific). The peptide samples were first loaded onto a trap column (C_{18} PepMap100, 300 μm × 5 mm, 5 μm, 100 Å, Thermo Scientific), and then separated on a PicoFrit analytical column (75 μm × 10 cm, 5 μm BetaBasic C_{18}, 150 Å, New Objective, MA) at a flow rate of 300 nL/min. For a 130 min LC-MS run, a linear gradient was applied from 100% solvent A to 35% solvent B (0.1% formic acid in acetonitrile) over 110 min, followed by a steeper gradient to 80% solvent B over 15 min. The column was re-equilibrated with solvent A for 5 min. For a 90 min LC-MS run, the linear gradient time extended over 70 min (from 0 to 35% solvent B). Eluting peptides were sprayed at a voltage of 2.0 kV and acquired in a MS data-dependent mode using XCalibur software (version 2.2, Thermo Scientific). Survey scans were acquired at a resolution of 70,000 over a mass range of m/z 250 to m/z 1,800 with an automatic gain control (AGC) target of 10^6. For each cycle, the ten most intense ions were subjected to fragmentation by higher energy collisional dissociation (HCD) with normalized collision energy of 27%. Peptide ion fragments from the MS/MS scans were acquired at a resolution of 17,500 with an AGC target of 5×10^4. Dynamic exclusion was enabled, as MS/MS ion scans were repeated once and then excluded from further analysis for 20 s. Unassigned ions and those with a charge of +1 were rejected from further analysis.

Protein Identification and Quantification Methods. The raw files acquired by the MS system were processed using the Proteome Discoverer platform (version 1.4, Thermo Scientific). An integrated workflow using the algorithms Sequest HT and Mascot (version 2.4, Matrix Science) was employed. Either a human UniProtKB database (Release 2013_6; 88 295 human sequences) or a database consisting of the aforementioned human proteins and all protein sequences derived from 21 microbial genomes (Supporting Information Table S-1) were used. The latter database was used to identify human and microbial proteins present in UP samples. MS search parameters similar to published previously[27] are described in detail in Supporting Information. For protein quantification of the data sets, the MaxQuant software suite (version 1.4.2) was used.[32] Most of the default settings provided in this software suite were accepted, and data were processed using both the label-free quantitation (LFQ) and the intensity-based absolute quantitation (iBAQ) tools. The LFQ algorithms provide relative quantification of the integrated MS^1 peak areas from the high resolution MS data. The iBAQ algorithms sum the integrated peak intensities of the peptide ions for a given protein divided by the number of theoretically observable peptides, which are calculated by in silico digestion of protein sequences including all fully tryptic peptides with a length of 6−30 amino acids.[33]

RESULTS AND DISCUSSION

96FASP Evaluation Compared to the Single-Filter FASP Method. To our best knowledge, only one type commercially available 96-well filter plate is suitable for FASP. It is equipped with Ultracel-10 membrane filters and has a 10 kDa MWCO. The FASP method with single-filter devices has been evaluated for filters with MWCO values of 3,[18] 10,[18] and 30 kDa.[25,34] The 30 kDa MWCO filter device was reported to facilitate sample preparation with shorter centrifugation times and to generate a larger quantity of peptides in a MW range suitable for MS analysis compared to the 10 kDa MWCO filters.[34] For the performance comparison (single-filter FASP versus 96FASP), the 30 kDa MWCO single-filter device was selected. The sample chosen to evaluate the quality of proteomic data including experimental repeatability of the 96FASP method was a soluble urinary concentrate (UC) derived from a donor with an apparent urinary tract infection. Experimental repeatability is defined as the repeated analysis of the proteome of a sample processed in the same laboratory, LC-MS instrument, and LC-MS methods, following guidelines proposed by Tabb et al.[35] We did not assess experimental reproducibility.

The question presented itself as to whether the 96FASP plate with the relatively low MWCO of 10 kDa and limited centrifugal forces applied in a plate-adapted centrifuge allowed ultrafiltration of high molarity solutions (8 M urea) in a

dx.doi.org/10.1021/ac5008317 | Anal. Chem. 2014, 86, 5470−5477

reasonable time frame. Completion of a 96FASP experiment included denaturation and concentration of a body fluid or cell lysate sample, protein reduction, protein alkylation, and intermittent wash steps at centrifugal forces of 2,600 × *g* prior to enzymatic digestion overnight (Figure 1). In contrast, the single-filter device for FASP allowed spins at centrifugal forces as high as 14 000 × *g*. Using UC samples estimated to contain 10−15 μg of protein, we determined that a volume reduction of 200 μL of UA buffer to less than 50 μL in wells of a 96FASP plate required centrifugation times of 45−60 min. This seems reasonable as indicated by a previous study that 10 kDa MWCO filters usually take three to four times longer than 30 kDa filters, which traditional FASP typically use.[34] The entire 96FASP procedure prior to overnight digestion, including urine concentrate denaturation, alkylation and intermittent centrifugation steps, requires 4−6 h depending on the UA buffer volume, the total protein amount loaded and possibly the presence of other macromolecular substances. While longer centrifugation times are a comparative weakness of 96FASP, the ability to parallelize pipetting steps, reduce sample handling and the prospects of automating the process on robotic platforms are significant advantages in comparison to traditional single filter-based FASP methods. Then we assessed whether removal of SDS applied in a 1% concentration was achieved during UA buffer wash steps. Indeed, SDS signals in LC-MS runs were not observed (Supporting Information Figure S-1). This was encouraging because a 1% SDS solution facilitates the lysis of microbial and mammalian cells, solubilization of proteins integrated in phospholipid membranes and extraction of proteins from tissues and other clinical samples.[36,37] 1% SDS also denatures proteins and thus improves the effectiveness of proteolysis at the cleavage sites expected for a given protease. Using the same 96-well filter plate source, Switzar et al. reported using a 0.1% SDS solution for HEK293T cell lysis.[30] A high molarity urea solution was not used during the subsequent wash steps to deplete SDS from the protein sample prior to enzymatic digestion. The omission of urea during the FASP wash steps may have resulted in decreased protease activity and lower proteome coverage.[34] Retention of SDS in a peptide mixture also impacts ion suppression during LC-MS analysis.[34,38]

Experimental Repeatability Assessed for the Entire Workflow. Approximately 10−15 μg of total protein from one UC sample was loaded into five different wells of a 96FASP plate and processed as shown in Figure 1, defined here as 96-well replicates. Duplicate LC-MS runs for each well injecting ∼2 μg of the digested peptide mixture with a simple 90 min gradient were performed, defined here as LC-MS replicates. On average, 3955 unique peptides (±241, SD; n = 5) and 852 (±7) protein groups were identified from each well at a 1% FDR. The analytical performance of 96FASP with approximately five peptides per protein group was in the expected range given that the stochastic sampling of proteomes with only a few highly abundant proteins by LC-MS/MS generally decreases peptide identifications per protein for many low and medium abundant proteins.[12] This is also illustrated in Figure 3. The percentages of shared peptide and protein identifications among the 96-well replicates were on average 82.1% (±5.8, SD; n = 20) and 75.4% (±2.2), respectively. Of the 852 protein groups, 60.3% were shared among all five 96-well replicates (Supporting Information Figure S-2A). Variability in peptide identifications is clearly associated with the stochastic sampling nature of a data-dependent MS^2 analysis, particularly in the context of low abundance proteins.[35] Indeed, proteins surveyed exclusively in a single replicate were represented to approximately 75% by a single unique peptide, thus supporting the notion of peptide identification variability in the low protein abundance range. As shown in Figure S-2B (Supporting Information), the average percentage of shared peptide and protein identifications among LC-MS replicates was 82.2% (±0.9, n = 5) and 75.9% (±1.5), respectively. These values were almost indistinguishable from those of the 96-well replicates. We conclude that replication of 96FASP analysis in the wells of a single plate introduces relatively low variability at the sample preparation stage. The LFQ analysis tool was used to demonstrate low quantitative differences for the 513 proteins shared among all five 96-well replicates. The LFQ algorithms sum the normalized peptide intensities for a given protein. The tool has been widely used for proteome-wide relative quantification.[28,39−41] The heat map for plate I displayed in Figure 2 (right panel) visualizes the overall high similarity of protein abundances across all five experiments and was confirmed by pairwise Pearson correlation analysis. The Pearson correlation R values ranged from 0.980 to 0.993 among the 96-well replicates (Supporting Information

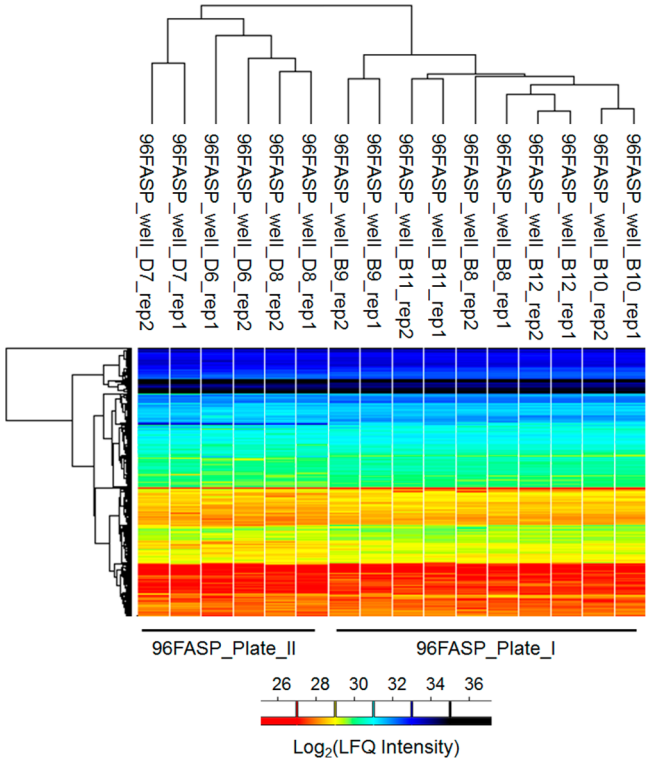


Figure 2. Unsupervised hierarchical clustering of LFQ intensities of urinary proteins identified in eight separate 96FASP wells. Two LC-MS replicates (rep1 and rep2) were acquired for each well. In the first experiment (plate I, right panel), five wells (B8−B12) were used to examine well-to-well repeatability. The experiment was repeated using three wells (D6−D8) from a different 96FASP (plate II, left panel). A high level of similarity in the abundances of matched proteins across experiments is visualized in the heap map. Pearson correlation analyses revealed average R values of 0.994 (±0.002, SD; n = 8) for LC-MS replicates, 0.985 (±0.005, n = 13) for 96-well replicates, and 0.967 for those 96-well replicates derived from different plates. For the LFQ analyses, the minimum number of unique peptides per protein used for quantitation was set at 2. Only those proteins quantified in all experiments (335 in total) were included in the clustering and correlation analyses.

dx.doi.org/10.1021/ac5008317 | *Anal. Chem.* 2014, 86, 5470−5477

Figure S-2C). In conclusion, we demonstrate that the analytic process starting with 96FASP features excellent well-to-well repeatability. Peptide/protein identification differences for the data from different wells appear to be primarily linked to variability at the LC-MS stage.

Experiments to assess 96FASP repeatability were continued using the same UC sample in a different 96-well plate (plate II, Figure 2, left panel). Regarding depth of coverage and repeatability of quantification of proteins (Supporting Information Figure S-3A), the proteomic profiles of three 96-well replicates were comparable to those of plate I. Importantly, the comparison of the data comparing individual wells from plate I versus plate II revealed equally low variability (Figure 2 and Supporting Information Figure S-3B). We suggest that experiments with a second 96-well filter plate, 3 weeks after completion of the first experiment, did not adversely affect repeatability. Another experiment was conducted to examine whether protein loading in a 96-well plate could increase without compromising sample processing times and data quality. UC samples with ~65 μg total protein were processed in five wells of a 96FASP plate. As expected, the centrifugation time to reduce the UA buffer volume from 200 μL to 50 μL increased to 70 to 90 min. With a 130 min LC gradient run, 5,810 unique peptides (±441, SD; n = 5) and 1,075 unique protein groups (±45) were identified on average at a 1% FDR. Quantitative assessments using LFQ-based intensity and Pearson correlation analyses revealed remarkably high R values (0.985 ± 0.006, n = 10; Supporting Information Figure S-4), when comparing the data among 96-well replicates. In summary, evaluations of experimental repeatability and urinary proteome coverage using the 96FASP method were encouraging with respect to interwell and interplate comparisons as well as the increase of protein loading amounts.

Data Comparison Using Single-Filter FASP Method. The same UC sample (10−15 μg protein) was processed using the single-filter device in triplicates followed by LC-MS, resulting in 7164 (±131, SD; n = 3) peptide identifications corresponding to 1063 (±15, SD) protein groups. In comparison with 96FASP (samples were prepared simultaneously), 4959 (±155) peptide identifications corresponding to 894 (±18) protein groups were obtained. The number of peptide and protein identifications employing 96FASP was 31% and 16% lower, respectively, compared to single-filter FASP (further illustrated in Supporting Information Figure S-5). The Pearson correlation coefficients ranged from 0.938 and 0.962 with an average R value of 0.947 (n = 9), comparing LFQ-based protein intensities for two data sets, one from *96FASP* and the other from single-filter FASP (Supporting Information Figure S-6). Likely causes of the moderately lower performance of 96FASP were the 3- to 4-fold longer centrifugation times in 96-well filter plates and differences in the material of the polypropylene-based collection plate versus the filter device. The plate material may adsorb more peptides and result in lower recovery compared to the single-filter collection device. We previously switched from a polystyrene-based collection plate which had revealed even higher peptide adsorption and low recovery assessed by LC-MS with a urinary protein load of 10−15 μg. Polystyrene-based lids may also compromise peptide recovery as suggested before.[30] We did not attempt to replace this lid with an alternative one. The 96-well plate lids also did not seal the plate. A buffer volume of at least 100−150 μL had to be added for the digestion step to prevent evaporation. In addition, as discussed in the first paragraph, differences in the MWCO of the membranes may influence urinary proteome coverage.

Represented Biological Functions in the UC Sample. Combining all protein identifications from 96FASP experiments performed with one UC sample, 10 974 identified unique peptide sequences corresponded to 2339 unique proteins (Supporting Information Table S-2). The 1247 protein groups (53.3%) were identified based on a single unique peptide. This result is consistent with data from a previous urinary proteome survey.[41] The urinary proteome has a high dynamic range of abundances (>10^5). The top 3 proteins accounted for almost 25% of the total protein mass, the top 20 proteins for approximately 50% of it. Therefore, lower abundance proteins are more challenging to identify unless further fractionation or immunodepletion techniques are employed. The average sequence coverage was 17.6%, and the average number of identified peptides per protein was 5. We assessed the dynamic range of protein abundances using the intensity-based absolute quantitation (iBAQ) algorithm.[27,33] This algorithm generates estimates of abundance for quantitative comparison of different proteins present in the same sample.[33] From five 96-well replicates, the median values of 854 proteins which passed iBAQ-integrated quality filters were calculated and plotted. As shown in Figure 3A, the

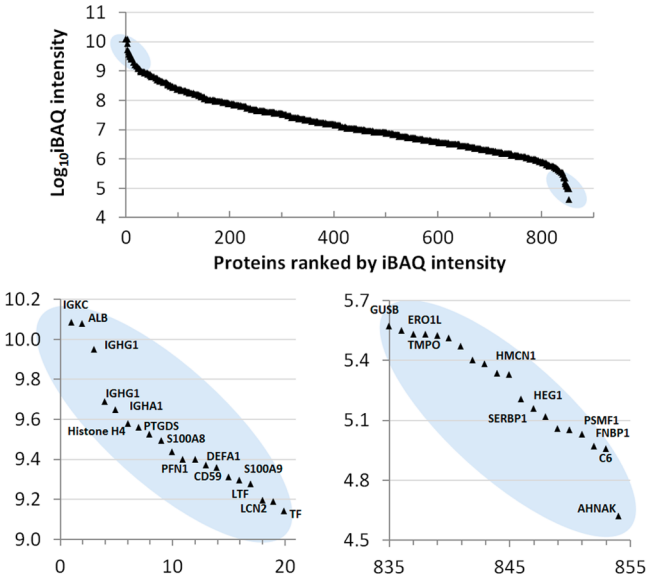


Figure 3. Dynamic protein abundance range for a UC sample associated with urinary tract infection. Median iBAQ values of five 96-well replicates were calculated for each of the 854 proteins identified with at least two unique peptides and plotted against the proteins' abundance rank. The highlighted areas show the most abundant and least abundant protein groups in the top graph, whereas the bottom two graphs depict these areas in a magnified view including the short names for 20 proteins in the two groups. IGK and IGH are immunoglobulin chains.

dynamic range of protein abundances in the urine sample was ~5.5 orders of magnitude. The 20 most abundant proteins made up 52.8% of the total protein mass. Immunoglobulin kappa chain (rank 1) and 37 other immunoglobulin subunits or isoforms contributed 26.9% to the total protein mass in the UC sample. Histones, proposed to have antimicrobial properties during infection,[42] contributed 5%. This included histones H4 (rank 5), H3.1 (rank 25), H2A (rank 28), and H2B (rank 54).

dx.doi.org/10.1021/ac5008317 | *Anal. Chem.* 2014, 86, 5470−5477

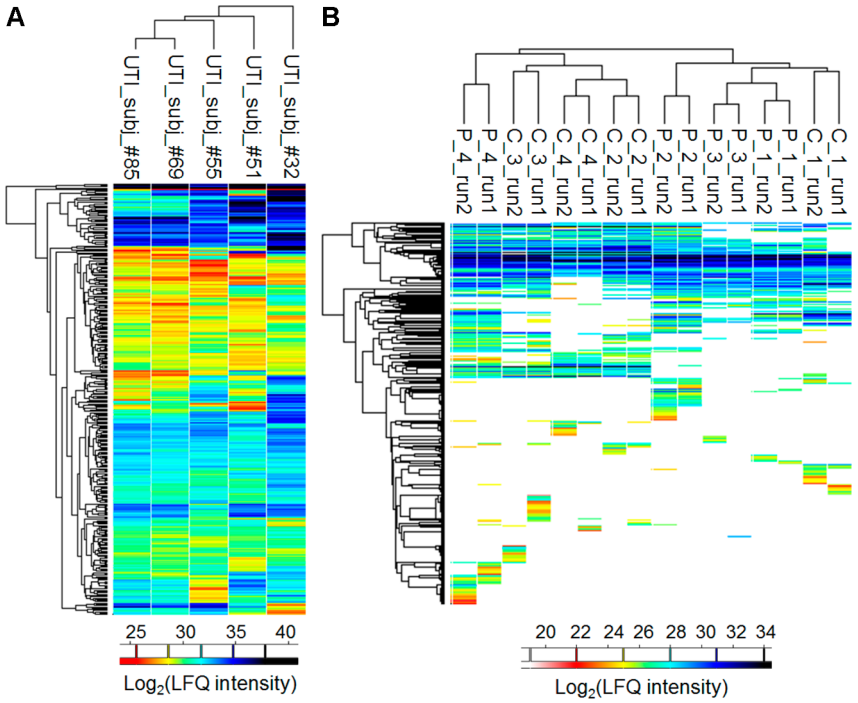


Figure 4. (A) Urinary pellets from five human subjects with potential urinary tract infections were analyzed by 96FASP and label free quantitation (LFQ). About 340 proteins were quantified in all five subjects and were used in the plot. (B) Heat map presenting differentially expressed proteins on the basis of LFQ-based quantitation from eight urine concentrate (UC) samples used in a Type 1 diabetes (T1D) project. About 1143 proteins were quantified by LFQ in at least one LC-MS replicate of the eight samples, and were used in the plot. Unsupervised hierarchical clustering generated two clusters. Each set of LC-MS replicates (run1 and run2) for a given sample clustered together. The data shows that, using highly parallel urine sample processing by 96FASP followed by shotgun proteomic analysis, quantitative protein profiles from technical replicates can be easily discerned from those originating from different human donors. P, T1D patient; C, control.

High histone quantities were not observed for other UC samples surveyed here and in two other urinary proteome surveys.[43,44] Increased quantities of leukocytes which infiltrate the urothelium, apoptose and lyse during a urinary tract infection explain the release of nuclear contents, including histones, into the urine. Urothelial cell exfoliation also occurs as a defense mechanism to wash invading bacteria out of the urinary tract.[10] Other proteins contributing to the innate immune defense, such as protein S100A8 (rank 9), neutrophil defensin 1 (rank 12), protein S100A9 (rank 15) and neutrophil gelatinase-associated lipocalin (rank 17) and lactotransferrin (rank 16), accounted for 7.4% of the total protein mass. A global functional analysis for this urinary proteome was performed using the DAVID bioinformatics resource with Gene Ontology (GO) Biological Process (BP) terms.[45] Significantly enriched BP terms for 1,200 proteins that were recognized by the DAVID tool from the UC sample were compared to a reference data set with a similar depth of coverage.[43] The latter pertained to the urinary proteome of acute appendicitis and control subjects. Our data set revealed enriched BP terms for acute inflammatory response (p-value of 3.66×10^{-24}) and response to wounding (p-value of 2.14×10^{-22}), with 48 and 120 protein identifications, respectively (Supporting Information Figure S-7). Since the UC sample was derived from a human subject with UTI symptoms, the analysis confirmed the functional importance of inflammatory and tissue regenerative processes in the urinary tract compared to a condition of distal inflammation (acute appendicitis).

96FASP for Urinary Tract Infection Diagnostics. Urinary pellet (UP) samples were isolated from specimens of human donors positive in at least two of three diagnostic tests for UTI (elevated leukocyte esterase activity; nitrite concentration; bacterial cell counts $>10^5$/mL urine). Five samples were analyzed using 96FASP. UP samples were resuspended in a denaturing solution, incubated and sonicated to achieve cell lysis. Considering higher phospholipid and lipid content in such lysates compared to UC samples that may clog filters, it was of interest to evaluate whether the 96FASP method permitted the use of equal loading aliquots of total protein (10–20 μg). Centrifugation times for the UP samples were not prolonged and digestions were equally efficient. An integrated database including the human proteins and protein sequences for microbial species causing~98% of all diagnosed UTIs was used for analysis of LC-MS data.[46] In four of them, microbial species *Escherichia coli* or *Klebsiella pneumoniae*, both are common causes of UTI, were confidently identified based on ≥10 bacteria proteins (Supporting Information Table S-3). Hierarchical clustering data shown in the heat map of Figure 4A revealed that the protein abundance patterns of subject number 85 and 69, each of which rich in *E. coli* proteins relative to all identified proteins clustered together. The corresponding Pearson correlation coefficient was 0.911 (Supporting Information Figure S-8). In conclusion, using 96FASP for the proteomic analysis of urinary pellets permitted the identification of pathogenic bacteria in urine. Quantitative protein profiles integrating human urinary and microbial proteins may be helpful to associate the presence of a specific infectious agent with the local immune response in the urinary tract. A metaproteomic approach aimed at discerning UTI from asymptomatic bacteriuria was also previously reported.[9]

Measuring Differences in Urinary Proteome Profiles Using 96FASP. To contrast the high repeatability of technical

dx.doi.org/10.1021/ac5008317 | *Anal. Chem.* 2014, 86, 5470–5477

replicates from a single UC sample with the considerable variability of urinary protein profiles derived from different human subjects, eight UC samples were prepared from a cohort related to a juvenile T1D project. Four specimens each belonged to the T1D and healthy sibling control cohorts. The purpose of the experiment was not to demonstrate that T1D biomarkers can be identified; rather, the intent was to show that urine samples analyzed by 96FASP and LC-MS followed by LFQ quantification result in protein abundance patterns clearly discerning the individuals from each other. Two LC-MS replicates were run for each sample. The numbers of protein identifications were 1247 and 1335 proteins on average for the T1D and healthy control cohort, respectively. Performing unsupervised hierarchical clustering (Figure 4B), all of the technical replicate sets cluster with each other. The Pearson correlation coefficients were much lower for LFQ data comparing different samples (R = 0.465 ± 0.185, n = 28) than for LC-MS replicates (R = 0.966 ± 0.017, n = 8). However, as recently reported, intraindividual variability in the urinary proteome can be high and presents additional challenges for a biomarkers discovery effort using urine.[47] This pilot study supports the notion that 96FASP, used in conjunction with a robust LC-MS method, is a credible approach to improve sample throughput without sacrificing data quality for a large-scale biomarker discovery project using urine.

CONCLUSION

Recently, Switzar et al.[30] published data processing protein samples of a HEK293T cell lysate in 96-well plates for shotgun proteomic analysis. We modified this method assessing its performance with a clinically relevant body fluid, quantitatively analyzed experimental repeatability including well-to-well and plate-to-plate variability, addressed questions of centrifugal speed and protein load capacity (loading range of 10−70 μg total protein) and applied the method to urinary pellets, a valuable sample source for UTI diagnostics. The method was successfully used to identify microbial species from several UP samples. Furthermore, a pilot project revealed a high level of clustering of quantitative urinary proteomic data derived from technical replicates. This was not the case when urine samples from different human subjects, including siblings, were compared. The experiments revealed the potential of the 96FASP urine sample processing to become a gold standard for high-throughput sample preparation in quantitative clinical proteomics investigations. Urine is collected noninvasively yielding plenty of protein,[48] can reveal evidence of renal and urogenital diseases[49] and diseases anatomically distant from the kidneys and urogenital tract. Examples are coronary artery disease,[50] acute appendicitis,[51] preeclampsia,[52] and Kawasaki disease.[53] The complexity of the human urinary proteome, the extensive post-translational processing of its proteins, the interindividual and intraindividual variability of protein content based on diet, exercise, sexual activity, and microbial colonization also render biomarker discovery projects more challenging than other sample sources.[5,47,48] Good disease and control cohort definitions, thorough metadata collection, efficient sample processing and the availability of large cohorts offer the best opportunities to discover and validate useful biomarkers.[48] The 96FASP method will contribute to such efforts, and its application is not limited to urine but includes other body fluids, extracellular matrix, complex tissues, and tumors.

ASSOCIATED CONTENT

Supporting Information

Additional material as described in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: yayu@jcvi.org. Tel: (301) 795-7868.
*E-mail: rpieper@jcvi.org. Tel: (301) 795-7605. Fax: (301) 795-7070.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported in part by Grant NIH-1R01GM103598 (National Institute of General Medical Sciences). We would like to thank Joseph Kramer and Susan Gigler (EMD Millipore Chemicals) for helpful technical discussions, Dr. Ramana Madupu (JCVI) for the provision of urine specimens from a Type 1 diabetes project (Grant NIH-1DP3DK094343-01, NIDDK), Dr. Nicholas Cacciabeve (Shady Grove Adventist Hospital, Rockville, MD), and Cynthia Bowman-Holston (Quest Diagnostics Nichols Institute) for the provision of urine specimens from a urinary tract infection study.

REFERENCES

(1) Hüttenhain, R.; Soste, M.; Selevsek, N.; Röst, H.; Sethi, A.; Carapito, C.; Farrah, T.; Deutsch, E. W.; Kusebauch, U.; Moritz, R. L.; Niméus-Malmström, E.; Rinner, O.; Aebersold, R. *Sci. Transl. Med.* **2012**, *4*, No. 142ra94.
(2) Shen, Y.; Jacobs, J. M.; Camp, D. G.; Fang, R.; Moore, R. J.; Smith, R. D.; Xiao, W.; Davis, R. W.; Tompkins, R. G. *Anal. Chem.* **2004**, *76*, 1134−44.
(3) Pisitkun, T.; Johnstone, R.; Knepper, M. A. *Mol. Cell. Proteomics* **2006**, *5*, 1760−71.
(4) Spahr, C. S.; Davis, M. T.; McGinley, M. D.; Robinson, J. H.; Bures, E. J.; Beierle, J.; Mort, J.; Courchesne, P. L.; Chen, K.; Wahl, R. C.; Yu, W.; Luethy, R.; Patterson, S. D. *Proteomics* **2001**, *1*, 93−107.
(5) Pieper, R.; Gatlin, C. L.; McGrath, A. M.; Makusky, A. J.; Mondal, M.; Seonarain, M.; Field, E.; Schatz, C. R.; Estock, M. A.; Ahmed, N.; Anderson, N. G.; Steiner, S. *Proteomics* **2004**, *4*, 1159−74.
(6) Adachi, J.; Kumar, C.; Zhang, Y.; Olsen, J.; Mann, M. *Genome Biol.* **2006**, *7*, R80.
(7) Cash, P. *Expert Rev. Proteomics* **2014**, *11*, 43−58.
(8) Hagan, E. C.; Mobley, H. L. T. *Infect. Immunol.* **2007**, *75*, 3941−9.
(9) Fouts, D.; Pieper, R.; Szpakowski, S.; Pohl, H.; Knoblach, S.; Suh, M.-J.; Huang, S.-T.; Ljungberg, I.; Sprague, B.; Lucas, S.; Torralba, M.; Nelson, K.; Groah, S. *J. Transl. Med.* **2012**, *10*, 174.
(10) Nielubowicz, G. R.; Mobley, H. L. T. *Nat. Rev. Urol.* **2010**, *7*, 430−41.
(11) Wolfe, A. J.; Toh, E.; Shibata, N.; Rong, R.; Kenton, K.; FitzGerald, M.; Mueller, E. R.; Schreckenberger, P.; Dong, Q.; Nelson, D. E.; Brubaker, L. *J. Clin. Microbiol.* **2012**, *50*, 1376−83.
(12) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R. *Chem. Rev.* **2013**, *113*, 2343−94.
(13) Yu, Y.-Q.; Gilar, M.; Lee, P. J.; Bouvier, E. S. P.; Gebler, J. C. *Anal. Chem.* **2003**, *75*, 6023−8.
(14) Nagaraj, N.; Lu, A.; Mann, M.; Wiśniewski, J. R. *J. Proteome Res.* **2008**, *7*, 5028−32.
(15) Sharma, R.; Dill, B. D.; Chourey, K.; Shah, M.; VerBerkmoes, N. C.; Hettich, R. L. *J. Proteome Res.* **2012**, *11*, 6008−18.
(16) Zhou, J.-Y.; Dann, G. P.; Shi, T.; Wang, L.; Gao, X.; Su, D.; Nicora, C. D.; Shukla, A. K.; Moore, R. J.; Liu, T.; Camp, D. G., II; Smith, R. D.; Qian, W.-J. *Anal. Chem.* **2012**, *84*, 2862−7.

(17) Manza, L. L.; Stamer, S. L.; Ham, A.-J. L.; Codreanu, S. G.; Liebler, D. C. *Proteomics* **2005**, *5*, 1742−5.
(18) Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. *Nat. Methods* **2009**, *6*, 359−62.
(19) Tang, Y.; Underwood, A.; Gielbert, A.; Woodward, M. J.; Petrovska, L. *Appl. Environ. Microbiol.* **2013**, DOI: 10.1128/AEM.02472-13.
(20) Skeie, J. M.; Mahajan, V. B. *PLoS One* **2013**, *8*, No. e82140.
(21) Shimwell, N. J.; Bryan, R. T.; Wei, W.; James, N. D.; Cheng, K. K.; Zeegers, M. P.; Johnson, P. J.; Martin, A.; Ward, D. G. *Br. J. Cancer* **2013**, *108*, 1854−61.
(22) Raimondo, F.; Corbetta, S.; Morosi, L.; Chinello, C.; Gianazza, E.; Castoldi, G.; Di Gioia, C.; Bombardi, C.; Stella, A.; Battaglia, C.; Bianchi, C.; Magni, F.; Pitto, M. *Mol. Biosyst.* **2013**, *9*, 1139−46.
(23) Welberry Smith, M. P.; Zougman, A.; Cairns, D. A.; Wilson, M.; Wind, T.; Wood, S. L.; Thompson, D.; Messenger, M. P.; Mooney, A.; Selby, P. J.; Lewington, A. J. P.; Banks, R. E. *Kidney Int.* **2013**, *84*, 1214−25.
(24) Nagaraj, N.; Wisniewski, J. R.; Geiger, T.; Cox, J.; Kircher, M.; Kelso, J.; Paabo, S.; Mann, M. *Mol. Syst. Biol.* **2011**, *7*, No. 548, DOI: 10.1038/msb.2011.81.
(25) Wiśniewski, J. R.; Zougman, A.; Mann, M. *J. Proteome Res.* **2009**, *8*, 5674−8.
(26) Han, D.; Moon, S.; Kim, Y.; Kim, J.; Jin, J.; Kim, Y. *Proteomics* **2013**, *13*, 2984−8.
(27) Geiger, T.; Wehner, A.; Schaab, C.; Cox, J.; Mann, M. *Mol. Cell. Proteomics* **2012**, *11*, No. M111.014050.
(28) Wisniewski, J. R.; Ostasiewicz, P.; Dus, K.; Zielinska, D. F.; Gnad, F.; Mann, M. *Mol. Syst. Biol.* **2012**, DOI: 10.1038/msb.2012.44.
(29) Mischak, H.; Kolch, W.; Aivaliotis, M.; Bouyssié, D.; Court, M.; Dihazi, H.; Dihazi, G. H.; Franke, J.; Garin, J.; de Peredo, A. G.; Iphöfer, A.; Jänsch, L.; Lacroix, C.; Makridakis, M.; Masselon, C.; Metzger, J.; Monsarrat, B.; Mrug, M.; Norling, M.; Novak, J.; Pich, A.; Pitt, A.; Bongcam-Rudloff, E.; Siwy, J.; Suzuki, H.; Thongboonkerd, V.; Wang, L.-S.; Zoidakis, J.; Zürbig, P.; Schanstra, J. P.; Vlahou, A. *Proteomics: Clin. Appl.* **2010**, *4*, 464−78.
(30) Switzar, L.; van Angeren, J.; Pinkse, M.; Kool, J.; Niessen, W. M. A. *Proteomics* **2013**, *13*, 2980−3.
(31) Chen, Y.-T.; Chen, H.-W.; Domanski, D.; Smith, D. S.; Liang, K.-H.; Wu, C.-C.; Chen, C.-L.; Chung, T.; Chen, M.-C.; Chang, Y.-S.; Parker, C. E.; Borchers, C. H.; Yu, J.-S. *J. Proteomics* **2012**, *75*, 3529−45.
(32) Cox, J.; Matic, I.; Hilger, M.; Nagaraj, N.; Selbach, M.; Olsen, J. V.; Mann, M. *Nat. Protoc.* **2009**, *4*, 698−705.
(33) Schwanhausser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. *Nature* **2011**, *473*, 337−42.
(34) Wiśniewski, J. R.; Zielinska, D. F.; Mann, M. *Anal. Biochem.* **2011**, *410*, 307−9.
(35) Tabb, D. L.; Vega-Montoto, L.; Rudnick, P. A.; Variyath, A. M.; Ham, A.-J. L.; Bunk, D. M.; Kilpatrick, L. E.; Billheimer, D. D.; Blackman, R. K.; Cardasis, H. L.; Carr, S. A.; Clauser, K. R.; Jaffe, J. D.; Kowalski, K. A.; Neubert, T. A.; Regnier, F. E.; Schilling, B.; Tegeler, T. J.; Wang, M.; Wang, P.; Whiteaker, J. R.; Zimmerman, L. J.; Fisher, S. J.; Gibson, B. W.; Kinsinger, C. R.; Mesri, M.; Rodriguez, H.; Stein, S. E.; Tempst, P.; Paulovich, A. G.; Liebler, D. C.; Spiegelman, C. *J. Proteome Res.* **2009**, *9*, 761−76.
(36) Wiśniewski, J. R.; Duś, K.; Mann, M. *Proteom. Clin. Appl.* **2013**, *7*, 225−33.
(37) Lu, A.; Wiśniewski, J. R.; Mann, M. *J. Proteome Res.* **2009**, *8*, 2418−25.
(38) Wisniewski, J. R.; Mann, M. *Nat. Methods* **2009**, *6*, 785−6.
(39) Hubner, N. C.; Bird, A. W.; Cox, J.; Splettstoesser, B.; Bandilla, P.; Poser, I.; Hyman, A.; Mann, M. *J. Cell Biol.* **2010**, *189*, 739−54.
(40) Smaczniak, C.; Li, N.; Boeren, S.; America, T.; van Dongen, W.; Goerdayal, S. S.; de Vries, S.; Angenent, G. C.; Kaufmann, K. *Nat. Protoc.* **2012**, *7*, 2144−58.
(41) Kosanam, H.; Prassas, I.; Chrystoja, C. C.; Soleas, I.; Chan, A.; Dimitromanolakis, A.; Blasutig, I. M.; Rückert, F.; Gruetzmann, R.; Pilarsky, C.; Maekawa, M.; Brand, R.; Diamandis, E. P. *Mol. Cell. Proteomics* **2013**, *12*, 2820−32.
(42) Brinkmann, V.; Reichard, U.; Goosmann, C.; Fauler, B.; Uhlemann, Y.; Weiss, D. S.; Weinrauch, Y.; Zychlinsky, A. *Science* **2004**, *303*, 1532−5.
(43) Kentsis, A.; Monigatti, F.; Dorff, K.; Campagne, F.; Bachur, R.; Steen, H. *Proteom. Clin. Appl.* **2009**, *3*, 1052−61.
(44) Zheng, J.; Liu, L.; Wang, J.; Jin, Q. *BMC Genomics* **2013**, *14*, 777.
(45) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. *Nat. Protoc.* **2008**, *4*, 44−57.
(46) Ronald, A. *Am. J. Med.* **2002**, *113*, 14−9.
(47) Nagaraj, N.; Mann, M. *J. Proteome Res.* **2010**, *10*, 637−45.
(48) Decramer, S.; de Peredo, A. G.; Breuil, B.; Mischak, H.; Monsarrat, B.; Bascands, J.-L.; Schanstra, J. P. *Mol. Cell. Proteomics* **2008**, *7*, 1850−62.
(49) Rodríguez-Suárez, E.; Siwy, J.; Zürbig, P.; Mischak, H. *Biochim. Biophys. Acta, Proteins Proteomics* **2014**, *1844* (5), 884−98.
(50) Zimmerli, L. U.; Schiffer, E.; Zürbig, P.; Good, D. M.; Kellmann, M.; Mouls, L.; Pitt, A. R.; Coon, J. J.; Schmieder, R. E.; Peter, K. H.; Mischak, H.; Kolch, W.; Delles, C.; Dominiczak, A. F. *Mol. Cell. Proteomics* **2008**, *7*, 290−8.
(51) Kentsis, A.; Lin, Y. Y.; Kurek, K.; Calicchio, M.; Wang, Y. Y.; Monigatti, F.; Campagne, F.; Lee, R.; Horwitz, B.; Steen, H.; Bachur, R. *Ann. Emerg. Med.* **2010**, *55*, 62−70.
(52) Carty, D. M.; Siwy, J.; Brennand, J. E.; Zürbig, P.; Mullen, W.; Franke, J.; McCulloch, J. W.; North, R. A.; Chappell, L. C.; Mischak, H.; Poston, L.; Dominiczak, A. F.; Delles, C. *Hypertension* **2011**, *57*, 561−9.
(53) Kentsis, A.; Shulman, A.; Ahmed, S.; Brennan, E.; Monuteaux, M. C.; Lee, Y.-H.; Lipsett, S.; Paulo, J. A.; Dedeoglu, F.; Fuhlbrigge, R.; Bachur, R.; Bradwin, G.; Arditi, M.; Sundel, R. P.; Newburger, J. W.; Steen, H.; Kim, S. *EMBO Mol. Med.* **2013**, *5*, 210−20.