Maximizing Peptide Identification Events in Proteomic Workflows Using Data-Dependent Acquisition (DDA)*

Nicholas W. Bateman‡¶¶, Scott P. Goulding‡§¶¶, Nicholas J. Shulman¶, Avinash K. Gadok‡, Karen K. Szumlinski¶, Michael J. MacCoss¶, and Christine C. Wu‡§**

Current analytical strategies for collecting proteomic data using data-dependent acquisition (DDA) are limited by the low analytical reproducibility of the method. Proteomic discovery efforts that exploit the benefits of DDA, such as providing peptide sequence information, but that enable improved analytical reproducibility, represent an ideal scenario for maximizing measureable peptide identifications in “shotgun”-type proteomic studies. Therefore, we propose an analytical workflow combining DDA with retention time aligned extracted ion chromatogram (XIC) areas obtained from high mass accuracy MS1 data acquired in parallel. We applied this workflow to the analyses of sample matrixes prepared from mouse blood plasma and brain tissues and observed increases in peptide detection of up to 30.5% due to the comparison of peptide MS1 XIC areas following retention time alignment of co-identified peptides. Furthermore, we show that the approach is quantitative using peptide standards diluted into a complex matrix. These data revealed that peptide MS1 XIC areas provide linear response of over three orders of magnitude down to low femtomole (fmol) levels. These findings argue that augmenting “shotgun” proteomic workflows with retention time alignment of peptide identifications and comparative analyses of corresponding peptide MS1 XIC areas improve the analytical performance of global proteomic discovery methods using DDA. Molecular & Cellular Proteomics 13: 10.1074/mcp.M112.026500, 329–338, 2014.

Label-free methods in mass spectrometry-based proteomics, such as those used in common “shotgun” proteomic studies, provide peptide sequence information as well as relative measurements of peptide abundance (1–3). A common data acquisition strategy is based on data-dependent acquisition (DDA)1 where the most abundant precursor ions are selected for tandem mass spectrometry (MS/MS) analysis (1–2). DDA attempts to minimize redundant peptide precursor selection and maximize the depth of proteome coverage (2). However, the analytical reproducibility of peptide identifications obtained using DDA-based methods result in <75% overlap between technical replicates (3–4). Comparisons of peptide identifications between replicate analyses have shown that the rate of new peptide identifications increases sharply following two replicate sample injections and gradually tapers off after approximately five replicate injections (4). This phenomenon is due, in part, to the semirandom sampling of peptides in a DDA experiment (5).

Alternate label-free methods focused on measuring the abundance of intact peptide ions, such as the accurate mass and time tag (AMT) approach (6–8, 42), are aimed at differential analyses of extracted ion chromatogram (XIC) areas integrated from high mass accuracy peptide precursor mass spectra (MS1 spectra) exhibiting discrete chromatographic elution times. This method is particularly powerful for the analysis of post-translationally modified (PTM) peptides as pairing the low abundance of PTM candidates with the variable nature of DDA complicates comparisons between samples (9, 43). However, label-free strategies focused on the analysis of peptide MS1 XIC areas are dependent on a priori knowledge of peptide ions and retention times (2–10). Thus, this phenomenon is due, in part, to the semirandom sampling of peptides in a DDA experiment (5).

Label-free methods in mass spectrometry-based proteomics, such as those used in common “shotgun” proteomic studies, provide peptide sequence information as well as relative measurements of peptide abundance (1–3). A common data acquisition strategy is based on data-dependent acquisition (DDA)1 where the most abundant precursor ions are selected for tandem mass spectrometry (MS/MS) analysis (1–2). DDA attempts to minimize redundant peptide precursor selection and maximize the depth of proteome coverage (2). However, the analytical reproducibility of peptide identifications obtained using DDA-based methods result in <75% overlap between technical replicates (3–4). Comparisons of peptide identifications between replicate analyses have shown that the rate of new peptide identifications increases sharply following two replicate sample injections and gradually tapers off after approximately five replicate injections (4). This phenomenon is due, in part, to the semirandom sampling of peptides in a DDA experiment (5).

Alternate label-free methods focused on measuring the abundance of intact peptide ions, such as the accurate mass and time tag (AMT) approach (6–8, 42), are aimed at differential analyses of extracted ion chromatogram (XIC) areas integrated from high mass accuracy peptide precursor mass spectra (MS1 spectra) exhibiting discrete chromatographic elution times. This method is particularly powerful for the analysis of post-translationally modified (PTM) peptides as pairing the low abundance of PTM candidates with the variable nature of DDA complicates comparisons between samples (9, 43). However, label-free strategies focused on the analysis of peptide MS1 XIC areas are dependent on a priori knowledge of peptide ions and retention times (2–10). Thus,

1 The abbreviations used are: DDA, data-dependent acquisition; XIC, extracted ion chromatogram; MS1, precursor mass spectra; MS/MS, tandem mass spectrometry; AMT, accurate mass and time tag; PTM, posttranslationally modified; v/v, volume/volume; IAA, 2-iodoacetamide; FA, formic acid; Hgb, hemoglobin; nLC-MS/MS, nano liquid chromatography tandem mass spectrometry; i.d., inner diameter; o.d., outer diameter; R, resolution; HCD, higher-energy collisional dissociation; ppm, parts per million; K, lysine; R, arginine; CV, coefficient of variation; PSD95, postsynaptic density protein-95; NMDA, N-methyl-D-aspartate; mGluR1A, metabotropic glutamate receptor 1 A; DGL-alpha, Sn1-specific diacylglycerol lipase alpha; Cp, ceruloplasmin; Cfd, complement factor D; ApoE, lipoprotein scavenger apolipoprotein E; KO, knockout; WT, wild-type; HL, hemolysis.
prospective analyses of samples are needed to assess peptides and their respective retention times. This prospective analysis may not be possible for reagent-limited samples. Further, the usage of previously established peptide features in the analysis of different sample types can be confounded by unknown matrix effects that can produce variable retention time characteristics and peptide ion suppression (2). Therefore, proteomic strategies that make use of DDA, to provide peptide sequence information and identify features within the sample, but that also use MS1 data for comparisons between samples, represent an ideal combination for maximizing measurable peptide identification events in “shotgun” proteomic discovery analyses.

Here we describe an analytical workflow that combines traditional DDA methods with the analysis of retention time aligned XIC areas extracted from high mass accuracy peptide precursor MS1 spectra. This method resulted in a 25.1% (±6.6%) increase in measurable peptide identification events across samples of diverse composition because of the inferential extraction of peptide MS1 XIC areas in sample sets lacking corresponding MS/MS events. These findings were observed in measurements of peptide MS1 XIC abundances using sample types ranging from tryptic digests of olfactory bulb tissues dissected from Homer2 knockout and wild-type mice to mouse blood plasma exhibiting differential levels of hemolysis. We further establish that this method is quantitative using a dilution series of known bovine standard peptide concentrations spiked into mouse blood plasma. These data show that comparative analysis between samples should be performed using peptide MS1 data as opposed to semirandomly sampled peptide MS/MS data. This approach maximizes the number of peptides that can be compared between samples.

**Experimental Procedures**

**Preparation of a Dilution Series of Bovine Peptide Standard in Mouse Blood Plasma Peptide Digests**—The concentration of protein in commercial mouse blood plasma (D408-04-0050, Rockland Immunochemicals, Inc. Gilbertsville, PA) was determined using the DC Protein Assay (Bio-Rad Laboratories, Hercules, CA) and equivalent amounts of protein were immunodepleted using a MARS-mouse 3 column (Agilent Technologies, Santa Clara, CA). Briefly, octoflary bulb tissues dissected from Homer2 knockout and wild-type mice to mouse blood plasma exhibiting differential levels of hemolysis. We further establish that this method is quantitative using a dilution series of known bovine standard peptide concentrations spiked into mouse blood plasma. These data show that comparative analysis between samples should be performed using peptide MS1 data as opposed to semirandomly sampled peptide MS/MS data. This approach maximizes the number of peptides that can be compared between samples.

**Preparation of a Synaptic Density Fraction From Olfactory Bulb Digests Harvested From Homer2 Knockout and Wild-type Mice**—Olfactory bulb tissue was dissected from three Homer2 knockout (129/Sv X C57BL/6) (12) and three wild-type male mice obtained from Dr. Karen K. Szumilinski (University of California, Santa Barbara) and subjected to a modified postsynaptic density enrichment strategy previously described by Phillips et al. (13–14). Briefly, olfactory bulbs were pooled according to genotype and homogenized using a 15 ml Potter-Elvejem tissue grinder with PTFE pestle (Wheaton Science Products, Millville, NJ) over wet ice in 3 ml of a 0.32 mM sucrose solution containing 0.1 mM CaCl₂, 1 mM MgCl₂, and 1X phosphatase inhibitor mixture (Halt Phosphatase Inhibitor Mixture, Thermo Scientific, Rockford, IL). The homogenate was brought to a working concentration of 1.25 mM sucrose via the addition of 17 ml of a sucrose solution containing 0.1 mM CaCl₂ and 1X phosphatase inhibitor mixture, overlaid with 10 ml of a 1.0 mM sucrose solution containing 0.1 mM CaCl₂, and centrifuged at 100,000 x g/4 °C for 3 h in a Sorvall WX90 Ultracentrifuge (Thermo Scientific, Rockford, IL). Synaptosomes (~1 ml) were collected from the 1.0/1.25 M interface and incubated for 30 min in 10 ml of 20 mM Tris-buffered saline containing 1% Triton X-100 at pH 6. Samples were then centrifuged at 40,000 x g/4 °C, supernatants were decanted, and pellets were incubated in 1 ml of a second buffer containing 20 mM Tris-buffered saline and 1% Triton X-100 at pH 8. Samples were centrifuged at 40,000 x g/4 °C in a Sorvall MTX 150 Micro-Ultracentrifuge (Thermo Scientific, Rockford, IL) to obtain enriched synaptic density pellets. Tryptic digests of synaptic density proteins were prepared identically to methods described above except samples were solubilized in 0.2% RapiGest containing 50 mM ammonium bicarbonate and subjected to a cleanup step using Agilent cleanup C-18 pipette tips (Agilent Technologies, Santa Clara, CA) according to manufacturer’s recommendations. Lyophilized peptides were resuspended at a final concentration of 0.25 μg/μl in 0.1% (v/v) FA. Further, peptide identifications (method detailed below) corresponding to cutaneous keratins were removed from this sample set resulting in a loss of 1.31% of total peptide identifications.

**Preparation of Hemolysed Blood Plasma Digests**—Hemolysates were prepared from a 1:4 mixture of whole blood obtained by cardiac puncture of a C57BL/6 mouse and commercial mouse plasma (Rockland Immunochemicals, Inc. Gilbertsville, PA). Blood plasma mixtures were subjected to (1) gentle shaking by hand for 30 s to emulsify excessive mixing of blood samples post-collection (15–16) (low hemolysis condition) or (2) sampling of the supernatant and a fraction of the blood cell pellet emulsifying poor separator barrier integrity (15) following centrifugation (high hemolysis condition). Blood plasma was prepared by centrifuging whole blood mixtures for 10 min at 50,000 rpm/4 °C in a 5417R refrigerated microcentrifuge (Eppendorf, Hauppauge, NY) and 0.1 mM phenylmethylsulfonyl fluoride (PMSF, Thermo Scientific, Rockford, IL) was then added to plasma supernatants before storage at −80 °C. The extent of hemolysis was determined by direct spectrophotometric measurement of hemoglobin (Hgb) concentration as per the all method detailed by Fairbanks et al. (17) using an Epoch Microplate Spectrophotometer (BioTek, Winooski, VT).
Plasma samples were immunodepleted on a MARS-mouse 3 column and quantitated by DC protein assay as described above. Twenty micrograms of depleted protein was run ~2.0 cm into a 5.0% bis-acrylamide stacking gel and processed for in-gel digestion as previously described(18) using 5 mM DTT in 25 mM ammonium bicarbonate and incubation at 60 °C and 15 mM IAA in 25 mM ammonium bicarbonate for reduction and alklylation of reactive cysteine residues, respectively. Lyophilized peptides were resuspended at a final concentration of 0.2 μg/μl in 0.1% (v/v) FA.

nLC-MS/MS Analyses—Peptide digests corresponding to the bovine/mouse plasma dilution series or Homer2 knockout and wild-type olfactory bulb tissues were resolved by nLC-MS/MS using an EASY-nLC II (Thermo Scientific, San Jose, CA) coupled online via electrospray ionization (ESI) to an Orbitrap Elite mass spectrometer (Thermo Scientific, San Jose, CA). Randomized, triplicate injections of 1.0 μg of peptide extracts were resolved on a 75 μm i.d. by 360 μm o.d. by 300 mm long fused silica capillary column (Polymer Microtechnologies, Phoenix, AZ) slurry-packed in-house with 5 μm particle size, 125 Å pore size C-18 silica-bonded stationary phase (phenomenex, Torrance, CA). After sample injection, peptides were eluted from the column using a gradient of 2% mobile phase B (99.9% (v/v) acetonitrile (ACN)/0.1% (v/v) FA) to 12% mobile phase B over 60 min, stepping up to 32% mobile phase B for an additional 60 min at a constant flow rate of 200 nL/min followed by a column wash consisting of 80% mobile phase B for an additional 6 min. The column used for olfactory bulb tissues was heated to 40 °C. The Orbitrap Elite mass spectrometer was configured to collect high resolution (R = 60,000 at m/z 400) broadband mass spectra (m/z 400–1400) from which the ten most abundant peptide molecular ions, dynamically determined from the MS1 scan, were selected for MS/MS using a relative collision-induced dissociation (CID) energy of 35%. Peptide digests from differentially hemolyzed blood plasma were resolved by nLC on a 75 μm i.d. by 360 μm o.d. by 250 mm long fused silica capillary column as described above, but coupled online via ESI to a Q Exactive mass spectrometer (Thermo Scientific, San Jose, CA). After sample injection, peptides were eluted from the column using a linear gradient of 2% to 42% mobile phase B over 120 min at a constant flow rate of 250 nL/min followed by a column wash consisting of 80% mobile phase B for an additional 10 min. The Q Exactive mass spectrometer was configured to collect high resolution (R = 70,000 at m/z 200) broadband mass spectra (m/z 400–1400) and MS/MS events (R = 17,500 at an automatic gain control target of 2.0E+5 and an underfill ratio of 20%) on the five most abundant peptide molecular ions dynamically determined from the MS1 scan using a relative higher-energy collisional dissociation (HCD) energy of 35%. A default dynamic exclusion setting was used to minimize redundant selection of peptides.

Peptide Identification and Differential Analysis of Peptide MS1 XIC Areas—Peptide identifications were obtained by searching nLC-MS/MS RAW file data using a pipeline that consisted of extraction of retention time windows of ± 5.0 min surrounding initial peptide identification events. Deconvolution scores were then calculated for integration events, reflecting the fit of theoretical peptide isotope distributions to observed experimental MS1 spectra (23, 26). Peptide MS1 XIC data that was successfully integrated, as defined by whether a MS1 peak corresponding to the theoretical peptide mass derived from the MS/MS identification was integrated within the retention time window surrounding said MS/MS event, were then exported for further analysis (supplemental Tables S2A–S2C). Individual peptide MS1 XIC areas for a given sample injection were normalized relative to the maximum sum of the total peptide MS1 XIC areas observed across all injections in a sample cohort. Significant, differentially abundant peptides were calculated using a two-tailed Student’s t test assuming equal variance. Resulting p-values were converted to q-values using the R software program TkQVALUE (27) and differential peptide abundances exhibiting q-values ≤ 0.05 that were successfully integrated across three replicate injections and a minimum of two injections in a comparison condition were considered significant.

RESULTS AND DISCUSSION

The data analysis workflow used is detailed in Fig. 1. All samples were subjected to a typical DDA workflow consisting of tryptic digestion of complex samples matrices, nLC-MS/MS analysis of peptide digests, and subsequent identification of peptides using established bioinformatic pipelines. XIC areas for peptide MS1 spectra corresponding to peptide identification events were then extracted from corresponding RAW data files using a two-step process (Fig. 1). First, sample sets were aligned using an iterative strategy fitting a straight line through retention times observed for peptides co-identified between nLC-MS/MS runs (Fig. 1A). The peptide furthest from the best fit linear fit of the two runs was removed and the
linear regression repeated. This process was performed iteratively until a Pearson score of $R^2 > 0.99$ was obtained. Second, peptide MS1 XIC areas were integrated within retention time windows of ±5.0 min surrounding a given MS/MS event and restricted to a 20 ppm mass tolerance relative to theoretical masses obtained from peptide identification results (Fig. 1B). Resulting peptide MS1 XIC areas were then statistically compared to assess significantly altered peptides.

Representative results of this pipeline when applied to triplicate replicate injections of highly hemolyzed mouse blood plasma are shown in Fig. 2. These results illustrate a 30.5% increase in the number of measureable peptide events identified between replicate sample injections based on successfully integrated, retention time aligned peptide MS1 XIC areas versus peptide identifications events alone ($p$-value $= 5.8 \times 10^{-5}$ for increases in peptides co-identified between technical replicates).

Quantitation of Bovine Standard Peptides Diluted in Mouse Blood Plasma—A six-point dilution curve of tryptic peptides derived from an equimolar mix of six bovine proteins was diluted into a trypptic digest of immunodepleted mouse plasma. Samples were subjected to nLC-MS/MS analysis in triplicate where total bovine peptides analyzed on-column ranged from 1.0 μg to 0.01 μg (equivalent to $\sim 250 - 2.5$ fmol) in a constant matrix of 1.0 μg bovine/mouse plasma digest mix. A total of 588 peptides were identified for all six bovine standard proteins across all sample runs and analysis of peptide MS1 XIC areas were performed for 560 of these candidates (reflecting a loss of 4.8% because of unsuccessful peak integration events). Peptide MS1 XIC areas calculated
for representative peptides corresponding to four bovine proteins, i.e. alpha casein (IPI00706094 & IPI00698843, 62 peptides identified, 58.9% sequence coverage), beta-lactoglobulin (gi 2194089, 70 peptides, 59.2% sequence coverage), carbonic anhydrase (IPI00716246, 143 peptides, 81.8% sequence coverage), and glutamate dehydrogenase (gi 118533, 105 peptides, 55.1% sequence coverage) are detailed in Fig. 3. These data revealed that peptide MS1 XIC areas can be detected linearly and quantitatively to low fmol levels (mean $R^2 = 0.9 \pm 0.33$ for all bovine peptides analyzed).

A. Alpha Casein
K.EGIHAQQKE

B. Beta-Lactoglobulin
K.IDALNENKV

C. Carbonic Anhydrase
K.DGPLTGTYR.L

D. Glutamate Dehydrogenase
K.YNLGLDLRT

Fig. 3. Quantitative analysis of bovine peptide MS1 XIC areas in a dilution series of six equimolar bovine standard proteins in immunodepleted mouse blood plasma. Plots of peptide MS1 XIC areas versus bovine standard concentration across a six-point dilution series for four representative bovine standard peptides corresponding to alpha casein (A), beta-lactoglobulin (B), carbonic anhydrase (C), and glutamate dehydrogenase (D). Results revealed that bovine standard peptides diluted in immunodepleted mouse blood plasma are quantifiable to low fmol ranges. Mean $R^2 = 0.9 \pm 0.33$ for all bovine peptides analyzed.

Differential Analysis of a Synaptic Density Fraction From Homer2 Knockout and Wild-type Olfactory Bulb Digests—This sample set provided a complex sample matrix that was used to test the robustness of this analytical workflow in identifying peptides with altered abundances in the absence of Homer2, a postsynaptic density scaffolding protein (28). Triplicate injections of pooled olfactory bulb tissues dissected from a cohort of Homer2 knockout and wild-type mice resulted in a total of 11,973 peptide identifications (7857 ± 395; CV = 5.03%, across technical replicate injections) with 8713 of these candidates being identified between sample sets. Comparison of the peptide MS1 XIC areas was possible for 11,971 of these events (11,847 ± 65; CV = 0.55%) and yielded a total of 11,953 events being detected between sample sets. This resulted in a 27.1% increase in measureable peptide identification events to be used for differential analyses across sample sets, corresponding to 99.8% of total peptides identified between samples (Fig. 4A).

A total of 529 peptides (4.4% of total identifications) were differentially abundant between knockout and wild-type sam-
Interestingly, only 4.8% of the peptides exclusively identified in Homer2 knockout or wild-type samples (1630 peptides /H11006
544) were among these proteins (79 /H11006
23), highlighting the value of comparing retention time aligned peptide MS1 XIC areas versus peptide identification events when considering presence/absence scenarios. The top ten significantly increased and decreased peptides in wild-type versus Homer2 knockout are detailed in Table I. Several Homer2 peptides were identified as being the most significantly abundant peptides in wild-type samples; representative peptide MS1 XIC for a Homer2 peptide identified in wild-type versus knockout samples is shown in Fig. 4B. Within the population of peptides exhibiting differential abundance, several candidates corresponding to previously identified Homer2 (29–30) interacting proteins involved in neurotransmission and dendritic spine morphology were observed. With respect to glutamatergic signaling, differential analysis revealed a decrease in the abundance of peptides corresponding to postsynaptic density protein-95 (PSD95, Q62108), an N-methyl-D-aspartate (NMDA) receptor anchoring protein in knockout samples. This analysis also revealed decreases in the metabotropic glutamate receptor 1 A (mGlur1A, P97772–1), an integral membrane protein and Homer2 binding partner responsible for several second messenger activities(31–33). The abundance of Sn1-specific diacylglycerol lipase alpha (DGL-alpha, Q6WQJ1), an enzyme im-
plicated in Homer2-mediated retrograde endocannabinoid signaling (34), was also decreased in knockout samples. These studies further revealed a change in abundance of peptides (q-value ≤ 0.05) corresponding to two Homer2-interacting proteins previously shown to govern dendritic spine morphogenesis, dynamin-3 (Q8BZ98) (32, 35) and drebrin (Q9QXS6). Together, these analyses showed that several peptides corresponding to Homer2 and known Homer2-interacting proteins are differentially abundant in wild-type versus knockout tissue. These results demonstrate the ability of our analytical pipeline to identify peptides candidates anticipated to be differentially abundant in complex sample matrix with altered abundances of Homer2.

Differential Analysis of Hemolyzed Mouse Blood Plasma—This sample set represents high complexity peptide matrixes differing in the levels of red blood cell hemolysis. We compared two plasma samples, one with low hemolysis (0.021 g/L hemoglobin (Hgb) and the other high hemolysis (0.182 g/L Hgb). Hemolysis is produced by the lysis of red blood cells and the subsequent release of Hgb resulting in reddish blood plasma (16). Hemolysis can occur in vivo because of disease processes, such as hemolytic anemia (16, 36), as well as in vitro because of sample collection or handling errors (15, 16).

Triplicate injections of each sample yielded a total of 3771 peptide identifications (2918; CV = 2.51%, across technical replicate injections) with 3088 of these candidates being co-identified between sample sets. Extraction of corresponding peptide MS1 XIC areas was successful for a total of 3768 of these events (3734 ± 7; CV = 0.19%) and yielded a total of 3755 events being integrated between sample sets. This resulted in a 17.8% increase in the number of measurable peptide events co-identified across replicate injections.

Comparative analysis revealed a total of 837 peptides that were differentially abundant between high and low hemolysis conditions.

**TABLE I**

**Top ten differentially abundant peptide MS1 XIC areas following comparative proteomic analyses of olfactory bulb synaptic density proteins harvested from Homer2 knockout and wild-type mice**

Positive and negative Log2 values denote peptides increased or decreased in wild-type samples, respectively. Table I is organized as follows: (Peptide), peptide sequence identified; (UniProt ID), UniProt protein accession; (Protein Name), protein description; and (Log2 WT:KO), details the log2 fold change ratios of peptide MS1 XIC abundances in wild-type versus knockout samples.

| Peptide       | UniProt ID | Protein Name                   | Log2 WT:KO |
|---------------|------------|--------------------------------|------------|
| R.SKIELEEGCSEINR.E | Q9QWW1  | Homer protein homolog 2        | 6.17       |
| R.LTALQESAASVEQK.WR | Q9QWW1  | Homer protein homolog 2        | 6.12       |
| R.RIELESEVR.D    | Q9QWW1  | Homer protein homolog 2        | 5.83       |
| K.SFLEVLDGK.I    | Q9QWW1  | Homer protein homolog 2        | 5.73       |
| K.EITQLQAETK.L   | O35668   | Huntington-associated protein 1 | 4.50       |
| R.IELESEVR.D     | Q9QWW1  | Homer protein homolog 2        | 3.21       |
| K.STVSEAAVEKESLK.L | P20357  | Microtubule-associated protein 2 | 2.16       |
| R.ICNQVLICR.K    | Q92219   | Succinyl-CoA ligase [ADP-forming] subunit beta, mitochondrial | 2.06       |
| R.SAQDLSDSTDEGIIPLR.N | Q62147  | Sorbin and SH3 domain-containing protein 1 | 1.96       |
| R.AHVFIQIDPSTK.K | Q9QWW1  | Homer protein homolog 2        | 1.94       |

**Fig. 5.** Peptide identification events versus successfully integrated peptide MS1 XIC areas in comparative proteomic analyses of hemolyzed blood plasma. Two-way Venn diagrams detailing overlapping peptide identification events (A) versus successfully integrated, retention time aligned peptide MS1 XIC areas (B) across three technical replicate injections of tryptic digests derived from high and low hemolyzed mouse blood plasma. These results revealed a 17.8% increase in the number of measurable peptide events co-identified across replicate injections.
conditions (q-value ≤ 0.05, 22% of total identifications) (supplemental Table S5). The large proportion of differentially abundant peptides observed emphasizes the stark differences between low and highly hemolyzed blood plasma proteomes. Additionally, only ~31% of the peptides initially identified exclusively in low or high hemolysis conditions alone (341 peptides) were among this significant population (106), further underscoring the value of comparing retention time aligned peptide MS1 XIC areas for presence/absence events. The top ten significantly increased and decreased peptides between high versus low hemolysis conditions are detailed in Table II. The majority of top peptides significantly increased in highly hemolyzed blood plasma correspond to different hemoglobin isoforms (Hba-a1, Q9CY10, Hbb-b1, A8DUK4, and Hba, P01942), a protein group comprising over 97% of the red blood cell proteome (37). The top ten peptides significantly decreased exhibited greater complexity and corresponded to proteins previously described as being localized to blood plasma, such as the copper-binding glycoprotein ceruloplasmin (Cp, Q61147) (38) a member of the immune-associated complement system, complement factor D (Cfd, P03953) (38–39), as well as the lipoprotein scavenger apolipoprotein E (ApoE, P08226) (38–39). Additionally, top significantly decreased peptides included candidates not previously described in blood, such as the novel anaphylotoxin-like domain containing protein (A2AS37). These data demonstrate the ability of this analytical pipeline to identify an array of differentially abundant peptides between highly complex sample matrices. These findings further highlight that differentially hemolyzed blood plasma samples exhibiting hemoglobin concentrations below the current clinical reference limits for borderline hemolysis (~0.5 g/L) (15, 40) revealed starkly different proteomes. These results support previous findings (41) that differential blood cell lysis represents an analytical variable in plasma proteomic workflows that significantly impacts sample complexity and thus, can potentially confound comparative plasma proteomic analyses. Therefore, these data further suggest that levels of accept-

### Table II

| Peptide | UniProt ID | Protein Name | Log2 High:Low HL |
|---------|------------|--------------|-----------------|
| R.NGCNPGEEGLPHWEDEKEGVLQGATTQQARL | P2935 | Liver carboxylesterase N | 4.95 |
| K.VNADEGYGEALGRLLVYPWTQR | A8DUK4 | Beta-globin | 3.50 |
| R.EYVPRPGEVTATITLGEPGTEYTVIALK | P11276 | Isoform 1 of Fibronectin | 3.44 |
| R.TPNFPPCK.N | P20918 | Plasminogen | 3.35 |
| K.AAWGKIGGIGAEGAEAL.R | P01942 | Hemoglobin subunit alpha | 3.11 |
| R.IDAPPSISVEWCR.K | A2AXJ5 | DENN/MADD domain containing 4C | 3.02 |
| K.VADALANAAGHLDDLPGALSDDLHAKLR | Q9CY10 | Putative uncharacterized protein | 2.73 |
| K.EAETFENPFDLTK.V | P24270 | Catalase | 2.70 |
| R.DAILFPSFIHSQK.R | P24270 | Catalase | 2.70 |
| R.LNMNVVSLGHLGKDFTPAAQAFQK.V | A8DUK2 | Beta-globin | 2.69 |
| K.RSCLHVPVCKDPEEK.K | P11680 | Properdin | -1.425128646 |
| K.IIEHVSDDLSSATLITYFLK.G | Q06770 | Corticosteroid-binding globulin | -1.617236553 |
| K.EEKDEEWYK | Q8R21 | Isotpe 1 of Protein Z-dependent protease inhibitor | -1.825093523 |
| K.IYYYIPEEKE.K | Q91N72 | Hemopexin | -1.931330998 |
| R.EITTVKPEATTMMATAVPCTPSLSAGHAN | Q4C0Q9 | Putative uncharacterized protein | -2.395243816 |
| R.LQGRDLRKL.R | P08226 | Apolipoprotein E | -2.700514892 |
| K.IEVEPTLCVDWAVGWVTHAGR.R | P03953 | Isoform 1 of Complement factor D | 2.813435393 |
| K.TGTYFVEK.E | A2AS37 | Novel anaphylatoxin-like domain containing protein | -3.346243638 |
| R.GDPEELHILGPGVPWAEGVDTIK.V | Q61147 | Cerulolespin | -3.819258046 |
| R.DLSLSDLSTASK.I | D3YW52 | Uncharacterized protein | -8.655422482 |

Maximizing Peptide Identification Events Using DDA
able hemolysis in blood plasma samples used for proteomic studies require reassessment relative to current clinical standards.

CONCLUSION

These results provide evidence to support a facile and robust workflow for the comparative proteomic analysis of peptide MS1 XIC areas directed by retention time alignment of peptides identified by MS/MS events across samples. We provide evidence that the strategy is quantitative using a six-point dilution curve of bovine standard peptides in mouse plasma where 88% (±3.2%) of bovine peptide MS1 XIC areas exhibited a linear response ($R^2 > 0.9$). This strategy was shown to increase the number of peptides detectable between samples by up to 30.5%. Thus, combining “shotgun” proteomic workflows for the identification of peptides with comparison of aligned peptide MS1 features can maximize “shotgun” proteomic discovery efforts and improve comparative sample analyses.

Acknowledgments—We would like to thank Dr. Kyle Orwig for donation of the C57BL/6 mice used in hemolysis studies, Dr. Rob Helton for olfactory bulb dissections in Homer2 knockout and wild-type mice, Dr. Jarrett Egertson for assistance in manuscript preparation, Dr. Nicolas Stewart for critical review of this manuscript, and Michael Riffle for establishing the Yeast Resource Center.

We would like to thank Dr. Kyle Orwig for donation of the C57BL/6 mice used in hemolysis studies, Dr. Rob Helton for olfactory bulb dissections in Homer2 knockout and wild-type mice, Dr. Jarrett Egertson for assistance in manuscript preparation, Dr. Nicolas Stewart for critical review of this manuscript, and Michael Riffle for establishing the Yeast Resource Center.

These results provide evidence to support a facile and robust workflow for the comparative proteomic analysis of peptide MS1 XIC areas directed by retention time alignment of peptides identified by MS/MS events across samples. We provide evidence that the strategy is quantitative using a six-point dilution curve of bovine standard peptides in mouse plasma where 88% (±3.2%) of bovine peptide MS1 XIC areas exhibited a linear response ($R^2 > 0.9$). This strategy was shown to increase the number of peptides detectable between samples by up to 30.5%. Thus, combining “shotgun” proteomic workflows for the identification of peptides with comparison of aligned peptide MS1 features can maximize “shotgun” proteomic discovery efforts and improve comparative sample analyses.

Acknowledgments—We would like to thank Dr. Kyle Orwig for donation of the C57BL/6 mice used in hemolysis studies, Dr. Rob Helton for olfactory bulb dissections in Homer2 knockout and wild-type mice, Dr. Jarrett Egertson for assistance in manuscript preparation, Dr. Nicolas Stewart for critical review of this manuscript, and Michael Riffle for establishing the Yeast Resource Center.

We would like to thank Dr. Kyle Orwig for donation of the C57BL/6 mice used in hemolysis studies, Dr. Rob Helton for olfactory bulb dissections in Homer2 knockout and wild-type mice, Dr. Jarrett Egertson for assistance in manuscript preparation, Dr. Nicolas Stewart for critical review of this manuscript, and Michael Riffle for establishing the Yeast Resource Center.
Maximizing Peptide Identification Events Using DDA

from plasma: considerations for proteomic discovery in sickle cell disease and other hemolytic processes. Proteomics Clin. Appl. 4, 926–930
37. Ringrose, J. H., van Solinge, W. W., Mohammed, S., O’Flaherty, M. C., van Wijk, R., Heck, A. J., and Sliper, M. (2008) Highly efficient depletion strategy for the two most abundant erythrocyte soluble proteins improves proteome coverage dramatically. J. Proteome Res. 7, 3060–3063
38. Anderson, N. L., Polanski, M., Pieper, R., Gattin, T., Tirumalai, R. S., Conrads, T. P., Veenstra, T. D., Adkins, J. N., Pounds, J. G., Fagan, R., and Lobley, A. (2004) The human plasma proteome: a nonredundant list developed by combination of four separate sources. Mol. Cell. Proteomics 3, 311–326
39. Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W., Menon, R., Herrmajorkob, H., Apweiler, R., Haab, B. B., Simpson, R. J., Eddes, J. S., Kapp, E. A., Moritz, R. L., Chan, D. W., Rai, A. J., Admon, A., Aebersold, R., Eng, J., Hancock, W. S., Hefta, S. A., Meyer, H., Paik, Y. K., Yoo, J. S., Ping, P., Pounds, J., Adkins, J., Qian, X., Wang, R., Wasinger, V., Wu, C. Y., Zhao, X., Zeng, R., Archakov, A., Tsugita, A., Beer, I., Pandey, A., Pisano, M., Andrews, P., Tammen, H., Speicher, D. W., and Hanash, S. M. (2005) Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. Proteomics 5, 3226–3245
40. Lippi, G., Luca Salvagno, G., Blanckaert, N., Giavarina, D., Green, S., Kitchen, S., Palicka, V., Vassault, A. J., and Plebani, M. (2008) Multi-center evaluation of the hemolysis index in automated clinical chemistry systems. Clin. Chem. Lab. Med. 47, 934–939
41. Hsieh, S. Y., Chen, R. K., Pan, Y. H., and Lee, H. L. (2006) Systematic evaluation of the effects of sample collection procedures on low-molecular-weight serum/plasma proteome profiling. Proteomics 6, 3189–3198
42. Cutillas, P. R., and Vanhaesebroeck, B. (2007) Quantitative profile of five murine core proteomes using label-free functional proteomics. Mol. Cell. Proteomics 6, 1560–1573
43. Alcolea, M. P., Casado, P., Rodriguez-Prados, J.-C., Vanhaesebroeck, B., and Cutillas, P. R. (2012) Phosphoproteomic analysis of leukemia cells under basal and drug-treated conditions identifies markers of kinase pathway activation and mechanisms of resistance. Mol. Cell. Proteomics 11, 453–486