Published in final edited form as:

J Proteome Res. 2008 November ; 7(11): 5055–5061. doi:10.1021/pr800239e.

High Quality Catalog of Proteotypic Peptides from Human Heart

Kelli G. Kline[†], Barbara Frewen[‡], Michael R. Bristow[§], Michael J. MacCoss[‡], and Christine C. Wu^{*,†}

Department of Pharmacology, University of Colorado School of Medicine, Aurora, Colorado 80045, Department of Genome Sciences, University of Washington, Seattle, Washington 98101, and Department of Cardiology, University of Colorado School of Medicine, Aurora, Colorado 80045

Abstract

Proteomics research is beginning to expand beyond the more traditional shotgun analysis of protein mixtures to include targeted analyses of specific proteins using mass spectrometry. Integral to the development of a robust assay based on targeted mass spectrometry is prior knowledge of which peptides provide an accurate and sensitive proxy of the originating gene product (i.e., proteotypic peptides). To develop a catalog of “proteotypic peptides” in human heart, TRIzol extracts of left-ventricular tissue from nonfailing and failing human heart explants were optimized for shotgun proteomic analysis using Multidimensional Protein Identification Technology (MudPIT). Ten replicate MudPIT analyses were performed on each tissue sample and resulted in the identification of 30 605 unique peptides with a q -value ≤ 0.01 , corresponding to 7138 unique human heart proteins. Experimental observation frequencies were assessed and used to select over 4476 proteotypic peptides for 2558 heart proteins. This human cardiac data set can serve as a public reference to guide the selection of proteotypic peptides for future targeted mass spectrometry experiments monitoring potential protein biomarkers of human heart diseases.

Keywords

proteotypic peptides; targeted mass spectrometry; human heart explant; dilated cardiomyopathy; MudPIT

Introduction

Genomic technologies that facilitate the measurement of global gene expression have begun to mature, whereas the equivalent proteomic technologies are not as routine or robust. The challenges associated with comprehensive measurement of the entire protein complement of a cell or tissue are more demanding as the dynamic range of protein expression spans greater than 6 orders of magnitude in simple eukaryote cells.¹ Furthermore, the broad biochemical heterogeneity of proteins exacerbates the complications of using a single approach to measure all proteins.

*To whom correspondence should be addressed. Christine C. Wu, Ph.D., Department of Pharmacology University of Colorado School of Medicine, Fitzsimons RC1 South L18-6117, P.O. Box 6511, Mail Stop 8303 Aurora, CO 80045. Phone: 303-724-3351. christine.wu@uchsc.edu.

[†]Department of Pharmacology, University of Colorado School of Medicine.

[‡]Department of Genome Sciences, University of Washington.

[§]Department of Cardiology, University of Colorado School of Medicine.

Data Availability. Data can be accessed at the proteotypic peptide database: http://proteome.gs.washington.edu/supplementary_data/.

Supporting Information Available: Tables of total proteins and counts; figure before and after optimization of pretreatment steps. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Most mass spectrometry-based pipelines have taken a “shotgun” approach to proteomics, where proteins are first digested into peptides because the biochemical diversity and challenges associated with the measurement of peptides are minimized relative to intact proteins.^{2–4} In shotgun proteomics based approaches, peptides are usually separated by nanoflow liquid chromatography, ionized and emitted into a tandem mass spectrometer, and the resulting peptide ions are sampled for tandem mass spectrometry by data-dependent acquisition. This general approach has become extremely powerful for determining the protein contents of a moderately complicated mixture. However, the ability to compare different samples is complicated by the semirandom sampling process of data-dependent acquisition. Some proteins of specific interest can go undetected in one or more samples that are being compared. Furthermore, to gain sufficient dynamic range to handle mixtures containing thousands of proteins, multiple dimensions of liquid chromatography need to be performed—an approach that extends the measurement peak capacity by increasing the overall analysis time. The additional dimension of liquid chromatography separation dramatically decreases the throughput despite limiting the practicality of the method to handle the quantity of data required to measure variance.

To address these limitations, there has been a shift toward the development and application of technologies for the targeted analysis of proteins within complex mixtures. Numerous derivations of targeted mass spectrometry using the specific acquisition of tandem mass spectra of peptides predicted *in silico* have been reported,^{5,6} and more recently these methods have been based on the use of selected reaction monitoring (SRM) on triple quadrupole mass spectrometers.^{7–10} The concept of monitoring specific peptides from proteins of interest is well-established. These methods have high specificity within a complex mixture and thus can be performed in a fraction of the instrument time relative to shotgun methods. Ultimately, targeted studies are intended to complement discovery based analysis.

An essential step of any targeted analysis is determining which peptides are most representative and most likely to be observed for a protein of interest. These peptides are commonly referred to as proteotypic peptides and can be used as a suitable proxy for the protein.^{11,12} To establish a targeted experimental assay, significant amounts of time and resources can be spent to produce quantitative internal standards such as synthetic peptides^{7,10,13,14} or recombinant proteins from concatenated peptide sequences,¹⁵ and/or developing immunoaffinity reagents for the enrichment of very low-abundance tryptic peptide.¹⁶ Instead, having experimental tandem MS data acquired for the sample matrix would be invaluable for minimizing the expense of developing reagents of limited utility.

Assuming that proteotypic peptides are those observed most frequently in replicate shotgun proteomics experiments, a onetime, large-scale, discovery-based analysis can be used to create a data source for future targeted experiments of a specific sample or tissue. For each peptide identified, we know that it exists in its unmodified form and can be detected from within the context of the experimental sample matrix. For abundant proteins, knowing how frequently the protein is identified in replicate analyses, and how many times the representative peptide has been sampled for tandem mass spectrometry, can provide a quantitative measure of the peptides that are most proteotypic. For low-abundance proteins, just having a single peptide spectrum match to a unique peptide can provide a starting point for future analyses.

Here, we report a high-quality catalog of human cardiac proteins generated from comprehensive MudPIT analyses of TRIzol precipitates extracted from human heart tissue explanted from two donors, one with a normal cardiac phenotype and the other diagnosed with idiopathic dilated cardiomyopathy (IDCM). For each sample, 10 replicate MudPIT analyses were performed, resulting in the acquisition of 3 490 763 total tandem mass spectra. Of these spectra, 144 349 were mapped to peptide identifications with a *q*-value ≤ 0.01 resulting in 30

605 unique peptide identifications, and corresponding to 7138 distinct proteins. From this data set, we report 4476 proteotypic peptides for 2558 human cardiac proteins. This human cardiac proteome data set will provide an invaluable resource for future targeted mass spectrometric analyses to monitor putative biomarker candidates of heart function.

Results and Discussion

We have developed a simple and robust approach for sampling total protein that is compatible with any human tissue. The single step approach of extracting total RNA from biological samples of different sources using an acidic solution of guanidinium thiocyanate, sodium acetate, phenol and chloroform followed by centrifugation was originally developed by Chomczynski and Sacchi¹⁷ and is based on the fact that under acidic conditions, total RNA can be extracted into an aqueous phase and away from other cellular components. Genomic DNA and proteins are denatured and can be recovered at the interface between the aqueous and organic phase. Subsequent extractions using methanol and chloroform to enrich for protein result in the separation and isolation of RNA, DNA, and protein. Because the protein denaturation is rapid, protein degradation by proteolysis is minimized. Importantly, total proteins (soluble and membrane proteins, cellular and extra-cellular proteins alike) are indiscriminately separated from human tissue. Because this approach is fast and simple and facilitates the extraction of RNA, protein, and genomic DNA in series, our approach opens the door to future studies comparing protein and gene expression from clinical tissue samples of the same origin.

Shotgun Proteomic Analysis of TRIzol Extracted Human Heart Explants

Left-ventricular tissue from nonfailing and failing heart explants were subjected to TRIzol extraction. The TRIzol reagent is an optimized commercial implementation of the phenol-guanidinium isothiocyanate method¹⁷. We started with the protein/DNA interface produced from this extraction protocol to ensure that our optimized methods would be compatible with concurrent gene expression studies. Three pretreatment steps were required for optimal analysis of the TRIzol interface by MudPIT. [Note: Incorporation of these 3 pretreatment steps resulted in a >20-fold increase in protein identifications in the resulting MudPIT analysis. (Supplementary Figure 1A–C in Supporting Information).] The experimental workflow is illustrated in Figure 1. First, the TRIzol interface, composed mostly of protein and DNA, was collected and extracted with MeOH-CHCl₃ as described by Wessel and Flugge¹⁸ to separate the nonprotein components and the residual pink TRIzol dye. The resulting protein precipitate was extremely dense and difficult to resuspend in digestion buffers. Therefore, the second step was to sonicate the precipitate in MeOH to produce a fine protein powder with decreased particle size and increased surface area to improve the exposure to denaturants in subsequent solubilization steps. Finally, the protein powder was solubilized using the mass spectrometry compatible acid-labile detergent Rapigest and digested with trypsin. Digested peptide samples were analyzed by MudPIT.

Generation of a Human Cardiac Proteotypic Peptide Database

Ten replicate MudPIT analyses were acquired for each heart sample using this analysis platform. Total unique protein identifications with the addition of each replicate are shown in Figure 2A. The plot begins to level off after six analyses showing a plateau of protein identifications at ~3500 proteins in each sample, indicating that we have saturated the number of protein identifications. Minimal gains in unique protein identifications were observed after the ninth replicate (Figure 2B). This reduction in new protein identifications suggests that new identifications contributed by additional analyses would be false positive protein identifications as frequently as true positive identifications. The sampling of new proteins using data-dependent acquisition follows similar sampling statistics as sampling expressed sequence

tags and serial analysis of gene expression in mRNA analyses and is a function of the distribution of protein levels within the sample and the total number of spectra that are acquired.¹⁹ We have sampled this data set to the point of producing <3% new proteins with each additional MudPIT. These small increases in protein numbers with additional analyses indicate saturation in the total number of protein identifications possible using these unfractionated heart samples; thus, even a very large increase in the total number of tandem mass spectra acquired would produce only minimal new protein identifications.

A total of 7138 proteins were identified in this study by combining the 3746 proteins identified from the nonfailing heart tissue sample and the 3818 proteins from the failing tissue sample (Supplementary Table 1 in Supporting Information). On average, $174\,538 \pm 16\,706$ (mean $\pm$ SD) tandem mass spectra were acquired per run. Figure 3A shows that the overlap of identified proteins between the nonfailing and failing heart tissues samples was 3669 (51.4%), indicating that while a majority of proteins were found in both samples, a large number of identifications were found in just one of the donor's tissue and not the other. The relatively poor overlap of protein identifications is somewhat expected in a tissue sample of this complexity with different genetic backgrounds. Figure 3B shows the cellular localization of those proteins identified in these analyses. A majority of proteins did not have associated gene ontology (GO) annotations as determined from the Pathway Studio 5.0 program (Ariadne Genomics, Rockville, MD) (60% unannotated)²⁰ However, of the proteins identified, the relative ratios of cell localizations seemed reasonable, with largest percent of proteins localized to the cytoplasm/cytoskeleton (25.9%) and nucleus (17.2%). Various organelles also were represented with the mitochondrial (9.9%) and ER (4.1%) proteins showing the highest representation. In addition, 13.8% of protein identifications were predicted to be localized to membranes (either the plasma membrane or other intracellular membranes) which correlates well with the percent of proteins predicted to have transmembrane domains (14.9% of the total protein identifications as predicted by TMHMM). These data indicate that, using this experimental workflow, we get protein identifications from all the cellular compartments with no apparent biases.

Multivariate Analyses and Multiple Testing Corrections in Establishing High Quality Peptide Identifications

Many database search algorithms return multiple output scores or features (e.g., Xcorr, Sp, and deltaCN for SEQUEST). Thus, most proteomics studies apply thresholds that need to be exceeded for multiple individual features for a database search result to be considered correct.^{21,22} However, because these features are considered individually, these simple threshold approaches ignore the multivariate gains that can be obtained if the scores are independent and are considered together. For example, a peptide spectrum match that has a SEQUEST XCorr well above the threshold will be discarded if the PSM deltaCN is even just marginally below its respective cutoff. To improve the number of peptide identifications and provide a statistical measure of significance for each peptide-spectrum match, we applied the database search postprocessor Percolator to our SEQUEST results. Percolator, like the commonly used algorithm PeptideProphet, takes database searching results from programs like SEQUEST and combines multiple scores to create a single score that improves the discrimination of correct and incorrect peptide identifications. Unique to Percolator is the ability to learn the weight of the individual scores directly from the respective data set.²³ By retraining the classifier on each data set, Percolator generalizes well between all types of data.

An often overlooked aspect of proteomics is the use of a multiple testing correction that accounts for the analysis of thousands and even millions of spectra in a single data set. Percolator performs multiple hypotheses testing automatically by reporting a *q*-value for each spectrum. Percolator computes the *q*-value using the distribution of scores from a duplicate collection of peptide identifications obtained by searching the same spectra against a decoy

database.^{24,25} Because the data in this manuscript are reported using *q*-values, we accurately report the number of target peptide spectrum matches at each score threshold that are actually false,^{24,25} making the proteotypic peptides reported in this analysis of high quality.

Proteotypic Peptides Are Ranked Using Experimental Observation Frequency

Not all of the possible tryptic peptides for a given protein are consistently observed.¹² In fact, it has been shown that some tryptic peptides are preferentially identified in mass spectrometry runs, while others are not seen at all.²⁶ Commonly observed peptides from a given protein are designated as “proteotypic” as they are most consistently identified. To identify the peptides that are most often detected from each protein, we grouped the 30 605 positively identified peptides from the combined 20 MudPIT analyses by observation frequency (OF, Supplementary Table 2 in Supporting Information). The OF was defined as the number of runs in which a peptide was observed divided by number of runs in which **any** peptide from the same protein is seen. As previously reported, a peptide was denoted as proteotypic if had an observation frequency of at least 0.5, meaning it was identified in at least 50% of the mass spectrometry runs where the corresponding protein was identified.²⁶ Peptides that occur in more than one protein can have more than one observation frequency. In these cases, the highest frequency was reported. The spectrum counts are the sum of spectra in each MudPIT run identified as a given peptide.

Selection of Proteotypic Peptide for Use in Targeted Proteomic Studies

Our shotgun proteomic evaluation of nonfailing and failing human heart explants yielded 4476 proteotypic peptides representing 2558 unique proteins. Each of these proteotypic peptides can potentially serve as a measure by proxy for the protein abundance within a complex sample. To illustrate this point, Figure 4A maps the experimentally identified proteotypic peptides of muscle creatine kinase to its primary sequence. Proteotypic peptides in this figure were grouped by observation frequency and color. Peptides observed with an observation frequency of 0.9–1.0 are shown in red, 0.5–0.89 are shown in orange, and <0.5, blue. Spectral counts for each peptide are also listed next to each peptide in parentheses. One of the peptides in the 0.9–1.0 observation frequency group was selected and monitored in SRM experiments in peptide mixtures prepared from unfractionated human heart explants during a 2 h reverse phase gradient. Each of the parent ion to y-product ion transitions were monitored and shown in Figure 4B.

Conclusions

Cardiovascular disease is the leading cause of death in the United States that is manifested in many different clinical sequelae. While shotgun proteomics is extremely powerful, it does not have the throughput or dynamic range to handle the quantitative measurement of protein abundance across many clinical samples. Because of this limitation, targeted analyses are being developed for the measurement of specific proteins with hypothesized functional relevance in human heart disease to be measured in tissue or the plasma proteome.⁸ An essential prerequisite of these experiments is prior knowledge of which peptides can be used for hypothesized proteins in targeted mass spectrometric assays.

To create a list of proteotypic peptides from heart proteins, we performed multiple MudPIT experiments on two different human samples. This list is an extensive human cardiac protein database, containing high quality proteotypic peptides selected for 2558 proteins. These peptides are potential candidates for quantitative targeted analyses, providing a high-throughput quantitative solution to proteomic comparisons between actual patient samples. These data will provide a powerful resource for the development of targeted mass spectrometry analysis of potential protein markers of human heart disease.

Materials and Methods

Heart Samples

Human nonfailing and failing left-ventricular free wall heart explants were obtained from the Heart Tissue Bank at the Division of Cardiology at the University of Colorado School of Medicine. All patient identifiers were removed from the samples except for age, gender, race, and clinical diagnosis. The nonfailing heart donor used in this study was a 20 year old white male, while the failing donor heart was a 36 year old Hispanic male diagnosed with IDCM.

MS Sample Preparation

Frozen heart explants (10 mg pieces) were homogenized in 800 μ L of ice-cold TRIzol reagent (Invitrogen, Carlsbad, CA) using a Tissue Tearor homogenizer with a 4.5 mm diameter probe (Biospec Products, Bartlesville, OK) for approximately 30 s. TRIzol extraction utilized the optimized protocol supplied by the manufacturer. After removal of RNA in the aqueous fraction, the protein/DNA interface fraction was collected and extracted with methanol-chloroform. Methanol-chloroform extractions were carried out as previously described with the following exception. Before removing all the methanol from the protein pellet following methanol-chloroform treatment, additional methanol was added to total protein precipitate to 500 μ L of total volume, and the pellet was sonicated using a Microson Ultrasonic cell disruptor (Misonix, Farmingdale, NY; 10 s, power 10). The fine protein powder was allowed to settle to the bottom of the tube and most of the methanol was carefully removed without drying the powder. The resulting protein powder in minimal amounts of MeOH was resolubilized in 0.2% RapiGest (Waters Corporation, Milford, MA) in 25 mM ammonium bicarbonate. Solubilization was assisted using probe sonication and passing solution through an insulin syringe approximately 10 times.

Samples were then diluted down to 0.1% RapiGest with 25 mM ammonium bicarbonate buffer and boiled for 5 min. Dithiothreitol (DTT) was added to 5 mM final concentration and incubated at 60 °C for 30 min to reduce protein disulfide bonds. After the sample was cooled to room temperature, iodoacetamide (IAA) was added to 15 mM final concentration and incubated at room temperature for 30 min in the dark to alkylate all reduced cysteines. Following reduction and alkylation reactions, CaCl_2 was added to 1 mM final concentration. Modified trypsin (Promega, Madison, WI) was added at a 1:100 enzyme/substrate ratio and incubated at 37 °C overnight using a Thermomixer (Eppendorf, Westbury, NY). Digestion reaction was stopped by storing the samples at -20 °C. Prior to mass spectrometry analysis, HCl was added to a final concentration of 100 mM, and the solution was incubated at 37 °C for 45 min using a Thermomixer, and then centrifuged at 20 800g at 4 °C for 15 min. The supernatant was collected and analyzed by MudPIT.

MudPIT

Peptide samples (volumes of digest equivalent to 75 μ g of predigested protein) were loaded off-line onto a fused-silica microcapillary desalting column (250 μ m i.d., Agilent Technologies, Palo Alto, CA) packed with 1 cm Aqua C18 material (5 μ m 125 Å, Phenomenex, Torrance, CA) using a high-pressure bomb (manufactured in-house). After the sample was desalted with Buffer A (95% H_2O , 5% acetonitrile, and 0.1% formic acid), the desalting column was attached to a biphasic microcapillary column (100 μ m i.d., Polymicro Tech., Phoenix, AZ) with a ~5 μ m tip and packed with 13 cm Aqua C18 material (5 μ m 125 Å, Phenomenex, Torrance, CA) and 3 cm of strong cation exchange material (Whatman, Inc., Florham Park, NJ) using an in-line filter assembly (Upchurch, Oak Harbor, WA). The biphasic column was set into a column heater (built in-house)²⁷ and placed in-line with an LTQ linear ion trap mass spectrometer (Thermo Fisher Scientific, Waltham, MA) interfaced with an Agilent 1100 binary HPLC and autosampler system (Agilent Technologies, Palo Alto, CA). The mobile phase buffers were

Buffer A (95% H₂O, 5% acetonitrile, and 0.1% formic acid) and Buffer B (95% acetonitrile, 5% H₂O, and 0.1% formic acid). Samples were analyzed with a 12-step MudPIT run at 40 °C during chromatography. The mobile phase gradient for Step 1 was 100% Buffer A-2 min, 100% Buffer A to 62.1% Buffer A/37.9% Buffer B-over 95 min, 62.1% Buffer A to 35% Buffer A/65% Buffer B-over 10 min, 100% Buffer A-15 min. The mobile phase gradient for Steps 2–10 was 100% Buffer A-8 min, 100% Buffer A to 70% Buffer A/30% Buffer B-over 87 min, 70% Buffer A to 36.8% Buffer A/63.2% Buffer B-over 10 min, 100% Buffer A-15 min. The mobile phase gradient for Step 11 was 100% Buffer A-8 min, 100% Buffer A to 50% Buffer A/50% Buffer B-over 87 min, 50% Buffer A to 20% Buffer A/80% Buffer B-over 10 min, 100% Buffer A-15 min. The mobile phase gradient for Step 12 was: 100% Buffer A-8 min 100% Buffer A to 70% Buffer A/30% Buffer B-over 87 min, 70% Buffer A/30% Buffer B to 0% Buffer A/100% Buffer B-over 5 min, 100% Buffer B-5 min, 100% Buffer A-20 min. Ammonium acetate salt pulses were injected by the autosampler at the beginning of each subsequent step (Step 2, 50 μ L 100 mM; Step 3, 50 μ L 200 mM; Step 4, 50 μ L 300 mM; Step 5, 50 μ L 400 mM; Step 6, 50 μ L 500 mM; Step 7, 50 μ L 600 mM; Step 8, 50 μ L 800 mM; Step 9, 50 μ L 900 mM; Step 10, 50 μ L 1 M; Step 11, 75 μ L 1 M; Step 12, 100 μ L 5 M). Tandem mass spectra were acquired using data-dependent acquisition with a single full mass scan followed by 5 MS/MS scans.

Data Analysis

MS/MS spectra from each analysis were searched using no enzyme specificity on a 96 node G5 Beowulf cluster against the a human protein database (downloaded January, 2006) appended to an *Escherichia coli* protein database (downloaded January, 2006) concatenated to a shuffled decoy database²⁸ using a normalized implementation of SEQUEST²⁹ and postprocessed with Percolator²³ to combine multiple scores into a single discriminant score and assign *q*-values to each peptide spectrum match. The resulting peptide identifications were assembled into proteins using DTASelect,³⁰ and thresholds were adjusted to include only peptide spectrum matches with a *q*-value ≤ 0.01 . The OF was calculated by taking the number of runs in which a peptide was observed divided by number of runs in which **any** peptide from the same protein was seen.

Selected Reaction Monitoring (SRM)

IDCM donor heart sample (5 μ g) was loaded onto a 75 μ m i.d. fused silica column packed with 15 cm of reverse phase material (Jupiter Proteo, 5 μ m, 90 Å, Phenomenex) using an autosampler as described previously.³¹ The column was then connected to the microspray ion source and placed in-line with a TSQ Quantum Access mass spectrometer (ThermoElectron Corp, San Jose, CA) coupled with a ThermoElectron Surveyor pump and MicroAS autosampler (ThermoElectron Corp, San Jose, CA). The HPLC gradient and mass spectrometer transitions were controlled by the Xcalibur software. The mobile phase buffers were Buffer A (95% H₂O, 5% acetonitrile, and 0.1% formic acid) and Buffer B (95% acetonitrile, 5% H₂O, and 0.1% formic acid). The mobile phase gradient was 100% Buffer A-15 min during sample loading, 100% Buffer A to 74.7% Buffer A/25.3% Buffer B-over 35 min, 74.7% Buffer A/25.3% Buffer B to 32.6% Buffer A/67.4% Buffer B-over 5 min, 100% Buffer A-10 min.

The doubly charged parent ion and all singly charged fragment y-ions (12 total transitions) of the unique proteotypic peptide, LSVEALNSLTGEFK (*m/z* 754.4) from the muscle creatine kinase protein (gi|21536288|) were monitored using selective reaction monitoring on the triple sector quadrupole mass spectrometer. All data were acquired with a Q1 and Q3 resolution of 0.7 *m/z*. Each transition was monitored with a dwell time of 80 ms resulting in a total cycle time of 1 s. The RF-only q2 collision cell was pressurized with 1 mTorr of argon gas and all transitions were monitored using a collision offset of 0.034 V.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support for this work was provided in part from National Institutes of Health grants F31-AA017341 (K.G.K.), R01-DK069386 (M.J.M.), P41-RR011823 (M.J.M.), R21-HL083360 (C.C.W.), U01-AA016653 (C.C.W.), and R01-AA016171 (C.C.W.).

References

1. Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS. Global analysis of protein expression in yeast. *Nature* 2003;425(6959):737–41. [PubMed: 14562106]
2. Gaucher SP, Taylor SW, Fahy E, Zhang B, Warnock DE, Ghosh SS, Gibson BW. Expanded coverage of the human heart mitochondrial proteome using multidimensional liquid chromatography coupled with tandem mass spectrometry. *J Proteome Res* 2004;3(3):495–505. [PubMed: 15253431]
3. McDonald WH, Yates JR III. Shotgun proteomics and biomarker discovery. *Dis Markers* 2002;18(2):99–105. [PubMed: 12364816]
4. Fournier ML, Gilmore JM, Martin-Brown SA, Washburn MP. Multidimensional separations-based shotgun proteomics. *Chem Rev* 2007;107(8):3654–86. [PubMed: 17649983]
5. Kalkum M, Lyon GJ, Chait BT. Detection of secreted peptides by using hypothesis-driven multistage mass spectrometry. *Proc Natl Acad Sci USA* 2003;100(5):2795–800. [PubMed: 12591958]
6. Arnott D, Kishiyama A, Luis EA, Ludlum SG, Marsters JC Jr, Stults JT. Selective detection of membrane proteins without antibodies: a mass spectrometric version of the Western blot. *Mol Cell Proteomics* 2002;1(2):148–56. [PubMed: 12096133]
7. Kirkpatrick DS, Gerber SA, Gygi SP. The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods* 2005;35(3):265–73. [PubMed: 15722223]
8. Anderson L, Hunter CL. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol Cell Proteomics* 2006;5(4):573–88. [PubMed: 16332733]
9. Barnidge DR, Goodmanson MK, Klee GG, Muddiman DC. Absolute quantification of the model biomarker prostate-specific antigen in serum by LC-MS/MS using protein cleavage and isotope dilution mass spectrometry. *J Proteome Res* 2004;3(3):644–52. [PubMed: 15253448]
10. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci USA* 2003;100(12):6940–5. [PubMed: 12771378]
11. Craig R, Cortens JP, Beavis RC. The use of proteotypic peptide libraries for protein identification. *Rapid Commun Mass Spectrom* 2005;19(13):1844–50. [PubMed: 15945033]
12. Kuster B, Schirle M, Mallick P, Aebersold R. Scoring proteomes with proteotypic peptide probes. *Nat Rev Mol Cell Biol* 2005;6(7):577–83. [PubMed: 15957003]
13. Wan H, Umstot ES, Szeto HH, Schiller PW, Desiderio DM. Quantitative analysis of [Dmt(1)]DALDA in ovine plasma by capillary liquid chromatography-nanospray ion-trap mass spectrometry. *J Chromatogr, B: Anal Technol Biomed Life Sci* 2004;803(1):83–90.
14. Desiderio DM, Zhu X. Quantitative analysis of methionine enkephalin and beta-endorphin in the pituitary by liquid secondary ion mass spectrometry and tandem mass spectrometry. *J Chromatogr, A* 1998 ;794(1–2):85–96. [PubMed: 9491558]
15. Rivers J, Simpson DM, Robertson DH, Gaskell SJ, Beynon RJ. Absolute multiplexed quantitative analysis of protein expression during muscle development using QconCAT. *Mol Cell Proteomics* 2007;6(8):1416–27. [PubMed: 17510050]
16. Anderson NL, Anderson NG, Haines LR, Hardie DB, Olafson RW, Pearson TW. Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *J Proteome Res* 2004;3(2):235–44. [PubMed: 15113099]

17. Chomczynski P, Sacchi N. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nat Protoc* 2006;1(2):581–5. [PubMed: 17406285]
18. Wessel D, Flugge UI. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* 1984;138(1):141–3. [PubMed: 6731838]
19. Liu H, Sadygov RG, Yates JR III. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 2004;76(14):4193–201. [PubMed: 15253663]
20. Nikitin A, Egorov S, Daraselia N, Mazo I. Pathway studio-the analysis and navigation of molecular networks. *Bioinformatics* 2003;19(16):2155–7. [PubMed: 14594725]
21. Washburn MP, Wolters D, Yates JR III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001;19(3):242–7. [PubMed: 11231557]
22. Wu CC, MacCoss MJ, Howell KE, Yates JR III. A method for the comprehensive proteomic analysis of membrane proteins. *Nat Biotechnol* 2003;21(5):532–8. [PubMed: 12692561]
23. Kall L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* 2007;4(11):923–5. [PubMed: 17952086]
24. Kall L, Storey JD, MacCoss MJ, Noble WS. Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J Proteome Res* 2008;7(1):29–34. [PubMed: 18067246]
25. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 2003;100(16):9440–5. [PubMed: 12883005]
26. Mallick P, Schirle M, Chen SS, Flory MR, Lee H, Martin D, Ranish J, Raught B, Schmitt R, Werner T, Kuster B, Aebersold R. Computational prediction of proteotypic peptides for quantitative proteomics. *Nat Biotechnol* 2007;25(1):125–31. [PubMed: 17195840]
27. Speers AE, Blackler AR, Wu CC. Shotgun analysis of integral membrane proteins facilitated by elevated temperature. *Anal Chem* 2007;79(12):4613–20. [PubMed: 17500534]
28. Finney, G.; Merrihew, G.; Klammer, A.; Frewen, B.; MacCoss, MJ. Protein False Discovery Rates from MS/MS Experiments: Decoy Databases and Normalized Cross-Correlation. *Proceedings of the 53rd ASMS Conference on Mass Spectrometry*; 2005.
29. MacCoss MJ, Wu CC, Yates JR III. Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal Chem* 2002;74(21):5593–9. [PubMed: 12433093]
30. Tabb DL, McDonald WH, Yates JR III. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* 2002;1(1):21–6. [PubMed: 12643522]
31. Klammer AA, MacCoss MJ. Effects of modified digestion schemes on the identification of proteins from complex mixtures. *J Proteome Res* 2006;5(3):695–700. [PubMed: 16512685]

Figure 1.

Pretreatment of TRIzol extracted precipitate is required for optimized shotgun proteomic analysis. Left-ventricular tissue from nonfailing and failing heart explants were subjected to TRIzol extraction. Three pretreatment steps were required for optimal analysis of the TRIzol protein precipitate by MudPIT: (1) the TRIzol protein/DNA interface was collected and extracted with MeOH-CHCl₃ to remove nonprotein components and residual TRIzol dye. (2) The resulting protein precipitate was sonicated in MeOH to produce a fine protein powder with decreased particle size and increased surface area to optimize subsequent solubilization step. (3) The protein powder was solubilized using an acid labile detergent and digested with trypsin. Digested peptide samples were analyzed by MudPIT.

Figure 2.

Ten replicate analyses were required for saturation of protein identifications using the MudPIT shotgun proteomic platform. (A) Cumulative unique protein identifications were plotted for each additional MudPIT analysis on both the nonfailing and failing heart explant samples. (B) The percent new protein identifications with each additional MudPIT analysis approached saturation with 10 replicates.

Figure 3. Characterization of total identified unique proteins from nonfailing and failing heart explants. (A) The total number of unique proteins identified from the combined analyses (7138 proteins) has 51.4% overlap between the nonfailing and failing heart samples. Proteins were identified with an FDR < 1%. (B) Gene ontology annotations for the cellular localization (and function) of identified proteins revealed no apparent bias for any category.

Figure 4.

Proteotypic peptides are ranked using experimental observation frequency and selected to determine parent ion-product ion transitions for MRM analyses. (A) The primary sequence of muscle creatine kinase (gi|21536288) is shown in black. Experimentally observed proteotypic peptides are mapped below the protein sequence with the following color coding for observation frequency (OF): red, OF 0.9–1.0, orange, OF 0.5–0.89, blue, OF < 0.5. Total spectral counts are presented next to each peptide in parentheses. (B) A proteotypic peptide was selected [LSVEALNSLTGEFK, OF = 0.9, shown boxed in A] and parent ion-product ion transitions were monitored in the failing heart sample using MRM analyses.