Characterization of Mouse Spleen Cells by Subtractive Proteomics

Francisco J. Dieguez-Acuna‡§, Scott A. Gerber‡, Shohta Kodama§, Joshua E. Elias‡, Sean A. Beausoleil‡, Denise Faustman§, and Steven P. Gygi‡¶

Major analytical challenges encountered by shotgun proteome analysis include both the diversity and dynamic range of protein expression. Often new instrumentation can provide breakthroughs in areas where other analytical improvements have not been successful. In the current study, we utilized new instrumentation (LTQ FT) to characterize complex protein samples by shotgun proteomics. Proteomic analyses were performed on murine spleen tissue separated by magnetic beads into distinct CD45− and CD45+ cell populations. Using shotgun protein analysis we identified ~2,000 proteins per cell group by over 12,000 peptides with mass deviations of less than 4.5 ppm. Datasets obtained by LTQ FT analysis provided a significant increase in the number of proteins identified and greater confidence in those identifications and improved reproducibility in replicate analyses. Because CD45− and not CD45+ cells are able to regenerate functional pancreatic islet cells in a mouse model of type 1 diabetes, protein expression was further compared by a subtractive proteomic approach in search of an exclusive protein expression profile in CD45− cells. Characterization of the proteins exclusively identified in CD45− cells was performed using gene ontology terms via the JavaScript GoMiner. The CD45− cell subset readily revealed proteins involved in development, suggesting the persistence of a fetal stem cell in an adult animal. Molecular & Cellular Proteomics 4:1459–1470, 2005.

The use of proteomic technologies for global characterization of proteins expressed in cells, tissues, and biological fluids is a key component in furthering our ability to understand biological processes in normal and diseased states. In many proteomic studies, cells or tissues are characterized by profiling and comparing proteins expressed in treated and untreated cells, cancerous versus normal tissues, or various cell populations. A seminal tool in this development has been the application of mass spectrometry technologies to identify proteins and post-translation modifications in complex protein mixtures. The term “shotgun proteomics” is used to describe mass spectrometry-based methods that enable the rapid identification of proteolytic peptides from complex protein mixtures in a data-dependent manner (1). Two major analytical challenges encountered by shotgun proteome analysis include the significant variability of protein abundance (dynamic expression range) and the diversity of protein expression (multiple protein forms) in protein mixtures (2), which combine to produce many hundreds of thousands of individual peptide species in the final analysis, often spanning 6 orders of magnitude in abundance. Such issues are only partially addressed with the simple expansion of chromatographic peak capacities in multidimensional separations (3, 4).

Due in part to the dynamic range challenge, recent proteomic studies have profiled proteins expressed in cellular organelles rather than whole cell extracts. By reducing initial mixture complexity, investigators have been able to identify lower abundance proteins involved in nuclear pore trafficking (5), centrosome function (6), and chromatin organization (7, 8) increasing our knowledge of basic cellular physiology. Because mass spectrometers are generally poor at measuring quantitative differences in peptide concentration purely by ion intensity, a number of methods have been developed to measure biological changes between samples. These include a diverse set of labeling methods including stable isotope labeling with amino acids in cell culture (SILAC)1 (9) and the ICAT (10) strategies. In addition, investigators have also proposed non-stable isotope-based methods to profile protein expression differences between complex mixtures including several based on peptide counting or protein coverage (11–13). In general, relative extent of protein coverage by detected peptides scales with the expression level of the observed protein.

Recently investigators have utilized subtractive proteomics to identify proteins from cellular organelles by subtracting out common protein contaminants. In one elegant example, Schirmer et al. (14) utilized a subtractive proteomic technique to identify nuclear envelope proteins with possible disease links. Thirteen known and 67 potentially new integral nuclear membrane proteins were described by removing common nuclear membrane preparation contaminants. Another study

---

1 The abbreviations used are: SILAC, stable isotope labeling with amino acids in cell culture; bis-Tris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; AGC, automatic gain control; FP, false positive.
Subtractive Proteomics of Spleen Cells

identified candidate substrates for sumoylation in yeast via large scale subtraction of proteomic datasets (15). Subtractive proteomics relies on the exclusive or differential detection of peptides from a given protein during a shotgun proteomic experiment. As might be expected, the effectiveness of subtractive proteomics is greatly influenced by the sampling rate of peptides in the complex mixture and the number of unique peptides identified for each protein. Characterization of the proteome of interest by subtractive proteomics depends primarily on obtaining a pure (approaching 100%) protein extract of at least one of the two samples to be compared.

In the present studies, we performed subtractive proteomic analysis with a new (LTO FT) and more traditional (LCQ Deca XP) instrumentation to characterize two populations of cells from mouse spleen. The LTQ FT is a novel hybrid mass spectrometer consisting of a linear ion trap and an FT ICR mass analyzer. Conventional three-dimensional (spherical trap) ion trap mass spectrometers exhibit lower mass accuracy and limited ion trapping efficiency. In contrast, the linear (two-dimensional) ion trap of the LTQ has a relatively high ion capacity, a feature that directly results in improved dynamic range. In addition, improvements to the ion source and transfer optics have allowed for increased sensitivity. Huge increases in both resolution and mass accuracy of precursor ions are achieved via mass analysis in the ICR cell (16, 17). With the improved mass accuracy provided by the ICR cell and the substantial increase in scan rate for the LTQ ion trap, this hybrid instrument has the potential to significantly improve shotgun analysis of very complex protein samples.

Protein mixtures were generated by whole cell fractions prepared from CD45− and CD45+ spleen cells. These cell populations were chosen due to the observation that CD45− cells were able to regenerate functional pancreatic islet cells in a non-obese mouse model of type I diabetes (18). Because the CD45− spleen cell population also contains the stem cells responsible for the regeneration, the stated goal of these experiments was to identify proteins expressed exclusively in CD45− cells through subtractive proteomics of the CD45+ cell fraction. We define subtractive proteomics as the set of proteins identified exclusively by at least two unique peptides in each cell population. When replicates are considered an exclusively identified protein should be identified by at least two unique peptides in any of the replicate studies but not in any of the replicate control samples. The assumption is that the exclusive expression of a protein in one cell population would result in the exclusive detection of peptides from that protein.

After LC-MS/MS analysis using identical chromatographic conditions on both instruments, proteins were identified by database searching using the SEQUEST algorithm (19). Proteins from CD45− and CD45+ cells were compared by a subtractive proteomic approach in search of exclusive expression in CD45− cells. To our knowledge no studies have utilized this subtractive proteomic approach to identify differences between complex tissue samples. Our goal was to evaluate proteins involved in the regeneration of pancreatic islet cells by subtractive proteomics with new and exciting instrumentation.

EXPERIMENTAL PROCEDURES

Mice—Male C57B1/6 mice from The Jackson Laboratory (Bar Harbor, Maine) were maintained under pathogen-free conditions until harvesting of splenocytes at 6 weeks of age for MS analysis.

Isolation of CD45− and CD45+ Cell Populations from C57B1/6 Mice—Splenocytes from seven mice were harvested from spleen tissue mechanically disrupted with forceps. Following the lysis of red blood cells (140 mM ammonium chloride in 100 mM Tris buffer, pH 7.5) CD45− and CD45+ spleen cells were separated using mouse-specific CD45 MicroBeads (Miltenyi Biotec, Auburn, CA) according to the manufacturer’s instructions. Briefly pooled splenocytes were counted, and 102 cells were resuspended in 10 μl of CD45 MicroBeads in 90 μl of 1× PBS, pH 7.2 and incubated for 15 min at 4 °C. Cells were washed with 1× PBS, centrifuged at 1,800 rpm for 5 min, and diluted in 3 ml of 1× PBS. Labeled cells were placed in a SuperMAGS column in the magnetic field and washed three times with 3 ml of 1× PBS to recover the unlabeled, negatively selected CD45− cells. The column was removed from the magnetic field, and the positively selected CD45+ cell fraction was collected in 5 ml of 1× PBS.

Peptide Preparation, Separation, and In-gel Digestion—Whole cell fractions were prepared from either pooled CD45− or CD45+ cell populations using RIPA buffer (1× phosphate-buffered saline (Invitrogen), 0.5% deoxycholic acid, 1% Nonidet P-40, 0.1% SDS, 0.02 mM phenylmethylsulfonyl fluoride, 2 mM dithiothreitol, and 20 mM sodium orthovanadate (Sigma) and stored at −80 °C. Protein concentrations were determined by the Bio-Rad DC protein assay (Bio-Rad), and a total of 1 mg of protein was resolved by one-dimensional polyacrylamide gel electrophoresis using a 4–12% bis-Tris gel on a Novex Mini-Cell (Invitrogen). Proteins prepared from CD45− and CD45+ cells were separated by molecular weight to ~5 cm from the origin in independent 1-well Invitrogen gels using 1× MES, SDS running buffer. Gels were removed from the cassette, stained with 0.1% Coomassie Brilliant Blue R250 (Pierce) for 2 min, and destained overnight in a solution of 10% acetic acid and 30% methanol. Gel bands were then excised and used to prepare independent in-gel trypsin (Promega, Madison, WI) digests for each gel as described previously (20). Peptides were extracted by washing gel pieces twice for 20 min at room temperature with a solution containing 5% formic acid and 50% acetonitrile, then dried to complete dryness by vacuum concentration, and stored at −20 °C until analysis by mass spectrometry.

Peptide Sample Preparation—Prior to analysis, peptide samples were prepared by using in-house nanocolumns to remove excess salt and particulates. Nanocolumns were prepared using Eppendorf Geloader™ tips (Brinkmann Instruments) pinched 1 mm from the end and filled to 1.5 cm with Oligo R3 reverse phase packing resin (PerSeptive Biosystems, Framingham, MA) in 100% 2-propanol with the aid of a 1-mL syringe with a flow of 1 μl/min without drying the resin. The column was washed with 20 μl of 2-propanol and 40 μl of elution buffer (97.4% acetonitrile, 2.5% H2O, and 0.1% formic acid) and conditioned with loading buffer (98% H2O, 0.1% TFA, and 2% acetonitrile). Peptide samples were diluted in 30 μl of loading buffer, incubated for 10 min at room temperature, and loaded onto the column. The column was washed with 40 μl of 0.1% TFA, and peptides were eluted with 20–30 μl of elution buffer. Samples were vacuum-dried and reconstituted in 100–200 μl of sample buffer A (95% H2O, 5% acetonitrile, and 0.5% formic acid), and 2% was loaded via autosampler for MS analysis.
LCQ Deca-XP Mass Spectrometry—LC-MS/MS was performed using an LCQ Deca XPPlus ion trap mass spectrometer (ThermoElectron, San Jose, CA). Samples were autosampled (Famos autosampler, LC Packings, Sunnyvale, CA) to a 125-μm-inner diameter fused silica C₁₈ capillary column packed to 14 cm with Magic (Michrom BioResources) C₁₈ resin (200-Å pore size, 5-μm diameter) using an Agilent 1100 series binary pump with an in-line flow splitter. Peptides were loaded onto the column for 15 min at 120 bars in buffer A (2.5% acetonitrile and 0.15% formic acid). Peptides were then resolved by applying a gradient of 5–33% buffer B (97.5% ACN and 0.15% formic acid) for 55 min at 60 bars. Five MS/MS spectra were acquired per cycle in a data-dependent manner (2).

LTQ FT Mass Spectrometry—LC-MS/MS was performed using an LTQ FT hybrid linear ion trap FT ICR MS system (ThermoElectron, San Jose, CA) in similar fashion to the XP with slight modifications. All aspects of the microcapillary separation were identical including the column, autosampler, sample amounts, HPLC pumps, and HPLC gradient formation. Within the LTQ, 10 MS/MS spectra were acquired per cycle in a data-dependent manner from a preceding FT scan (400–1,800 m/z at a resolution setting of 10⁵) with an automatic gain control (AGC) setting of 2 × 10⁶. Charge state screening was used such that singly charged peptides were not selected, and a threshold of 500 counts was required to trigger MS/MS spectra. Where possible, the instrument operated in a parallel processing mode where the LTQ and ICR cell were both detecting ions.

Database Searching—Raw MS/MS data were searched against the mouse NCBI non-redundant database with no enzyme constraint using SEQUEST (19) (version 27, revision 9). Parameters included a precursor mass tolerance of 1.08 and 2.0 Da for LTQ FT and LCQ XP mass spectrometers, respectively. Fragment ion tolerance was set at the default, and dynamic modification to methionine (+15.9949) was allowed. Cysteines were searched with a static modification (+71.0370). Only fully

**Fig. 1.** Schematic representation of methods. Spleens were removed from C57B1/6 mice and divided into CD45⁺ and CD45⁻ populations by CD45 microbead separation as described under “Experimental Procedures.” Whole cell protein fractions were prepared and resolved by SDS-PAGE. Fourteen regions were subjected to in-gel digestion with trypsin digests, and identical aliquots were analyzed by LC-MS/MS techniques on both LTO FT and LCO XP mass spectrometers. The goal was to compare the utility of new and existing mass spectrometry instrumentation to identify proteins expressed in two spleen cell populations. Proteins were identified using the SEQUEST algorithm (19) on a Linux cluster. Proteins identified by LC-MS/MS were characterized by subtractive protein analysis of the CD45⁻ and CD45⁺ cell populations, by UniProt and other publicly available databases, and by gene ontology assignment using the GoMiner application.
tryptic peptides were considered for further processing with Xcorr and mass accuracy thresholds as described in the text. A simple target/decoy database approach was used to estimate false positive rates through distraction of random hits and to establish threshold criteria (3) such that <1% false positives were included in the peptide list. An Excel spreadsheet file is available (see supplemental tables) that contains all MCP-required information concerning peptide identifications for more than 30,000 identified peptides.

RESULTS

To assess the ability to quickly identify proteins of interest from primary tissue samples by subtractive proteins analysis, we performed experiments using a recently developed LTQ FT hybrid ion trap mass spectrometer and a traditional LCQ Deca XPplus instrument. We analyzed spleen tissues pooled from up to seven C57B1/6 mice and separated by magnetic beads into distinct CD45− and CD45+ cell populations (Fig. 1). CD45 magnetic bead separation has been performed routinely in our laboratory, and we typically observe only a small (1–3%) CD45− contamination rate in the CD45+ cell fraction as determined by flow cytometry analysis (data not shown). In the present study, 1 mg of total cell lysate from CD45− or CD45+ cells was resolved by one-dimensional SDS-PAGE,
and 14 independent in-gel trypsin digests of each cell population were prepared for LC-MS/MS analysis. A total of 28 gel slices were prepared, and a small aliquot was analyzed by both LTQ FT and XP mass spectrometers. Proteins were identified by SEQUEST using only fully tryptic peptides as a starting point for matches. In addition, each dataset was required to have only a 1% false positive rate as estimated via a target/decoy database approach (3) to allow for a maximum of 1% false positive rate at the peptide level (see Fig. 4). N/A, not applicable.

Table I presents a summary of the entire experiment. In total, 56 samples (28 on each instrument) were analyzed. Database searching using a composite target/decoy database, where mouse proteins were present in both a forward and a reversed orientation (3), was used to estimate false positive rates from each dataset and to provide final filtering criteria. Each dataset was allowed a maximum estimated false positive rate of 1%. The final filtering criteria for the LTQ FT dataset utilized mass accuracy as an additional constraint, which allowed for lower XCorr values to be used. Although fewer proteins were ultimately identified in the CD45 – samples, this was not due to sample loading because for most proteins similar numbers of CD45 – and CD45 + peptides were detected (Fig. 3). In fact, more total peptides were actually acquired five MS/MS scans/cycle with an average scan time of 7.3 s. Approximately 450 MS and 2,300 MS/MS spectra were acquired per gel slice. In the example shown, 442 unique peptides (155 proteins) were identified from gel slice 7.

Table I presents a summary of proteomic analyses of spleen cell samples analyzed by LTQ FT and LCQ XP.

| Spleen cell population a, b | LTQ FT | LCQ XP |
|-----------------------------|--------|--------|
| CD45 – | CD45 + | CD45 – | CD45 + |
| Protein for SDS-PAGE (mg)  | 1      | 1      | 1      | 1      |
| Gel slices analyzed         | 14     | 14     | 14     | 14     |
| HPLC gradient time (min)    | 55     | 55     | 55     | 55     |
| MS scans                    | 7,987  | 7,554  | 6,097  | 6,109  |
| MS/MS scans                 | 78,870 | 75,539 | 30,484 | 30,543 |
| XCorr (1 +, 2 –, 3 –, 4 –) c| 1, 1, 1, 1 | 1, 1, 1, 1 | 1.9, 2.2, N/A | 1.9, 2.2, N/A |
| ΔCorr e                      | None   | None   | None   | None   |
| ppm cutoff f                 | 4.5    | 4.5    | None   | None   |
| Unique proteins identified   | 2,425  | 1,948  | 1,147  | 989    |
| Unique peptides identified   | 12,175 | 12,466 | 3,721  | 3,823  |

a One milligram of whole cell protein fraction was prepared from CD45 – and CD45 + spleen cells.
b A total of 28 gel slices were prepared by in-gel trypsin digestion and analyzed by LTQ FT and LCQ XP over a 55-min HPLC gradient for each sample.
c Peptides were identified with SEQUEST (19) using no enzyme specificity. Only fully tryptic peptide matches were considered. Due to two very powerful constraints (search nontryptic → require tryptic; search 1.1 Da → require 4.5 ppm) almost no other thresholding was required to achieve <1% false positives on the LTQ FT. Thresholds for XCorr, ΔCorr, and ppm cutoff were determined via a target/decoy database approach (3) to allow for a maximum of 1% false positive rate at the peptide level (see Fig. 4). N/A, not applicable.
d See Supplemental Tables 5–8 for a list of peptides identified by LTQ FT and LCQ XP analyses.
but false positives are equally split as matches to either orientation. Not using mass accuracy as a filter provided 75 reversed hits of 1,307 total peptides for an unacceptable false positive (FP) rate of 11.5% (2 × reverse hits/total hits × 100%). Using the mass accuracy filter (4.5 ppm) resulted in a FP rate of only 0.7% (Fig. 4A, inset). In an analysis of the peptides identified in all CD45− samples combined (14 gel slices) by LTQ FT, ~10,000 peptides had mass accuracies of <1 ppm (Fig. 4B). The distribution of all peptides identified within 10 ppm is shown in Fig. 4B. To contain >99% of the correct answers, 4.5 ppm was used as the final cutoff for correct matches. The average absolute mass deviation was 0.84 ppm for all accepted peptides. It should be noted that only a minimal XCorr cutoff (1.0) was needed when both a tryptic peptide requirement and mass accuracy were used. When a mass accuracy cutoff was not used to identify correct answers, we were required to increase both XCorr and ΔCorr to maintain a false positive rates of <1%, significantly reducing the total number of peptides and proteins identified (Fig. 4B, inset).

We next used subtractive proteomics to discover differences between CD45− and CD45+ cell populations. We compared results from both instruments of all identified proteins (all protein hits) and proteins identified by two or more pep-
tides in one sample (excluding one-hit proteins) (Fig. 5). As anticipated, significantly more proteins were identified by the LTQ FT for either all protein hits or proteins identified with two or more peptides per protein. We also observed an ~10% increase in the number of overlapping proteins between cell groups when using the LTQ FT. In an analysis of the most abundant differences between cell groups (excluding one-hit proteins), relatively few proteins were identified exclusively in

Fig. 4. The effect of mass accuracy on peptide identification by LTQ FT analysis. A, distribution of peptide spectral matches from SEQUEST analysis of an LC-MS/MS analysis of a single representative gel region using a composite target/decoy mouse database. The effect of XCorr and mass accuracy (ppm) is shown. The only requirement was that peptide spectral matches be fully tryptic. Peptide spectral matches derived from the decoy (reversed) database are shown in red, and those derived from the target (forward) database are shown in blue. The estimated FP rate was 11.5% for this dataset. Further filtering for correct matches concentrated in the ppm region between ±4.5 ppm resulted in an estimated FP rate of 0.7%. B, mass accuracy (ppm) distribution (between 0 and 10 ppm) for all CD45 samples (14 gel regions) analyzed by LTQ FT using only tryptic peptides and an XCorr cutoff of 1.0. The inset shows total unique peptides and proteins in 14 CD45 gel regions at <1% false positive peptides detected with and without ppm cutoff. When a ppm cutoff was not used, higher Xcorr and ΔCorr were required to maintain a low false positive rate, reducing the total number of peptides and proteins identified.
CD45<sup>+</sup> samples by both the LTQ FT (31 proteins) and LCQ XP (23 proteins) likely due to the higher purity of that preparation (Fig. 3C). Although the potential biological significance of the differences in proteins identified in spleen cells will be presented elsewhere, a preliminary description of the 220 and 31 most abundant proteins exclusively identified in CD45<sup>−</sup> and CD45<sup>+</sup> cells, respectively, is presented in Supplemental Tables 1 and 2.

In an effort to better determine the reproducibility of shotgun subtractive proteomics, we compared all exclusive CD45<sup>−</sup> proteins identified by the XP and LTQ FT from two independent analyses by number of peptides per protein (Tables II and III). The entire experiment including harvesting, SDS-PAGE, gel region excision, and LC-MS/MS analysis was repeated as the first experiment. Exclusive CD45<sup>−</sup> cell proteins obtained in experiment 1 (XP<sub>1</sub>-CD45<sup>−</sup>) by subtractive analysis were listed by the number of peptides identified for each protein and compared with experiment 2 (XP<sub>2</sub>-CD45<sup>−</sup>). Table II shows that few exclusive CD45<sup>−</sup> cell proteins were identified by XP analysis by four or more peptides. With such low numbers it was difficult to estimate the reproducibility of the results in a replicate XP analysis of the samples. With proteins identified by two to four peptides, ∼30% were also identified by subtractive analysis in experiment 2 (XP<sub>2</sub>-CD45<sup>−</sup>). Exclusive CD45<sup>−</sup> cell proteins identified by only one peptide were observed in a replicate analysis only 17% of the time (59 of 340 proteins) with many of the proteins not identified in either CD45<sup>−</sup> or CD45<sup>+</sup> cells. A similar comparison of

---

**TABLE II**

Comparison of exclusive CD45<sup>−</sup> cell proteins identified by LCQ XP in two independent analyses of 14 gel regions

| Proteins exclusively identified in Exp. 1 (XP<sub>1</sub>-CD45<sup>−</sup>) | Proteins exclusively identified in both Exp. 1 and Exp. 2 (XP<sub>2</sub>-CD45<sup>−</sup>) |
|---------------------------------------------------------------|---------------------------------------------------------------|
| No. of proteine propeptide/protein | No. of proteins from CD45<sup>−</sup> | No. of proteins identified in both<sup>a</sup> | Percent reproducible<sup>b</sup> |
| 7+ | 4 | 2 | 50 |
| 6 | 1 | 1 | 100 |
| 5 | 1 | 0 | 0 |
| 4 | 3 | 1 | 33 |
| 3 | 15 | 4 | 27 |
| 2 | 62 | 20 | 32 |
| 1 | 340 | 59 | 17 |

<sup>a</sup> Protein identifications in experiment 2 were exclusive but not necessarily identified by the same number of peptides.

<sup>b</sup> For results to be considered reproducible proteins had to be identified in replicated studies and identified as exclusive to CD45<sup>−</sup>.

---

**TABLE III**

Comparison of proteins identified by subtractive proteomics in CD45<sup>−</sup> samples in LTQ FT in two independent analyses of 14 gel regions

| Proteins exclusively identified in Exp. 1 (FT<sub>1</sub>-CD45<sup>−</sup>) | Proteins exclusively identified in both Exp. 1 and Exp. 2 (FT<sub>2</sub>-CD45<sup>−</sup>) |
|---------------------------------------------------------------|---------------------------------------------------------------|
| No. of peptide/protein | No. of proteins from CD45<sup>−</sup> | No. of proteins identified in both<sup>a</sup> | Percent reproducible<sup>b</sup> |
| 8+ | 6 | 5 | 83 |
| 6 | 7 | 5 | 71 |
| 5 | 8 | 4 | 50 |
| 4 | 20 | 10 | 50 |
| 3 | 30 | 13 | 43 |
| 2 | 149 | 66 | 44 |
| 1 | 589 | 173 | 29 |

<sup>a</sup> Protein identifications in experiment 2 were exclusive but not necessarily identified by the same number of peptides.

<sup>b</sup> For results to be considered reproducible proteins had to be identified in replicated studies and identified as exclusive to CD45<sup>−</sup>.
FIG. 6. Comparison of exclusive CD45\(^{−}\) cell proteins characterized by replicate LCQ XP and LTQ FT studies. A, the top 35 proteins identified by LCQ XP in CD45\(^{−}\) samples in experiment 1 (XP\(_{1}\)-CD45\(^{−}\)) that were exclusive to CD45\(^{−}\) cells (not present in CD45\(^{+}\) samples) are compared in two completely independent analyses. Replicate LC-MS/MS analysis showed 17 of the top 35 proteins that were exclusive to CD45\(^{−}\) samples in XP\(_{1}\)-CD45\(^{−}\) were identified as exclusive to CD45\(^{−}\) samples or not detected in either sample in experiment 2 (XP\(_{2}\)-CD45\(^{−}\) and XP\(_{2}\)-CD45\(^{+}\)). Of the 18 proteins that were identified in CD45\(^{−}\) samples in experiment 2, seven proteins were identified with less total unique peptides in CD45\(^{−}\) samples, indicating that the relative abundance is higher in CD45\(^{−}\) samples although not exclusive. B, the top 35 proteins identified by LTQ FT in CD45\(^{−}\) samples (FT\(_{1}\)-CD45\(^{−}\)) exclusive to CD45\(^{−}\) samples were compared in two independent analyses. LC-MS/MS analysis of replicate samples by LTQ FT showed 24 of the top 35 proteins that were exclusive to CD45\(^{−}\) samples in FT\(_{1}\)-CD45\(^{−}\) were also exclusive in FT\(_{2}\)-CD45\(^{−}\) samples or were not detected in either sample in experiment 2. Of the 11 proteins that were not exclusive in experiment 2, eight were identified with less total unique peptides in CD45\(^{−}\) samples. ANK1, ankyrin 1; VWF, von Willebrand factor; FGB, fibrogen B-\(\beta\) chain; EHD4, EH domain-containing protein 4; PPOX, protoporphyrinogen oxidase; CA1, carbonic anhydrase I; FGA, fibrogen A-\(\alpha\) chain; PARVB, \(\beta\)-parvin; ALOX12, arachidonate 12-lipoxygenase; CASP3, caspase 3; PBEF1, pre-B-cell colony-enhancing factor; LMNA, lamin A; TUBB1, \(\beta\) tubulin 1; SPTA1, spectrin \(\alpha\) chain; SPTB, spectrin \(\beta\) chain; MPP1, murine proliferation-associated protein 1; ADA, adenosine-deaminase; NAPG, \(\gamma\)-soluble NSF attachment protein.
Subtractive Proteomics of Spleen Cells

LTQ FT results showing exclusive CD45<sup>−</sup> cell proteins identified in experiment 1 (FT<sub>1</sub>-CD45<sup>−</sup>) with results obtained in experiment 2 (FT<sub>2</sub>-CD45<sup>−</sup>) showed improved reproducibility and many more proteins identified (Table III). As expected, reproducibility decreased with decreasing number of peptides identified. Fig. 6 shows a partial list of exclusive CD45<sup>−</sup> cell proteins characterized by replicate LCQ XP and LTQ FT studies.

Using the LTQ FT, we identified 220 proteins by two or more peptides exclusively in CD45<sup>−</sup> cells through subtractive proteomics. As a preliminary assessment and comparison of the performance of the two instruments only exclusive proteins identified by LTQ FT analysis in experiment 1 were characterized using a Swiss-Prot/TrEMBL/UniProt database search (us.expasy.org and www.pir.uniprot.org). A list of the 220 proteins is shown in Supplemental Table 1 by the number of peptides identified for each protein in LTQ FT experiment 1. The number of peptides identified in subsequent analysis by LTQ FT and LCQ XP is also shown. Of the 220 exclusive proteins identified in LTQ FT experiment 1, 126 proteins were consistently absent in CD45<sup>−</sup> cell samples after repeated analysis. A partial list of the most abundant protein differences identified in CD45<sup>−</sup> cells that were consistent across studies is shown in Table IV. Some of the most abundant differences observed between cell types were due to a small population of erythrocyte and platelet cells. Red blood cells are the dominate cell type in CD45<sup>−</sup> cell fractions within the spleen population, and although most (>98%) are removed during the initial cell lysis of spleen tissue, small amounts remained. Platelets also do not express CD45 and were identified within the initial cell lysis of spleen tissue, small amounts remained. Platelets also do not express CD45 and were identified within the initial cell lysis of spleen tissue, small amounts remained.

Gene ontology characterization using the Javascript GoMiner (21) provided a way to further categorize these identified gene products by function. Of particular interest to this study were stem cell proteins that might be involved in the regeneration process of pancreatic cells already described (18). GoMiner was used to apply all known gene ontology

### Table IV

| Gene<sup>a,b</sup> | CD45<sup>−</sup>/CD45<sup>+</sup> | Protein name | Function |
|-------------------|----------------------|-------------|----------|
| ANK1              | 22/0                  | Ankyrin 1   | Attach integral membrane proteins to cytoskeletal elements |
| SPTA1             | 21/0                  | Spectrin α chain, erythrocyte | Cytoskeletal structure erythrocyte plasma membrane |
| SPTB              | 15/0                  | Spectrin β chain, erythrocyte | Cytoskeletal structure erythrocyte plasma membrane |
| TUBB1             | 9/0                   | β-tubulin 1, class VI | Microtubule polymerization |
| FGB               | 8/0                   | Fibrinogen B-β chain | Cofactor in platelet aggregation |
| IFG2              | 8/0                   | Integrin α-IIb precursor | Receptor for fibrinectin involved in platelet aggregation |
| CA1               | 6/0                   | Carbonic anhydrase I | Reversible hydration of carbon dioxide |
| VWF               | 6/0                   | von Willebrand factor | Promotes adhesion of platelets to the sites of vascular injury |
| EHD4              | 6/0                   | EH domain-containing protein 4 | Unknown |
| ALOX12            | 6/0                   | Arachidonate 12-lipoxygenase | Has oxygenase and 14,15-leukotriene A4 synthase activity |
| PARVB             | 5/0                   | β-Parvin | Regulation of cell adhesion and cytoskeleton organization |
| FGA               | 5/0                   | Fibrogen A-α chain | Involved in platelet aggregation |
| EBP42             | 5/0                   | Erythrocyte membrane protein band | Regulation of erythrocyte shape |
| PRSS34            | 5/0                   | Mast cell protease-11 | Unknown; possible serine-type peptidase activity |
| LMNA              | 4/0                   | Lamin A | Component of the nuclear lamina |
| PSAP              | 4/0                   | Sulfated glycoprotein 1 | Unknown |
| PPOX              | 4/0                   | Protoporphyrinogen oxidase | Catalyzes the six-electron oxidation of protoporphyrinogen IX |
| PBEF1             | 4/0                   | Pre-B-cell colony-enhancing factor | Unknown |
| CASP3             | 4/0                   | Caspase 3 | Activation cascade of caspases responsible for apoptosis execution |
| PSMB5             | 4/0                   | Proteasome subunit β type 5 | The proteasome has an ATP-dependent proteolytic activity |
| NAPG              | 4/0                   | γ-Soluble NSF<sup>c</sup> attachment | Required for vesicular transport between the ER<sup>c</sup> and the Golgi |
| Q9J2P0            | 4/0                   | Fibrinogen γ-chain | Cofactor in platelet aggregation |
| XRN2              | 3/0                   | 5′-3′ exoribonuclease 2 | Possesses 5′-3′ exoribonuclease; involved in homologous recombination |
| AP2M1             | 3/0                   | Clathrin coat assembly protein AP50 | Component of the adaptor complexes that link clathrin to receptors |
| FECH              | 3/0                   | Ferrochelatase | Catalyzes the ferrous insertion into protoporphyrin IX |

<sup>a</sup> Top 25 proteins identified by subtractive proteomics by LTO FT analysis.

<sup>b</sup> Final list of genes expressed in CD45<sup>−</sup> cells was produced from exclusive CD45<sup>−</sup> genes identified in LTQ FT experiment 1 compared to a replicate LTQ FT study and two replicated LCQ XP analyses.

<sup>c</sup> Total number of unique peptides identified for each gene in CD45<sup>−</sup> and CD45<sup>+</sup> cells in LTQ FT experiment 1.

<sup>d</sup> N-Ethylmaleimide-sensitive factor.

<sup>e</sup> Endoplasmic reticulum.
categories associated with the 126 proteins exclusive to CD45− spleen cells in an automated fashion. Gene ontology terms (as defined by the gene ontology consortium, www.geneontology.org/) were assigned to ~90% of the proteins identified by subtractive proteomic analysis. A partial list of spleen cell proteins described by seven different gene ontology terms (cell adhesion, cell cycle, development, DNA binding, extracellular space, integral to membrane, and signal transduction) that may identify relevant CD45− extracellular space, integral to membrane, and signal transduction are seven proteins involved in development. An interesting observation is that ~50 of the proteins described in Supplemental Table 3 (shaded proteins) have been observed previously in mouse tissue at various stages of development, and many proteins are involved in spermatogenesis and other biological processes relevant to stem cell biology (www.informatics.jax.org/). A list of all the gene ontology terms associated with the 126 exclusive CD45− proteins is shown in Supplemental Table 4.

**DISCUSSION**

A recent comparison between human and mouse genomes showed gene homology approaching 99% (22), indicating that the diversity between such far removed mammals is largely dependent on differences in protein regulation and expression of strikingly similar genes. This further highlights the importance of developing tools to evaluate biological diversity at a more global protein level. To meet this challenge, new methods and instrumentation are being developed for a more comprehensive, systems-level understanding of biology. In the current studies we used new instrumentation (LTQ FT) to characterize a complex protein mixture by subtractive proteomics. In preparations of whole cell fractions from two diverse spleen cell populations we identified ~2,000 proteins per preparation from more than 12,000 peptides by LTQ FT analysis and, by comparison, ~1,000 proteins (~3,700 peptides) by LCQ XP analysis. Each of these peptide datasets had a false positive rate estimated to be <1% based on a target/decoy database searching strategy (3). A subtractive proteomic comparison of proteins expressed by CD45− and CD45+ cells was also improved by replicating the entire experiment.

The field of quantitative proteomics provides a discipline for differential protein expression profiling. Stable isotope labeling is the surest way to precisely measure differences in protein abundance. However, the object of some proteomic experiments is to identify proteins that are exclusively expressed in one state but not the other. Two recent studies have provided useful experience with subtractive proteomics as a quantitative proteomic technique (14, 15). Our interest was in finding stem cell-specific proteins that were exclusively expressed in the CD45− cell population. Some evidence of the usefulness of the technique comes from the finding of red blood cell- and platelet-specific proteins exclusively in the CD45− cells. This is because a very small amount of red blood cells and platelets are not completely removed during cell preparations and remain within CD45− cell populations. Twenty-four proteins of 126 from subtracted CD45− datasets were either red blood or platelet cell proteins.

We believe the idea of using shotgun sequencing of whole proteomes followed by subtractive analysis will be useful for many specific proteomic applications because it (i) can be used on primary tissue, (ii) makes use of mature technologies (peptide sequencing) that require little refinement, (iii) can be improved by simple replication of the experiment to provide more significant differences, and (iv) relies on instrumentation that now can provide much deeper analysis of protein sequences because of increased scan rates and mass accuracy.

One stated purpose of this project was to initiate the protein profiling of a newly identified stem cell population in the spleen of adult mice (18). This stem cell population is contained within the capsular regions of the CD45− regions of the spleen and has been proposed to be a remnant of an embryonic stem cell region called the aorta gonad mesoderm (23). Mass spectrometry analysis of the CD45− cell subset readily revealed proteins involved in development, indicative of the persistence of a fetal stem cell in an adult animal. Subtractive proteomics showed development-specific proteins that control the formation of the fetal nervous system (cerebellum, neurogenesis, and axon guidance axonogenesis), blood vessels, muscle, skin, and gonads (gametogenesis and spermatogenesis). Although many of the proteins exclusive to CD45− samples had unknown function, the majority are abundantly expressed on day 10–11 of gestation (Supplemental Table 1). Therefore, this preliminary analysis of the CD45− fractions of the spleen is consistent with a stem cell population that might represent a frozen fetal cluster of a midstage murine embryo (24). These interesting candidate proteins were (i) identified by two or more unique peptides, (ii) exclusive to CD45− samples, (iii) detected by LTQ FT analysis, and (iv) consistent across replicate studies. These results indicate that the improved performance of the LTQ FT significantly aided our ability to characterize complex protein mixtures by shotgun proteomics. Our primary focus for future analysis will be proteins implicated in development and exclusive integral plasma membrane proteins that maybe used to more specifically isolate potential stem cells from the CD45− cell population.

Shotgun proteome analysis by LC-MS/MS with new instru-
Subtractive Proteomics of Spleen Cells

mentation provided significantly larger datasets for characterizing protein expression differences. Yet a number of concerns remain in particular with respect to the reproducibility of the results and with data processing. The reproducibility was considerably improved by LTQ FT analysis (Table III) but remained low for the majority of the proteins, which were typically identified by five or less peptides. This can be somewhat misleading for a truly subtractive approach. It is likely that many proteins are expressed differentially between the two populations but not exclusively. If a 10-fold difference in protein expression results in a peptide being detected in the control sample of one replicate but not in the other, this protein would not be included in the subtracted list but may be of great interest to the investigator due to the 10-fold difference. Repeated sample analysis can be used to provide more confidence that the protein is truly not present in a cell population or that it is present at significantly less abundant quantities. To improve the reliability of our final data we only accepted subtractive differences that were consistent in multiple analyses.

Gene ontology classification of the proteins identified was used to characterize proteins in an automated fashion to improve data processing. In the present study most of the proteins were associated with at least one gene ontology term, but ~10% were completely unknown. Furthermore the evidence for term assignments vary by protein, and because many proteins have unknown functions, the gene ontology database is incomplete. Studies to characterize the biological relevance of differences observed in these experiments are currently being performed.

Acknowledgments—We thank members of the Gygi and Faustman laboratories for fruitful discussions.

This work was funded by National Institutes of Health Grants HG003456 (to S. P. G.) and TgT32DK07028 (to D. F.) and by the Iacocca Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The on-line version of this article (available at http://www.mcpjournal.org) contains supplemental material.

To whom correspondence should be addressed. Tel.: 617-432-3155; Fax: 617-432-1144.

REFERENCES

1. Wolters, D. A., Washburn, M. P., and Yates, J. R., III (2001). An automated multidimensional protein identification technology for shotgun proteomics. Anal. Chem. 73, 5683–5690
2. Peng, J., and Gygi, S. P. (2001) Proteomics: the move to mixtures. J. Mass Spectrom. 36, 1083–1091
3. Peng, J., Elias, J. E., Thoren, C. C., Licklider, L. J., and Gygi, S. P. (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J. Proteome Res. 2, 43–50
4. Wohlschlegel, J. A., and Yates, J. R. (2003) Proteomics: where’s Waldo in yeast? Nature 425, 671–672
5. Cronshaw, J. M., Krutchinsky, A. N., Zhang, W., Chait, B. T., and Matunis, M. J. (2002) Proxemic analysis of the mammalian nuclear pore complex. J. Cell Biol. 156, 915–927
6. Andersen, J. S., Wilkinson, C. J., Mayor, T., Mortensen, P., Nigg, E. A., and Mann, M. (2003) Proxemic characterization of the human centrosome by protein correlation profiling. Nature 426, 570–574
7. Shiio, Y., Eisenman, R. N., Yi, E. C., Donohoe, S., Goodlett, D. R., and Aebersold, R. (2003) Quantitative proxemic analysis of chromatin-associated factors. J. Am. Soc. Mass Spectrom. 14, 696–703
8. Garcia, B. A., Busby, S. A., Barber, C. M., Shabanowitz, J., Allis, C. D., and Hunt, D. F. (2004) Characterization of phosphorylation sites on histone H1 isoforms by tandem mass spectrometry. J. Proteome Res. 3, 1219–1227
9. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proxemomics. Mol. Cell. Proteomics 1, 376–386
10. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. 17, 994–999
11. Sanders, S. L., Jennings, J., Canutescu, A., Link, A. J., and Well, P. A. (2002) Proteomics of the eukaryotic transcription machinery: identification of proteins associated with components of yeast TFIID by multidimensional mass spectrometry. Mol. Cell. Biol. 22, 4723–4738
12. Rappsilber, J., Ryder, U., Lamond, A. I., and Mann, M. (2002) Large-scale proxemic analysis of the human spliceosome. Genome Res. 12, 1231–1245
13. Hitchcock, A. L., Auld, K., Gygi, S. P., and Silver, P. A. (2003) A subset of membrane-associated proteins is ubiquitinated in response to mutations in the endoplasmic reticulum degradation machinery. Proc. Natl. Acad. Sci. U. S. A. 100, 12735–12740
14. Schirmer, E. C., Florers, L., Guan, T., Yates, J. R., III, and Gerace, L. (2003) Nuclear membrane proteins with potential disease links found by subtractive proteomics. Science 301, 1380–1382
15. Wohlschlegel, J. A., Johnson, E. S., Reed, S. I., and Yates, J. R., III (2003) Global analysis of protein sumoylation in Saccharomyces cerevisiae. J. Biol. Chem. 278, 45662–45668
16. Schwartz, J. C., Senko, M. W., and Syka, J. E. (2002) A two-dimensional quadrupole ion trap mass spectrometer. J. Am. Soc. Mass Spectrom. 13, 659–669
17. Syka, J. E., Marto, J. A., Bai, D. L., Horning, S., Senko, M. W., Schwartz, J. C., Ueberheide, B., Garcia, B., Busby, S., Muratore, T., Shabanowitz, J., and Hunt, D. F. (2004) Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. J. Proteome Res. 3, 621–626
18. Kodama, S., Kuhthreiber, W., Fujimura, S., Dale, E. A., and Faustman, D. L. (2000) The regeneration during the reversal of autoimmune diabetes in NOD mice. Science 302, 1223–1227
19. Yates, J. R., III, Eng, J. K., McCormack, A. L., and Schiltz, D. (1995) Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. Anal. Chem. 67, 1426–1436
20. Baliff, B. A., Villen, J., Beausoleil, S. A., Schwartz, D., and Gygi, S. P. (2004) Phosphoproteomic analysis of the developing mouse brain. Mol. Cell. Proteomics 3, 1093–1101
21. Zeeberg, B. R., Feng, W., Wang, G., Wang, M. D., Fojo, A. T., Sunshine, M., Narasimhan, S., Kane, D. W., Reinhold, W. C., Lababidi, S., Bussey, K. J., Riss, J., Barrett, J. C., and Weinstein, J. N. (2003) GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol. 4, R28
22. Huang, H., Winter, E. E., Wang, H., Weinstock, K. G., Xing, H., Goodstadt, L., Stenson, P. D., Cooper, D. N., Smith, D., Alba, M. M., Ponting, C. P., and Fechtel, K. (2004) Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes. Genome Biol. 5, R47
23. Kodama, S., Davis, M., and Faustman, D. L. (2005) Diabetes and stem cell research turns to the lowly spleen. Science 302, 671–672
24. Yamada, H., Takete, K., and Takete, K. (2002) In vivo differentiation of stem cells in the aorta-gonad-mesonephros region of mouse embryo and adult bone marrow. Exp. Hematol. 30, 957–966

1470 Molecular & Cellular Proteomics 4.10