Label-Free Quantitation for Clinical Proteomics

Robert Moulder, Ph.D.
Senior Research Scientist
University of Turku
rmoulder@btk.fi

Young Ah Goo, Ph.D.
Research Assistant Professor
Associate Director, Mass Spectrometry Center
University of Maryland - Baltimore
ygoo@rx.umaryland.edu

*David R. Goodlett, Ph.D.
Isaac E. Emerson Professor of Pharmaceutical Sciences
Director, Mass Spectrometry Center
20 North Pine Street
Baltimore, MD 21201 USA
University of Maryland - Baltimore
dgoodlett@rx.umaryland.edu

* Corresponding Author

Abstract

Label free quantification (LFQ) has emerged as viable option for quantitative LC-MS/MS-based proteomic analyses for use on the scale of hundreds of samples such as are encountered in clinical analysis. Notably, sample preparation, sample loading, HPLC separations and mass spectrometric performance must be highly reproducible for this approach to be effective. The following protocols describe the key steps in the methods related to sample preparation and analysis for LC-MS/MS based label-free quantitation using standard data-dependent acquisition.

Key Words

Label-free quantification, proteomics, mass spectrometry, area under the curve (AUC),

1. Introduction

In combination with chromatographic separation, electrospray ionization mass spectrometry (ESI-MS) provides a means of characterization and comparison of highly complex mixtures in a concentration dependent manner. For ionizable analytes in general and peptides in particular, the intensity of the signal produced by ESI-MS will be proportional to its concentration in the eluting chromatographic peak *(1)* and thus depends on the peak volume. In circumstances where the quantity of sample is limited, analyte detectability can be enhanced by using separation columns with a smaller radial dimension. In such instances the benefit stems from the reduced flow rates and physical volume of the eluting peaks. In this manner the gain in sensitivity/peak intensity can be estimated from the ratio of the square of the column diameter *(2)*; e.g. from a 2 mm to a 75 μm i.d. column the estimated gain in signal intensity is 700 fold. The use of so called nano-flow systems has been particularly efficacious in applications with limited sample amounts, where flow rates of 50 to 300 nl/min have been typically used with columns of 50 to 75 μm i.d. In the past twenty years such nano-flow separations have grown as the mainstay of

qualitative and quantitative proteomics experiments. Moreover, with the maturation of liquid chromatographic systems capable of providing reproducible separation gradients at flow rates in the order of hundreds of nanolitres per minute, proteomics profiles can, with appropriate attention to detail, be produced in a reproducible manner suitable for quantitative comparisons *(3)*.

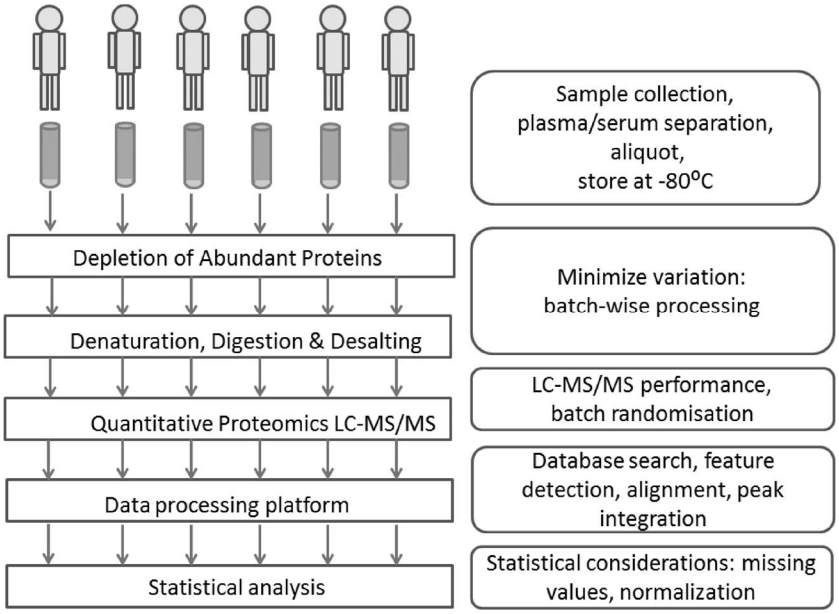
From LC-MS/MS analysis, the chromatographic profiles of the precursor/isotope envelope of identified peptides can be determined from precursor ion scans and integrated as an area under the curve (AUC) measurement proportional to the abundance of the analyte. Software has been developed for the alignment, integration and normalization of LC-MS profiles, which can be subsequently compared at the protein or peptide level. Amongst the common platforms that have been used, Progenesis (NonLinear Dynamics) and MaxQuant currently remain popular *(4)*.

At this stage the key procedures involve defining the peak profile for integration whilst not including nearby interferences from neighboring isotope clusters. With the Progenesis software the integrated peaks can be previewed and modified to remove errors/overlap in the integration. The collected peptide features for each protein are summed. Whilst the software may provide an option to use all peptides associated with each protein, it is important to base the integration on peptides that are unique to each protein, which means culling homologous peptide matches to more than one parent protein. For normalization of the data several options are possible, depending on whether you have used a spiked standard or prefer some *housekeeping* or reference protein. For a well characterized sample, where the injected amount has been matched, the use of total ion intensity of the identified proteins is a suitable choice.

For the analysis of plasma or serum samples a depletion step can be beneficial for extending the detectable dynamic range for the purpose of discovery work flows, as shown in Figure 1 from collection to depletion and data analysis. The workflow is thus broken down into sections: depletion (3.1); concentration of the depleted proteins (3.2); digestion (3.3); desalting (3.4); sample concentration adjustment (3.5);

LC-MS-MS and data analysis (3.6). Additional suggestions and alternatives are included under Notes in section 4.

Figure 1. A schematic of the sample preparation and analysis pipeline for label free quantification of depleted serum.



2. Materials

All reagents should be HPLC grade purity or at least 99%. This is specifically required for the following: MilliQ water, ammonium bicarbonate (NH_4HCO_3), 1,4-dithiotreitol (DTT), Iodoacetamide (IAA), Trypsin (Modified Sequencing Grade, e.g. from Promega V5111), and Urea.

2.1 Denaturing and Digestion Reagents

1. 50 mM NH_4HCO_3 buffer: Dissolve 350 mg NH_4HCO_3 in 100 ml water.
2. 8M Urea: Dissolve 24 g of urea in 50.0 ml of 50 mM NH_4HCO_3 solution.
3. Reducing reagent: 200 mM DTT in NH_4HCO_3 as above.
4. Alkylating reagent 200 mM IAA in NH_4HCO as above.
5. Trypsin solution: Select a sufficient quantity of trypsin to digest at a ratio of 1:30. For Promega sequence grade trypsin each vial contains 20 µg of trypsin,

mix multiple vials such they are digested with an equivalent batch. Typically for 30 μg of protein add 10 μl of a 0.1 μg/μl trypsin solution.

2.2 HPLC Buffers

1. Mobile phase buffer A: 0.2% formic acid in 98 % water 2% acetonitrile
2. Mobile phase buffer B: 0.2% formic acid in 5 % water 95% acetonitrile

3. Methods

3.1. Depletion

Depletion of the top 12 most abundant serum proteins with Pierce spin columns. (Prod # 85164, 85165). These are single use depletion cartridges that can be used to achieve parallel processing of multiple samples in a reproducible cost effective fashion. Make sure that you have sufficient columns that have been produced in the same batch / lot number when planning your experiments in case there is significant lot-to-lot variations that will affect quantitative results.

Proteins targeted for depletion:

α 1-acid glycoprotein (P02763), Fibrinogen (P02761), α 1-antitrypsin(P01009), Haptoglobin (P00738), α 2-macroglobulin (P01023), IgA (P01876), Albumin (P02768), IgG (P01857, P01859-61; all major subclasses of gamma globulin), Apolipoprotein A-I (P02647), IgM (P01871), Apolipoprotein A-II (P02652), Transferrin (P02787).

Although the column capacity is 10 μl (~600 μg of protein), the use of 8 μl is recommended. Columns should be stored at 4 ºC.

Depletion Protocol

1. Equilibrate depletion spin columns (Pierce) to room temperature.
2. Add 8 μl of serum or plasma into each column.
3. Close the caps and manually invert the columns until the resin is completely suspended in solution.

4. Place the columns to an Eppendorf-tube rotator and rotate for 1h at room temperature.
5. Once their end closures have been twisted off, place the columns into 2 ml Eppendorf tubes. Loosen the caps of the columns
6. Centrifuge the columns at 1000 g for 2 min (at room temperature). The collected filtrates are the depleted fraction (V ~ 500 µl).
7. The cartridges contain the bound fraction, which can, if required, be saved for further analysis.

3.2. Concentration of the sera after depletion:

To facilitate the handling and implementation of the denaturing and digestion protocols for the depleted serum fraction it is necessary to change the buffer composition and volume (see Note 1). As an alternative to the preferred ultracentrifugation-buffer exchange method please see Note 2.

Ultracentrifugation-Buffer exchange

1. Rinse a sufficient number (include extra) of Ultrafiltration spin columns for the samples (Sartorius-Stedim, Vivaspin, 4ml, 5kDa cut-off) with 1 ml of buffer equivalent to the Pierce slurry solution for the depletion cartridge (10mM PBS, 0.15M NaCl, 0.02% sodium azide, pH 7.4): +4 ºC, 3000 x g, ~20 minutes. Check the remaining volume and remove any columns that have performed at a slower rate. It is recommended that sufficient cartridges be purchased at the start of the project to conduct all the intended work as there can be batch-to-batch variations which can cause anomalies in results.
2. Concentrate depleted serum samples to a volume of ~100µl using the washed Ultrafiltration spin columns (+4 ºC, 3000 x g, ~20 min).
3. Perform buffer exchange with 8M Urea in 50 mM NH_4HCO_3
    1. 1200 µl of 8M Urea: +4 ºC, 3000 x g, 35 min.
    2. 500 µl of 8M Urea: +4 ºC, 3000 x g, 35 min.
    3. 500 µl of 8M Urea: +4 ºC, 3000 x g, 30 min.

    Final volume of sample should be ~100 µl.

4. To ensure that the concentrated proteins are in solution, ultrasonicate the spin columns for 5 min on ice and withdraw the liquid with a pipette.
5. To reduce the losses from transfer of the concentrate, wash the spin columns with 50 µl of 8M Urea by sonicating 5 min on ice. Combine this with the rest of the concentrated sample.

3.3. Digestion (See Note 1)

1. Add 5 µl of reducing solution to reach the final concentration of ~5mM DTT. Incubate 1h at +37 ºC.
2. Add 10 µl of alkylating solution to reach the final concentration of ~13 mM IAA. Vortex, incubate for 30 min in the dark at room temperature without an added IAA quenching step.
3. Add 850 µl of 50 mM NH_4HCO_3 to the samples to dilute the urea concentration below 1.5M before trypsin digestion.
4. Add 10 µl of trypsin (~1:30) to each sample. Incubate at +37 ºC overnight (16 h).

3.4. Desalting

Desalting with Sep Pak 50mg cartridge (part number:WAT054955). Dried tryptic digested peptides are reconstituted with 1 ml of 1% trifluoroacetic acid (TFA). Check the pH is acidic. See Note 3 concerning potential sources of contamination and interference.

1. Wet the column with 1 ml of 100 % methanol.
2. Equilibrate with 1 ml of 80% Acetonitrile + 0.1% TFA.
3. Equilibration with 2 x 1 ml of 0.1% TFA.
4. Apply sample. Repeat with flow through.
5. Wash the cartridge with 3 x 1 ml of 0.1 % formic acid (FA) in 2% acetonitrile.
6. Elute the peptides with 1 ml of 80% acetonitrile + 0.1% FA.

3.5. Sample concentration adjustment

1. Speed vac to dryness.
2. Reconstitute with 20 μl of 2% formic acid + 2% acetonitrile.
3. Use a Nanodrop detector (Thermo Scientific) to determine the UV absorbance spectrum (200 to 350 nm).
    a. Observe the estimated concentration based on the generalization that 1 absorbance unit is equivalent to 1 μg/μl of protein.
    b. Note the 260/280 nm ratio, this should be in the order of 0.7 for proteins/peptides not contaminated with nucleotides.
4. Spiking iRT peptides (Biognosis) 30:1 in the LC-MS/MS analyzed solution.
5. Using a 20 μl sample loop for 5 μl injections for a total of 200 ng, the target concentration should be 40 ng/μl. As the Nanodrop detector only gives reliable measurements down to 200 ng/μl, the dilutions should be made on the basis of solution at least at this concentration.

3.6. LC-MS/MS (see Note 3)

1. With an EasyNano-LC: A 20 x 0.1 mm i.d. pre-column packed with 5 μm Magic C18 (Michrom) silica connected by a New objective two-way union together with a 75 μm x 150 mm analytical column packed with 5 μm Magic C18 (Michrom).
2. A separation gradient from 5%B (95%A) to 35%B (65%A) in 65 min. at a flow rate of 300 μl/min.
3. Autosampler loop size and injection size: 20 μl sample loop for 5 μl injections.
4. To remove the influence of injection order, the samples are randomized for batches of single injections (each sample separated by a 15 min blank), with three/four replicate injections in total (three batches) with the system performance monitored between batches using a lab standard sample. A pool of the samples in the batch is analysed at the start of each batch. The

maintenance of constant/accountable instrument performance is essential for a successful LFQ experiment.

5. Using an Orbitrap-Velos Pro, to perform data dependent MS/MS data acquisition, the following are typical for a proteomics analysis: Ionization in positive ion mode with CID of the 15 most intense ions (m/z 300-2000, charge states > 1+). Dynamic exclusion *30 seconds. Orbitrap precursor ion scan resolution 60,000 (at m/z 400), with a target value of 1,000,000 ions and a maximum injection time of 100 milliseconds. For the ion trap the target values and maximum injection time values are set to 500,000 and 50 milliseconds *(5, 6)*. When making the selection of the top "n" most intense ions it is important to consider the associated duty cycle and the width of the *chromatographic peak, i.e. ensuring that there are sufficient MS1 data points to describe the peak elution profile.

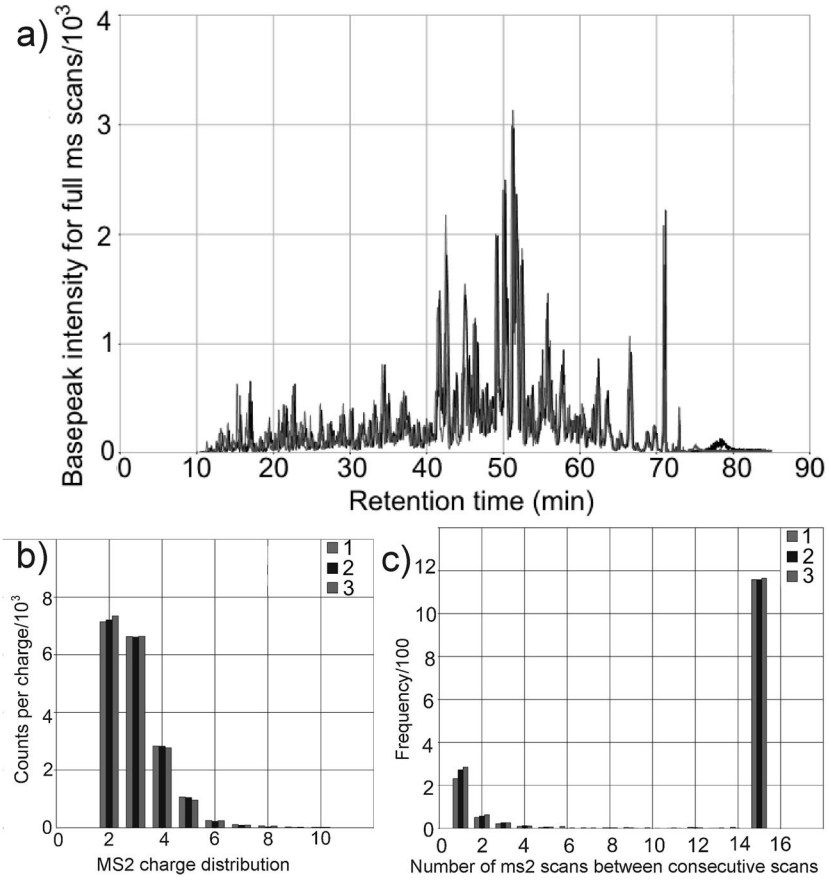
3.7. Data Analysis

The following describes a standard workflow built around the Proteome Discoverer and Progenesis software. See Note 4 addressing an alternative approach using the open source platform MaxQuant.

1. General data evaluation: Even though proprietary data analysis software is in constant evolution, in addition to Xcalibur and Proteome Discoverer (Thermo Scientific data accusation and analysis software), the free software *RawMeat* (Vast Scientific) provides a quick and useful tool to gain an overview of the data attributes and the suitability of the sample and applied method (Figure 2). For example, for a "top n" method it provides an indication of the suitability of the selected method in terms of sample complexity, although note that it is important for quantification to aim to have sufficient MS1 scans to define the elution profile of the chromatographic peaks (*vide supra*). In terms of the efficiency of the trypsin digestion the charge distribution provides a good indication of the general success: The doubly charged

precursors should be the most frequent for a tryptic digest. The current version of Progenesis, Progenesis QI includes similar QC metrics for sample preparation and instrument performance (see step 3 j)

Figure 2. Data evaluation using the RawMeat software. (a) Overlaid base peak chromatograms for three replicate LC-MS/MS analyses. (b) Charge distribution, the horizontal axis is precursor charge and the vertical axis is counts per charge. (c) top n usage: number of ms2 scans in between consecutive scans, shown for a top 15 method with CID fragmentation.



2. Using ProteomeDiscoverer (V. 1.4) with Mascot 2.1.

a) *Database Selection:* Select the most recent release of the human Swissprot database. If part of an extended study, then be sure to use the same version of the database for the duration of the study (i.e. avoid automatic updates which

will change search results). Whilst isoform databases can be very informative, the overlap due to differences in protein inference can be problematic when combining data from many searches. Include with the database a contaminant list (e.g. http://www.crapome.org/), then create a concatenated forward and reverse database so that the estimates of the false discovery rate can be accountable after the search has been conducted.

b) *Search parameters*: carboamidyl methylation fixed, methionine oxidation variable, suitable mass tolerances: e.g. 6 ppm precursor tolerance, 0.6 Da fragment tolerance, one missed cleavage, fragmentation type = ESI trap. With a concatenated database and Proteome Discover use the *Fixed value PSM validator* for false positive estimation.

c) *Multi-consensus report*: When the searches are complete use Proteome Discoverer to create a multi-consensus report of the collected search results, export this to Excel format including protein group and PSM results (layer 1 and 2).

3. Alignment, normalization and integration using Progenesis:

a) Create an experiment and load the RAW data files to the program as described in the Progenesis operation instructions.

b) Observe the ion maps of the loaded files and pay attention to any irregularities.

c) Select the alignment file or allow Progenesis to do this automatically. You might choose from a pooled standard sample that you have used, which gives a good representation of all the detected peaks.

d) When the alignment is complete confirm that the selected vectors are appropriate. Figure 3a displays a representation of the detected features with the vectors used for alignment. Progenesis provides a color-coded quality assessment of the alignment results, indicating regions that are in need of attention.

e) Filtering: choose which region of the chromatogram and mass range to use in the quantitative comparisons and select the precursor charge states to consider, e.g. 1 to 5.

f) Check the normalization results, noting any irregularities in the normalization factors; a lower intensity chromatogram will produce larger values. Be wary of files with large differences.

g) Design the experiment: select which analyses are replicates and define groups for comparison.

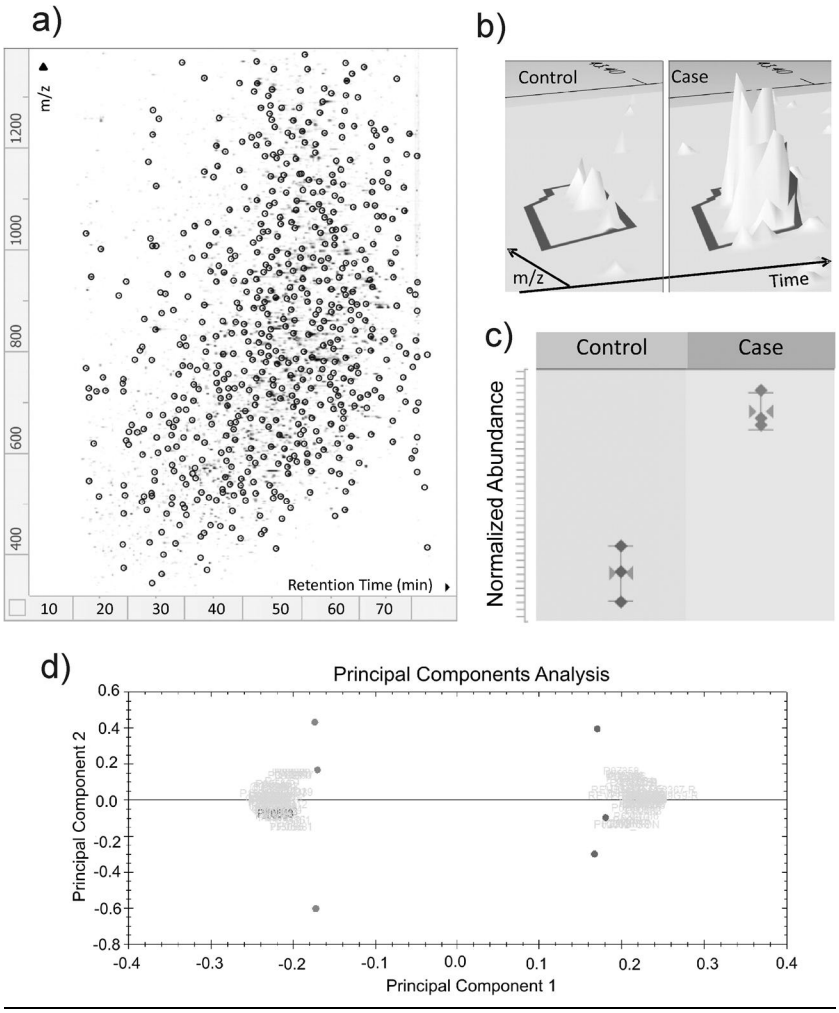
h) Check the integration of significant peaks, this is particularly critical when dealing with large fold differences and limited quantitative evidence. Figure 3b is taken from a 3D montage view of a differential peak that was detected from triplicate analyses of serum sample from two individuals.

i) To assign identifications to the features, upload the multi-consensus identification report created with Proteome discoverer.

j) At this stage in the workflow the quality control (QC) metrics are available. These can be used to confirm the quality of analysis in terms of differences that might occur during sample preparation and in instrument performance. For example, the charge distribution, number of missed cleavages and modification frequency between samples are reported, as are the chromatographic peak widths, mass accuracy and the scan rates.

k) The Progenesis statistics include PCA analysis and hierarchical clustering to identify differentially abundant features/proteins. Figure 3d shows the PCA separation displayed by Progenesis for data from the triplicate analysis of two distinct serum samples. Examples of a differentially abundant feature and a protein are shown in Figures 3b & c, respectively. Lists describing the protein and peptide intensities (normalized and un-normalized) may be exported in csv format so that more advance statistical analyses and comparisons can be made.

Figure 3. Examples of the different steps displayed in data analysis using Progenesis. (a) Visualization of an ion map, (b) Display of the precursor intensity and chromatographic profiles of a pair of differential features. (c) Difference detected between two proteins d) PCA analysis (illustrated for two samples).



3.8. Time Line for a LFQ serum proteomics experiment

Here follows a time line for a typical experiment that could be based on the preparation of a sub batch of 20 samples.

Day 1:

1. Deplete
2. Precipitate / concentrate

Day 2:

3. Reconstitute sample in approximately 150 µl of 8.0 M urea in a 1.5 mL polypropylene centrifuge tube.
4. Add 5 µl of Reducing Reagent and mix the sample by gentle vortex.
5. Reduce the mixture for 1 hour at room temperature or in an oven at 37 ºC.
6. Add 10 µl of Alkylating Reagent and alkylate for 0.5 hour at room temperature in the dark (use aluminum foil to cover the sample).
7. Add 850 µl of NH_4HCO_3 solution to dilute the urea before digesting it with trypsin.
8. Add trypsin in appropriate ratio (1:30) to approximate amount of protein by weight. Digest over night at 37 °C.

Day 3:

9. Adjust pH for desalting
10. Desalt
11. Dry, re-constitute and prepared solution for analysis
12. Design randomization experiment and start batch analyses.

4. Notes

1. *Volume reduction*. Starting with the processing of 8 µl of serum, the flow through fraction of interest is isolated at the expense of about 60-fold volume expansion, i.e. to 500 µl in PBS. To accommodate the digestion protocol, changes of the buffer composition and volume are needed. Precipitation can be performed in a “hands off” fashion, but may suffer from differences in performance and requires some degree of visual judgement. We have found that the buffer exchange approach is more reproducible. However, we have at times encountered problems with membranes that limit the passage of the liquid such that the procedure is slowed considerably.

2. *Alternative to the ultracentrifugation-buffer exchange*. As an alternative to the method described under section 3.2 one may use the following protein precipitation method on the depleted serum flow through:

1. Add 2 ml of cold acetone (-20 ºC) to the sample (4 volumes).
2. Invert several times and keep at -20 ºC for at least 4 hours (overnight).
3. Centrifuge for 15 min at 4 ºC at 1300 x g.
4. Remove the supernatant and dry the sample at room temperature.
5. Add 150 µl of 8M urea to the sample, vortex to dissolve.

3. *Instrument performance Ion suppression*. One of the largest threats to a successful LFQ experiment is the occurrence of ion suppression. This may arise from the contamination of the sample, buffers or originate from the LC system. Importantly, one should avoid detergents and sources of plasticisers. If you use facility labware that is routinely washed, in-house detergents may occur as harmful residues in these. Be sure to rinse with appropriate solvents and make sure that all components of the container are compatible with the solvents and acids used. Note that concentrated acetonitrile and/or formic acid can release residues from low quality/inappropriate labware. Internet discussion groups, such as the ABRF, frequently discuss and advise on such issues (http://www.abrf.org/).

It is essential to monitor instrument operations (both chromatographic and mass spectrometric) during the data acquisition. For a commercial or in house standard digest, the ion intensities, peak areas, sequence coverage and proteins/peptides identified can be used to gauge success or failure. Similarly, the retention times of a simple mixture are easily spotted and compared.

Trifluoroacetic acid can form ion pairs with peptide ions and improve their retention to the reverse phase column during desalting. It can, however, also cause ion suppression. For instance, if this is carried throughout the protocol and used in the elution of the peptides during desalting. We have previously associated such usage with the occurrence of ion suppression in the LC-MS/MS analyses.

4. *Alternative search and alignment strategies.* The previous examples have been described mostly in the context of using Progenesis. As an alternative, MaxQuant is an open source quantitative proteomics software package that built around the Andromeda search engine *(7, 8)*. The platform facilitates alignment and AUC quantification and has recently been developed to include visualization capabilities *(9)*. The output can be analysed with the Perseus module that was developed for bioinformatics analysis of the MaxQuant and Andromeda proteomics data.

References

1. Ikonomou M.G., Blades A.T., Kebarle P. (1991) Electrospray-ion spray: a comparison of mechanisms and performance. *Anal Chem* 63(18):1989-1998.

2. Abian J., Oosterkamp A.J., Gelpi E. (1999) Comparison of conventional, narrow-bore and capillary liquid chromatography / mass spectrometry for electrospray ionization mass spectrometry: practical considerations. *J Mass Spectrometry* 34(4):244–254.

3. Foss E.J., Radulovic D., Shaffer S.A. et al. (2007) Genetic basis of proteome variation in yeast. *Nat Genet* 39(11):1369-1375.

4. Nahnsen S., Bielow C., Reinert K. et al. (2013) Tools for Label-free Peptide Quantification. Molecular & Cellular Proteomics. *Mol Cell Proteomics* 12(3):549-556.

5. Kalli A., Smith G.T., Sweredoski M.J. et al. (2013) Evaluation and optimization of mass spectrometric settings during data-dependent acquisition mode: focus on LTQ-Orbitrap mass analyzers. *J Proteome Res* 12(7):3071-3086.

6. Moulder R., Bhosale S.D., Erkkila T. et al. (2015) Serum Proteomes Distinguish Children Developing Type 1 Diabetes in a Cohort with HLA-Conferred Susceptibility. *Diabetes* 64(6):2265-2278.

7. Cox J., Mann M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotech* 26(12):1367-1372.

8. Cox J., Neuhauser N., Michalski A. et al. (2011) Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J Proteome Res* 10(4):1794-1805.

9. Tyanova S., Temu T., Carlson A. et al. (2015) Visualization of LC-MS / MS proteomics data in MaxQuant. *Proteomics* 15(8):1453-1456.