Journal of Chromatography B, 877 (2009) 1306–1316

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

# Robustness and accuracy of high speed LC–MS separations for global peptide quantitation and biomarker discovery☆

Johan Lengqvist[a,b,*], Jorge Andrade[a,b], Yang Yang[c], Gunvor Alvelius[d], Rolf Lewensohn[a,b], Janne Lehtiö[a,b]

[a] Karolinska Biomics Center, Karolinska University Hospital, Stockholm, Sweden
[b] Department of Oncology/Pathology, Karolinska Institutet, Stockholm, Sweden
[c] Analytical Development, AstraZeneca R & D, Södertälje, Sweden
[d] Dept. of Medical Biochemistry & Biophysics, Karolinska Institutet, Stockholm, Sweden

ARTICLE INFO

*Article history:*
Received 1 July 2008
Accepted 17 February 2009
Available online 28 February 2009

*Keywords:*
Peptide separation
Biomarker discovery
Quantitative accuracy
Retention time alignment

ABSTRACT

The present work investigates qualitative and quantitative variability for rapid liquid chromatography–mass spectrometry (LC–MS) analyses of tryptic digests of total mammalian cell lysate. Experimental variability is characterized in a global manner across technical replicates using label-free quantification software (DeCyder MS 2.0). The effects of a novel time-alignment algorithm are described. Although effective to correct for retention time shifts the time-alignment tool adds only minute benefits for ultra performance (UP)LC–MS data. In differential display experiments, quantitative changes down to 3.5% of the experimental dynamic range could be accurately detected ($p < 0.001$). In 17 min analyses, almost 8500 peptide features were detected following injection of ~1 μg of total cytoplasmic digest. Ion intensity coefficient of variance were <15% for the majority (89%) of all detected peaks. Carry-over of double-charged (tryptic peptide) species was very low (<0.26%). Although the number of peptides detected was highly consistent across replicates, only 58% could be matched to all runs ($n = 5$). However, 90% matched to ≥3/5 runs, indicating the importance of replicate runs. The method presented allows high sample throughput which is essential for clinical biomarker discovery. The results are highly encouraging, especially in light of the dynamic range improvements that are presently becoming available on quadrupole time-of-flight instruments.



## 1. Introduction

In mass spectrometry-based discovery proteomics, the task of protein identification is far more mature than performing reproducible and accurate quantification across a large number of replicate samples. In contemporary proteomics research, one aim is to achieve quantification of a high portion of all expressed proteins across different sample groups such as e.g. treated–control samples, healthy–diseased conditions or drug sensitive–resistant phenotypes. The traditional method for this so-called differential display experiment has been 2D gel-electrophoresis (2DE [1], for review see e.g. [2]). In extreme cases coverage using this method can reach up to about 10,000 protein entities resolved in a single gel [3]. Quantification is based on the use of various protein staining methods including fluorescent dyes that allow for standardization across samples [4] or multiplexed analysis [5]. However, the method is labor intensive and requires considerable skill both in sample processing and data analysis. The analysis time for even a limited sample set can easily be counted in weeks or even months.

Unlike in 2DE, protein quantification in mass spectrometry (MS) analyses is often done by measuring peptide levels and from that inferring protein abundance. Mass spectrometry-based quantification is commonly done by relating analyte ion intensity to that of a stable isotope labeled internal standard. Methods for absolute protein quantification have been established such as AQUA [6], QconCAT [7] as well as absolute SILAC (Stable-isotope labeling by amino acids in cell culture) [8] and absolute iTRAQ (isobaric tag for relative and absolute quantification) protocols [9] but for biomarker discovery purposes relative quantification is often employed. Relative quantification approaches are often performed as differential stable isotope labeling (coding) of each sample in a multiplexed analysis. Such labeling can be either chemical (for review see e.g. [10]), enzymatic [11] or metabolic [12]. The quantitative information is obtained either from the full scan mass spectrum or the fragment ion spectrum [13,14]. Drawbacks of stable isotope labeling

☆ This paper is part of the special issue "Quantitative Analysis of Biomarkers by LC–MS/MS", J. Cummings, R.D. Unwin and T. Veenstra (Guest Editors).

* Corresponding author at: Molecular Toxicology, Safety Assessment, AstraZeneca R&D, SE-151 85, Södertälje, Sweden. Tel.: +46 8 55252398; fax: +46 8 55258823.
*E-mail address:* Johan.Lengqvist@astrazeneca.com (J. Lengqvist).

1570-0232/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2009.02.052

approaches include considerable expense and questions of labeling efficiency. Further, depending on the method, the isotope coding will increase the complexity of the mass spectrum (effectively doubly or triply depending on the degree of multiplexing). The iTRAQ method is based on peptide fragmentation but precursor selection and peptide fragmentation on a chromatographic time-scale may be subject to under-sampling [15].

Quantification approaches that do not use stable isotope labeling are sometimes referred to as "label-free" methods. For reviews of label-free methods see e.g. [16,17]. Such methods include the spectral counts [18] and EmPAI approaches [19]. These methods do not derive quantitative information through integration of mass spectral or chromatographic peak areas. Rather they are based on comparing the number of occurrences of a peptide (or peptides) of a given protein, either in the set of fragment ion spectra or the number of peptides identified per protein. As with iTRAQ, for some methods quantitative information is linked to the peptide fragmentation-database search process and thus suffer similar limitations as outlined above.

Other label-free approaches attempt to directly correlate ion intensities across analyses and across samples and such a method is presented herein. As the final separation step prior to mass spectrometry analysis is often on-line coupled liquid chromatography (LC), the output data can be represented by a three-dimensional plot of $m/z$ (mass), retention time and ion intensity. Many software packages include functions for visualization of data in this format, e.g. MSight [20] and MZmine [21]. Data analysis (i.e. peak picking, alignment and matching, differential display) can then be performed using tools based on image analysis. This is in analogy with analysis of 2D gel data and indeed some of the MS-software packages are based on 2D gel analysis software (MSight, Decyder MS). In recent years, many software solutions have been developed to perform this type of differential display of LC–MS runs, each with their respective benefits. As these methods cover all peaks resolved during the LC–MS run, they give a more holistic view of the experiment and can thus also be used to assess the experimental data quality.

Lately significant improvements have occurred in LC instrumentation with the advent commercial systems capable of operating pressure in excess of 10,000 psi [22], small (<2 μm) particles and various column formats for rapid separations of biomolecules such as monolithic columns and superficially porous particles capable of very high peak capacities in peptide separations [23]. Such rapid and powerful separations have been widely adopted for the analysis of small molecules [24] but less for peptide applications [25]. In terms of mass spectrometry instrumentation, scan rates have increased to allow data collection on the timescale of rapid LC-separations. However, as pointed out in [26] using e.g. elevated temperature separations peak times have a risk of becoming too short for accurate sampling using mass spectrometry.

The retention time is the parameter which is (most likely) subject to the largest fluctuations, a time alignment function is included in the DeCyder MS 2.0 software used here. The aim of the time alignment function is to tie all LC–MS runs in an experiment to a common time scale. This is achieved in practice by aligning them to one of the LC–MS runs selected to be the time alignment reference. The time alignment function works by comparing individual spectra, merging or repeating them as necessary (similar to inserts and gaps in genomic sequence alignment). A dynamic programming algorithm ensures that the best possible solution is found. The method is working on individual spectra instead of blocks of spectra. Given that the data is sufficiently oversampled, this method is expected to work better with a combination of large shifts and short regions of high variability commonly seen in LC–MS data. The advantage of using time alignment is that the time tolerance for matching detected peptides between different LC–MS runs can be set tighter and in theory leading to fewer mismatches while at the same time keeping the number of incorrectly missed matches low. The effects of time alignment are investigated for UPLC– and HPLC–MS datasets (ultra performance and high performance LC, respectively).

The quality of the experimental data is a result of the combined performance of the LC system, the LC–MS coupling and the mass spectrometry instrumentation. For an electrospray interface, sensitivity is dependent on liquid flow rate as shown by Wilm and Mann [27]. However, although automated nanospray devices are present [28,29] and show good spray stability, LC-systems and electrospray interfaces operating at higher flow-rates (≥∼50 μL/min) are generally more robust and easier to maintain (for a review of common problems associated with maintaining nanoLC-system see [30]). Further, rapid separations as discussed below are mainly implemented using higher flow-rate systems. Although potentially sacrificing ultimate sensitivity and requiring more sample material, the trade-off may be beneficial in terms of robustness and increased results validity based on improved statistics (higher sample throughput). The present study aims to investigate the applicability of ultra performance liquid chromatography time-of-flight mass spectrometry (UPLC-TOF MS, ACQUITY with LCT Premier XE, Waters Corp.) for discovery proteomics studies. Thus the terms "UPLC" and "UPLC-TOF" in the following refers to that specific LC–MS system.

In the current study technical variability in terms of retention time, mass accuracy and ion intensity is investigated for UPLC-TOF analysis of highly complex cell lysate samples. The final outcome of data collection, normalization and analysis is the quantitative changes observed between treated and control samples in the differential display experiment. The degree of change that can be accurately measured is addressed below. To our knowledge, few applications of the UPLC-TOF system to biomarker discovery research using complex whole cell digests have been reported.

## 2. Experimental

### 2.1. Chemicals and reagents

The bovine serum albumin digest analyzed was obtained from Waters (MassPrep protein digest standards, Waters Corp, Milford, MA, USA). All solvents used for chromatographic separation and mass spectrometry were of HPLC grade or better.

### 2.2. Sample preparation

Human HCT116 colon cancer cells were subjected to 8 Grey ionizing radiation (IR sample) while control cells were mock-exposed. Five hours post-IR, $10 \times 10^6$ cells were collected using scraping without trypsination. Cells were lysed and one cytosolic and one DNA-binding protein fraction prepared using the Qproteome Nuclear Protein kit (QIAGEN Inc, Valencia, CA, USA). Protein concentrations were determined (Bradford assay) using bovine serum albumin (BSA) as the standard and 200 μg of protein was precipitated using acetone. To facilitate digestion, 0.2% of an acid labile detergent (Rapigest, Waters Corp, Milford, MA, USA) was added and reduction (dithiothreitol) and alkylation (iodoacetamide) was performed as per the manufacturer's instructions before incubation with trypsin (Promega Sequencing grade, Promega Corp, Madison, WI, USA) at 37 °C overnight.

Non-digested material was removed using 10 kDa molecular weight cut-off spin filters (NanoSep 10 K, Omega membrane, Pall Corp, East Hills, NY, USA) and samples were frozen at −20 °C until analysis. Prior to analysis, the sample was diluted 20 times in buffer A (2% acetonitrile, 0.2% formic acid) to a concentration of

~0.125 μg/μL and 10 μL was loaded on column (corresponding to a maximal theoretical loading of ~1.25 μg).

For the HPLC dataset (sample denoted as CapLC) three pooled fractions from an isoelectric focusing (IEF) separation (performed as described in [31]) were analyzed in replicates ($n = 5$) using a nanoLC-quadrupole TOF instrument (CapLC XE, see below).

It should be observed that the whole cell digest analyzed by UPLC–MS was significantly more complex than the DNA-binding protein sample in the HPLC-analysis.

### 2.3. LC–MS conditions

The instrument parameters used are described briefly below.

The CapLC sample was analyzed using a quadrupole time-of-flight (QTOF) Ultima API instrument with the CapLC XE system both from Waters. This system is referred to hereafter interchangeably as "CapLC". Reversed phase separation parameters were initially 5% B (3 min) followed by a linear gradient from 5 to 28% B over 37 min and subsequently a ramp from 28 to 80% B over 5 min. After a hold at 80% B for 7 min, and a ramping step down to 5% B over 1 min, the system was re-equilibrated at 5% B for 10 min. The solvents were (A) 2% acetonitrile (HPLC grade S, Rathburn, Walkerburn, U.K.) in Milli-Q grade water with 0.1% formic acid and (B) 95% acetonitrile in water with 0.1% formic acid. The column used was an Atlantis dC18 NanoEase column (75 μm × 150 mm, 3 μm particles, 300 Å, Waters Corp). The nanoelectrospray was achieved using Picotip spraying tips (New Objective Inc, Woburn, MA, USA) and the flow rate delivered to the electrospray interface was ~300 nL/min. The mass range was scanned from 400 to 1700 $m/z$.

For the UPLC analyses, an LCT-Premier instrument with the ACQUITY UPLC system (both from Waters) were used. The gradient was much shorter than for the CapLC analyses. Specifically, initial conditions were, 0% B followed by a linear gradient to 30% B over 15 min, a hold at 30% B for 1 min followed by a ramp to 95% B over 30 s. After a hold at 95% B for 1.5 min the system was ramped down to 0% B over 6 s. The A and B solvents were 2% acetonitrile in Milli-Q grade water with 0.2% formic acid and 97.5% acetonitrile in water with 0.2% formic acid, respectively. The flow rate was 200 μL/min. The column was a 1.0 mm × 100 mm bridged ethane hybrid (BEH) C18 column with 1.7 μm particles. The column was maintained at 60 °C for the separation.

For the analyses, the electrospray interface was used fitted with the standard stainless steel capillary. Interface parameters were capillary voltage 3800 V, source temperature 130 °C and the desolvation gas temperature was 320 °C. The mass range scanned was 400–1800 $m/z$ for the BSA standard and $m/z$ 300–1500 for the cellular protein digest samples.

### 2.4. Import of raw data to the DeCyder MS software

The original raw data files from the two Waters LC–MS systems were converted to ASCII-files using the Databridge function of the MassLynx software package. These files were then imported (i.e. converted to the intensity map file format used by the DeCyder MS software suite version 2.0, GE Healthcare, Uppsala, Sweden) using the DeCyder MS import tool. Raw data was imported without any cropping in either the time or $m/z$ dimension using the default parameter settings.

### 2.5. Peak detection in the DeCyder MS software

In the following, parameter settings are given in the format they are specified by the user in the DeCyder MS software.

Peak detection parameters were optimised to allow sensitive detection with a low degree of erroneous peak assignment as determined empirically.

The peak detection parameters were set to allow charge states as indicated in Figures and legends (unless otherwise indicated the allowed charge states are 1+ to 10+). The only peak detection parameter that was changed between the CapLC and the UPLC systems was the peak width (0.6 min for CapLC and 0.08 min for UPLC). In all else the detection parameters were as stated below.

The "TOF resolution" parameter was set to 5500 as determined empirically. The "advanced detection" parameters were optimised to; "background model type": uniform, "signal-to-background detection threshold": 5.0 with the background subtracted quantitation function enabled. Further, the "charge assignment from two peaks" parameter set to "limited". The "Charge state matching" parameters were set to; "LC peak shape tolerance" 20%, $m/z$ shift tolerance 0.05 Da and "$m/z$ shape tolerance" to 5%.

After detection, filtering was applied such that the following detected peptides were removed; peptides below S/N 2.0 (also peptides of unspecified charge below S/N 2.0) and overlapping peptides. Further, the "Remove peptides with low quality LC peaks" parameter was enabled and the "Spectrum vs. model tolerance" was set to 15%.

### 2.6. Time alignment

Alignment of intensity maps was always performed using the map holding the highest number of detected peptides as the reference. Time alignment was done using the default parameter settings. Briefly, the "Max stretch/compress" parameter was set at 2 and "Max leader and trailer" at 10% with a "stretch/compress penalty" of 0.10.

### 2.7. Peak matching

Matching of detected peptides across intensity maps was done varying the time and $m/z$ constraints as indicated in Figures and legends using aligned or non-aligned retention times as indicated. Otherwise the default settings were used in that the "Correct inconsistent charge states" was enabled and charge states were included if present in at least two intensity maps. Further, unconfirmed peptides were included (this is simply to allow the inclusion of detected peptides that have not been manually confirmed as in this study peak detections were not manually curated).

### 2.8. Data analysis

Each set of matched peptides (sample groups) were then analyzed using Student's $t$-test with or without applying a normalization based on the measured intensity distribution (total-ion-current, TIC-normalization) as indicated in Figures and legends. The match results were then exported for as text-files for subsequent analyses. Graphs were created from images in the DeCyder MS 2.0 software.

## 3. Results

### 3.1. Analysis of bovine serum albumin digest

Although the ability of the DeCyder MS software to accurately reflect quantitative changes in LC–MS data has been demonstrated elsewhere [32], software performance was tested using a low complexity sample. Therefore, a digest of bovine serum albumin was analyzed with a 10 min gradient on the UPLC–MS system to assess technical reproducibility. The dataset generated consisted of a dilution series over the concentration range 0.05, 0.1 and 0.2 μg (each point performed in triplicate).

The software can be set to consider only certain charge states in the peak detection step. This allows the inclusion or exclusion of

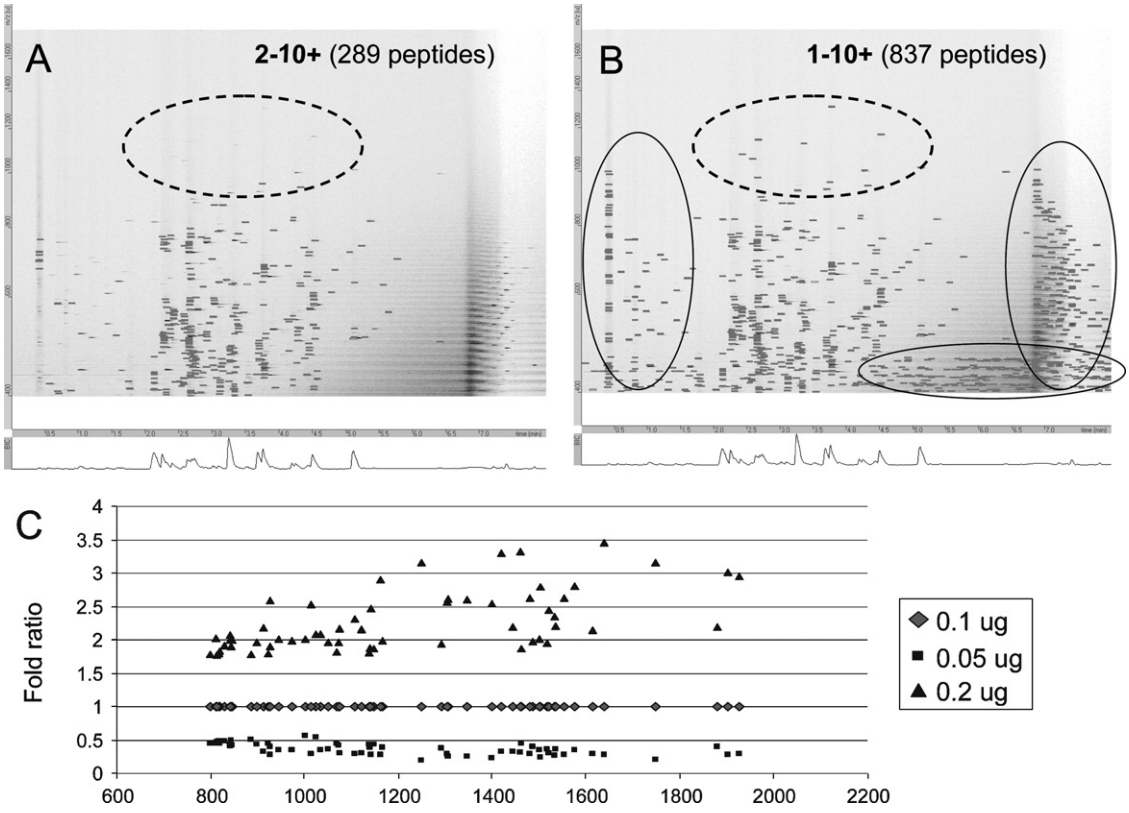


**Fig. 1.** UPLC–MS analysis of a BSA digest. In (A) is shown the peak intensity map of one injection of 0.2 μg of BSA digest after detection of 2–10+ charged peaks. The same intensity map after detection of 1–10+ species is shown in (B). In (C) is shown the ion intensities for all peaks matched to all nine runs. Values are shown normalized to the middle concentration (0.1 μg BSA injected, values for the 0.05 μg, 0.1 μg and 0.2 μg are indicated by black squares, grey diamonds and black triangles, respectively).

certain charge states (e.g. single chargers). Initially the $[M+2H^+]^{2+}$ to $[M+10H^+]^{10+}$ (hereafter "2–10+") charge states were allowed in the peak detection. Fig. 1A shows the intensity map from a UPLC–MS run of the BSA digest (0.2 μg injected) in which 289 features (hereafter referred to interchangeably as "peptides") were detected. If 1–10+ charge states were instead considered, 837 species were detected (Fig. 1B), a near threefold increase. However, the majority of additional peaks are found either in the injection front, in the final wash phase or as low mass streaking phenomena (circled using solid lines in Fig. 1B). Thus these peaks most likely constitute a mix of singly charged low mass impurities. As tryptic peptides are often doubly charged in electrospray experiments, excluding singly charged peaks increases the likelihood that only peptides are considered in the analysis. Thus, based on the result for the BSA digest analysis (Fig. 1A and B) singly charged peaks are rightly excluded from analysis. However, excluding 1+ charge states was not fully appropriate for more complex peptide mixtures as information may be lost for peptides occurring as single chargers as discussed below.

For the BSA digests, on average 84.7, 198 and 281.3 peptides were detected after injection of 0.05, 0.1 and 0.2 μg, respectively (2–10+ detection, std dev 4.2%, Table 1). This indicates that the UPLC-TOF MS instrument has very high precision in terms of the total number of peptides detected for a low complexity peptide mixture. As expected, the number of peptides observed increased with the amount of BSA digest injected.

Further, it was evident that the bulk of detected peptides could be matched across runs with high reproducibility. Out of all detected peaks, 63–75% could be matched to 3/3 replicates and 90–94% to ≥2/3 runs (Table 1).

Further, the BSA dilution series experiment was analyzed for quantitative reproducibility. The coefficient of variation (CV) of ion intensity was calculated for all peaks matched to ≥2/3 runs at each concentration. A very high portion of the matched peptides were measured with high quantitative accuracy. For the 0.05 μg injections, 68.4% of all peptides matched (to ≥2/3 runs) showed ion intensity CVs <5%. For the 0.1 and 0.2 μg injections, the numbers were 69.4% and 82.5%, respectively (Table 1). Indeed 93.4, 93.0 and 92.2% of the peptides had CVs <10% indicating a very high degree of quantitative precision. The results are shown after normalization based on the total ion current (TIC) across replicates at each concentration level. The change in the number of peptides with ion intensity CVs <5% was small (at most 3.6%) between normalized and non-normalized BSA data (data not shown). Notably, the normalization step was not in all cases reducing the CVs measured. In terms of dynamic range, ion intensities recorded ranged from ~10 to 4500, effectively covering about 2.5 orders of magnitude (data not shown).

**Table 1**
Dilution series of BSA digest analyzed using a 10 min gradient on a UPLC–MS instrument. Peak matching parameters were 0.5 min and 0.05 Da windows using time-aligned data. Each concentration was run in triplicate.

| Amount (μg) | Peptides detected (average) | Std. dev. | Matched to 3/3 runs | Matched to ≥2/3 runs | CV <5% (intensity) | CV <10% (intensity) |
|---|---|---|---|---|---|---|
| 0.05 | 84.7 | 2.1 (2.5%) | 64 (76%) | 76 (90%) | 68.4% | 93.4% |
| 0.1 | 198.0 | 8.3 (4.2%) | 126 (63%) | 186 (94%) | 69.4% | 93.0% |
| 0.2 | 281.3 | 6.1 (2.2%) | 206 (73%) | 262 (93%) | 74.0% | 92.2% |

For differential display, features have to be matched (and normalized) not just across technical replicates but also across different experimental conditions (e.g. sample/control runs). The BSA data set includes three replicates at three different concentrations giving a total of nine runs to be matched and normalized. After time alignment and matching as above, 60 peptides could be matched to all 9 runs. This is close to the number of peptides (64) matched to all runs at the lowest concentration (0.05 μg). An additional 61 peptides could be matched to ≥6/9 runs. The sum (60 + 61 = 121) is very close to the number matched to all runs for the 0.1 μg injections (126 peptides, Table 1). This shows that alignment and matching is satisfactory also when performed across different concentration levels in the face of varying number of total detected peptides.

For the 0.05 μg injections, 76 peptide features matched to ≥2/3 runs. Among these, 35 match the masses of theoretical tryptic BSA peptides (1 missed cleavage site allowed) corresponding to a sequence coverage of 57%. For the 0.1 μg and 0.2 μg injections, the numbers of matched theoretical BSA peptides increase to 47 and 54, respectively, bringing sequence coverage to 75.4 and finally 86.2%. It should be remembered that peptide assignments are based on correspondence in mass only as no fragmentation data was available from the UPLC-TOF system. However, except for two cases, all peptides assigned at lower concentration levels were also found at higher concentration. This indicates a high degree of correct peptide assignments. One additional peak ($m/z$ 1024.60) is tentatively assigned as the $Na^{2+}$-adduct of a BSA peptide observed at $m/z$ 1002.61 (LVVSTQTALA, theoretical $m/z$ 1002.58) due to the observed +22 Da mass difference and identical retention times of the two peaks (4.29 min, data not shown).

In Fig. 1C is shown the average ion intensities for the 60 peptides detected and matched across all 3 different concentrations (0.05 μg peaks are indicated by black squares, 0.1 μg by grey diamonds and 0.2 μg by black triangles (Fig. 1C)). In the graph, ion intensity values are plotted against peptide mass. Intensities are shown as average values normalized to the 0.1 μg intensities (these values are thus all equal to 1, Fig. 1C). It appears that peaks in the mass range 800–1200 Da approximately double in intensity with each increase in amount loaded (Fig. 1C). However, at higher mass, responses tend to increase disproportionately at the 0.2 μg level, the reason for which is unclear.

In conclusion, UPLC–MS-based analyses of low complexity peptide mixtures show very high qualitative and quantitative reproducibility. In the concentration range tested, a linear correlation between ion intensity and peptide concentration was observed.

### 3.2. Peak detection for total cellular tryptic digests

To assess performance of the UPLC–MS system for a highly complex sample, tryptic digests of total cytoplasmic protein from human HCT116 colon cancer cells were analyzed. The total gradient time used was just 17 min for this complex sample. As the aim was to investigate performance in a differential display type of experiment, one irradiated sample ("IR" sample) and one control sample ("Ctrl"/"control") were analyzed with five repeat injections per sample.

The number of peptides detected in each run is given in Table 2A for different detection settings as indicated. For 1–10+ detection, on average 8472 (std dev: 33 or 0.4%) and 7972 peptides (std dev: 27, 0.4%) were observed in IR and control replicates, respectively (Table 2A). Detection of 2–10+ charged species gave on average 7759 and 7245 peptides (Table 2A). The low decrease in the total number of features detected (∼10%) indicates that the majority of peptides are detected as 2+ charge states or higher. The low standard deviations for the number of peptides detected (<1.5%) indicate high reproducibility of sample loading, separation and MS detection.

Blank runs were included in order to assess the degree of carry-over (Table 2A, the position in the table indicate the run order of the blanks and samples). In the blank runs 12, 19 and 13 peptides were observed (2–10+ detection), indicating a minimal degree of carry-over (<0.26%). However, 842, 597 and 649 singly charged species were detected (detection of 1+ charge states only, average 696, std

**Table 2**
The number of peptides detected using different detection settings. In (A) results are given for the UPLC samples and in (B) for the CapLC instrument.

| | 1–10+ | | | 2–10+ | | | 1+ | | |
|---|---|---|---|---|---|---|---|---|---|
| Sample | Detected | Mean | Std. dev. | Detected | Mean | Std. dev. | Detected | Mean | Std. dev. |
| (A) UPLC samples | | | | | | | | | |
| Blank_01 | 846 | | | 12 | | | 842 | | |
| IR | | 8472 | 33 | | 7759 | 38 | | 3062 | 46 |
| 1_a | 8524 | (100%) | (0.4%) | 7807 | (91.6%) | (0.5%) | 3129 | (36.1%) | (1.5%) |
| 1_b | 8483 | | | 7804 | | | 3080 | | |
| 1_c | 8446 | | | 7727 | | | 3082 | | |
| 1_d | 8442 | | | 7723 | | | 3011 | | |
| 1_e | 8465 | | | 7732 | | | 3009 | | |
| Blank_02 | 608 | | | 19 | | | 597 | | |
| Ctrl | | 7972 | 29 | | 7245 | 40 | | 2781 | 24 |
| 2_a | 7971 | (100%) | (0.4%) | 7298 | (90.9%) | (0.6%) | 2793 | (34.9%) | (0.9%) |
| 2_b | 8021 | | | 7245 | | | 2771 | | |
| 2_c | 7996 | | | 7268 | | | 2820 | | |
| 2_d | 7944 | | | 7236 | | | 2768 | | |
| 2_e | 7926 | | | 7176 | | | 2751 | | |
| Blank_03 | 654 | | | 13 | | | 649 | | |
| (B) CapLC | | 4350 | 152 | | 3560 | 155 | | 2391 | 83 |
| 1_a | 4604 | (100%) | (3.5%) | 3817 | (81.8%) | (4.4%) | 2498 | (55.0%) | (3.5%) |
| 1_b | 4379 | | | 3609 | | | 2393 | | |
| 1_c | 4378 | | | 3576 | | | 2467 | | |
| 1_d | 4182 | | | 3419 | | | 2293 | | |
| 1_e | 4205 | | | 3381 | | | 2303 | | |
| Blank_a | 1340 | | | 167 | | | 1067 | | |
| Blank_b | 1271 | | | 124 | | | 1207 | | |
| Blank_c | 1161 | | | 134 | | | 1165 | | |

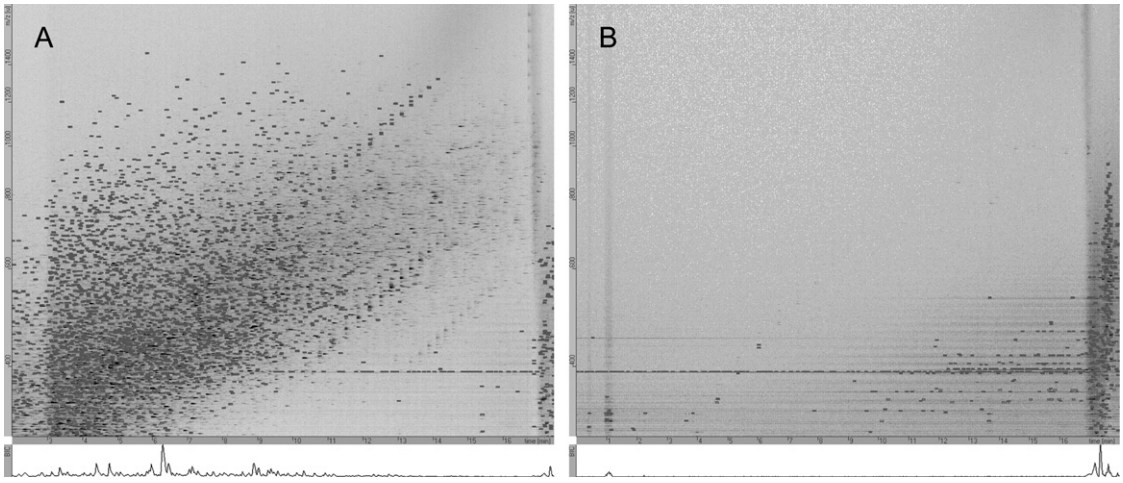


**Fig. 2.** One-third of the peptide population is present also as single chargers. LC–MS peak intensity maps of one HCT control sample run (A) and one blank run (B). The boxes indicate a detected peak (1+ only detection).

dev: 105). This finding would support the notion of excluding singly charged species from the analysis.

As shown the increase in the number of detected peptides is only ~10% for 1–10+ detection compared to 2–10+ detection. However, the situation is more complex than at first apparent. In the IR and control sample runs, on average 3062 and 2781 singly charged species are detected (36.1 and 34.9% of the total number detected respectively, 1+ only detection). Thus a substantial portion of the peptide population exists also as single charges. A representative detection workspace for 1+ only detection is shown in Fig. 2A for the control sample. The detected single chargers show a chromatographic behaviour typical of peptides rather than background peaks, i.e. an even distribution over retention time space and an increase in mass with increased retention time (Fig. 2A, compare to Fig. 2B showing a blank run). So, a substantial amount of potentially useful quantitative information will be missed if singly charged species are excluded. Based on these results, 1–10+ detection was used in subsequent analyses.

### 3.3. Retention time alignment and peak matching

After detection, peaks in each run have to be matched to the corresponding peaks (if present) in other runs. Matching is done using retention time and mass windows either with or without applying time alignment algorithms to correct for retention time shifts. Of the two dimensions in an LC–MS experiment retention time often shows the largest variation while the mass scale is relatively stable. For the TOF-based electrospray instruments used in the current study, mass deviations were <0.02 Da (c.f. Fig. 3A and B, peptides matched to 5/5 replicates are shown).

Relatively minor retention time shifts were observed for the UPLC system (Fig. 3C). To challenge the time alignment algorithm

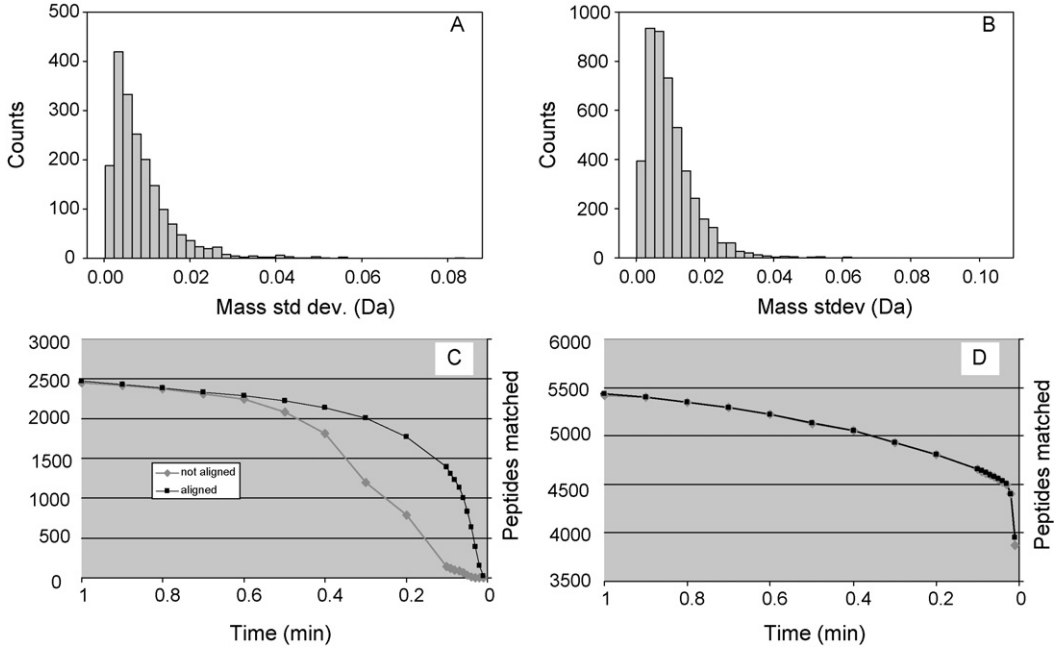


**Fig. 3.** Efficiency of time-alignment for two different LC–MS instruments. In (A) and (B) are shown histograms of the mass accuracy for the HPLC and UPLC–MS instruments, respectively. Further (in C and D) are shown the number of peaks matched as a function of the "time window" allowed in matching without (filled light grey diamonds) and with time alignment (filled black squares) for the HPLC–MS (C) or UPLC–MS instrument (D). It should be noted that the peptide digest analyzed in (D) is far more complex than in (C) and that the total separation time in (D) is only 17 min compared to 65 min in (C).

**Table 3**
Number of peptides matched in the presence or absence of time-alignment for each sample and LC–MS system. Matching conditions were as indicated.

| Sample | Detected (mean) | Matched (−align) | Matched (+align) | Matched (−align) | Matched (+align) |
|---|---|---|---|---|---|
| # of replicates | | 5/5 | 5/5 | 3/5 | 3/5 |
| Matching conditions | | 0.075 min, 0.05 Da | 0.075 min, 0.05 Da | 0.075 min, 0.05 Da | 0.075 min, 0.05 Da |
| IR ($n$ = 5) | 8472 | 4924 (58%) | 4933 (58%) | 7599 (90%) | 7600 (90%) |
| Ctrl ($n$ = 5) | 7972 | 4601 (58%) | 4604 (58%) | 7148 (90%) | 7154 (90%) |
| # of replicates | | 10/10 | 10/10 | ≥3/5 + ≥3/5 | ≥3/5 + ≥3/5 |
| IR + Ctrl ($n$ = 10) | 8222 | 3776 (47%) | 3776 (47%) | 5848 (73%) | 5848 (73%) |
| Matching conditions | | 0.25 min, 0.05 Da | 0.25 min, 0.05 Da | 0.25 min, 0.05 Da | 0.25 min, 0.05 Da |
| # of replicates | | 5/5 | 5/5 | ≥3/5 | ≥3/5 |
| CapLC ($n$ = 5) | 4350 | 1012 (23%) | 1903 (44%) | 3725 (86%) | 3842 (88%) |

of the DeCyder MS 2.0 software, data generated on an HPLC-type system was included for analysis (CapLC-QTOF system, technical replicates analysis of one sample, $n$ = 5). The CapLC sample was prepared by pooling three neighbouring fractions from a peptide IEF separation of a complex tryptic digest from a DNA-binding protein fraction. The number of peptides detected for this sample with the CapLC system is given in Table 2B similar to Table 2A for the UPLC analyses. As shown, standard deviations for the total number of peptides detected are higher for the CapLC system.

The number of peptides matched to all replicates as a function of the width of the time window used is plotted in Fig. 3C and D (CapLC and UPLC samples, respectively). A mass window of 0.05 Da was chosen for matching as the bulk of the peptide populations had standard deviations <0.02 Da (Fig. 3A and B). As the time window was decreased the number of matched peptides decreased (Fig. 3C). The left-hand curve in Fig. 3C depicts the retention time accuracy of the CapLC instrument (the numbers of matched peptides in the absence of time alignment is indicated by filled light grey diamonds). About half of the peptides were still matched at a window of 0.3 min, giving a crude estimate of retention time accuracy. This indicated a need for some form of alignment of the time axis across multiple runs. In peptide separations, retention time shifts may be complex and non-linear [33]. Non-linear shifts were present in the CapLC dataset and were readily observed when comparing replicate LC–MS peak intensity maps (not shown).

To compensate for retention time shifts, a time alignment function based on time-warping has been included in DeCyder MS 2.0 software. In Fig. 3C and D are shown the number of peaks matched before and after alignment (filled black squares for CapLC and UPLC systems, Fig. 3C and D). The right shift of the curve (Fig. 3C) shows that a higher number of peptides are matched for smaller time windows indicating successful alignment. A minor improvement is seen for the UPLC data, but is hard to be distinguished from the curve for the non-aligned situation (Fig. 3D). Thus the time-alignment function reduces retention time variation which increases the number of peptides matched across replicates, which is especially apparent in the HPLC dataset (c.f. Fig. 3C and D).

Based on the results shown in Fig. 3C and D, time windows for matching were selected to be as small as possible while retaining a high number of matched peptides. Time windows of 0.25 min and 0.075 min were chosen for CapLC and UPLC, respectively and used in subsequent experiments. In Table 3 are shown the numbers of peptides matched across runs for different matching conditions and samples, with and without time alignment as indicated.

The CapLC system showed lower numbers of matched peptides and here time alignment can effectively increase the number of matched peptides. Only 23% of all detected peptides were matched without alignment (CapLC, Table 3). After time alignment, this effectively doubled to 44% matched. For the less strict criteria (≥3/5), the numbers were 85.6 and 88.3% before and after alignment.

For the UPLC replicates, about 58% of all detected peptides could be matched to all replicates and about 90% to ≥3/5 runs (Table 3). For the UPLC dataset, time alignment provided only minor improvements (Fig. 3C and Table 3).

In the UPLC dataset matching has to be done of all IR and control replicates ($n$ = 5 + 5) for the differential display experiment. In this setting, only 47% of all peptides could be matched to 10/10 runs (3776 out of 7972 peptides, i.e. the average number detected in the ctrl replicate set). However, for the less strict matching (tolerances ≥3/5 runs in IR and ≥3/5 runs in the control) 5848 peptides (73%) were detected and matched (Table 3).

In conclusion, the time alignment function allows matching an increased number of peptides both across replicates and across samples for data sets where retention time fluctuations are present. For UPLC–MS data however, these improvements are only minor, indicating a very high retention time precision.

### 3.4. Quantitative variation

In addition to simply comparing the number of peptides matched across runs, the quality of the quantitative result is highly important. When coefficient of variation (CV, relative standard deviation) was calculated for the CapLC sample, the bulk population of peptides matched to all runs show CVs <40% (Fig. 4A, histogram plots of intensity CVs in the absence of TIC normalization "−"). For the UPLC–MS system the variation is considerably lower with bulk CVs <10% (IR sample, Fig. 4B). This indicates a high degree of reproducibility for the UPLC–MS system in the absence of intensity normalization.

A common way of normalizing intensities between runs is to use the total measured ion intensity (total ion current, TIC). If TIC-based normalization is applied, a significant improvement is observed for the CapLC data with bulk populations CVs <20% (Fig. 4C). An improvement is observed also for the UPLC system, not as a shift of the whole population, but rather as an increase in the number of peptides having the lowest CVs (0–1.67%) from ∼950 to ∼1300 (c.f. the left-hand most columns in Fig. 4B and D). As expected, TIC-based normalization can decrease quantitative variation between replicates LC–MS analyses.

### 3.5. Differential display

The goal of many proteomics approaches is to measure and compare peptide/protein levels across different conditions and samples in a global manner. For differential display experiments, peptide peak intensities have to be accurately measured across LC–MS runs and compared across samples and cohorts.

One often used measure of statistical significance between populations is calculating the $p$-value (performed using Student's $t$-test in the DeCyder MS 2.0 software). To test the relevance of the $p$-value for a given dataset the $p$-value distribution can be plotted (shown in the histogram plots in Fig. 5). The UPLC data set has five technical replicates per sample. This allows "mock comparison" within replicate groups (i.e. comparing 2 + 3 IR and 2 + 3 control replicates) in addition to comparisons across sample and control groups. His-

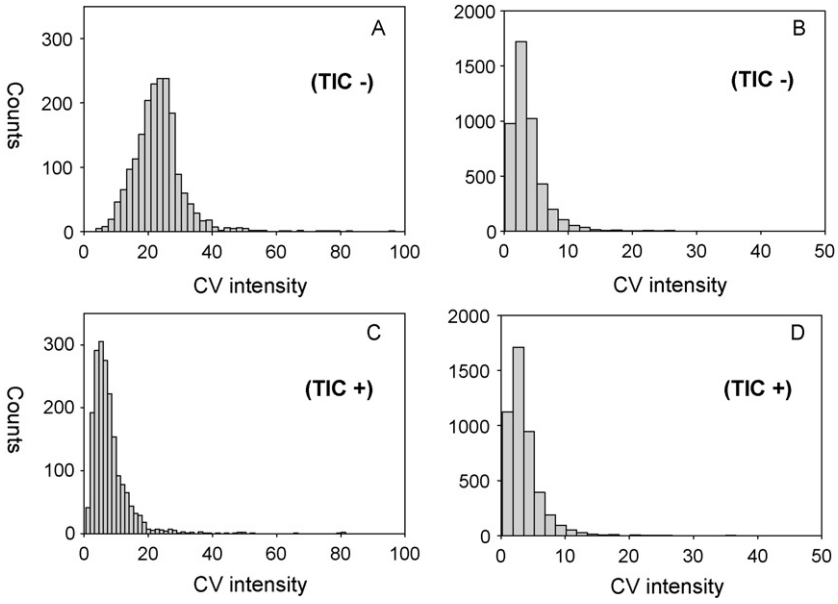


**Fig. 4.** Quantitative reproducibility in HPLC and UPLC-generated data. Histogram plots of the ion intensity CVs ($n = 5$) for the LC–MS analysis of the CapLC sample (A and C) and one UPLC sample (B and D). Plots are without (A and B) with TIC-normalization (C and D) as indicated by "−" and "+", respectively.

togram plots of the *p*-value distributions are shown in Fig. 5 in the presence or absence of TIC-based normalization. In each plot, the left-hand most columns correspond to *p*-values <0.033, the next to *p*-values from 0.0331 to 0.066 and so on. The plots show that for the comparison within replicate groups, ∼110–290 peptides (1.4–3.4%) have *p*-values below 0.033 (Fig. 5A and B).

The slightly skewed *p*-value distribution in Fig. 5B compared to 5A (almost 300 peptides have *p*-values <0.033) indicates that the IR replicates are less homogenous than the control replicates. As can be observed, the differences decrease when TIC-normalization is applied (Fig. 5B). In Fig. 5C is shown the comparison of the full dataset and is clear that statistically significant differences are observed for a high portion of the peptide population. Based on non-normalized ion intensities, 3697 peptides have *p*-values <0.05. When TIC-normalization is applied, the number of peptides with *p*-values <0.05 drop to 3010 (Fig. 5C). Thus about 50% of all detected peptides are flagged as significantly changed. It should be emphasized that this is only an analysis of a single treated and control sample. In such a setting, distinguishing sample load variability from intervention induced (biological) changes is not feasible.

### 3.6. The degree of change detectable using UPLC-TOF MS for total cellular digests

As shown above reproducibility in terms of peptide detection and quantitation is high when using UPLC-TOF-MS system for complex proteomic samples.

We then aimed to investigate what magnitude of change could be accurately measured with high statistical significance in a differential display experiment. To determine the experimental dynamic range of the system, ion intensities for all peptides matched to 8/10 runs in the differential display experiment were plotted on a logarithmic scale (Fig. 5D). The bulk population spanned a range of ∼1.9 orders of magnitude from ∼25 to 1000 (indicated by the arrowheads). The quantitative difference between the IR and control sample was plotted for each peptide as a function of *p*-value (Fig. 5E, the *x*-axis is on a logarithmic scale). On the *y*-axis, the quantitative difference for each peptide is plotted as a percentage of the total experimental dynamic range (i.e. the difference between the lowest and the highest peak intensity recorded, ∼1.5 orders of magnitude). For clarity, different *p*-value intervals are indicated by grey triangles ($p > 0.5$), black squares ($p < 0.001$) and intermediate *p*-values are indicated by grey circles (Fig. 5D). It is immediately evident that the distribution is not centered around zero. This indicates a systematic error, most likely corresponding to unequal protein concentrations between the two samples. The dotted lines (+4.16 and −2.5% change) have been added to indicate possible cut-off levels. Given the dynamic range of the experiment, a 5%-change corresponds to a twofold increase as indicated (Fig. 5E).

### 3.7. The presence/absence situation

Potentially the most interesting biomarker candidates are peptides (proteins) that are regulated in an on/off fashion. To investigate the presence/absence situation, the UPLC dataset was analyzed for features detected in ≥3/5 runs among either IR or control replicates but which did not occur in the other set. This gave a total of 675 and 452 peptides detected uniquely in IR and control sample, respectively. Peptides regulated in an on/off fashion were observed to be distributed throughout the *m*/*z* and time space (data not shown).

## 4. Discussion

Discovery proteomics should in theory be completely non-biased and cover the entire proteome. However, reproducible repeated sampling of a large number of proteins across many samples is still challenging, effectively limiting study cohorts. Contributing factors include the extensive fractionation required to obtain depth of analysis and a degree of randomness in sampling using e.g. data-dependent LC–MS/MS data-acquisition

In this study, we aimed to investigate the qualitative performance of UPLC-TOF for analysis of complex protein digests using a 1.0 mm inner diameter column. The resolving power proved to be very high with about 8000 peptides resolved using a gradient cycle time of only 17 min. Further, replicate analyses of total cellular protein digests demonstrated low standard deviations, low ion intensity CVs and low carryover. Together, this underscores the stable performance and power of the analysis, i.e. the combined performance of the LC-system and the electrospray ionization

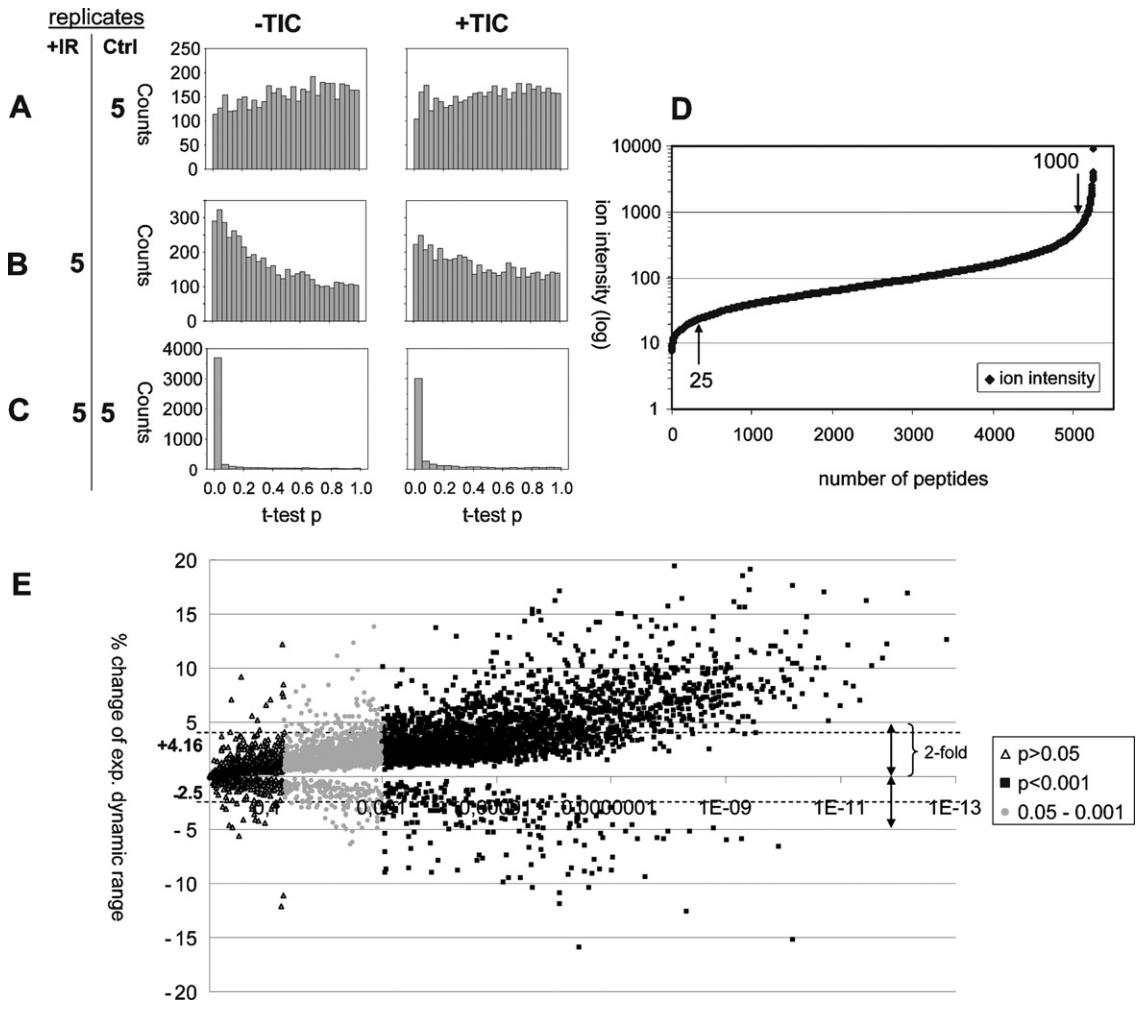


**Fig. 5.** Measuring statistical significance in a differential display experiment. Histograms of the *p*-value distribution are shown for the comparison within replicates (A) and (B) as well as between treated and control sample (C). The bins in the histograms correspond to 0.033 units (A and B) and 0.05 units (C). Plots are in the presence and absence of TIC-normalization as indicated. (D) shows the dynamic range covered in the differential display experiment (~1.9 orders of magnitude) with the ion intensities of 25 and 1000 for clarity. In (D) are plotted the observed difference for each peptide detected in ≥8/10 runs as a function of the associated *p*-value. Differences are shown as percentages of the experimental dynamic range on the *x*-axis. *P*-values >0.05 are indicated by open non-filled grey triangles and *p*-values <0.001 are shown by black squares. Intermediate values are shown as light grey circles.

interface. Raw data quality was sufficient that neither normalization (using TIC) nor retention time alignment provided any major benefits for peak matching or quantitative accuracy.

The current study also illustrates the necessity of replicate analysis for highest coverage. As shown, only about 60% of all detected features in a replicate analysis set ($n = 5$) could be matched to all runs. However, 90% were able to be matched to at least 3/5 replicates. Including technical replicates further increases the demand on throughput. The time-frame of analysis shown here (17 min) is far more suited to high throughput (in the face of large sample cohorts or extensive fractionation efforts) than the >60 min separations more commonly seen in proteomic analyses. It should be remembered that in the analysis of less complex (fractionated) samples, gradient times may possibly be shortened significantly.

Recently Zubarev and Mann highlighted the benefit of reporting the mass accuracy obtained in the LC–MS/MS experiment (i.e. in the protein identification result list) rather than giving general figures of merit e.g. for a certain instrument type [34]. There is no reason why a similar approach should not be applicable to the reporting of quantitative proteomic information as well. This could for instance include specifying the dynamic range covered in a quantitative proteomics experiment or the observed experimental variability in quantitative data. In the present study, quantitative variability of the raw data is described and accuracy is related to the experimental dynamic range.

Another aspect which is essential for the final result is sample pre-treatment (for in-depth discussion see e.g. [35]). In the present study, total cellular protein was digested with trypsin in the presence off acid-labile surfactant for denaturing protein structure to aid digestion. Further, the digest was cleared using a molecular-weight cut-off filter (10 kDa) to remove non-digested protein as well as the proteolytic enzyme before sample injection. These precautions were included in the workflow as, for larger studies, care must be taken to maintain column performance over multiple injections. As indicated by the results, five replicate injections (back-to-back, no blank injections interspaced) were possible with no drastic effects on separation performance.

One aspect of the proteomics workflow which has undergone drastic improvement in recent years is peptide fractionation. Emerging separation methods using isoelectric focusing offer several major benefits which should be considered in the context of the present study. First of all, they offer high resolution and are capable of highly reproducible separations. Second, they can tolerate high sample loads which will help to overcome any sensitivity issues of the final analysis. In the present study, ~1 μg of protein digest was loaded on column. To obtain such quantities in individ-

ual IEF fractions, requires loading in the range of 400 μg to 1 mg using the method as presented in [36]. IEF separation is also compatible with both label-free (as demonstrated herein for the CapLC sample analysis) and stable isotope methods of quantitation [31]. Finally, great depth of proteome analysis can be achieved especially together with protein-level fractionation [36,37].

However, one potential problem with using pre-fractionation is the matching of fractions across multiple samples. For the UPLC-TOF and any other LC-based analysis, it is highly beneficial for optimal retention time accuracy if all samples injected are overall similar. This is something that speaks for analyzing non-fractionated samples. However, gas-phase fractionation methods such as the ion-mobility separation method discussed below can be introduced to the mass spectrometer source region. This approach can circumvent the problem of matching fractions. For traditional off-line peptide fractionation, peptide IEF has several advantages over other methods. It allows the inclusion of fluorescent isoelectric point (pI)-markers to aid matching fractions across experiments [38,39] and adds the experimental peptide pI information of each identified peptide. Indeed Cargile and co-workers have shown the utility of peptide pI in combination with accurate mass alone and tandem MS-based approaches [40].

The present study is performed on an instrument lacking MS/MS capability. Thus potential biomarker peptides cannot be directly identified. The most straightforward approach may be to analyze the sample on another (MS/MS) instrument with the peptide of interest targeted based on its' determined mass. Additionally, the UPLC system can be used to collect fractions which may alleviate the need for LC–MS/MS analysis as the fractions could be analyzed directly (e.g. on a MALDI TOF/TOF system).

However, an alternate instrument may not always be available. In such a case, an approach similar to the "LC–$MS^E$"-method [41] may be applicable. First of all the gradient should be prolonged and perhaps extended to longer columns such that base-line separation of the peptide of interest may be achieved. Then an alternating scan method (high versus low cone voltage) should be set-up to attempt source fragmentation of the peptide of interest. Fragments observed can then be correlated with the elution profile of the peptide of interest. At least in theory, peptide sequencing may be attempted from such data, especially if blank runs have been included to allow removal of contaminant (fragment) ions.

Targeted proteomics approaches including multiple-reaction-monitoring (MRM) are currently gaining interest and can be seen as an attempt to address the issue of reproducible high throughput analysis of the proteome (for review see e.g. [42]). The here used method demonstrating retention time robustness and quantitative accuracy can be seen as one way to perform directed proteomics analyses with large patient cohorts. Additionally, the performance of the LC-method would justify combination with triple quadrupole mass spectrometry for robust targeted MRM-based proteomics.

Today much interest is directed towards the use of top-range mass spectrometry instrumentation capable of ultimate resolution and optimal mass accuracy. However, one limitation of such instruments is high acquirement cost. Another viable approach to proteome investigation is development of robust workflows employing powerful orthogonal separation methods together with targeted analysis.

This approach ensures depth of analysis in combination with minimizing the degree of missing values in larger sample sets. The need for the highest mass accuracy may be reduced for lower complexity (pre-fractionated) samples especially with the powerful reversed-phase separation and retention time accuracy shown here. Further, high-flow rate LC–MS systems offer robust quantitation as shown here in combination with high throughput. Both triple quadrupole and QTOF instruments are readily compatible with conventional electrospray flow-rates. Triple quadrupoles have excellent dynamic range and modern QTOF-instruments are continuously improving in terms of dynamic range and mass accuracy.

However, instead of triple quadrupoles and QTOF-instruments, many laboratories employ ion-trap based instruments. As these instruments have a finite capacity (ion filling of the trap) the dynamic range may be affected, especially for high complexity samples with a wide peptide concentration range. Thus low abundance ions may be inaccurately measured if they co-elute with high abundance peptides. Another problem with such systems is that space-charge effects may reduce mass accuracy as the trap is approaching saturation. Recently, one approach to overcome these limitations was published by Canterbury and co-workers for a (linear) ion trap [43]. The method described is based on including an ion-mobility (High-Field Asymmetric Waveform Ion Mobility Spectrometry, FAIMS) device as pre-fractionation step intermediate between the electrospray interface and the mass analyzer. The authors could demonstrate an improved dynamic range of analysis (>5-fold increase) and increased peak capacity (>8-fold) for a high-complexity sample (a yeast full proteome digest). Importantly, this was achieved on a chromatographic time scale where the average peak width was 0.47 min. The FAIMS separation consisted of five steps with each step requiring 100 ms and thus the total cycle was 0.5 s. In the present study peak times of ~12 s where observed. Such a peak time would then give 24 data-points across a peak which would allow the use of a FAIMS approach to increase dynamic range on trap instruments. However, the quantitative accuracy of such an approach remains to be determined.

## 5. Conclusions

With the continued mapping of the human proteome, fast and reproducible approaches will become increasingly important to gain knowledge on dynamic changes in biological systems. Here studied approaches using UPLC–MS allow increased throughput, which is important e.g. in clinical proteomics efforts where natural biological variability is high. Over time LC–MS instrumentation has become increasingly automated and user-friendly. Thus robust and powerful separation and analyses are becoming increasingly available for addressing scientific problems. The development and use of automated systems for robust discovery proteomics in connection with current developments on multiplexed targeted protein quantification using MRM pose an important area of further investigation.

## Acknowledgements

The study was supported by the following grants; CHEMORES and APO-SYS EU projects, Cancer society in Stockholm and Stockholm county council. Lennart Björkesten and Anders Kaplan of GE Healthcare, Uppsala, Sweden are thanked for fruitful discussions and contributions.

## References

[1] K.G. Kenrick, J. Margolis, Anal. Biochem. 33 (1970) 204.
[2] A. Gorg, W. Weiss, M.J. Dunn, Proteomics 4 (2004) 3665.
[3] J. Klose, U. Kobalz, Electrophoresis 16 (1995) 1034.
[4] A.M. Wheelock, D. Morin, M. Bartosiewicz, A.R. Buckpitt, Proteomics 6 (2006) 1385.
[5] M. Unlu, M.E. Morgan, J.S. Minden, Electrophoresis 18 (1997) 2071.
[6] S.A. Gerber, J. Rush, O. Stemman, M.W. Kirschner, S.P. Gygi, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 6940.
[7] J.M. Pratt, D.M. Simpson, M.K. Doherty, J. Rivers, S.J. Gaskell, R.J. Beynon, Nat. Protoc. 1 (2006) 1029.
[8] S. Hanke, H.X.S. Besir, D. Oesterhelt, M. Mann, J. Proteome Res. 7 (2008) 1118.
[9] M. Quaglia, C. Pritchard, Z. Hall, G. O'Connor, Anal. Biochem. 379 (2008) 164.
[10] A. Leitner, W. Lindner, J. Chromatogr. B: Anal. Technol. Biomed. Life Sci. 813 (2004) 1.

[11] M. Heller, H. Mattou, C. Menzel, X. Yao, J. Am. Soc. Mass Spectrom. 14 (2003) 704.
[12] S.-E. Ong, B. Blagoev, I. Kratchmarova, D.B. Kristensen, H. Steen, A. Pandey, M. Mann, Mol. Cell. Proteomics 1 (2002) 376.
[13] P.L. Ross, Y.N. Huang, J.N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, D.J. Pappin, Mol. Cell. Proteomics 3 (2004) 1154.
[14] L. Dayon, A. Hainard, V. Licker, N. Turck, K. Kuhn, D.F. Hochstrasser, P.R. Burkhard, J.C. Sanchez, Anal. Chem. 80 (2008) 2921.
[15] A. Schmidt, N. Gehlenborg, B. Bodenmiller, L.N. Mueller, D. Campbell, M. Mueller, R. Aebersold, B. Domon, Mol. Cell. Proteomics (2008).
[16] W.M. Old, K. Meyer-Arendt, L. Aveline-Wolf, K.G. Pierce, A. Mendoza, J.R. Sevinsky, K.A. Resing, N.G. Ahn, Mol. Cell. Proteomics 4 (2005) 1487.
[17] A.H. America, J.H. Cordewener, Proteomics 8 (2008) 731.
[18] H. Liu, R.G. Sadygov, J.R. Yates III, Anal. Chem. 76 (2004) 4193.
[19] Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, M. Mann, Mol. Cell. Proteomics 4 (2005) 1265.
[20] P.M. Palagi, D. Walther, M. Quadroni, S. Catherinet, J. Burgess, C.G. Zimmermann-Ivol, J.C. Sanchez, P.A. Binz, D.F. Hochstrasser, R.D. Appel, Proteomics 5 (2005) 2381.
[21] M. Katajamaa, J. Miettinen, M. Oresic, Bioinformatics 22 (2006) 634.
[22] R. Plumb, J. Castro-Perez, J. Granger, I. Beattie, K. Joncour, A. Wright, Rapid Commun. Mass Spectrom. 18 (2004) 2331.
[23] N. Marchetti, G. Guiochon, J. Chromatogr. A 1176 (2007) 206.
[24] A. Nordstrom, G. O'Maille, C. Qin, G. Siuzdak, Anal. Chem. 78 (2006) 3289.
[25] A. Motoyama, J.D. Venable, C.I. Ruse, J.R. Yates, Anal. Chem. 78 (2006) 5109.
[26] R. Plumb, J.R. Mazzeo, E.S. Grumbach, P. Rainville, M. Jones, T. Wheat, U.D. Neue, B. Smith, K.A. Johnson, J. Sep. Sci. 30 (2007) 1158.
[27] M. Wilm, M. Mann, Anal. Chem. 68 (1996) 1.
[28] T. Wachs, J. Henion, Anal. Chem. 73 (2001) 632.
[29] H. Yin, K. Killeen, R. Brennen, D. Sobek, M. Werlich, T. van de Goor, Anal. Chem. 77 (2005) 527.
[30] M. Noga, F. Sucharski, P. Suder, J. Silberring, J. Sep. Sci. 30 (2007) 2179.
[31] J. Lengqvist, K. Uhlen, J. Lehtio, Proteomics 7 (2007) 1746.
[32] A. Kaplan, M. Soderstrom, D. Fenyo, A. Nilsson, M. Falth, K. Skold, M. Svensson, H. Pettersen, S. Lindqvist, P. Svenningsson, P.E. Andren, L. Bjorkesten, J. Proteome Res. 6 (2007) 2888.
[33] A.H. America, J.H. Cordewener, M.H. van Geffen, A. Lommen, J.P. Vissers, R.J. Bino, R.D. Hall, Proteomics 6 (2006) 641.
[34] R. Zubarev, M. Mann, Mol. Cell. Proteomics 6 (2007) 377.
[35] A. Bodzon-Kulakowska, A. Bierczynska-Krzysik, T. Dylag, A. Drabik, P. Suder, M. Noga, J. Jarzebinska, J. Silberring, J. Chromatogr. B: Anal. Technol. Biomed. Life Sci. 849 (2007) 1.
[36] L.M. Orre, M. Pernemalm, J. Lengqvist, R. Lewensohn, J. Lehtio, Mol. Cell. Proteomics (2007).
[37] H. Eriksson, J. Lengqvist, J. Hedlund, K. Uhlen, L.M. Orre, B. Bjellqvist, B. Persson, J. Lehtio, P.J. Jakobsson, Proteomics 8 (2008) 3008.
[38] J. Chmelik, K. Mazanec, K. Slais, Electrophoresis 28 (2007) 3315.
[39] K. Mazanec, K. Slais, J. Chmelik, J. Mass Spectrom. 41 (2006) 1570.
[40] B.J. Cargile, J.L. Stephenson Jr., Anal. Chem. 76 (2004) 267.
[41] J.C. Silva, M.V. Gorenstein, G.Z. Li, J.P. Vissers, S.J. Geromanos, Mol. Cell. Proteomics 5 (2006) 144.
[42] V. Lange, P. Picotti, B. Domon, R. Aebersold, Mol. Syst. Biol. 4 (2008) 222.
[43] J.D. Canterbury, X. Yi, M.R. Hoopmann, M.J. MacCoss, Anal. Chem. 80 (2008) 6888.