Multiparameter optimization of an LC-MS/MS shotgun proteomics experiment was performed without any hardware or software modification of the commercial instrument. Under the optimized experimental conditions, with a 50-cm-long separation column and a 4-h LC-MS run (including a 3-h optimized gradient), 4,825 protein groups and 37,550 peptides were identified in a single run and 5,354 protein groups and 56,390 peptides in a triplicate analysis of the A375 human cell line, for approximately 50% coverage of the expressed proteome. The major steps enabling such performance included optimization of the cell lysis and protein extraction, digestion of even insoluble cell debris, tailoring the LC gradient profile, and choosing the optimal dynamic exclusion window in data-dependent MS/MS, as well as the optimal m/z scan window. Molecular & Cellular Proteomics 12: 10.1074/mcp.O113.028787, 3330–3338, 2013.

LC-MS-based proteomics has by now become an analytical method of choice in biological studies that demand deep proteome coverage (1–3). In order to increase the number of identified proteins, LC-MS analysis is commonly preceded by sample fractionation on the level of proteins or proteolytic peptides, or both (e.g. using two-dimensional gel electrophoresis, strong anion exchange, or isoelectric focusing) (4–7). These multidimensional approaches greatly reduce the complexity of the protein or peptide mixture in each fraction prior to MS detection, which enables comprehensive analysis of nearly the entire human proteome (>10,000 proteins) (6). The reverse side of the coin is the substantial operational cost, sample consumption (up to milligrams), and integral instrument time spent in these analyses (typically several days or longer). This puts severe limitations on high-throughput biological and clinical research.

In recent years, the power of the core analytical methods employed in proteomics, liquid chromatography and mass spectrometry, has sizably increased. Owing to the technological developments in packing materials of analytical columns and coupling interfaces, LC is now entering the era of ultra-high-pressure liquid chromatography (UPLC) characterized by unparalleled peak capacity and speed of separation (8–11). High-resolution MS is progressing at a fast rate with regard to sequencing capabilities and sensitivity of detection (12–15). Apart from that, notable improvements have been achieved in related areas, such as sample preparation methods and MS data processing (16–21).

The improving performance of shotgun LC-MS proteomics reduces the gap between the analytical capabilities of one-dimensional and multidimensional approaches. This trend is likely to continue in the near future, in view of the ongoing rapid technology developments. Considering the evident advantages of one-dimensional proteomics (i.e. the ease and speed of operation, lower sample consumption, and lower cost per run), it may regain the dominant position in many biological and clinical applications that it lost with the advent of multidimensional strategies. A wide selection of one-dimensional LC-MS platforms is commercially available nowadays for routine protein analyses with complete automation of the operational workflow, allowing large arrays of biological samples to be screened without attendance. In contrast, multidimensional analyses often involve interruptions in the experimental procedure for important steps that need to be performed manually by experienced personnel.

Most recent one-dimensional proteomics studies employing the combination of UPLC separation and high-resolution MS detection demonstrate remarkable progress in protein coverage, as well as in sensitivity and speed of analysis. In a very recent study, Nagaraj et al. reported an average of 3,923 protein groups identified in a single 4-h LC-MS analysis of 4 μg of yeast cell lysate (22). Combined analysis of six single runs increased the number of identifications to more than 4,000, which is close to the total number of proteins expressed in yeast under normal conditions. The median coverage of proteins in pathways with at least 10 members in the Kyoto Encyclopedia of Genes and Genomes was 88%, and the pathways that were not covered have not been expected to be active under the conditions used (22). But relative to the yeast proteome, the comprehensive analysis of the human proteome is considerably more challenging in view of its greater complexity and large dynamic range (at least 7 orders
of magnitude, compared with 4 orders of magnitude for yeast). Nonetheless, significant progress has recently been achieved in the field of one-dimensional LC-MS shotgun human proteomics. For example, in a single 8-h LC-MS run of proteolytic digest from a human cancer cell line, Cristobal et al. identified over 4,500 proteins and more than 26,000 unique peptides from as little as 1 μg of loaded sample (23). Thakur et al. reported an average of 4,695 proteins in a single LC-MS run of a human embryonic kidney cell line (HEK293) with a 480-min gradient time, and 5,376 proteins after a combined triplicate analysis (~1 day of total MS time). The identified proteins covered in total 173 out of the 200 metabolic and signaling pathways in the Kyoto Encyclopedia of Genes and Genomes (24).

In this study, we set out to investigate, and if possible extend, the current limits of one-dimensional shotgun human proteomics using the advanced commercial LC-MS instrumentation available today. All experiments utilized multiparameter optimization, without any hardware or software modification of the vendor-provided installation.

In a single UPLC-MS run of proteolytic digest from the A375 cancer line (3 μg) with a 3-h gradient time, we were able to identify 37,554 peptides and 4,825 protein groups. These numbers increased to 56,390 peptides and 5,354 proteins in a triplicate analysis, which is likely to be over 50% of the expressed cellular proteome. To the best of our knowledge, this is the deepest proteome coverage ever reported for such a low amount of loaded material and such a short replicate analysis time.

This level of analytical performance required careful optimization of the entire experimental workflow, including sample preparation, LC separation, MS detection, and data processing. Here we describe the optimization steps that made the greatest contribution to the ultimate analytical performance and discuss the venues remaining for further improvement in one-dimensional human proteomics.

**EXPERIMENTAL PROCEDURES**

The experimental workflow employed in this study included consecutive steps of A375 cell culturing, protein extraction and digestion, sample purification, UPLC-MS analysis of the proteolytic peptide mixture, and data processing (Fig. 1). The procedure is detailed below.

**Sample Preparation**—The A375 cell culture was grown with Dulbecco’s modified Eagle’s medium supplemented with 10% fetal bovine serum and 1% antibiotics (penicillin/streptomycin) (Switzerland). The cells were harvested after 70% confluence had been reached.

**Preparation of Lysis Solution**—The three different lysis solutions used in this study were prepared as follows. Aqueous ammonium bicarbonate (50 mM) was mixed with acetonitrile at a ratio of 9:1 (v:v). To prepare lysis solutions containing ProteaseMAX™ Surfactant (Promega, Madison, WI) and sodium deoxycholate (SDC) (Sigma Aldrich), 1 ml of the mixture was spiked with 1 mg of corresponding detergent powder (0.1% w/v). The third lysis solution contained 8 M urea in 50 mM aqueous ammonium bicarbonate (50 mM).

**Cell Lysis and Protein Extraction**—A cell pellet containing ~10^7 cells was resuspended in 1 ml lysis solution. Cell lysis was done over 10 min with the aid of rigorous vortexing. The lysate was left at 95 °C for 5 min and then subjected to 15 min sonication (30% amplitude, 3:3 pulse) with a Branson sonicator. The precipitate was either preserved or discarded, as described in main text, following sample centrifugation at 14,000 rpm over 7 min at room temperature. The total concentration of proteins was measured using a bicinchoninic acid assay (BCA assay kit, Pierce/Thermo Fisher Scientific). Prior to digestion, a urea-containing sample was diluted with aqueous ammonium bicarbonate (50 mM) to achieve a final urea concentration of 1 M.

**In-solution Digestion**—Proteins were reduced and alkylated via incubation with equal amounts of DTT and iodoacetamide, respectively, to a final concentration of 10 mM. Proteins (80 μg) were digested with trypsin added at a ratio of 1:40 (w/w) and incubated at 37 °C for 9 h. The digest was rigorously vortexed over 5 min. Trypsinization was terminated by the addition of 5% acetic acid (vol.).

**Sample Clean-up**—Samples that were prepared in ProteaseMAX or SDC solution were heated and shaken (500 rpm) for 30 min at 45 °C in order to precipitate the detergent. All peptide mixtures were purified using spin filtration (Pall Nanosep® 10 kDa with Omega membrane). The urea sample was cleaned from detergent using ZipTip® Pipette Tips C18 (Bedford, MA). ProteaseMAX and SDC samples were dried using a SpeedVac and resuspended in water with 0.1% formic acid.

**LC-MS/MS Analysis**—LC separation was done with three different Thermo Scientific EASY-Spray columns (PepMap® RSLC, C18, 100 Å, 3-μm-bead-packed 15-cm column, 2-μm-bead-packed 25-cm column, and 2-μm-bead-packed 50-cm column) connected to an Easy-nLC 1000 pump (Proxeon Biosystems, Odense, Denmark, now part of Thermo Fisher Scientific). In order to reduce backpressure, separation was conducted at 60 °C. The sample injection speed was limited by the backpressure value of 800 bar. In each single LC-MS run, the amount of loaded peptides was ~3 μg. Samples were loaded onto the column with buffer A (99.9% water, 0.1% formic acid) and eluted with a 120-, 180-, 240-, or 300-min gradient time, as described in the main text, with 2% to 30% buffer B (99.9% acetonitrile, 0.1% formic acid) at a flow rate of 250 nl min<sup>−1</sup>. Each of the gradients was followed by 1 h of column washing with the following steps: from 30% to 98% buffer B in 25 min, then 8 min in 98% buffer B, followed by a sharp decrease to 2% buffer B in 2 min, and finally 25 min in 2% buffer B. Typical backpressure values during separation were <550 bar.

Mass spectra were acquired with an Orbitrap Q Exactive mass spectrometer (Thermo Fisher Scientific) in a data-dependent manner, with automatic switching between MS and MS/MS scans using a top-20 method. MS spectra were acquired at a resolution of 70,000 with a target mass of 3 × 10<sup>6</sup> ions or a maximum integration time of 250 ms. The scan range was limited from 300 to 1650 m/z or from 400 to 1200 m/z, as explained in the main text. Peptide fragmentation was performed via higher-energy collision dissociation (HCD) with the energy set at 25 NCE (12). The ion selection abundance threshold was set at 0.1% with charge exclusion of z = 1 ions. The MS/MS spectra were acquired at a resolution of 17,500, with a target value of 2 × 10<sup>5</sup> ions or a maximum integration time of 120 ms. The fixed first m/z was 100, and the isolation window was 4.0 m/z.

**Data Analysis**—The raw data acquired were entered into MaxQuant, version 1.3.0.5, for peptide and protein identification (20). The Andromeda search engine (25) was run against the International Protein Index (human version 3.87). An MS/MS database search was performed with a 20-ppm mass tolerance for precursor ions in the initial search and a 6-ppm tolerance in the main search. Cysteine carbamidomethylation was selected as a fixed modification, and N-acetylation of protein and oxidation of methionine were selected as variable modifications. Trypsin/P was selected as the protease, with up to two missed cleavages allowed. Results were filtered by a 1%
false discovery rate at both protein and peptide levels (considering co-fragmented (second) peptides) (20). The minimum length of acceptable identified peptides was set as six amino acids. The variability of the number of identified peptides and protein groups was calculated as the standard deviation of the corresponding numbers between the replicates (at least three replicates for each sample). Subcellular location was analyzed via mapping of the identified protein groups onto UniProtKB/Swiss-Prot (v2013–01) with a database reference to Gene Ontology (GO) terms for cellular components.

RESULTS AND DISCUSSION

The integral workflow involves a large number of variables that need to be carefully tuned in order for deep proteome coverage to be achieved in the shortest possible time with minimal sample consumption. These variables include the choice of detergent for cell lysis and protein extraction, the profile and duration of the LC gradient, the relevant MS settings, and the parameters of the database protein search. Here, we provide an overview of the step-by-step optimization procedure for the experimental conditions, following the chronological order of the workflow shown in Fig. 1. In the end, the results of an experiment are reported in which all the optimized steps are combined in one analysis.

In all the experiments, we used A375 cells, a malignant melanoma cell line. A375 has been widely used in cytokine research (mostly IL1 response studies), as well as melanoma cancer research (26). There have been several transcriptomics and proteomics studies of the cell line (27, 28); however, until now there has been no report published on the deep proteome profile of this cell line.

Sample Preparation—

Cell Lysis Buffer—The composition of the lysis solution in which cell lysis is conducted determines the overall efficiency of the process. The buffer composition affects the dissociation rate of protein complexes, the extraction efficiency, and the solubility of extracted proteins. Furthermore, some buffer components can be difficult to remove after the lysis procedure, which can be detrimental to the UPLC-MS performance.

Three commonly used cell lysis buffers were evaluated in this study based on the number of proteins and peptides identified in shotgun UPLC-MS analysis. Employed detergents included urea, SDC, and ProteaseMAX. All the experiments were conducted under the same conditions. Protein digestion was done with trypsin, the LC gradient time was 3 h, and the data were analyzed using MaxQuant search.

The results are summarized in Table I. The greatest number of identifications for both peptides and protein groups was obtained using the buffer based on ProteaseMAX. According to the GO analysis, the increase in the number of identified proteins was mostly due to the higher extraction of membrane, nucleus, and cytosolic fractions (Fig. 2).

ProteaseMAX, a hydrophobic anionic sulfonate with the chemical formula sodium 3-((1-(furan-2-yl)undecyloxy)carbonylamino)propane-1-sulfonate and a molecular weight of 425.51 Da, is a rather mild detergent and can be easily separated from the lysate via acidic precipitation, which makes it highly compatible with protein digestion and LC-MS. The addition of acetonitrile (10% v:v) to the ProteaseMAX solution enables better solubilization of hydrophobic proteins. Based on the results of our comparison study, a ProteaseMAX-based buffer was chosen for cell lysis in further analyses.

### Table I

| Extraction-buffer detergent | Number of peptides | Number of protein groups |
|----------------------------|--------------------|-------------------------|
| Urea                       | 17,024 ± 148       | 3,326 ± 20              |
| SDC                        | 22,171 ± 403       | 3,698 ± 18              |
| ProteaseMAX                | 29,884 ± 228       | 4,465 ± 100             |
| Including cell debris      | 33,098 ± 283       | 4,655 ± 51              |

---

**Fig. 1.** UPLC-MS proteome analysis of A375 cell line. A375 cancer cells were cultured, harvested, lysed, and digested; digestion was followed by ZipTip® purification of generated peptides. The peptide mixture (3 μg) was separated on an Easy Spray column and analyzed using an Orbitrap Q Exactive.
Cell Debris Filtration—Cell lysis results in partial precipitation of the organelle debris, observed as a pellet after centrifugation. This pellet is commonly removed prior to protein digestion. The pellet may also contain proteins not present in the supernatant (e.g., because of their poor solubility or low extraction efficiency). This is particularly relevant for membrane protein complexes and nucleus proteins.

Here, we investigated whether more proteins can be observed with UPLC-MS if precipitated organelles are maintained in lysate buffer during digestion. In the first test experiment, cell lysate in ProteaseMAX solution was vortexed and sonicated, after which cell debris was removed via centrifugation. The supernatant was then subjected to trypsin digestion. In the second test experiment, digestion was conducted without the removal of cell debris. In both cases, digested lysate was filtered with a 10-kDa size-exclusion cutoff. Fully digested peptides passed through the filter, and both undigested proteins and cell debris were collected on the membrane and discarded. Peptides were analyzed via UPLC-MS with a 3-h LC gradient.

The results summarized in Table I demonstrate that preserving precipitated organelles during protein digestion allowed a notable increase in the total identification number (~7%). Including cell debris in the protein digestion with ProteaseMAX allowed a further increase in proteome coverage, most notably from the nucleus and membranes (Fig. 2, comparison between red and blue bars).

Proteins that are located in membranes or bound to DNA components generally have the lowest extraction efficiency. Therefore, the removal of cell debris, which mostly consists of membranes, organelles, and nucleus, results in the loss of proteins attached to them. Preserving cell debris during digestion allows cleavage of these proteins into fully or partially soluble peptides, which is reflected in the higher number of identified proteins (Table I).

In all the experiments described further below, cell debris was preserved in the cell lysate for enzymatic digestion.

UPLC Parameters—The speed of peptide sequencing in UPLC-MS experiments is limited by the acquisition rate of the mass spectrometer. The Q Exactive is the fastest among the currently available Orbitrap MS instruments. When operated in data-dependent MS/MS mode, it allows the sequencing of over 400 different ion precursors per minute. Given the 50% efficiency (on average) of peptide identification from a single HCD MS/MS spectrum (29), the maximal sequencing speed of Q Exactive MS can be estimated as at least 200 peptides per minute. If peptides are eluting at a higher rate, then some of them are likely to be missed by the data-dependent MS/MS. At a low elution rate of analytes, higher sequencing efficiency can be reached, but the instrument’s speed is not fully employed in that case.

Both of these scenarios were observed during the UPLC-MS run with a conventional linear gradient (Fig. 3, blue traces). Most of the peptides eluted within the first 100 min of the gradient, resulting in “saturation” of the MS/MS mode (~230 identified peptides per minute), whereas the elution density of peptides in the second half of the run was significantly lower than the instrument’s sequencing capacity. The analyte distribution in LC is heavily dependent on the gradient profile of the organic solvent composition in the mobile phase. Therefore, proper tuning of the gradient profile can enable a more uniform distribution of peptides throughout the entire experiment.
run and higher sequencing efficiency. The increment in the identification number would result from the peptides that were “transferred” by the gradient change from the saturated zone to a later stage of the run.

We experimentally tailored the profile of the LC gradient so as to reach a nearly constant number of unique peptides identified per minute of retention time. In order to fully exploit the analytical capacity of the instrument during the entire gradient time, a linear gradient (from 2% to 30% buffer B) was made shallower at a higher number of eluted peptides and sharper at a lower number. This resulted in the identification of a greater number of unique peptides on average per minute of gradient (Fig. 3, red traces). It should be emphasized that the total ion current profile of an LC run does not directly reflect the number of eluting unique peptides. For example, hydrophobic peptides eluting at the end of the gradient are typically less abundant than the earlier eluting hydrophilic ones. As a result, the total ion current profile can have very low intensity at the end of the gradient, while the number of eluting peptides can be quite large. The optimized profile starts at 2% acetonitrile and increases linearly to 5% over 19 min, then to 19% over 133 min, and finally to 30% in 28 min. It can be seen that when the optimized gradient is used, the instrument sequencing capacity is fully employed throughout the entire run. In total, this allowed the identification of 35,623 unique peptides in a 3-h peptide separation gradient within a 4-h LC-MS run, outnumbering the corresponding linear-gradient run by an average of 2,000 identifications.
The optimized profile was used in all the experiments reported further in this study. In longer LC runs, the gradient was expanded in the LC domain proportional to the relative increase in analysis time.

**MS Parameters—**

**Dynamic Exclusion Window—** In shotgun proteomics experiments, each full MS scan is followed by a series of MS/MS of the most abundant peptide ions detected by full MS (normally, 10 MS/MS per one full MS spectrum). In order to identify as many peptides as possible, multiple sequencing of the same precursor ions should be reduced to a minimum.

In the Orbitrap Q Exactive, sequenced precursors are automatically uploaded to the dynamic exclusion list. The same precursor ion can be selected for MS/MS analysis again only after a certain time interval. This interval is specified by an operator prior to the analysis and is referred to as the dynamic exclusion window (DEW). We analyzed the influence of the DEW setting on the number of peptide identifications in our experiments (Table II).

The observations suggest the existence of an optimal DEW value—around 15 s—that affords the largest number of identifications. This time should be a little shorter than the typical LC peak width. Thus, at lower DEWs (e.g. 5 s), abundant precursor ions are probably oversampled, which leads to the corresponding undersampling of low-abundance species and therefore less efficient identification.

Alternatively, when the DEW significantly exceeds the LC peak width (e.g. 90 s), precursor ions cannot be sequenced more than once. Because the average identification reliability based on single-shot MS/MS is approximately 50% (29), a large number of ions remain misidentified in this case. Thus, the number of peptide identifications for DEW = 90 s was ~3,000 less than for DEW = 15 s. The optimal DEW value is directly linked to the LC parameters, particularly the gradient time, and needs to be adjusted accordingly when different LC conditions are used. For example, when the gradient time of an LC run is decreased from 3 h to 1 h, the average LC peak width decreases only slightly, whereas the average density of eluting analytes undergoes a proportional 3-fold increase and will exceed the sequencing capacity of the instrument. We found that the optimal strategy is to attempt MS/MS of every eluting peptide only once, which can be achieved by increasing the DEW above the average LC peak width.

With regard to the efficiency of single-shot MS/MS sequencing, it is beneficial that precursor ions are sampled at the time point of the LC peak when they have the highest abundance. With the Orbitrap Q Exactive it is possible to follow the intensity of precursor ions and submit them for MS/MS as soon their intensity starts to decrease. Hereinafter, we refer to this option as the “trace-peak method.” Although the trace-peak method affords higher reliability of identifications per single MS/MS scan (Table II), it is quite inefficient for low-abundance species, which are only observed in full MS within a few scans. In our experience, activating the trace-peak method setting yielded a lower number of identified peptides than the optimized DEW method (Table II).

A harmonic combination of the fast sequencing speed of modern mass spectrometers with a proper DEW leads to the identification of more precursors. The DEW seems to be critical in reducing the time wasted as a result of unnecessary repeated sequencing, while at the same time providing enough replicate MS/MS datasets for reliable sequencing of the target.

The optimized 15-s DEW was used in all the experiments reported further in this paper.

**m/z Detection Window—** The maximum number of trapped ions is normally limited in the Orbitrap in order to avoid undesirable space charge effects during ion detection. For the sensitive detection of analyzed species, it is important that their relative number in the trap compared with background ions be as high as possible. The sources of background ions include solvent and salt clusters, LC stationary-phase components, and ubiquitous contaminants (e.g. siloxanes or phthalic acids (30)). Both tryptic peptides and chemical interferences exhibit nonuniform distributions in the m/z domain. Therefore, the sensitivity and specificity of peptide analysis can be improved by optimizing the m/z window of MS detection.

Fig. 4 compares the spectral distributions of identified peptides (red bars) with a corresponding distribution of unidentified species (blue bars) in the UPLC-MS analysis of the A375 cell line. Here, the term “unidentified” refers to MS/MS spectra that did not give matches with tryptic peptides in a MaxQuant search. The data show that the identified peptides exhibit a bell-shaped spectral distribution, with a peak around m/z 600. In contrast, unidentified species were more evenly distributed in MS. A plateau region at m/z 300–600 is followed by a gradual decrease at higher m/z values.

We constrained the original detection window (m/z 300–1650) down to the range in which the ratio of identified peptides to unidentified species is the highest (m/z 400–1200).

### Table II

| DEW (s) | Total number of scans | Number of MS/MS scans | Number of peptides |
|--------|-----------------------|-----------------------|--------------------|
| 90     | 75,576                | 57,767                | 35,623             |
| 30     | 83,460                | 67,717                | 38,617             |
| 15     | 91,849                | 77,570                | 38,788             |
| 5      | 111,281               | 97,479                | 30,606             |
| TPM    | 70,009                | 50,727                | 32,979             |

1 The abbreviations used are: DEW, dynamic exclusion window; GO, Gene Ontology; HCD, higher-energy collision dissociation; SDC, sodium deoxycholate; UPLC, ultra-high-pressure liquid chromatography.

The maximum number of trapped ions in the Orbitrap is approximately 50% (29), a large number of ions remain misidentified in this case. Thus, the number of peptide identifications for DEW = 90 s was ~3,000 less than for DEW = 15 s. The optimal DEW value is directly linked to the LC parameters, particularly the gradient time, and needs to be adjusted accordingly when different LC conditions are used. For example, when the gradient time of an LC run is decreased from 3 h to 1 h, the average LC peak width decreases only slightly, whereas the average density of eluting analytes undergoes a proportional 3-fold increase and will exceed the sequencing capacity of the instrument. We found that the optimal strategy is to attempt MS/MS of every eluting peptide only once, which can be achieved by increasing the DEW above the average LC peak width.

With regard to the efficiency of single-shot MS/MS sequencing, it is beneficial that precursor ions are sampled at the time point of the LC peak when they have the highest abundance. With the Orbitrap Q Exactive it is possible to follow the intensity of precursor ions and submit them for MS/MS as soon their intensity starts to decrease. Hereinafter, we refer to this option as the “trace-peak method.” Although the trace-peak method affords higher reliability of identifications per single MS/MS scan (Table II), it is quite inefficient for low-abundance species, which are only observed in full MS within a few scans. In our experience, activating the trace-peak method setting yielded a lower number of identified peptides than the optimized DEW method (Table II).

A harmonic combination of the fast sequencing speed of modern mass spectrometers with a proper DEW leads to the identification of more precursors. The DEW seems to be critical in reducing the time wasted as a result of unnecessary repeated sequencing, while at the same time providing enough replicate MS/MS datasets for reliable sequencing of the target.

The optimized 15-s DEW was used in all the experiments reported further in this paper.

**m/z Detection Window—** The maximum number of trapped ions is normally limited in the Orbitrap in order to avoid undesirable space charge effects during ion detection. For the sensitive detection of analyzed species, it is important that their relative number in the trap compared with background ions be as high as possible. The sources of background ions include solvent and salt clusters, LC stationary-phase components, and ubiquitous contaminants (e.g. siloxanes or phthalic acids (30)). Both tryptic peptides and chemical interferences exhibit nonuniform distributions in the m/z domain. Therefore, the sensitivity and specificity of peptide analysis can be improved by optimizing the m/z window of MS detection.

Fig. 4 compares the spectral distributions of identified peptides (red bars) with a corresponding distribution of unidentified species (blue bars) in the UPLC-MS analysis of the A375 cell line. Here, the term “unidentified” refers to MS/MS spectra that did not give matches with tryptic peptides in a MaxQuant search. The data show that the identified peptides exhibit a bell-shaped spectral distribution, with a peak around m/z 600. In contrast, unidentified species were more evenly distributed in MS. A plateau region at m/z 300–600 is followed by a gradual decrease at higher m/z values.

We constrained the original detection window (m/z 300–1650) down to the range in which the ratio of identified peptides to unidentified species is the highest (m/z 400–1200).
The corresponding UPLC-MS analysis of the same sample gave a higher number of identified peptides and protein groups (Table III).

This observation can be explained as follows. The shortest peptides in tryptic digest are normally the least informative for protein identification because of the poor sequence specificity. The longest peptides are the least useful for MS/MS sequencing because of the often insufficient bond cleavage that HCD MS/MS produces. The medium-sized peptides fragment well and are specific enough, so their sequencing is the most efficient for deep proteomics.

Despite the fact that some tryptic peptides remain undetected after \( m/z \) window reduction, the benefits associated with this reduction turn out to be more important for greater proteome coverage. These benefits include the higher scores of identified peptides and the increased number of proteins identified with more than two peptides (Table III). Therefore, reducing the \( m/z \) detection window can be suggested as a means to achieve higher protein coverage in shotgun proteomics.

Despite the fact that some tryptic peptides remain undetected after \( m/z \) window reduction, the benefits associated with this reduction turn out to be more important for greater proteome coverage. These benefits include the higher scores of identified peptides and the increased number of proteins identified with more than two peptides (Table III). Therefore, reducing the \( m/z \) detection window can be suggested as a means to achieve higher protein coverage in shotgun proteomics. It is worth noting that while the relative increase in the number of identified protein groups is quite moderate, it accounts for the least abundant proteins. Low-abundance proteins are the most probable potential biomarkers in deep proteomics studies, because they tend to mediate highly specific cellular functions (4, 31).

High-content Proteomics—

Gradient Time and Length of the LC Column—The resolving power of LC separation is directly proportional to the total gradient time. Extending the gradient time is warranted so long as the associated benefits outweigh the increased instrument time. As the analysis time is increased, the number of identifications gradually reaches a plateau, from which point further increases in gradient time become rather inefficient in terms of the content rate (i.e. the number of proteins identified per unit time). “High-content” proteomics strives to identify and quantify at least 1,000 proteins per hour of LC gradient, and ultimately 10,000 proteins in 10 h (the “10k10” goal) (32).

We performed four experiments with increasingly long gradient times (from two to five hours). The corresponding numbers of identifications are presented in Fig. 5. It can be seen that the longer the length of the column, the greater the total number of identifications for both proteins (Fig. 5A) and peptides (Fig. 5B). However, starting from 3 h, the increase becomes less pronounced.

In a single 4-h LC-MS run with a 3-h gradient time, we identified 37,554 peptides and 4,825 protein groups. In a triplicate analysis, these numbers increased to 56,390 peptides and 5,354 protein groups. The triplicate experiment was quantified via the MaxQuant LFQ method. In a single replicate, 3,777, 3,854, and 4,029 proteins were quantified, with 2,922 proteins quantified in all three runs. For these proteins, excellent accordance between the replicates was achieved (R² ≥ 0.977 for the logarithmic scale; Fig. 6).

Given the same number of identified proteins and the same instrument time, performing more replicates is more benefi-
cial, because it provides data for a statistical evaluation of the results. Thus, under the same conditions, \( n \) short-gradient experiments are preferable to a single LC/MS run \( n \) times longer. As demonstrated in Fig. 6, the optimized single LC-MS run with a 3-h peptide separation gradient time and 50-cm column covered \( \sim 50\% \) of the expressed human proteome in the A375 cancer cell line, which should meet the demands of many biological applications.

Further increases in the LC gradient time and length of the column should yield even more protein identifications. Based on the results of our study and some recent reports (33), the number of identified protein groups \( (N) \) depends on the length of the column \( (L, \text{cm}) \) and the gradient time \( (T, \text{h}) \) approximately as

\[
N = 1434 \times L^{0.25} \times T^{0.2} \tag{1}
\]

If this formula is extrapolated to higher values of \( L \) and \( T \), it can be expected that as many as 6,500 protein groups can be identified within 6 h with the use of a 1-m column ("6.5K6"). Finally, and much more speculatively, it could be possible to reveal the entire expressed proteome of a human cell (\( \sim 10,000 \) groups) in a single-shot 10-h LC/MS analysis ("10K10") with a 4-m column.

**CONCLUSION**

Today’s “deep proteome” research largely relies on the multidimensional separation of proteins and/or peptides prior to mass spectrometry. This is a highly demanding approach in terms of the operational cost, time, and amount of sample required. The alternative “one-shot” one-dimensional shotgun analysis is considerably easier, cheaper, faster, and more sensitive, but it is justifiably criticized for its lesser depth of analysis. However, most recent one-dimensional human proteome studies, including that described here, unambiguously suggest that the gap between these two competing approaches is rapidly narrowing.

Further optimization of different LC-MS/MS aspects should lead to even greater identification and coverage of proteins in single-run analyses. Particularly, the lysis buffer composition is an important variable for further enhancement of protein extraction and digestion. Emerging packing material technologies, such as core shell beads, may increase the peak capacity of LC columns. Continuous increases in the dynamic range of commercial mass spectrometers and the resolution of detection are also likely to contribute to the depth of a single-run analysis. Overall, we expect that there is sufficient room for technological improvement that might make it pos-
sible to achieve deeper human proteome coverage via one-dimensional LC-MS/MS in the near future.

However, a deeper understanding of the underlying molecular and cell biology requires not only proper identification and quantification of the protein groups in proteomics experiment, but also deeper sequencing within the group for studying such perturbations as mutations, splicing variants, posttranslational modifications, etc. This still leaves a lot of room for multidimensional proteomics.

* This work was supported by the Knut and Alice Wallenberg Foundation, the VINNOVA Foundation, Alzheimersfonden, and the Swedish Research council.

To whom correspondence should be addressed: Tel.: 46-8-524-87594; E-mail: roman.zubarev@ki.se.

REFERENCES

1. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198–207
2. Cravatt, B. F., Simon, G. M., and Yates, J. R. (2007) The biological impact of mass-spectrometry-based proteomics. *Nat. Rev. Mol. Cell. Biol.* **11**, 427–439
3. Choudhary, C., and Mann, M. (2010) Decoding signalling networks by mass spectrometry-based proteomics. *Nat. Rev. Mol. Cell. Biol.* **11**, 175–186
4. Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymonska, A., Herzig, F., Rinner, O., Ellenberg, J., and Aebersold, R. (2012) The quantitative proteome of a human cancer cell line. *Mol. Syst. Biol.*, 7, 549
5. Geiger, T., Wehner, A., Schaab, C., Cox, J., and Mann, M. (2012) Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics* **11**, M111.014050
6. Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Plabé, S., and Mann, M. (2011) Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.*, 7, 548
7. Lundberg, E., Fagerberg, L., Klevbrink, D., Matic, I., Geiger, T., Cox, J., Alenius, C., Lundeberg, J., Mann, M., and Utthin, M. (2010) Defining the transcriptome and proteome in three functionally different human cell lines. *Mol. Syst. Biol.*, 6, 450
8. MacNair, J. E., Lewis, K. C., and Jorgenson, J. W. (1997) Ultrahigh pressure reversed-phase liquid chromatography in packed capillary columns. *Anal. Chem.*, **69**, 983–989
9. MacNair, J. E., Patel, K. D., and Jorgenson, J. W. (1999) Ultrahigh pressure reversed-phase liquid chromatography: isotropic and gradient elution using columns packed with 1.0-μm particles. *Anal. Chem.*, **71**, 700–708
10. Swartz, M. E. (2005) UPLC™: an introduction and review. *J. Liq. Chromatogr. Relat. Technol.* **28**, 1253–1263
11. Shen, Y. F., Zhang, R., Moore, R. J., Kim, J., Metz, T. O., Hixson, K. K., Zhao, R., Livesey, E. A., Udseth, H. R., and Smith, R. D. (2005) Automated 20 kapi RPLC-MS and MS/MS with chromatographic peak capacities of 1000–1500 and capabilities in proteomics and metabolomics. *Anal. Chem.*, **77**, 3090–3100
12. Michalski, A., Damoc, E., Hauschild, J. P., Lange, O., Wieg haus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., and Horning, S. (2011) Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol. Cell. Proteomics* **10**, M111.011015
13. Makarov, A., Denisov, E., and Lange, O. (2009) Performance evaluation of a high-field Orbitrap mass analyzer. *J. Am. Soc. Mass Spectrom.* **20**, 1391–1396
14. Andrews, G. L., Simons, B. L., Young, J. B., Hawkridge, A. M., and Mudiman, D. C. (2011) Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). *Anal. Chem.*, **83**, 5442–5446
15. Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E. R., Senko, M., Makarov, A., Mann, M., and Horning, S. (2009) A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* **8**, 2759–2769
16. Wisniewski, J. R., Zougman, A., and Mann, M. (2009) Combination of FASP and stage tip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J. Proteome Res.* **8**, 5674–5678
17. Wisniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, U359–U360
18. Geiger, T., Wisniewski, J. R., Cox, J., Zanivan, S., Kruger, M., Ishihama, Y., and Mann, M. (2011) Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nat. Protoc.* **6**, 147–157
19. Cox, J., and Mann, M. (2011) Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* **80**, 273–299
20. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372
21. Beck, M., Claassen, M., and Aebersold, R. (2011) Comprehensive proteomics. *Curr. Opin. Biotechnol.* **22**, 3–8
22. Nagaraj, N., Kulak, N. A., Cox, J., Neuhauser, N., Mayr, K., Hoening, O., Vorn, O., and Mann, M. (2012) System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Mol. Cell. Proteomics* **11**, M111.013722
23. Cristobal, A., Henrich, M. L., Gianpianti, P., Goerdalay, S. S., Heck, A. J. R., and Mohammed, S. (2012) In-house construction of a UHPLC system enabling the identification of over 4000 protein groups in a single analysis. *Analyst* **137**, 3541–3548
24. Thakur, S. S., Geiger, T., Chatterjee, B., Bandilla, P., Fröhlich, F., Cox, J., and Mann, M. (2011) Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Mol. Cell. Proteomics* **10**, M110.003699
25. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805
26. Moreb, J., and Zucali, J. R. (1992) The therapeutic potential of interleukin-1 and tumor necrosis factor on hematopoietic stem cells. *Leuk. Lymphoma* **8**, 267–275
27. Kaplanov, P., Cawley, S. E., Drenkow, J., Bekiranov, S., Strausburg, R. L., Fodor, S. P., and Gingeras, T. R. (2002) Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**, 916–919
28. Caputo, E., Maiorana, L., Vasta, V., Pezzino, F. M., Sunkara, S., Wynne, K., Elia, G., Marincola, F. M., McCubrey, J. A., Libra, M., Travali, S., and Cooper, R. E. (2000) Characterization of human melanoma cell lines and melanogeneses by proteome analysis. *Cell. Cycle* **9**, 2924–2936
29. Neuhauser, N., Michalski, A., Cox, J., and Mann, M. (2012) Expert system for computer-assisted annotation of MS/MS spectra. *Mol. Cell. Proteomics* **11**, 1500–1509
30. Keller, B. O., Suj, J., Young, A. B., and Whittal, R. M. (2008) Interferences and contaminants encountered in modern mass spectrometry. *Anal. Chem.* **80**, 71–81
31. Schwannhauser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011) Global quantification of mammalian gene expression control. *Nature* **473**, 337–342
32. Köcher, T., Swart, R., and Mechtler, K. (2011) Ultra-high-pressure RPLC hyphenated to an LTQ-Orbitrap Velos reveals a linear relation between peak capacity and number of identified peptides. *Anal. Chem.* **83**, 2699–2704
33. Zubarev, R. A. (2013) The challenge of the proteome dynamic range and its implications for in-depth proteomics. *Proteomics* **5**, 723–726