

A Double-Barrel Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) System to Quantify 96 Interactomes per Day*[§]

Fabian Hosp[‡]||, Richard A. Scheltema[‡]||, H. Christian Eberl[‡]||[¶], Nils A. Kulak[‡]||, Eva C. Keilhauer[‡], Korbinian Mayr[‡], and Matthias Mann[‡]§

The field of proteomics has evolved hand-in-hand with technological advances in LC-MS/MS systems, now enabling the analysis of very deep proteomes in a reasonable time. However, most applications do not deal with full cell or tissue proteomes but rather with restricted subproteomes relevant for the research context at hand or resulting from extensive fractionation. At the same time, investigation of many conditions or perturbations puts a strain on measurement capacity. Here, we develop a high-throughput workflow capable of dealing with large numbers of low or medium complexity samples and specifically aim at the analysis of 96-well plates in a single day (15 min per sample). We combine parallel sample processing with a modified liquid chromatography platform driving two analytical columns in tandem, which are coupled to a quadrupole Orbitrap mass spectrometer (Q Exactive HF). The modified LC platform eliminates idle time between measurements, and the high sequencing speed of the Q Exactive HF reduces required measurement time. We apply the pipeline to the yeast chromatin remodeling landscape and demonstrate quantification of 96 pull-downs of chromatin complexes in about 1 day. This is achieved with only 500 μ g input material, enabling yeast cultivation in a 96-well format. Our system retrieved known complex-members and the high throughput allowed probing with many bait proteins. Even alternative complex compositions were detectable in these very short gradients. Thus, sample throughput, sensitivity and LC/MS-MS duty cycle are improved severalfold compared with established workflows. The pipeline can be extended to different types of interaction studies and to other me-

dium complexity proteomes. *Molecular & Cellular Proteomics* 14: 10.1074/mcp.O115.049460, 2030–2041, 2015.

Shotgun proteomics is concerned with the identification and quantification of proteins (1–3). Prior to analysis, the proteins are digested into peptides, resulting in highly complex mixtures. To deal with this complexity, the peptides are separated by liquid chromatography followed by online analysis with mass spectrometry (MS), today facilitating the characterization of almost complete cell line proteomes in a short time (3–5). In addition to the characterization of entire proteomes, there is also a great demand for analyzing low or medium complexity samples. Given the trend toward a systems biology view, relatively large sets of samples often have to be measured. One such category of lower complexity protein mixtures occurs in the determination of physical interaction partners of a protein of interest, which requires the identification and quantification of the proteins “pulled-down” or immunoprecipitated via a bait protein. Protein interactions are essential for almost all biological processes and orchestrate a cell’s behavior by regulating enzymes, forming macromolecular assemblies and functionalizing multiprotein complexes that are capable of more complex behavior than the sum of their parts. The human genome has almost 20,000 protein encoding genes, and it has been estimated that 80% of the proteins engage in complex interactions and that 130,000 to 650,000 protein interactions can take place in a human cell (6, 7). These numbers demonstrate a clear need for systematic and high-throughput mapping of protein-protein interactions (PPIs) to understand these complexes.

The introduction of generic methods to detect PPIs, such as the yeast two-hybrid screen (Y2H) (8) or affinity purification combined with mass spectrometry (AP-MS)¹ (9), have revolutionized the protein interactomics field. AP-MS in particular has emerged as an important tool to catalogue interactions with the aim of better understanding basic biochemical mech-

From the [‡]Department of Proteomics and Signal Transduction, Max-Planck Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany

* Author's Choice—Final version free via Creative Commons CC-BY license.

Received February 27, 2015, and in revised form, February 27, 2015

Published, MCP Papers in Press April 17, 2015, DOI 10.1074/mcp.O115.049460

Author contributions: F.H., R.A.S., C.E., and M.M. designed the research; F.H., R.A.S., and C.E. performed the research; F.H., R.A.S., N.A.K., E.C.K., and K.M. contributed new reagents or analytic tools; F.H., R.A.S., C.E., N.A.K., E.C.K., and M.M. analyzed the data; and F.H., R.A.S., and M.M. wrote the paper.

¹ The abbreviations used are: AP-MS, affinity purification mass spectrometry; FDR, false discovery rate; GFP, green fluorescent protein; MaxLFQ, MaxQuant label-free quantification; PPI, protein-protein interaction; TMT, tandem mass tag; Y2H, yeast two-hybrid.

organisms in many different organisms (10–17). It can be performed under near-physiological conditions and is capable of identifying functional protein complexes (18). In addition, the combination of affinity purification with quantitative mass spectrometry has greatly improved the discrimination of true interactors from unspecific background binders, a long-standing challenge in the AP-MS field (19–21). Nowadays, quantitative AP-MS is employed to address many different biological questions, such as detection of dynamic changes in PPIs upon perturbation (22–25) or the impact of posttranslational signaling on PPIs (26, 27). Recent developments even make it possible to provide abundances and stoichiometry information of the bait and prey proteins under study, combined with quantitative data from very deep cellular proteomes. Furthermore, sample preparation in AP-MS can now be performed in high-throughput formats capable of producing hundreds of samples per day. With such throughput in sample generation, the LC-MS/MS part of the AP-MS pipeline has become a major bottleneck for large studies, limiting throughput to a small fraction of the available samples. In principle, this limitation could be circumvented by multiplexing analysis via isotope-labeling strategies (28, 29) or by drastically reducing the measurement time per sample (30–32). The former strategy requires exquisite control of the processing steps and has not been widely implemented yet. The latter strategy depends on mass spectrometers with sufficiently high sequencing speed to deal with the pull-down in a very short time. Since its introduction about 10 years ago (33), the Orbitrap mass spectrometer has featured ever-faster sequencing capabilities, with the Q Exactive HF now reaching a peptide sequencing speed of up to 17 Hz (34). This should now make it feasible to substantially lower the amount of time spent per measurement.

Although very short LC-MS/MS runs can in principle be used for high-throughput analyses, they usually lead to a drop in LC-MS duty cycle. This is because each sample needs initial washing, loading, and equilibration steps, independent of gradient time, which takes a substantial percentage for most LC setups - typically at least 15–20 min. To achieve a more efficient LC-MS duty cycle, while maintaining high sensitivity, a second analytical column can be introduced. This enables the parallelization of several steps related to sample loading and to the LC operating steps, including valve switching. Such dual analytical column or “double-barrel” setups have been described for various applications and platforms (30, 35–39).

Starting from the reported performance and throughput of workflows that are standard today (16, 21, 40–42), we asked if it would be possible to obtain a severalfold increase in both sample throughput and sensitivity, as well as a considerable reduction in overall wet lab costs and working time. Specifically, our goal was to quantify 96 medium complexity samples in a single day. Such a number of samples can be processed with a 96-well plate, which currently is the format of choice for

highly parallelized sample preparation workflows, often with a high degree of automation. We investigated which advances were needed in sample preparation, liquid chromatography, and mass spectrometry. Based on our findings, we developed a parallelized platform for high-throughput sample preparation and LC-MS/MS analysis, which we applied to pull-down samples from the yeast chromatin remodeling landscape. The extent of retrieval of known complex members served as a quality control of the developed pipeline.

EXPERIMENTAL PROCEDURES

Preparation of Yeast Lysates—GFP-tagged yeast strains from the *Saccharomyces cerevisiae* GFP Clone Collection (43), the parental strain BY4741 and the control strain pHis3-GFP (21) were cultured in YPD liquid medium in 96-deep well plates (Sarstedt, Nümbrecht, Germany) at standard conditions. We used 32 distinct yeast strains in biological triplicates, resulting in 96 experimental samples. Yeast cells were grown until they reached an Optical Density_{600 nm} of around 1, followed by harvesting culture volumes equaling 2 ODs per well. Yeast cell pellets were dissolved in 300 μ l lysis buffer (150 mM NaCl, 50 mM Tris-HCl (pH 8.0), 1 mM MgCl₂, 5% glycerol, 1% IGEPAL CA-630 (Sigma-Aldrich, Schnellendorf, Germany), complete protease inhibitors (Roche, Mannheim, Germany), 1% benzonase (Merck, Darmstadt, Germany)), transferred into FastPrep tubes (MP Biomedicals, Eschwege, Germany) containing 1 mm silica spheres (lysing matrix C, MP Biomedicals), and lysed in a FastPrep24 instrument (MP Biomedicals) for 6 \times 1 min at maximum speed. Lysates were cleared by centrifugation at 16,100 \times g for 10 min at 4 $^{\circ}$ C.

Affinity Purification—Each well of a GFP-multiTrap plate (Chromotek, Martinsried, Germany) was washed three times with 200 μ l buffer 1 (150 mM NaCl, 50 mM Tris-HCl, pH 8.0) and then incubated with the cleared yeast cell lysate (500 μ g total protein extract) with gentle shaking at 100 rpm for 60 min at 4 $^{\circ}$ C. Next, each well was washed twice with 200 μ l buffer 2 (150 mM NaCl, 50 mM Tris-HCl (pH 8.0), 0.25% IGEPAL CA-630) and four times with 200 μ l buffer 1 before incubation with 25 μ l elution buffer (2 M urea, 20 mM Tris-HCl (pH 8.0), 1 mM DTT, 100 ng sequence-grade modified trypsin (Promega, Madison, WI, USA) at room temperature for 90 min. Subsequently, the resulting peptides were alkylated with 25 μ l alkylation buffer (2 M urea, 20 mM Tris HCl (pH 8.0), 5 mM iodoacetamide) and finally washed once with 50 μ l urea buffer (2 M urea, 20 mM Tris HCl, pH 8.0) for 10 min, respectively. The supernatants from the elution, alkylation and washing step were collected after each step and combined in a clean 96-well plate. This plate was incubated overnight at room temperature to ensure a complete digest. The next morning, the digest was stopped by addition of 10 μ l 10% TFA per well. The acidified peptides were purified on StageTips (44) containing two layers of Poly(styrenedivinylbenzene)-Reversed-Phase Sulfonate (Empore 2241, 3 M, Neuss, Germany) material to desalt and purify the peptides. Samples were eluted from the StageTips with 60 μ l elution buffer (80% acetonitrile, 1% ammonium hydroxide) and evaporated in a SpeedVac concentrator for 30 min. The remaining peptide solution volume was adjusted to 4 μ l with buffer A* (2% ACN, 0.1% formic acid).

LC-MS/MS Analysis—Online chromatography was performed with a modified Thermo EASY-nLC 1000 UHPLC system (Thermo Fisher Scientific, Bremen, Germany) coupled online to the Q Exactive HF instrument with a nano-electrospray ion source (Thermo Fisher Scientific). Two analytical columns (15 cm long, 75 μ m inner diameter) were packed in-house with ReproSil-Pur C₁₈ AQ 1.9 μ m reversed phase resin (Dr. Maisch GmbH, Ammerbuch, Germany) in buffer A (0.5% formic acid) and matched with regard to back-pressure to

ensure intercolumn reproducibility. During online analysis, the analytical columns were placed in a modified column heater (Sonation GmbH, Biberach, Germany) regulated to a temperature of 55 °C. Modifications to both systems are described in RESULTS. Peptides were loaded onto the analytical columns with buffer A at a back pressure of 650 bar (generally resulting in a flow rate of 500 nL/min) and separated with two distinct linear gradients of 8–30% buffer B (80% ACN and 0.5% formic acid) at a flow rate of 450 nL/min controlled by IntelliFlow technology over 10 min and 22 min, respectively (generally at a back pressure of around 500 bar). Online quality control was performed with SprayQc (45), which was extended with an additional plugin to support a high-voltage switch controlling the spray voltage for the analytical columns (RESULTS). MS data were acquired with a Q Exactive Plus (27 min gradients) and a Q Exactive HF (14 min gradients) instrument, as the latter has been found to be up to twice as fast (34) and thus capable of dealing with the fast chromatography of the 14 min gradient. The instruments were programmed with a data-dependent top 5 and top 10 method, respectively, dynamically choosing the most abundant not yet sequenced precursor ions from the survey scans (300–1,650 Th). Instruments were controlled using Tune 2.5 and Xcalibur 3.0.63. At a maximum ion inject time of 45 ms for both instruments, the cycle time was ~800 ms, sufficient for generating a median of 16 data points (14 min) or 25 data points (27 min) over the observed elution peaks (RESULTS). Further settings were chosen according to their previously determined optimal values (34). Sequencing was done with higher-energy collisional dissociation fragmentation with a target value of 1e5 ions determined with predictive automatic gain control, for which the isolation of precursors was performed with a window of 1.4 Th. Survey scans were acquired at a resolution of 70,000 and 60,000, respectively, at m/z 200 and the resolution for HCD spectra was set to 17,500 and 15,000, respectively, at m/z 200. Normalized collision energy was set to 27 and the “underfill ratio,” specifying the minimum percentage of the target ion value likely to be reached at maximum fill time was defined as 10% (27 min) and 40% (14 min). The elevated sequencing threshold ensured that, with the reduced complexity of samples, the fragmentation scans are of higher quality. Furthermore, the S-lens radio frequency level was set to 60, which gave optimal transmission of the m/z region occupied by the peptides from our digest (34). We excluded precursor ions with unassigned, single, or five and higher charge states from fragmentation selection.

Data Analysis—All data were analyzed with the MaxQuant proteomics data analysis workflow version 1.4.3.14 (46). The false discovery rate (FDR) cut off was set to 1% for protein and peptide spectrum matches. Peptides were required to have a minimum length of seven amino acids and a maximum mass of 4,600 Da. MaxQuant was used to score fragmentation scans for identification based on a search with an initial allowed mass deviation of the precursor ion of a maximum of 4.5 ppm after time-dependent mass calibration. The allowed fragment mass deviation was 20 ppm. Fragmentation spectra were identified using the UniprotKB *S. cerevisiae* database (based on 2014–07 release; 6,643 entries) combined with 262 common contaminants by the integrated Andromeda search engine (47). Enzyme specificity was set as C-terminal to arginine and lysine, also allowing cleavage before proline, and a maximum of two missed cleavages. Carbamidomethylation of cysteine was set as fixed modification and N-terminal protein acetylation and methionine oxidation as variable modifications. Both “match between runs,” with a maximum time difference of 30 s, and label-free quantification (LFQ) with standard settings, were enabled (48). Additional metadata stored in the RAW files (e.g. ion inject time, noise level, etc.) were extracted using MS-FileReader (Thermo Scientific) with in-house-developed tools.

Further data analysis with the goal of assigning the interactors was performed with the R scripting and statistical environment (49) using

ggplot (50) for data visualization. Briefly, LFQ intensity values were base10 logarithmized, resulting in a normal distribution. Missing values were imputed by randomly selecting from a normal distribution centered on the lower edge of the intensity values (for this normal distribution the shift was set to 1.8 standard deviations from the mean and the width to 0.3 standard deviations; see histograms describing placement in Figs. S8 and S9). Proteins were excluded in subsequent steps for baits with less than two valid values in the triplicate for the bait (mostly presented as significantly depleted proteins due to the imputed character of the intensity values). The fold enrichment was calculated as the mean ratio between the bait measurements and the proteome measurements of the parental strain (conforming to the mean used in the consequent t test). For the fold enrichment, the standard error of the mean was additionally determined. Permutation-based FDR-controlled t test p values were calculated for each protein between the bait triplicate and the parental strain triplicate (employing 250 permutations). The p value was adjusted using a scaling factor s_0 with a value of 1 prior to FDR control, which magnifies the importance of the difference of the mean (51). Furthermore, the correlation of each protein's LFQ intensity profile (consisting of all the measured intensity values for that protein) to the LFQ intensity profile of the bait was calculated (21), and the resulting correlation p values were adjusted to 1% FDR using the Benjamini and Hochberg procedure. Interactor classes were assigned based on the following rules: (A) only <1% FDR t test significance, (A+) both <1% FDR t test significance, and <1% FDR correlation significance, (B+) both <5% FDR t test significance and <1% FDR correlation significance and (B) only <5% FDR t test significance. Known interactors from the *Saccharomyces* Genome Database (www.yeastgenome.org) mainly fell in classes A+, A, and B+. Therefore, we conducted follow-up analyses solely on these classes. For each significant outlier, we also introduced a single significance value, based on the s_0 scaling introduced in the t test, which combines the enrichment value and the t test statistic. This is calculated as the distance in log-space from the origin. The higher this value, the better the data quality and experimental success of that particular interactor. Stoichiometry information was determined in two ways. The first, termed interaction stoichiometry, is the ratio between the calculated intensity-based absolute quantification values (determining the copy numbers from the acquired mass spectrometry data) of the interactors to the bait (52). The second, termed abundance stoichiometry, is the ratio between the normal cellular copy numbers of the interactors to the bait.

RESULTS

Reducing the LC-MS/MS Analysis Time—First, we aimed to establish optimal conditions for reducing the LC gradient length. Both the flow rate and gradient starting percentage require adaptations to ensure that the signal of each peptide does not degrade and to maximize the spread of peptides over the gradient. To achieve this, we tested the effect of flow rate (ranging from 200 to 500 nL/min) and gradient length (from 15 to 120 min) on the chromatographic peak-width with a standard HeLa digest on the Q Exactive HF (34). By far, the largest effect on peak width was shortening the gradient length as this provided a reduction of ~75% on the width, while the flow rate reduced it only by ~4% (Fig. 1A). With regard to overall proteome depth, we were able to identify about 740 proteins with a standard HeLa digest using the shortest gradient length of 15 min with the Q Exactive HF (Fig. 1B). Hence, the complexity of protein samples should not exceed such a number when high sample throughput is en-

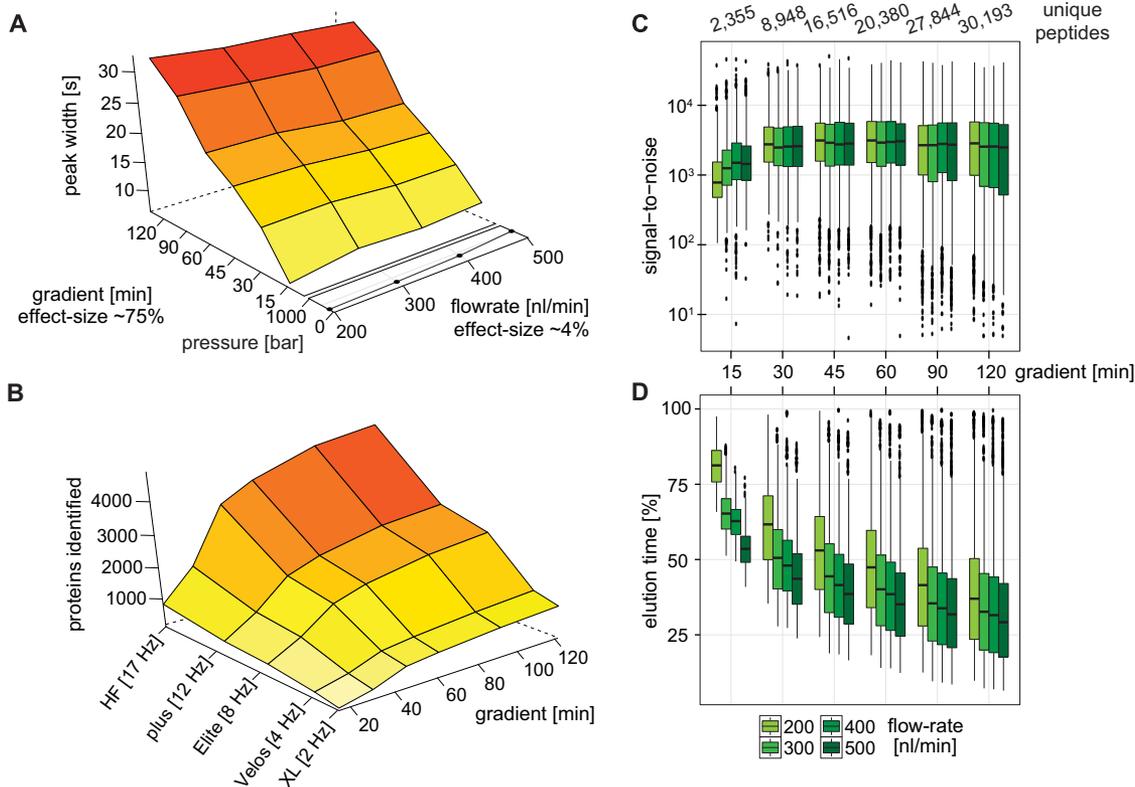


FIG. 1. Chromatography optimization for very short gradients. (A) Peak-width as a function of gradient length and flow rate. Effect size is the calculation of the reduction compared with the largest change in peak-width. (B) Extrapolation of protein identifications as a function of gradient length and scan speed of various MS platforms (Q Exactive HF and plus, Orbitrap Elite, and Velos, LTQ Orbitrap XL, respectively). (C) Effect of flow rate on the signal-to-noise for a set of 750 unique isotope patterns identified in all measurements and spread out over the entire gradient. (D) Elution time shift induced by higher flow rates, normalized to the gradient length.

visioned. We also determined protein identifications for lower sequencing speed (Fig. 1B). Notably, even platforms with lower sequencing speed like the Orbitrap XL identified about 1,000 proteins with a 120 min gradient, suggesting that already this machine generation had the potential to identify all proteins of a lower complexity sample given sufficiently long gradients.

Higher flow rates could have a detrimental effect on the signal-to-noise due to the higher dilution of peptides in the buffer, which we investigated by extracting the signal-to-noise values for a set of 750 isotope patterns identified in all the runs and spread out over the full retention time range. For the longest gradient length of 120 min, we observe a slight decrease in signal-to-noise for the higher flow rates, whereas unexpectedly higher flow rates partially improve the signal-to-noise for the shortest gradient. For the intermediate gradient lengths, the flow rate does not appreciably affect the signal-to-noise ratio. Between the two shortest gradients of 30 and 15 min, we observe a drop in signal-to-noise, which we attribute to imprecision of the buffer delivery by the LC (Fig. 1C). Given that it takes time for the buffer mixture to arrive from the mixing T connection to the tip of the analytical column, and therefore for the peptides to elute, the shorter gradients

suffer in terms of gradient occupancy (percentage of the gradient occupied by peptides) when using lower flow rates. This is mostly improved by forcing the peptides to elute earlier with higher flow rates. For the shortest gradient lengths, we were able to move the start of peptide elution from 60% in the gradient (at 9 min) to 40% in the gradient (at 6 min), improving the spread of the peptides over the complete gradient and providing better chromatographic resolution. For the 30 min gradient, the first elution was moved from 10 min (35% of the gradient time) to 7 min (25%) (Fig. 1D).

Based on these findings, we determined the optimal gradient time to be 27 min with a flow rate of 450 nl/min, which kept the backpressure of the LC pumps at an acceptable level of around 500 bar. This, however, still results in 2 days of measurements for 96 samples. The 12 min gradient at the same flow rate necessary for exactly 24 h of measurement for the same number of samples is expected to have reduced chromatographic performance compared with the 27 min gradient. This period is also too short to transfer the peptides onto the analytical column in parallel. We therefore increased the gradient time to 14 min and activated the loading pump during the intersample preparation time, which reliably loaded all the

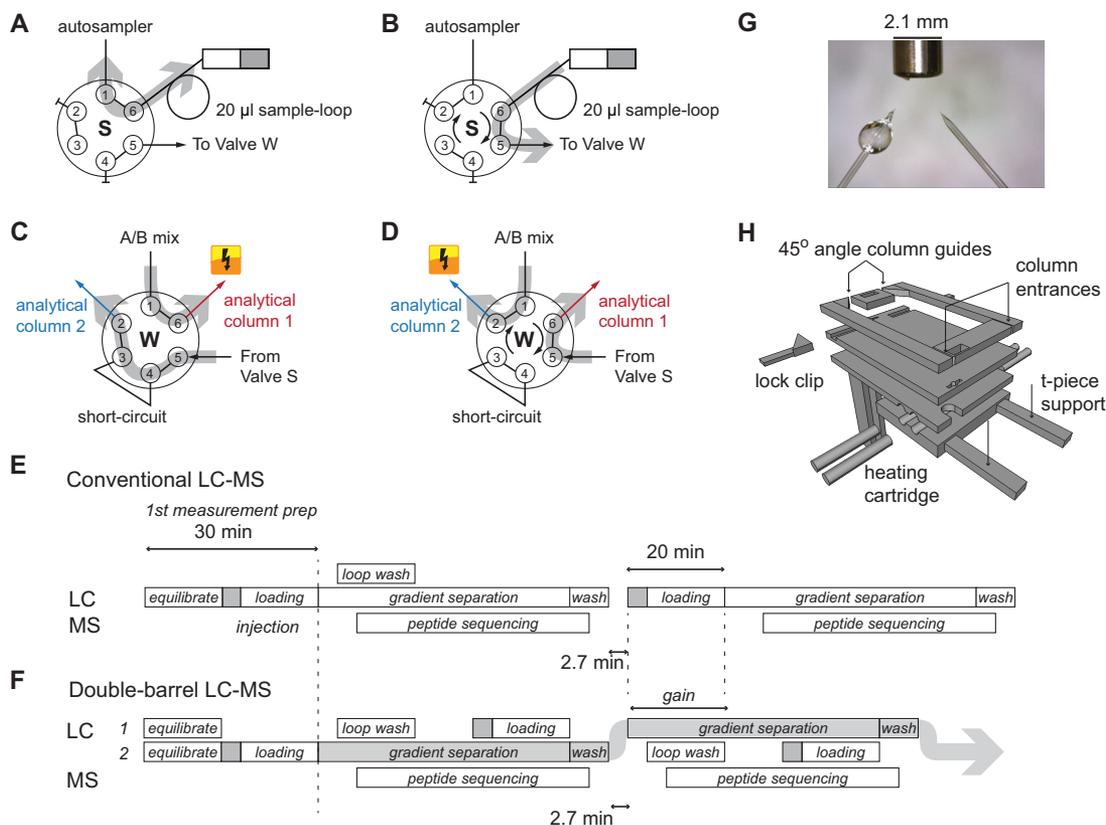


FIG. 2. Parallel UHPLC operation with two analytical columns. (A) In this position of valve S, the sample pump can fill the sample loop. (B) By switching valve S, the contents of the sample loop can be loaded onto one of the analytical columns. (C) In this position of valve W, the analytical column 1 can be eluted with the mobile phase, while analytical column 2 is loaded. (D) By switching the position of valve W, this behavior is inverted. (E) In the conventional setup, the mass spectrometer is not sequencing while the HPLC is loading a new sample. The light gray arrow indicates where the mobile phase is active. (F) With the double-barrel setup, this idle-time is circumvented, enabling almost continuous operation. (G) Positioning of the analytical columns in reference to the inlet of the mass spectrometer. (H) Redesign of the column oven for two analytical columns.

peptides onto the analytical column. Additionally, we increased the starting acetonitrile percentage of the gradient from 2% to 8% (EXPERIMENTAL PROCEDURES) to start the peptide elution at an earlier point of the gradient. Collectively, this resulted in a time frame for peptide elution of 8 min and 18 min, representing 60 and 75%, respectively, of the total measurement time for the 14 and 27 min gradients. At these conditions, the median peak-width (base-to-base) was 6 s (14 min) and 11 s (27 min), respectively.

Double-Barrel Chromatography on the EASY-nLC—Next, we set out to develop a double-barrel chromatography system in order to reduce the idling time of the mass spectrometer during loading of the peptides to the LC column. Unfortunately, no such setup has been described for the Thermo EASY-nLC 1000 UHPLC systems (Thermo Fisher Scientific) that we employ and that are widely used with the Orbitrap-family of mass spectrometers. To address this, we modified the liquid pathway of the EASY-nLC 1000 UHPLC system (Figs. 2A–2D). In brief, we placed the sample loop directly between the pump S and valve S, allowing the system to utilize pump S as both the sample pickup as well as the

sample-loading pump (in the original setup, pump A is used as sample-loading pump). The valve S is connected to valve W (in the original setup this valve is connected to a waste line used for rapid evacuation of the buffers from the lines), which connects to the buffer A and B mixing-T connection and the two analytical columns through standard sample lines. This setup allows loading of one sample onto one of the analytical columns while the other is eluted.

To make use of this new liquid pathway and to drive two analytical columns in parallel, we also modified the “business logic” controlling the UHPLC system. The normally sequential steps in the analysis process (Fig. 2E) were altered to work in parallel with each other (Fig. 2F). As soon as the preparation for the currently active analytical column has finished, the initiation phase and the valve W has switched to elute the loaded peptides, the inactive analytical column is prepared in parallel for the next sample. This is done in three consecutive steps: First, the sample loop is washed, then the new sample is loaded into the sample loop, and finally the sample is loaded from the sample loop onto the analytical column. With the above described arrangement of the pumps and valves,

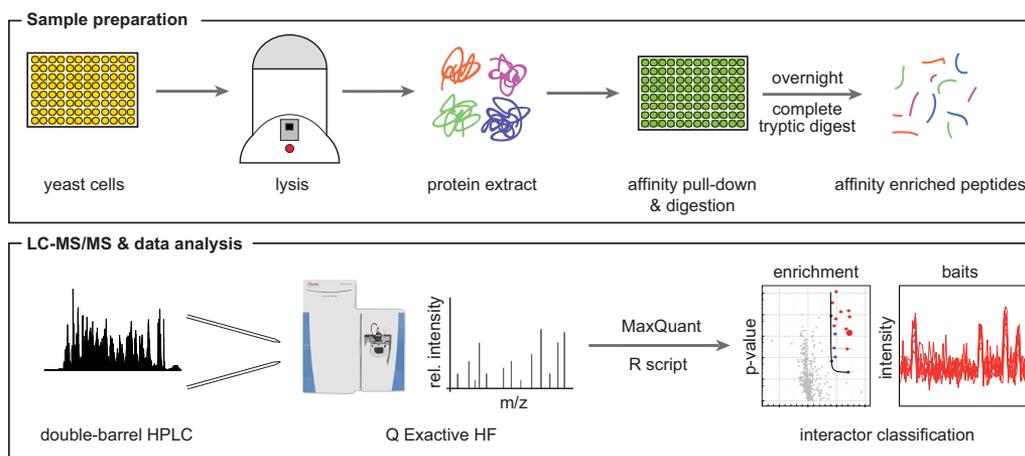


FIG. 3. **Workflow of the high-throughput LC-MS/MS protein interaction analysis pipeline.** Both culturing of yeast cells and affinity purification are performed in 96-well plate format, thus parallelizing sample preparation and minimizing handling errors. LC-MS/MS analysis of 96 pull-down samples in 1 day is achieved through a double-barrel chromatography setup and the increased sequencing speed of the Q Exactive HF mass spectrometer.

these operations can be performed independently for each of the two analytical columns. The intermeasurement time for the double-barrel system was clocked at a maximum of 160 s (Figs. 2E and 2F), which cannot be further reduced on this particular system due to the necessity of refilling the syringe-based pumps and bringing them back up to pressure (Supplemental Fig. S1).

Finally, we modified our standard analytical column heater (33) to accommodate the two analytical columns. The two columns are now pointing sideways toward the mass spectrometer inlet at a fixed angle of 45 degrees at a distance of roughly 2 mm from each other at the tip ends (equaling the width of the heated capillary mounted on Orbitrap platforms; Fig. 2G). As we utilize a fixed setup for the analytical columns, we cannot supply the spray voltage in parallel (Fig. 2H). To shift the voltage between the analytical columns, we additionally developed a high-voltage switch capable of supplying electricity to a single analytical column, controllable through a universal serial bus connection (Supplemental Fig. S2). A plugin module that we developed for the SprayQc environment (45) monitors the current position of the valve W and switches the spray voltage to the eluting analytical column according to a user-definable setting.

A Parallel Workflow for Analyzing 96 Pull-Down Samples within a Single Day—A high-throughput platform should be able to prepare samples in a parallelized format and subsequently measure all of them within a very short time period. Here, we developed an analysis pipeline for pull-down samples that is capable of achieving this goal on pull-down samples (Fig. 3). To facilitate a streamlined workflow necessary for achieving high-throughput processing of pull-down samples, we used GFP-tagged yeast strains originating from the yeast GFP clone collection (43). Further improvements were gained by combining both the cultivation of the yeast and the pull-downs in a 96-well format. Each well yields ~50 million yeast

cells, equal to 500 μ g of protein lysate, which turned out to be sufficient for the pull-down experiments.

Mass Spectrometry Platform Performance on Pull-Down Samples—Using the transcriptional adapter protein ADA2 as a bait, we compared the performance of the Q Exactive HF to that of the LTQ-Orbitrap XL, an instrument introduced about 9 years ago with a sequencing speed of 2 Hz that is frequently used for pull-down analyses. Notably, both instruments were able to identify all known members of the reconstituted ADA2 complex within the commonly used measurement time of 2 h (Supplemental Fig. S3A). This suggests that protein interaction data acquired with older Orbitrap generations over the last 10 years would generally gain little by remeasurement as long as extended LC-MS/MS gradients have been used. However, we note that the protein sequence coverage and, consequently, enrichment of the preys (calculated by dividing the MaxLFQ intensity of the interactors by the median of all MaxLFQ intensities) was somewhat improved with the Q Exactive HF, making the setup slightly more sensitive in detecting interactors (Supplemental Fig. S3B). Clearly, these gradient times are not making effective use of the superior sequencing speed of the Q Exactive HF. By lowering the measurement time to as low as 15 min, the identification performance of the older platform started to suffer while that of the Q Exactive HF still allowed capturing all the expected interactors (Supplemental Fig. S3A). The major difference between the systems was in the sequence coverage per protein, which for the Q Exactive HF remains constant up to 30 min and slightly degrades at 15 min, while it degrades dramatically for the Orbitrap XL (Supplemental Fig. S3C). The decreased sequence coverage negatively impacts the ability to accurately quantify proteins as label-free quantification improves with the number of peptides associated to a given protein (48). This is reflected in the measured enrichment ratios, which for the Orbitrap XL made the bait interactors nearly

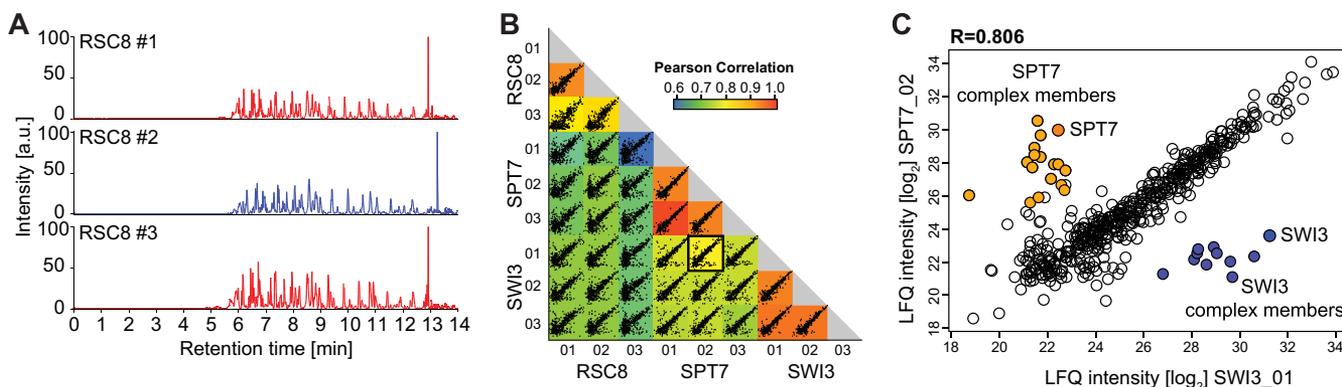


FIG. 4. **Double-barrel chromatography with 14 min gradients on three pull-downs.** (A) Base-peak chromatogram of a biological triplicate RSC8 pull-down run on the double-barrel LC-MS/MS setup. Chromatography in all cases is very reproducible. (B) Comparison of RSC8, SPT7 and SWI3 pull-downs; all measured in triplicates. The matrix of 36 correlation plots reveals high correlations between MaxLFQ intensities within triplicates. (C) Zoom into SPT7_02 versus the SWI3_01 correlation plot. While most proteins were detected with very similar MaxLFQ intensities, the two outlier populations marked in orange (SPT7) and blue (SWI3) represent the different complex members of the distinct protein complexes.

indistinguishable from the background, while for the Q Exactive HF it remained superior even at 30 min when comparing to 2 h (Supplemental Fig. S3B). Overall, as expected, the Q Exactive HF outperformed the Orbitrap XL for all measurement times tested in terms of prey enrichment, sequence coverage, and isotopic features (Supplemental Fig. S3B-S3D). While we observed a decrease in obtained sequence information in the 15 min Q Exactive HF methods, these very short runs still yielded sufficiently high sequence coverage to identify the members of the complex under investigation. In conclusion, these results show that mass spectrometers with relatively low sequencing speed can perform equivalently at long gradients for protein interaction studies, whereas very high sequencing speeds are required for high-throughput identification.

Reproducibility of the Data Acquisition System—To investigate the reproducibility of protein quantification between different measurements, we acquired PPI data for the yeast chromatin remodelers RSC8, SPT7, and SWI3 with our workflow. Visual inspection of the chromatograms for the RSC8 pull-down, measured in triplicates, already shows a high degree of technical reproducibility for the double barrel system with back pressure matched analytical columns (Fig. 4A). In modern PPI experiments, the number of background binders can be in the thousands as opposed to only a few true interactors. We take advantage of these unspecific binders to estimate reproducibility by calculating the correlation between each pair of the measurements where only the generally small number of true interactors degrade the correlation (21). Most of the detected unspecific binders were indeed reproducibly quantified in all three samples. There was one exception with a slightly reduced Pearson correlation coefficient for the RSC8 pull-down (Fig. 4B), for which we concluded based on the large number of imputed values that the enrichment was not completely successful. A small outlier

population observed for each bait protein indeed represented the expected interaction partners (Fig. 4C and Supplemental Fig. S4). Collectively, these results indicate that our double-barrel setup can be operated with very low MS idling time between two independent measurements and achieves high reproducibility at the same time.

PPI Data Quality from Very Short Gradients—To identify preys of a given bait protein, we classified all interactors into four distinct classes essentially as described (21) and improved on that concept by making it completely data driven (EXPERIMENTAL PROCEDURES). Distinction of specific from unspecific binders was achieved by a permutation-based false-discovery rate approach operating on a *t* test and enrichment with two distinct stringencies (EXPERIMENTAL PROCEDURES; Supplemental Fig. S5A). Proteins passing the stringent cutoff represent highly enriched interactors, whereas proteins only passing the less stringent cutoff are characterized as mildly enriched interactors. All other proteins were considered to be unspecific binders. In addition, we used Benjamini–Hochberg-corrected intensity profile correlation of potential interactors compared with the bait protein to minimize false-positive identifications of mildly enriched interactors (EXPERIMENTAL PROCEDURES; Supplemental Figs. S5E and S5F) (21). With these criteria, interactors were grouped into confidence classes A+, A, B+, and B (Supplemental Figs. S4C and S4G). Absolute quantification data from whole yeast proteome experiments (53) allowed us to also estimate interaction and abundance stoichiometries for every protein complex under investigation (Supplemental Fig. S5D).

To assess the quality control of both the LC-MS/MS measurements and the subsequent interactor classification given our large throughput, we employed three distinct layers. The first layer consists of the real-time validation provided by SprayQc (45). Besides the logic for the voltage switch, this software implements automatic warnings via E-mail to the

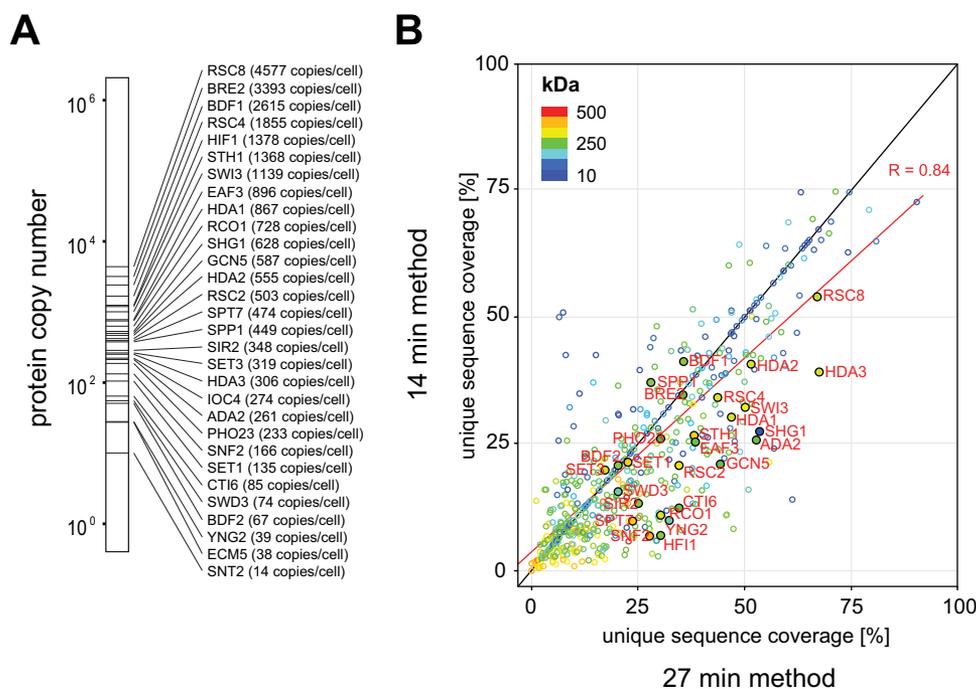


FIG. 5. Bait and prey characteristics comparing 14 versus 27 min gradient methods. (A) The 30 bait proteins selected for the pull-down experiments span several orders of protein expression abundance in *S. cerevisiae*, including several very low abundant proteins (<100 copies per cell). (B) Unique sequence coverage for all identified proteins decreases for the 14 min method compared with the 27 min method. Bait proteins are labeled in red.

operator for a large number of components involved in the measurement and reports meta-data for these components (EXPERIMENTAL PROCEDURES). The second layer consists of verification of the sample preparation and LC-MS/MS measurement success by the number of identified proteins per measurement. Given the preponderance of background proteins, this value should be roughly equal for all pull-downs. The histograms displaying the imputed values provide a simple visual guide in the form of the peaks for the imputed proteins (EXPERIMENTAL PROCEDURES). The third layer is the data-driven determination of what constitutes a successful pull-down experiment. For this, we used the information from the volcano plots, specifically the significance value as described (EXPERIMENTAL PROCEDURES). For all the pull-downs, we combine this value for all the baits to determine a valid range for the baits. Anything falling outside this range is flagged as potentially unreliable.

A Snapshot of the *S. cerevisiae* Chromatin Remodeling Landscape—The data obtained from our very short LC-MS/MS measurements operated with double barrel chromatography demonstrated that AP-MS screens of sufficient quality can be performed in a high-throughput format (Fig. 4). To investigate our workflow on a set of protein complexes involved in a particular biological pathway, we selected 30 distinct bait proteins that are part of the yeast chromatin remodeling landscape. In addition, we also used a GFP-expressing control and the haploid parental strain (EXPERIMENTAL PROCEDURES). Our bait selection spans three or-

ders of expression abundance over the whole yeast proteome (Fig. 5A) and includes several baits with very low abundance (<100 copies per cell). We found that the protein input amount of 500 μ g, which is much lower than that traditionally used, was sufficient to identify the bait proteins and to retrieve known interactors, even for lowest expressed bait proteins (Supplementary Material_14min and Supplementary Material_27min). Additionally, where possible, we selected multiple baits per protein complex in an attempt to characterize the complex as thoroughly as possible. This collection covers 21 distinct protein complexes subdivided into four enzyme classes: histone acetyltransferase, chromatin remodeling, histone methyltransferase, and histone deacetylase complexes. For the 32 distinct yeast strains, we performed pull-down experiments in biological triplicates, resulting in 96 samples. Each of these pull-down samples was measured with both the 14 and the 27 min LC-MS/MS methods, respectively. Together, the interactomes of 96 pull-down samples were measured in either 47.5 h (27 min method) or 26.7 h (14 min method) of start-to-end complete measurement time, including all overhead. As expected, we found that the sequence coverage of bait proteins and specific interactors was reduced for almost every protein in the 14 compared with the 27 min method (Fig. 5B). Nevertheless, the sequence information acquired in the 14 min runs was still sufficient to identify enriched baits and their corresponding preys. We did not experience problems with regard to the bioinformatic enrichment value based on the LFQ intensities, as had been the case for the short gradi-

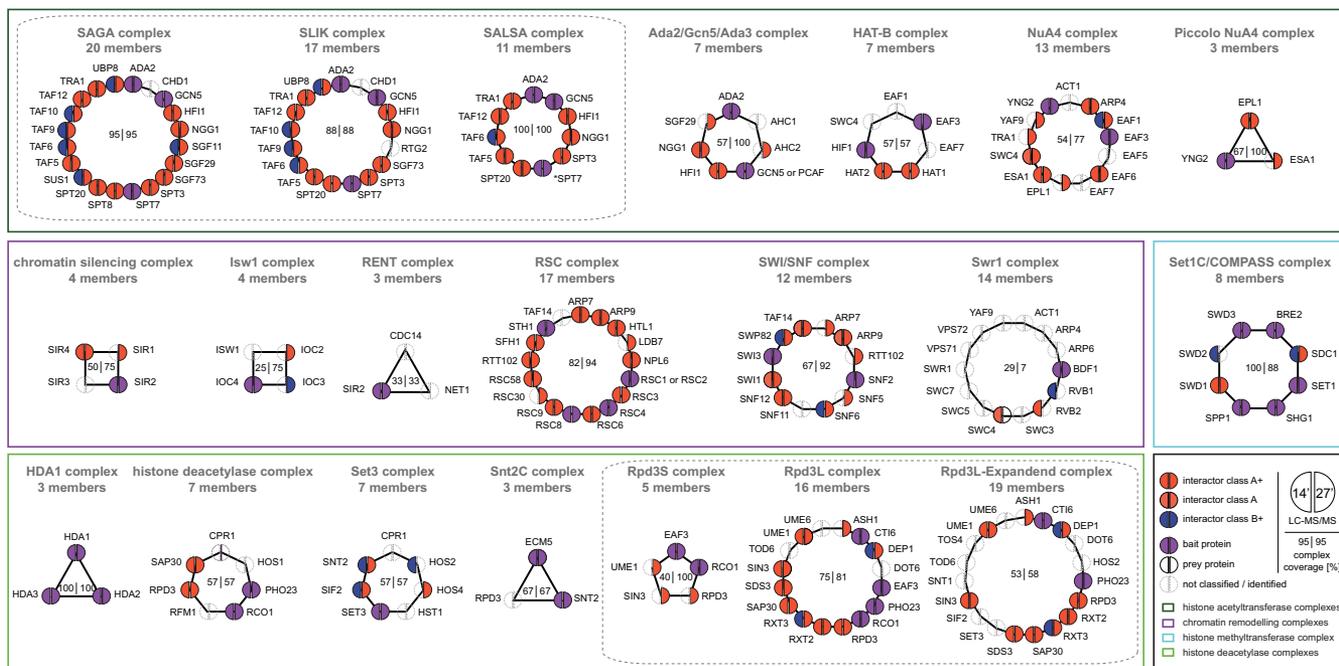


FIG. 6. **Topology network of all interactors.** Global overview of the measured complexes and the success-rate achieved with the 14 versus 27 min gradients. Each protein is depicted as a circle, where the left half corresponds to the 14 min and the right half to the 27 min run results. Color coding refers to the different interactor classes and selected bait proteins. Numbers in the center of each complex represent the percentile coverage of the total complex composition as identified by 14 min (left) or 27 min (right) runs. Colored rectangles group the complexes into their distinct biological functionalities.

ents on the older platforms (see Supplemental Figs. S6 and S7, Supplementary Material_14min and Supplementary Material_27min).

To provide an overview of our identified PPIs, we created a topology network of all interactors assigned to one of the defined prey classifications. While the overall interactor class ranking was slightly reduced, we found only small variations in the final complex coverage even though the LC-MS/MS gradient was nearly halved when comparing the 14 to the 27 min method (Fig. 6). Out of 21 protein complexes analyzed, both run times performed equally well in nine cases, whereas the 27 min outperforms the 14 min in ten cases. Conversely, the 14 min runs were better in two cases. The 27 min method allowed a high retrieval of known interactors even for several very low abundant baits with less than 100 copies per cells. While the 14 min method identified less preys of baits with very low abundance, its superior speed allowed throughput of the same sample set in almost half the time.

Remarkably, we could even validate the presence of two different RSC nucleosome-remodeling complexes. The RSC complex is present in two distinct isoforms with distinct roles in the DNA damage response, as defined by the presence of either RSC1 or RSC2 (54, 55). While performing pull-downs on either RSC4 or RSC8, we identified both RSC1 and RSC2 as interactors, demonstrating that RSC4 and RSC8 are part of both RSC complex isoforms (Supplementary Material_27min). In contrast, pulling down RSC2 only resulted in RSC2 but not RSC1 as complex members. These results demon-

strate that our workflow is capable of identifying distinct complex compositions in a rapid manner.

Discussion and Outlook—In this study, we have described advances for analyzing up to 96 proteomes with lower complexity in about 1 day of LC-MS/MS data acquisition, including all overhead. Our interaction workflow employs parallelized sample generation in a 96-well format together with a modified LC setup and mass spectrometers with very high sequencing speed. With this combination, we demonstrated a severalfold increase in sample processing throughput and sensitivity, as well as in the LC-MS duty cycle.

Including the preceding yeast cultivation and sample preparation steps, processing of 96 pull-down experiments can be achieved within 48 h. However, several 96 samples could be handled in parallel, allowing nesting upstream sample preparation and downstream LC-MS/MS analysis. This in principle would allow a sustained workflow with a capacity of 96 distinct samples per day. The data presented here were acquired through manual sample preparation. However, the majority of sample preparation steps in our workflow only require liquid handling and are thus easily automated using robotic sample preparation systems.

LC-MS/MS data acquisition within 14 min per sample pushes both the LC and MS systems to their current limits. Consequently, the 14 min runs yielded reduced chromatographic quality compared with the 27 min runs. Although this was still sufficient to yield almost the same complex coverage, the 14 min runs did result in lower sequence coverage for both

bait and prey proteins (Fig 5B). This adversely affects analyses and more importantly reduces the enrichment values, making it harder to pinpoint interactors (Fig 2B). Potential optimization could be obtained in an improved experimental design. In this study, we focused on the reproducibility of the complete workflow and chose to perform all steps and measurements in a consecutive series of steps. However, randomizing the measurements, while ensuring that all the replicates of one particular pull-down are always run on the same column, should further improve higher data quality and statistical significance for the interaction determination.

The implementation of double-barrel systems opens up interesting possibilities. On the technological side, it enables automatic detection of a break down in one of the columns due to clogging and reacting to this by using the other column, instead of stopping further analysis. To detect this situation, the software tracks the amount of pressure during the gradient and the flow rate achieved during loading. When the pressure during the gradient or the flow rate during loading exceed critical parameters the system automatically stops operations on this particular column. Further operation is then continued as a single-barrel system. This simple mechanism has the potential to drastically extend the effective up-time and enable almost 24/7 operation of the mass spectrometer. A second technological possibility is the automatic determination of the optimal time for sample loading. The flow rate achieved during loading of the previous sample on the particular analytical column can be used to estimate the required loading time for the current sample. The software then automatically determines the delay required before loading the sample, for instance with a 10 min overhead to ensure that the sample is completely loaded irrespective of fluctuations in the flow rate. This is particularly important for double-barrel-based LC setups as during long gradients it is conceivable that it would be detrimental for the sample to be loaded at the start of the gradient of the other analytical column and then remain at the elevated temperature conditions of the analytical column heater. Third, the described setup could be further extended by using two completely independent UHPLC systems. Even though such a concept is not straightforward to implement on our current system due to software-related issues, the extra redundancy of hardware components would enable troubleshooting of an erroneous UHPLC while the other system maintains measuring. In this way, genuine 24/7 operation of LC-MS/MS data acquisition would be feasible.

Recently, we have reported a high-performance affinity enrichment-mass spectrometry method (21) that uses accurate quantitation of background and unspecific binders for retrieval of true protein complexes. We propose to combine both strategies to allow both the confident retrieval of binding partners and a high throughput. This should be a powerful strategy, especially when a high sequence coverage is not essential (56). Moreover, our results also show that AP-MS can be performed with protein input amounts as low as 500

μg per pull-down and probably much lower in the future, which is considerably less than previously described (21, 42). This increase in sensitivity strongly promotes parallelization and thus throughput efforts. Currently, our pipeline permits a maximum throughput of 96 samples in about 1 day. Employment of other quantification strategies with higher multiplexing, such as TMT labeling for instance, would drastically increase throughput even further.

While we have demonstrated the workflow for protein-protein interactions, our pipeline is generic and can be extended to any kind of protein-based interaction studies in which there is an effective immobilization of the bait material as affinity matrix. We envision other baits such as peptides, DNA, RNA, lipids, or small molecules will greatly facilitate large-scale screening and elucidate drug targets, changes in protein complex formation upon perturbation, and the intertwined relationship between proteins and DNA or RNA.

Finally, the advances described here for the LC-MS/MS part of the workflow can also be extended to the analysis of whole proteomes. For example, biochemical fractionation of whole cell lysates is a routine procedure in mass-spectrometry-based proteomics as it enables much deeper characterization (57, 58). The concomitant increase in LC-MS/MS measurement time caused by the larger number of fractions could be mitigated by using our optimized LC-MS/MS setup. Here, we demonstrated that our very short gradients of 15 min are still able to identify about 700 proteins in a standard HeLa digest (Fig. 1B). If such a complexity is not exceeded, high-throughput analysis can be performed even for fractionated whole proteomes of cell lines, small model organisms, or clinical samples. Finally, given the exponential progress in proteomics related technology, it should only be a matter of time until entire proteomes can be measured in minutes.

Acknowledgments—We thank our colleagues at the Max Planck Institute, especially Marco Hein for help and fruitful discussions; Christian Schmid, Martin Wied, Daniel Vik, and Gabriele Sowa for technical assistance; Jochen Rech for providing the GFP yeast strains; scientists at Thermo Scientific, especially Ole Vorm and Ole Hoerning; and Georg Völkle and Wolfgang Schrader at Sonation GmbH. Last, but not least, we thank Skunkworks for inspiring a highly effective collaborative working model.

* The research leading to these results has received funding from the European Commission's 7th Framework Programme (grant agreement HEALTH-F4-2008-201648/PROSPECTS).

§ To whom correspondence should be addressed: mmann@biochem.mpg.de.

§ This article contains supplemental material Figs. S1 to S9.

|| These authors contributed equally.

¶ Current address: Cellzome GmbH, Molecular Discovery Research, GlaxoSmithKline, Meyerhofstraße 1, D-69117 Heidelberg, Germany.

Data availability: Supplementary data is available with this publication at the MCP web site. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (59) (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository (60) with the dataset identifier PXD001695.

REFERENCES

- Wolters, D. A., Washburn, M. P., and Yates, J. R., 3rd (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **73**, 5683–5690
- Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198–207
- Alteelaar, A. F., Munoz, J., and Heck, A. J. (2013) Next-generation proteomics: Towards an integrative view of proteome dynamics. *Nature Rev. Genetics* **14**, 35–48
- Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Pääbo, S., and Mann, M. (2011) Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7**, 548
- Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Herzog, F., Rinner, O., Ellenberg, J., and Aebersold, R. (2011) The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **7**, 549
- Stumpf, M. P., Thorne, T., de Silva, E., Stewart, R., An, H. J., Lappe, M., and Wiuf, C. (2008) Estimating the size of the human interactome. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6959–6964
- Venkatesan, K., Rual, J. F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K. I., Yildirim, M. A., Simonis, N., Heinzmann, K., Gebreab, F., Sahalie, J. M., Cevik, S., Simon, C., de Smet, A. S., Dann, E., Smolyar, A., Vinayagam, A., Yu, H., Szeto, D., Borick, H., Dricot, A., Klitgord, N., Murray, R. R., Lin, C., Lalowski, M., Timm, J., Rau, K., Boone, C., Braun, P., Cusick, M. E., Roth, F. P., Hill, D. E., Tavernier, J., Wanker, E. E., Barabási, A. L., and Vidal, M. (2009) An empirical framework for binary interactome mapping. *Nature Meth.* **6**, 83–90
- Fields, S., and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245–246
- Gingras, A. C., Gstaiger, M., Raught, B., and Aebersold, R. (2007) Analysis of protein complexes using mass spectrometry. *Nature Rev. Mol. Cell Biol.* **8**, 645–654
- Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A. R., Sassi, H., Nielsen, P. A., Rasmussen, K. J., Andersen, J. R., Johansen, L. E., Hansen, L. H., Jepsersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B. D., Matthiesen, J., Hendrickson, R. C., Gleeson, F., Pawson, T., Moran, M. F., Durocher, D., Mann, M., Hogue, C. W., Figeys, D., and Tyers, M. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183
- Gavin, A. C., Bösch, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Höfert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M. A., Copley, R. R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147
- Butland, G., Peregrin-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., Davey, M., Parkinson, J., Greenblatt, J., and Emili, A. (2005) Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **433**, 531–537
- Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dümpelfeld, B., Edelmann, A., Heurtier, M. A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A. M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J. M., Kuster, B., Bork, P., Russell, R. B., and Superti-Furga, G. (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636
- Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., Punna, T., Peregrin-Alvarez, J. M., Shales, M., Zhang, X., Davey, M., Robinson, M. D., Paccanaro, A., Bray, J. E., Sheung, A., Beattie, B., Richards, D. P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M. M., Vlasblom, J., Wu, S., Orsi, C., Collins, S. R., Chandran, S., Haw, R., Riilstone, J. J., Gandi, K., Thompson, N. J., Musso, G., St Onge, P., Ghanny, S., Lam, M. H., Butland, G., Altaf-Ul, A. M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J. S., Ingles, C. J., Hughes, T. R., Parkinson, J., Gerstein, M., Wodak, S. J., Emili, A., and Greenblatt, J. F. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643
- Ewing, R. M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., McBroom-Cerajewski, L., Robinson, M. D., O'Connor, L., Li, M., Taylor, R., Dharsee, M., Ho, Y., Heilbut, A., Moore, L., Zhang, S., Ornatsky, O., Bukhman, Y. V., Ethier, M., Sheng, Y., Vasilescu, J., Abu-Farha, M., Lambert, J. P., Duewel, H. S., Stewart, II, Kuehl, B., Hogue, K., Colwill, K., Gladwish, K., Muskat, B., Kinach, R., Adams, S. L., Moran, M. F., Morin, G. B., Topaloglou, T., and Figeys, D. (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol. Syst. Biol.* **3**, 89
- Guruharsha, K. G., Rual, J. F., Zhai, B., Mintseris, J., Vaidya, P., Vaidya, N., Beekman, C., Wong, C., Rhee, D. Y., Cenaj, O., McKillip, E., Shah, S., Stapleton, M., Wan, K. H., Yu, C., Parsa, B., Carlson, J. W., Chen, X., Kapadia, B., VijayRaghavan, K., Gygi, S. P., Celniker, S. E., Obar, R. A., and Artavanis-Tsakonas, S. (2011) A protein complex network of *Drosophila melanogaster*. *Cell* **147**, 690–703
- Malovannaya, A., Lanz, R. B., Jung, S. Y., Bulynko, Y., Le, N. T., Chan, D. W., Ding, C., Shi, Y., Yucer, N., Krenciute, G., Kim, B. J., Li, C., Chen, R., Li, W., Wang, Y., O'Malley, B. W., and Qin, J. (2011) Analysis of the human endogenous coregulator complexome. *Cell* **145**, 787–799
- Gavin, A. C., Maeda, K., and Kühner, S. (2011) Recent advances in charting protein-protein interaction: Mass spectrometry-based approaches. *Curr. Opin. Biotechnol.* **22**, 42–49
- Vermeulen, M., Hubner, N. C., and Mann, M. (2008) High confidence determination of specific protein-protein interactions using quantitative mass spectrometry. *Curr. Opin. Biotechnol.* **19**, 331–337
- Paul, F. E., Hosp, F., and Selbach, M. (2011) Analyzing protein-protein interactions by quantitative mass spectrometry. *Methods* **54**, 387–395
- Keilhauer, E. C., Hein, M. Y., and Mann, M. (2014) Accurate protein complex retrieval by affinity enrichment MS rather than affinity purification MS. *Mol. Cell. Proteomics*
- Blagoev, B., Kratchmarova, I., Ong, S. E., Nielsen, M., Foster, L. J., and Mann, M. (2003) A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nature Biotechnol.* **21**, 315–318
- Bantscheff, M., Eberhard, D., Abraham, Y., Bastuck, S., Boesche, M., Hobson, S., Mathieson, T., Perrin, J., Raida, M., Rau, C., Reader, V., Sweetman, G., Bauer, A., Bouwmeester, T., Hopf, C., Kruse, U., Neubauer, G., Ramsden, N., Rick, J., Kuster, B., and Drewes, G. (2007) Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nature Biotechnol.* **25**, 1035–1044
- Rinner, O., Mueller, L. N., Hubálek, M., Muller, M., Gstaiger, M., and Aebersold, R. (2007) An integrated mass spectrometric and computational framework for the analysis of protein interaction networks. *Nature Biotechnol.* **25**, 345–352
- Mousson, F., Kolkman, A., Pijnappel, W. W., Timmers, H. T., and Heck, A. J. (2008) Quantitative proteomics reveals regulation of dynamic components within TATA-binding protein (TBP) transcription complexes. *Mol. Cell. Proteomics* **7**, 845–852
- Selbach, M., Paul, F. E., Brandt, S., Guye, P., Daumke, O., Backert, S., Dehio, C., and Mann, M. (2009) Host cell interactome of tyrosine-phosphorylated bacterial proteins. *Cell Host Microbe* **5**, 397–403
- Vermeulen, M., Eberl, H. C., Matarese, F., Marks, H., Denisov, S., Butter, F., Lee, K. K., Olsen, J. V., Hyman, A. A., Stunnenberg, H. G., and Mann, M. (2010) Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **142**, 967–980
- McAlister, G. C., Huttlin, E. L., Haas, W., Ting, L., Jedrychowski, M. P., Rogers, J. C., Kuhn, K., Pike, I., Grothe, R. A., Blethrow, J. D., and Gygi, S. P. (2012) Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Anal. Chem.* **84**, 7469–7478
- Werner, T., Becher, I., Sweetman, G., Doce, C., Savitski, M. M., and Bantscheff, M. (2012) High-resolution enabled TMT 8-plexing. *Anal. Chem.* **84**, 7188–7194
- Baker, E. S., Livesay, E. A., Orton, D. J., Moore, R. J., Danielson, W. F., 3rd, Prior, D. C., Ibrahim, Y. M., LaMarche, B. L., Mayampurath, A. M., Schepmoes, A. A., Hopkins, D. F., Tang, K., Smith, R. D., and Belov, M. E. (2010) An LC-IMS-MS platform providing increased dynamic range

- for high-throughput proteomic studies. *J. Proteome Res.* **9**, 997–1006
31. Falkenby, L. G., Such-Sanmartin, G., Larsen, M. R., Vorm, O., Bache, N., and Jensen, O. N. (2014) Integrated solid-phase extraction-capillary liquid chromatography (speLC) interfaced to ESI-MS/MS for fast characterization and quantification of protein and proteomes. *J. Proteome Res.* **13**, 6169–6175
 32. Binai, N. A., Marino, F., Soendergaard, P., Bache, N., Mohammed, S., and Heck, A. J. (2015) Rapid analyses of proteomes and interactomes using an integrated solid-phase extraction-liquid chromatography-MS/MS system. *J. Proteome Res.* **14**, 977–985
 33. 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
 34. Scheltema, R. A., Hauschild, J. P., Lange, O., Hornburg, D., Denisov, E., Damoc, E., Kuehn, A., Makarov, A., and Mann, M. (2014) The Q Exactive HF, a benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. *Mol. Cell. Proteomics* **13**, 3698–3708
 35. Shen, Y., Tolić, N., Zhao, R., Pasa-Tolić, L., Li, L., Berger, S. J., Harkewicz, R., Anderson, G. A., Belov, M. E., and Smith, R. D. (2001) High-throughput proteomics using high-efficiency multiple-capillary liquid chromatography with on-line high-performance ESI FTICR mass spectrometry. *Anal. Chem.* **73**, 3011–3021
 36. Belov, M. E., Anderson, G. A., Wingerd, M. A., Udseth, H. R., Tang, K., Prior, D. C., Swanson, K. R., Buschbach, M. A., Strittmatter, E. F., Moore, R. J., and Smith, R. D. (2004) An automated high performance capillary liquid chromatography-Fourier transform ion cyclotron resonance mass spectrometer for high-throughput proteomics. *J. Amer. Soc. Mass Spec.* **15**, 212–232
 37. Bonneil, E., Tessier, S., Carrier, A., and Thibault, P. (2005) Multiplex multi-dimensional nanoLC-MS system for targeted proteomic analyses. *Electrophoresis* **26**, 4575–4589
 38. Livesay, E. A., Tang, K., Taylor, B. K., Buschbach, M. A., Hopkins, D. F., LaMarche, B. L., Zhao, R., Shen, Y., Orton, D. J., Moore, R. J., Kelly, R. T., Udseth, H. R., and Smith, R. D. (2008) Fully automated four-column capillary LC-MS system for maximizing throughput in proteomic analyses. *Anal. Chem.* **80**, 294–302
 39. Orton, D. J., Wall, M. J., and Doucette, A. A. (2013) Dual LC-MS platform for high-throughput proteome analysis. *J. Proteome Res.* **12**, 5963–5970
 40. Behrends, C., Sowa, M. E., Gygi, S. P., and Harper, J. W. (2010) Network organization of the human autophagy system. *Nature* **466**, 68–76
 41. Collins, B. C., Gillet, L. C., Rosenberger, G., Röst, H. L., Vichalkovski, A., Gstaiger, M., and Aebersold, R. (2013) Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14–3-3 system. *Nature Meth.* **10**, 1246–1253
 42. Poulsen, J. W., Madsen, C. T., Young, C., Poulsen, F. M., and Nielsen, M. L. (2013) Using guanidine-hydrochloride for fast and efficient protein digestion and single-step affinity-purification mass spectrometry. *J. Proteome Res.* **12**, 1020–1030
 43. Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., and O’Shea, E. K. (2003) Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691
 44. Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670
 45. Scheltema, R. A., and Mann, M. (2012) SprayQc: a real-time LC-MS/MS quality monitoring system to maximize uptime using off the shelf components. *J. Proteome Res.* **11**, 3458–3466
 46. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnol.* **26**, 1367–1372
 47. 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
 48. Cox, J., Hein, M. Y., Lubner, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526
 49. Ihaka, R., and Gentleman, R. (1996) R: A language for data analysis and graphics. *J. Comput. Graph. Stat.* **5**, 299–314
 50. Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer, New York.
 51. Tusher, V. G., Tibshirani, R., and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5116–5121
 52. Schwänhauser, 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
 53. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014) Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature Meth.* **11**, 319–324
 54. Cairns, B. R., Schlichter, A., Erdjument-Bromage, H., Tempst, P., Kornberg, R. D., and Winston, F. (1999) Two functionally distinct forms of the RSC nucleosome-remodeling complex, containing essential AT hook, BAH, and bromodomains. *Molecular Cell* **4**, 715–723
 55. Chambers, A. L., Brownlee, P. M., Durlley, S. C., Beacham, T., Kent, N. A., and Downs, J. A. (2012) The two different isoforms of the RSC chromatin remodeling complex play distinct roles in DNA damage responses. *PLoS One* **7**, e32016
 56. Ong, S. E., Li, X., Schenone, M., Schreiber, S. L., and Carr, S. A. (2012) Identifying cellular targets of small-molecule probes and drugs with biochemical enrichment and SILAC. *Meth. Mol. Biol.* **803**, 129–140
 57. Yang, F., Shen, Y., Camp, D. G., 2nd, and Smith, R. D. (2012) High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. *Expert Rev. Proteomics* **9**, 129–134
 58. Kelstrup, C. D., Jersie-Christensen, R. R., Bath, T. S., Arrey, T. N., Kuehn, A., Kellmann, M., and Olsen, J. V. (2014) Rapid and deep proteomes by faster sequencing on a benchtop quadrupole ultra-high-field Orbitrap mass spectrometer. *J. Proteome Res.* **13**, 6187–6195
 59. Vizcaíno, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Rios, D., Dianes, J. A., Sun, Z., Farrah, T., Bandeira, N., Binz, P. A., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R. J., Kraus, H. J., Albar, J. P., Martinez-Bartolomé, S., Apweiler, R., Omenn, G. S., Martens, L., Jones, A. R., and Hermjakob, H. (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature Biotechnol.* **32**, 223–226
 60. Vizcaíno, J. A., Côté, R. G., Csordas, A., Dianes, J. A., Fabregat, A., Foster, J. M., Griss, J., Alpi, E., Birim, M., Contell, J., O’Kelly, G., Schoenegger, A., Ovelheiro, D., Perez-Riverol, Y., Reisinger, F., Rios, D., Wang, R., and Hermjakob, H. (2013) The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41**, D1063–1069