analytical chemistry

ARTICLE

pubs.acs.org/ac

Online Nanoflow Reversed Phase-Strong Anion Exchange-Reversed Phase Liquid Chromatography−Tandem Mass Spectrometry Platform for Efficient and In-Depth Proteome Sequence Analysis of Complex Organisms

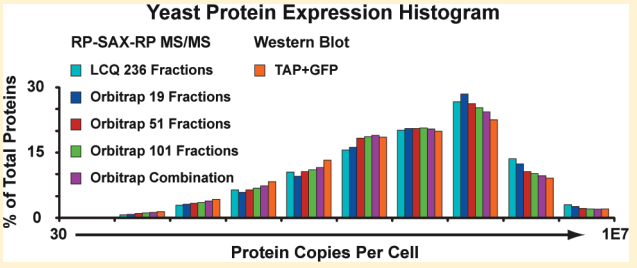
Feng Zhou,[§,‡] Timothy W. Sikorski,[§,‡] Scott B. Ficarro,[§,†,‡] James T. Webber,[§,†] and Jarrod A. Marto*[,†,§,‡]

[§]Department of Cancer Biology and [†]Blais Proteomics Center, Dana-Farber Cancer Institute, Boston, Massachusetts 02215-5450, United States

[‡]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115-6084, United States

Supporting Information

ABSTRACT: The dynamic range of protein expression in complex organisms coupled with the stochastic nature of discovery-driven tandem mass spectrometry (MS/MS) analysis continues to impede comprehensive sequence analysis and often provides only limited information for low-abundance proteins. High-performance fractionation of proteins or peptides prior to mass spectrometry analysis can mitigate these effects, though achieving an optimal combination of automation, reproducibility, separation peak capacity, and sample yield remains a significant challenge. Here we demonstrate an automated nanoflow 3-D liquid chromatography (LC)−MS/MS platform based on high-pH reversed phase (RP), strong anion exchange (SAX), and low-pH reversed phase (RP) separation stages for analysis of complex proteomes. We observed that RP-SAX-RP outperformed RP-RP for analysis of tryptic peptides derived from *Escherichia coli* and enabled identification of proteins present at a level of 50 copies per cell in *Saccharomyces cerevisiae*, corresponding to an estimated detection limit of 500 amol, from 40 μg of total lysate on a low-resolution 3-D ion trap mass spectrometer. A similar study performed on a LTQ-Orbitrap yielded over 4000 unique proteins from 5 μg of total yeast lysate analyzed in a single, 101 fraction RP-SAX-RP LC−MS/MS acquisition, providing an estimated detection limit of 65 amol for proteins expressed at 50 copies per cell.



Mass spectrometry-based proteomics is now well-established for characterization of proteins across a wide range of experimental contexts: recombinant protein expression systems, analysis of proteins resolved via sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE),[1] multicomponent protein complexes isolated through affinity purification,[2] post-translational modifications (phosphorylation,[3,4] glycosylation,[5−8] sulfation,[9] sumoylation,[10,11] and ubiquitination[12,13]), protein subdomains and termini,[14,15] subcellular compartments,[16,17] and finally whole proteomes.[18,19] While advances in instrumentation continue to improve the discovery power of mass spectrometry in biomedical applications,[20−26] it is nevertheless true that data quality often varies inversely with sample complexity. The deleterious effects observed during analysis of peptides present in complex matrixes manifest in diminished performance of both mass spectrometry (MS) and tandem mass spectrometry (MS/MS) scans and include signal suppression of low-basicity peptides[27] and undersampling of low-abundance peptides during MS/MS analysis.[28] In addition, overlapping isotope profiles of coeluting peptides can lead to errors in mass assignment, quantification, and precursor selection.[29,30] Collectively these phenomena diminish the reproducibility of discovery-based experiments and inhibit characterization of low-abundance proteins.[31,32] Biochemical fractionation of subcellular compartments and organelles or affinity purification based on sequence tags or specific post-translational modifications are effective strategies for analyte enrichment and reduction of sample complexity prior to LC−MS/MS analysis.[16,33−37] Although these techniques have proven reliable and robust, the protein and peptide concentration dynamic range of enriched samples often exceeds the analytical capabilities of current mass spectrometry platforms. In these cases, chromatographic prefractionation of proteins or peptides can significantly improve LC−MS/MS analysis. This is perhaps best exemplified by the traditional approach of gel-based protein separation, followed by in-gel digestion, peptide extraction, and LC−MS/MS analysis.[38,39] Although these "Gel-LC" methods have been refined,[18,40,41] they are typically labor-intensive and low throughput. These

Received: March 30, 2011
Accepted: July 21, 2011
Published: August 18, 2011

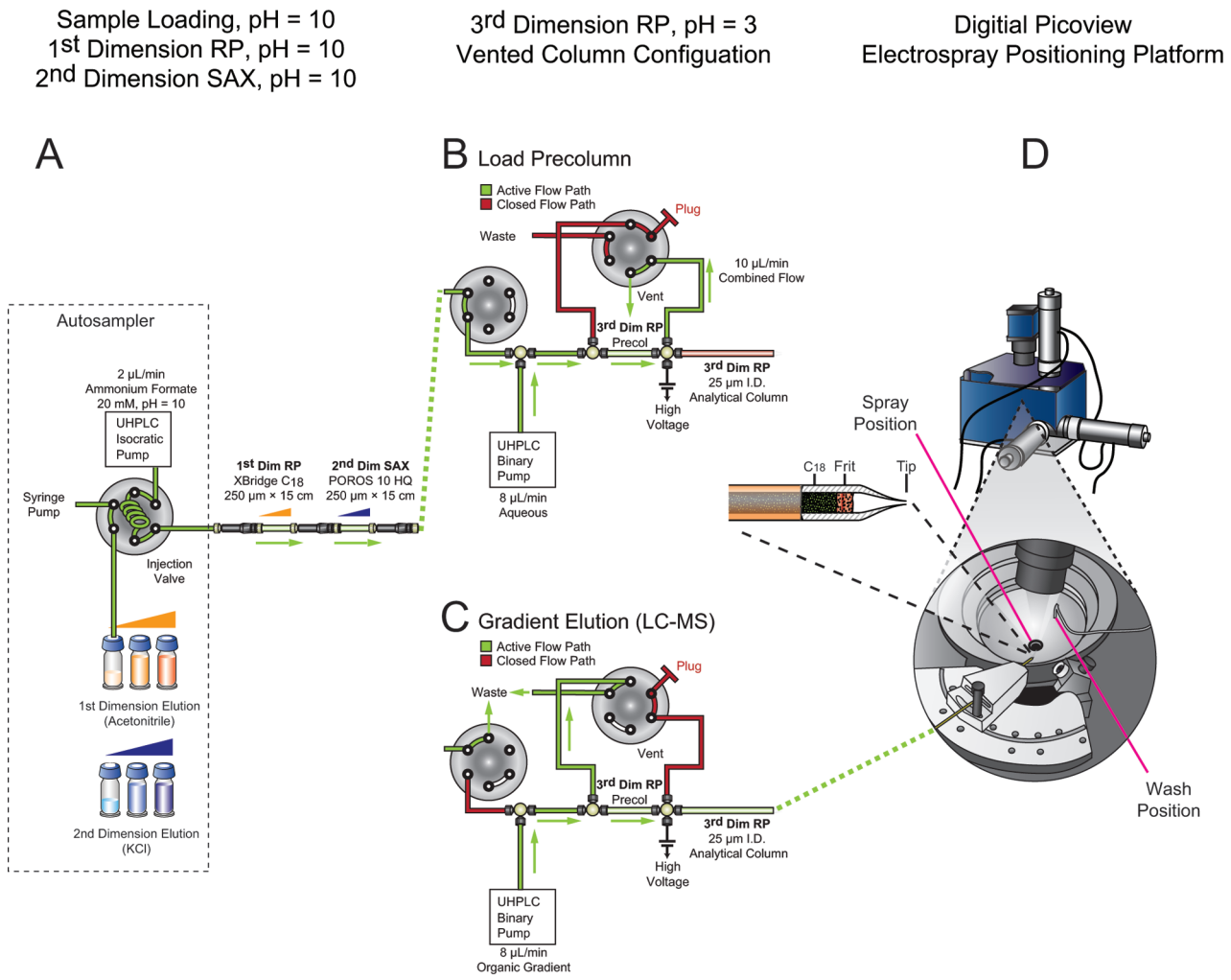


Figure 1. Schematic diagram of automated, online nanoflow RP-SAX-RP platform. (A) The autosampler first loads sample and then is used to inject first- (acetonitrile, orange) or second- (KCl, blue) dimension eluents, respectively. (B, C) An additional six-port, two-position valve provides a vented third dimension column configuration and ensures efficient capture of peptides on the low-pH RP precolumn. (B) Second dimension fractions are diluted (4:1) and acidified with reversed phase solvent A (0.1% formic acid, 3% acetonitrile) introduced by an ultrahigh-pressure binary pump. (C) An organic gradient is delivered by the binary pump to elute peptides from the third dimension column for MS/MS analysis. The vent valve generates a precolumn effluent split of ≈1000:1 to provide a stable column/ESI flow rate of ≈5−10 nL/min. Active solvent flow paths are highlighted in green. (D) A computer-controlled positioning platform (Digital PicoView) automatically moves the emitter tip between "electrospray" and "wash" positions during data acquisition and sample loading, respectively.

limitations catalyzed efforts to implement orthogonal peptide-based fractionation strategies, beginning with the combination of strong cation exchange (SCX) and reversed phase (RP) chromatography, termed "multidimensional protein identification technology" (MudPIT).[42,43] The success of the MudPIT approach has led to a plethora of alternative fractionation geometries, with the overall goal of improving separation power and proteome sequence coverage: offline/online SCX-RP,[44,45] offline immobilized pH gradient gels (IPG) coupled with RP-LC−MS/MS,[46,47] offline solution-based isoelectric focusing (IEF) (offgel) coupled with RP-LC−MS/MS,[48,49] offline high-pH RP coupled with low-pH RP-LC−MS/MS,[45,50−52] offline mixed-mode high pH RP-SAX coupled with low pH RP-LC−MS/MS,[53] and online HILIC-SCX-RP.[54] Generally speaking the offline formats can accommodate large sample quantities (>100 μg) but are susceptible to losses resulting from additional sample handling, lyophilization, and nonspecific adsorption to tube or other surfaces.[45,46] Online systems offer the advantage of efficient peptide recovery and transfer between all separation stages but are typically limited in total sample capacity as compared to larger-scale offline formats. For example, Dowell and colleagues reported that online SCX-RP detected twice the number of peptides as compared to an equivalent offline SCX-RP format.[45] Similarly, work by Slebos et al.[46] suggested that although the separation power of gel-based IEF fractionation of peptides as a first dimension is largely independent of sample quantity, the observed recovery and subsequent identification by LC−MS/MS varied inversely with total starting material. We recently demonstrated an automated, online multidimension fractionation platform that provided for direct comparison of SCX-RP and RP-RP configurations coupled to true nanoflow LC−MS/MS.[55] Although RP-RP exhibited superior analytical figures of merit as compared to SCX-RP under all conditions tested, we observed that fraction-to-fraction overlap of peptides began to increase beyond an analysis depth of 40, first dimension fractions. These results are consistent with fundamental descriptions

of peak capacity for multidimension separations in general[56] and previous reports of RP-RP fractionation in particular.[45,50,51,57] Collectively these observations and results ultimately limit the separation peak capacity and total proteome sequence coverage that can be obtained from 2-D platforms.

Here we expand our automated nanoflow RP-RP-LC−MS/MS platform[55] to include a stage of high pH SAX separation providing an integrated 3-D (RP-SAX-RP) fractionation system. On the basis of the analysis of peptides derived from an *Escherichia coli* standard, we found that RP-SAX-RP provided for significantly improved sequence identification at a fractionation depth beyond which RP-RP previously exhibited diminished peak capacity. With the use of biochemical data from a recent, large-scale protein expression study in *Saccharomyces cerevisiae* as a benchmark, we found that RP-SAX-RP provided for MS/MS-based identification of peptides and proteins across a wide abundance range on both low resolution 3-D ion trap and high performance Orbitrap instruments. Interestingly our data suggest an approximate 100-fold performance difference between previous generation and state-of-the-art mass spectrometers. Moreover, our results indicate that efficient, online fractionation strategies provide large-scale proteome sequence coverage from only a few micrograms of tryptic peptides. More generally, our data provide compelling evidence that improved capillary-based separation systems can augment the rapid pace of improvements in state-of-the-art mass spectrometry platforms.

EXPERIMENTAL SECTION

Because of space considerations, experimental methods related to cell culture, sample preparation, mass spectrometry, and data analysis are provided in the Supporting Information.

Data Availability. Lists of proteins and peptides associated with Figures 2 and 3 and Supplementary Figures S1 and S2 are provided in the Supporting Information. In addition, all native mass spectrometry files and mzResults files which include annotated MS/MS spectra are available for download from http://mzserver.blaisproteomics.org/doi/10.1002/ac200639v.

3-D RP-SAX-RP System. The 3-D RP-SAX-RP platform (Figure 1) consisted of Waters UHPLC binary and isocratic pumps, an autosampler (Waters Corp., Milford, MA), and an additional six-port, two-position valve (Valco Inc., Austin, TX). The first dimension reversed phase (RP) column consisted of a 250 μm i.d. capillary packed with 15 cm of 5 μm diameter XBridge C18 resin (Waters Corp., Milford, MA). An anion exchange (SAX) column (250 μm i.d. × 15 cm of 10 μm diameter POROS10HQ [AB Sciex, Foster City, CA] resin) was connected directly to the outlet of the first dimension RP column with a union. The third dimension was constructed as described previously.[58,59] The isocratic pump delivered either peptide samples or first (acetonitrile in 20 mM ammonium formate, pH 10) and second (KCl in 20 mM ammonium formate, pH 10) dimension eluents at 2 μL/min through the sample loop. Discrete eluent concentrations used for all experiments are provided in Supplementary Table S1 in the Supporting Information. The binary pump delivered 0.1% formic acid at 8 μL/min to dilute the organic content and acidify the first/second dimension effluent prior to the final dimension precolumn or provided for gradient elution (2−30% B in 45 min, A = 0.2 M acetic acid, B = acetonitrile with 0.2 M acetic acid) of peptides from the third dimension reversed phase columns for LC−MS/MS analysis at a flow rate of ≈10 nL/min.[60] A Digital PicoView electrospray source platform (New Objective, Woburn, MA; model DPV-550 for the Orbitrap XL) was used to automatically position the emitter tip at the heated metal capillary inlet during LC−MS/MS acquisition or beneath a gravity-driven drip station during injection of peptide samples or first/second dimension eluents.

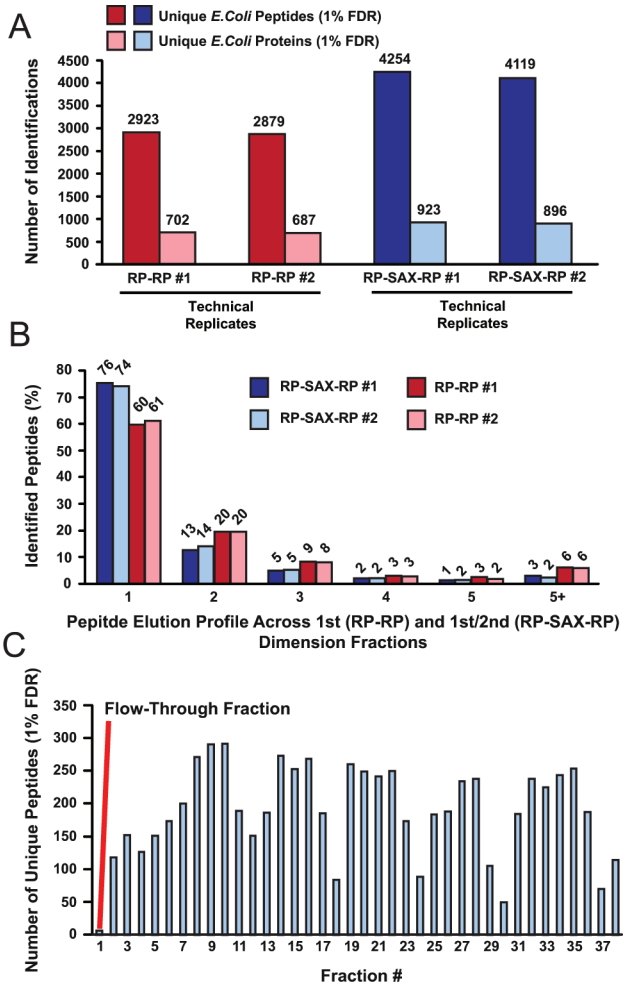


Figure 2. (A) Replicate analyses of tryptic peptides derived from *E. coli* lysate by RP-RP (left, red) and RP-SAX-RP (right, blue) fractionation on a LCQ Deca mass spectrometer. RP-SAX-RP provided for higher numbers of identified peptides and proteins as compared to RP-RP for independent experiments spanning 37 and 40 fractions, respectively. (B) Peptide elution profiles, defined as the number of first (high-pH RP) or first and second (high-pH RP and SAX) fractions spanned by peptides detected in part A. (C) Distribution of unique peptide identifications in a 37 fraction RP-SAX-RP experiment.

Safety Considerations. Safety glasses should be worn at all times during construction and use of fused silica based capillaries. In addition, a lab coat and gloves should be worn when handling other volatile organics in a chemical fume hood.

RESULTS AND DISCUSSION

We recently performed a detailed comparison of automated, online 2-D RP-RP and SCX-RP fractionation platforms for analysis of peptides derived from complex cell lysates and affinity purified multicomponent protein complexes.[55] Given the high degree of orthogonality that we observed between reversed phase separations performed at high- and low-pH, respectively, we

Table 1. Detection of Low-Expression Level Proteins[a]

low copy number proteins (<128 copies/cell)	detected in 19 fraction RP-SAX-RP (Orbitrap)	detected in 51 fraction RP-SAX-RP (Orbitrap)	detected in 101 fraction RP-SAX-RP (Orbitrap)	detected in 236 fraction RP-SAX-RP (LCQ)
YLL040C	Y	Y	Y	Y
YNL014W	N	Y	Y	Y
YML109W	N	Y	Y	Y
YIL092W	N	Y	Y	N
YIL084C	Y	Y	Y	N
YKL145W	Y	Y	Y	Y
YKL075C	N	Y	Y	N
YIL002C	Y	Y	N	N
YHR015W	N	N	N	N
YPL008W	N	N	Y	N
YGL006W	N	Y	Y	Y
YKR031C	Y	Y	N	Y
YLR035C	N	N	N	N
YNR067C	Y	Y	Y	Y
YOR093C	Y	Y	Y	Y
success rate	7/15	12/15	11/15	8/15

[a] Each "Y" entry provides an embedded link to an annotated MS/MS spectrum in the online manuscript or accessed directly at: http://mzserver.blaisproteomics.org/doi/10.1002/ac200639v/table1.html.

asked whether addition of a third dimension would provide significantly improved fractionation power. On the basis of the buffer conditions used in the first dimension of our RP-RP platform (20 mM ammonium formate, pH 10) we chose strong anion exchange as the second dimension in a (now) 3-D RP-SAX-RP configuration (Figure 1). Briefly, peptides are loaded through the autosampler (Figure 1, dashed box) in ammonium formate buffer (pH 10) and captured on the first dimension RP column. Peptides poorly retained on reversed phase at high pH are trapped on either the second (SAX, pH 10) or third (RP, pH 3.0) dimension columns. The low pH pre- and analytical columns constitute the third dimension separation and are assembled in a vented[58−60] configuration to facilitate loading of samples or first (acetonitrile, 20 mM ammonium formate, pH 10) and second (KCl in 20 mM ammonium formate, pH 10) dimension eluents, respectively. The system provides complete, online fractionation, meaning that once peptides are injected from the autosampler well onto the first dimension column, they are automatically transferred to the second and third dimensions and are only exposed to the fused silica tubing, PEEK LC fittings, and flow paths of the six-port valves.

Our previous analytical characterization studies[55] suggested that for LC−MS/MS analysis of *E. coli* peptides in a complex background of whole cell lysate, RP-RP performance appeared to peak at a depth of between 30 and 40 fractions. On the basis of this observation we asked whether RP-SAX-RP would provide for improved performance at a fractionation depth beyond which RP-RP previously exhibited diminishing returns. Toward this end we performed replicate RP-RP (40 fractions) and RP-SAX-RP (37 fractions) analyses of *E. coli* tryptic peptides. Figure 2A shows that while each fractionation technique provided reproducible data, the RP-SAX-RP platform identified 44% and 31% more peptides and proteins, respectively, as compared to RP-RP. As a surrogate analysis for separation peak capacity, we next asked whether the number of fractions across which discrete peptides were identified varied between the two separation platforms. Indeed, we observed that 75% of all peptide identifications spanned a single RP-SAX-RP fraction, while in the case of RP-RP (Figure 2B) this figure of merit was reduced to 60%. Consistent with these data, we observed that 34 of 37 RP-SAX-RP fractions contained between 100 and 300 unique peptide identifications (Figure 2C), suggesting that this 3-D platform provided for uniform fractionation of *E. coli* peptides across the entire separation space.

Given the results above, we next asked whether RP-SAX-RP could provide sufficient fractionation depth to identify proteins expressed across a wide concentration dynamic range in the context of a shotgun, data-dependent type acquisition. In order to facilitate comparison of our results with other large-scale proteomic efforts, we switched to yeast as our model system. *S. cerevisae* has a well-annotated genome and has been extensively studied at the whole proteome level using biochemical[61] as well as shotgun[18,43,44] and targeted[62] mass spectrometry-based approaches. With these studies as a reference point, we next sought to establish an extreme limit for automated, online RP-SAX-RP fractionation. Toward this end we loaded 40 μg of peptides derived from yeast whole cell lysate and acquired 236 LC−MS/MS fractions on a LCQ Deca XP ion trap mass spectrometer (Supplementary Figure S1 in the Supporting Information). This experiment required nearly 18 days of uninterrupted acquisition time and yielded 18 359 unique peptides (1% false discovery rate (FDR)) that mapped to 2902 proteins (see the Supplementary Experimental Section in the Supporting Information for details on protein inference). Although these data are based on somewhat dated mass spectrometry technology, it is nonetheless interesting to note that our results compare very favorably to a previous study that utilized 80 fractions of offline SCX fractionation coupled to LC−MS/MS on the same mass spectrometry platform (1504 proteins identified from 1 mg of protein lysate[44]). This comparison suggests that continued improvements in fractionation can compensate for limitations in the stochastic nature of data-dependent MS/MS acquisition. Consistent with

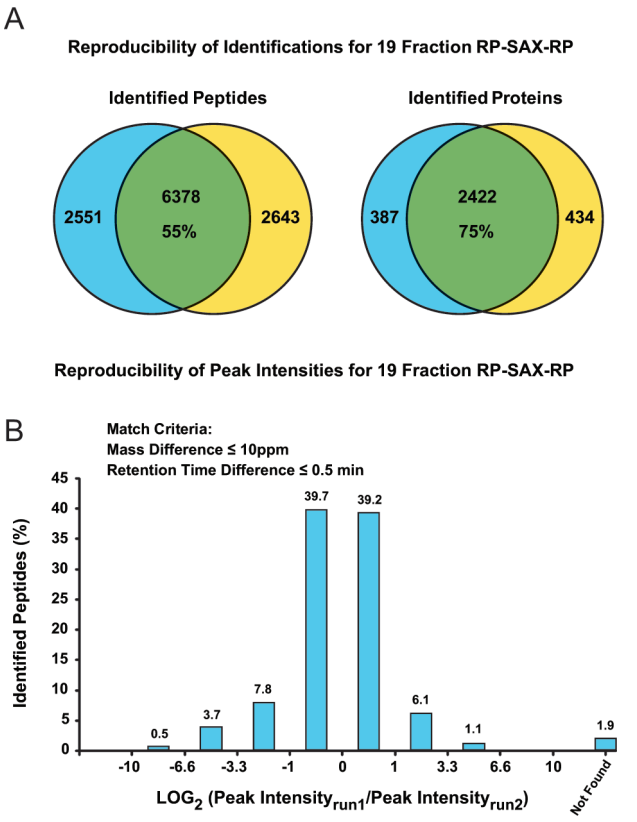


Figure 3. Technical replicate 19 fraction RP-SAX-RP analyses of *S. cerevisiae* tryptic peptides on a LTQ-Orbitrap mass spectrometer. (A) Venn diagrams illustrate the reproducibility of peptide (≈55%) and protein (≈75%) identifications. (B) Histogram distribution of $\log_2$ ratios for peptide precursors matched by unique amino acid sequence or a combination of intact peptide mass (±10 ppm) and third dimension LC elution time (±0.5 min).

this hypothesis, we observed a surprisingly uniform distribution of unique peptide identifications across all 236 RP-SAX-RP fractions.

Several proteins identified in this analysis have been previously characterized with expression levels at or below 128 copies per cell (Table 1), including YOR093C (41 copies/cell), YBL063W (49 copies/cell), and YJL084C (57 copies/cell).[61] For comparison, Table 1 also lists low abundance proteins (≤128 copies per cell) recently characterized by triple quadrupole-based MRM assays.[62] Interestingly we identified 8 out of these 15 proteins in our 236 fraction RP-SAX-RP analysis, suggesting that efficient fractionation can offset, in-part, the discrepancies in detection between discovery- and targeted-mode MS/MS analysis. On the basis of these data, we estimate a detection limit of 500 amol (≈50 protein copies per cell, 7×10^6 cells analyzed) and a dynamic range of 3.2×10^4 (YKL096W-a 1.59×10^6 copies/cell, 7×10^6 cells analyzed) for data-dependent analysis of total tryptic peptides on a 3-D ion trap instrument.

As an extension of the studies above, we next coupled our RP-SAX-RP separation system to a hybrid linear ion trap orbitrap (Orbitrap XL) mass spectrometer. As we expected significantly improved performance as compared to the 3-D ion trap instrument, we first reduced the sample quantity by nearly an order of magnitude (5 μg versus 40 μg) and then sought to establish the number of RP-SAX-RP fractions required to obtain yeast proteome coverage roughly equivalent to that observed with the LCQ platform (236 fractions, 18 359 unique peptides, 2 902 proteins). We found that 19 fractions yielded 2 809 protein identifications based on 8 929 unique peptide sequences (Figure 3A, blue). Acquisition of these data required approximately 2 days of LC−MS/MS time. We next performed a second 19 fraction RP-SAX-RP analysis (Figure 3A, yellow) to establish a baseline for reproducibility of our separation platform. Here we identified 2 856 proteins from 9 021 peptides. Across both analyses we identified 2 422 proteins in common (≈75% reproducibility) and 3 243 proteins in total. We also observed good agreement between signals of individual peptides identified in both runs (Figure 3B), with ≈80% of all intensity ratios within ±2-fold of the mean value. We next tested whether the peptide content of individual fractions reproduced across these replicate runs by first requiring that peptide sequences were identified in the same fraction across both analyses; under these conditions, we observed that ≈48% of all peptides were reproducibly identified across discrete fractions in these two RP-SAX-RP analyses. However, the stochastic nature of MS/MS likely makes this result the lower bound of "reproducibility." Hence we next used a more permissive approach which included as "reproduced" those MS-level precursor peaks that corresponded in mass (±10 ppm) and (third dimension) retention time (±0.5 min) to peptide sequences identified in only one of the two analyses. On the basis of these criteria we found that potentially 98% of all identified peptides were reproducibly detected.

We next asked whether additional fractionation depth would provide higher proteome sequence coverage. In independent experiments we loaded 5 μg of yeast tryptic peptides and acquired MS/MS data from 19, 51, and 101 RP-SAX-RP fractions, respectively. Encouragingly, at each fractionation depth we obtained data that constituted a significant superset of the preceding experiment, with overlaps of at least 70% and 92% at the peptide (Supplementary Figure S2A in the Supporting Information) and protein (Supplementary Figure 2SB in the Supporting Information) level, respectively. Moreover we observed that at least 62% of unique peptide identifications were confined to a single RP-SAX-RP fraction, even for the most in-depth analysis (101 fractions, Supplementary Figure S2C in the Supporting Information). As with the experiments performed on the low resolution instrument (Supplementary Figure S1 in the Supporting Information), we readily detected low abundance proteins (≈50 copies per cell) at each fractionation depth, albeit at total acquisition times (e.g., number of fractions) far less than were required on the 3-D ion trap. On the basis of the sample quantity consumed per analysis (5 μg), we estimate a detection limit of 65 amol and a dynamic range of 3.2×10^4 (YKL096W-a 1.59×10^6 copies/cell, 8×10^5 cells analyzed) for data-dependent analysis of total yeast tryptic peptides on a LTQ-Orbitrap XL. Collectively these results suggest that our 3-D RP-SAX-RP platform provides for reproducible separation and sufficient peak capacity to enable identification of low-abundance proteins in complex mixtures on low- and high-performance mass spectrometers.

For the sample quantity analyzed on the LTQ-Orbitrap (5 μg), it appeared that our 3-D separation platform reached a point of diminishing returns between 50 and 100 fractions, as we only identified an additional 184 unique proteins (1377 unique peptides) by extending the analysis to 101 fractions. To determine whether additional fractionation primarily impacted protein sequence coverage (e.g., more peptides identified per protein) or protein dynamic range (e.g., identification of proteins at low expression levels), we next plotted our fractionation data from Supplementary Figures S1 and S2 in the Supporting Information

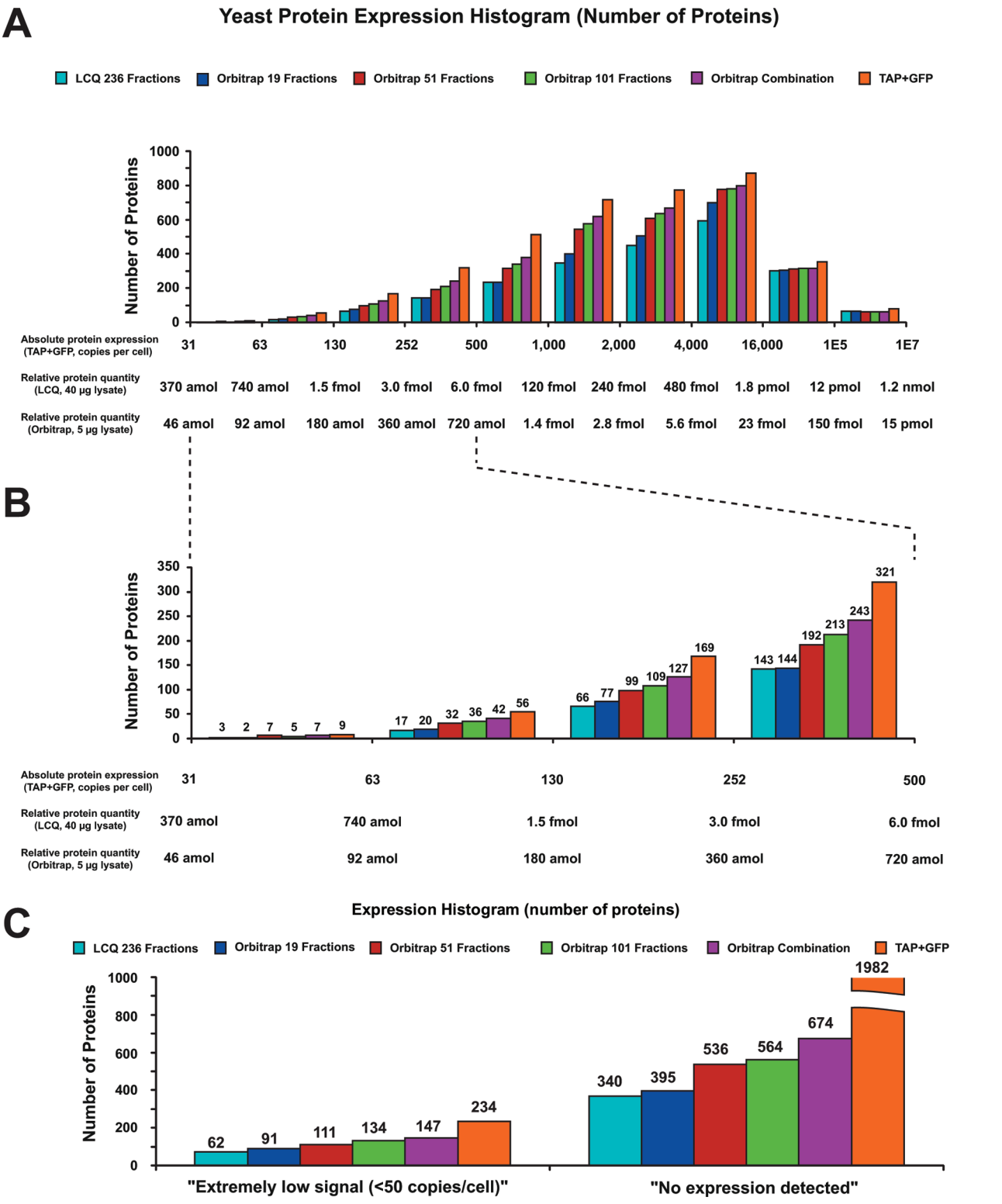


Figure 4. (A, B) Proteome coverage for *S. cerevisiae* as a function fractionation depth (19, 51, 101, 236 fractions) and mass spectrometry platform (LCQ Deca or LTQ-Orbitrap XL) as compared to biochemical-based quantification of protein expression.[61] Western blot data (orange) are used as a reference set within each expression level bin. The count of reference proteins detected in each RP-SAX-RP experiment is plotted with the color scheme as indicated in the legend. The *x*-axis is labeled based on absolute expression level (copies per cell as determined from TAP-GFP data[61]) or relative protein quantity calculated from absolute expression level[61] and total lysate analyzed by each RP-SAX-RP experiment (LCQ, 40 μg; LTQ-Orbitrap, 5 μg). (C) Expression histograms plotted as in parts A and B for protein groups from Ghememaghami et al. "extremely low signal (<50 copies/cell)" and "no expression detected,"[61] not amenable to absolute quantification by Western blot.

dx.doi.org/10.1021/ac200639v |*Anal. Chem.* 2011, 83, 6996–7005

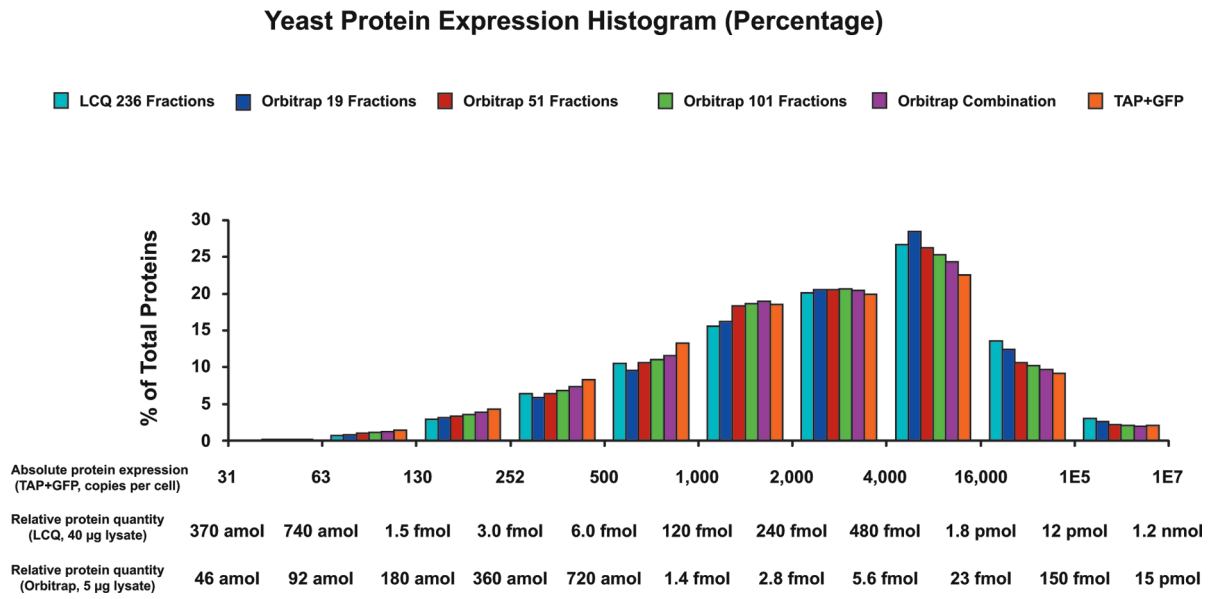


Figure 5. Expression histogram as in Figure 4A, plotted as a percentage of the total protein count detected in each experiment. TAP-GFP data (orange) serves as the reference protein set in each expression bin. The *x*-axis is labeled based on absolute expression level (copies per cell as determined from TAP-GFP data[61]) or relative protein quantity calculated from absolute expression level and total lysate analyzed by each RP-SAX-RP experiment (LCQ, 40 μg; LTQ-Orbitrap, 5 μg).

along with that from a recent large-scale biochemical analysis[61] of protein expression in yeast (Figure 4). We also included the combined set of unique protein identifications resulting from our 19, 51, and 101 fraction experiments. For each expression level bin, we used the quantitative Western blot data as a reference and then plotted the subset of these proteins identified in our RP-SAX-RP analyses. Between a range of ≈60 and ≈4 000 protein copies per cell, we observed a consistent trend in which the number of protein identifications increased systematically as a function of mass spectrometry technology (LCQ vs LTQ-Orbitrap) and fractionation depth (19, 51, and 101 fractions or the combination thereof). However, at the lowest expression level (Figure 4B, left, ≈30−60 copies per cell), the mass spectrometry data appeared somewhat stochastic, suggesting that either this range represented the practical detection limit under the conditions tested, and hence fractionation provided little benefit, or the number of proteins (n = 9) quantified by Western blot was too small to serve as a reference. To explore this question further we next plotted two other expression categories from Ghememaghami et al.,[61] "extremely low signal (<50 copies per cell)" and "no expression detected," along with the corresponding proteins detected in our RP-SAX-RP analyses (Figure 4C). With larger sets of reference proteins (234 and 1982 for each category, respectively) we again observed a systematic increase in the number of identified proteins as a function of mass spectrometry technology and fractionation depth. Interestingly, we observed that the number of protein identifications more than doubled (from 62 to 134) for the "extremely low signal (<50 copies per cell)" category moving from the LCQ (236 fractions) to the LTQ-Orbitrap (101 fractions).

Finally in order to identify overall bias in protein identification as a function of expression level, we replotted the data in Figure 4A as a relative percentage of all proteins identified by each approach (Figure 5). Not surprisingly, both the LCQ (236 fractions) and Orbitrap (19 fractions) data exhibited a bias toward highly expressed proteins. However the combined data set (Orbitrap 19, 51, and 101 fractions) correlated very well with the Western blot data across the majority of expression bins. Collectively our results suggest that increased fractionation provided for improved proteome coverage across the entire range of protein expression in yeast.

While cross-lab comparisons based on protein lists are admittedly wrought with caveats, it is nonetheless informative to compare our results with other analytically focused studies designed to maximize the discovery capacity of LC−MS/MS. For example, one recent report identified 3313 unique proteins in the context of a triplicate, 12 fraction offline SCX-RP experiments from 2 mg of yeast lysate.[19] Similarly a combination of chromatographic and SDS-PAGE protein fractionation, followed by LC−MS/MS (offline 3-D), generated 75 fractions in total and identified 3639 proteins cumulatively across triplicate analyses from 1.5 mg of cell lysate.[18] Finally, a triplicate, 24 fraction experiment based on offline isoelectric focusing peptide fractionation (OFFGEL) followed by LC−MS/MS (offline IEF-RP) identified 3987 proteins from 0.3 mg of cell lysate.[18] When compared to our data, in which we detected over 4000 unique yeast proteins from only 10% of equivalent input material, these results strongly suggest that high-performance fractionation can significantly augment the performance capabilities of state-of-the-art mass spectrometry instrumentation. This observation is further corroborated by the data in Table 1, which compares the ability of our RP-SAX-RP platform to detect a set of low-abundance proteins (<128 copies per cells) that were recently quantified by targeted MRM-type mass spectrometry assays. Cumulatively across our three analyses on the Orbitrap (19, 51, and 101 fractions), we detected 12 out of the 15 proteins, from approximately 2% of the input material as compared to the work of Picotti et al.[62]

Although the above comparisons provide useful insight, a recent, multilab report[64] illustrated the utility of considering numerous metrics when evaluating discovery-based proteomic methods. For example, it is informative to consider the overall analysis efficiency or rate of data production along with the total number of peptide and protein identifications. Toward this end, Table 2

Table 2. Performance Metrics for 2-D (RP-RP) and 3-D (RP-SAX-RP) LC−MS/MS Platforms

organism	platform	instrument	no. fractions	time (h)	μg	unique peptides	unique proteins	peptides/min	proteins/min
S. cerevisiae	RP-SAX-RP	Orbitrap	1 fraction	2	5	2 125	981	17.71	8.18
S. cerevisiae	RP-SAX-RP	Orbitrap	19 fractions	38	5	14 576	3 111	6.39	1.36
S. cerevisiae	RP-SAX-RP	Orbitrap	51 fractions	102	5	25 091	3 821	4.10	0.62
S. cerevisiae	RP-SAX-RP	Orbitrap	101 fractions	202	5	26 468	4 004	2.18	0.33
S. cerevisiae	RP-SAX-RP	LCQ	236 fractions	572	40	18 359	2 902	0.45	0.08
E. coli	RP-SAX-RP	LCQ	37 fractions	74	20	4 254	923	0.96	0.21
E. coli	RP-RP	LCQ	40 fractions	80	20	2 923	702	0.61	0.15

provides detection efficiencies for the 2-D/3-D experiments on the Orbitrap (1, 19, 51, and 101 fractions) and the LCQ (37, 40, and 236 fractions) instruments, respectively. These data demonstrate the inevitable compromise between experiment time and the depth of proteome coverage but nonetheless provide useful benchmarks for evaluation of overall system performance. It is also worth noting that while the capillary format and online geometry of our nanoflow RP-SAX-RP fractionation platform provides for fully automated LC−MS/MS analyses, there is a practical trade-off in terms of limited sample loading capacity. While it is true that offline configurations provide for analysis of larger analyte quantities, recent reports have noted that sample handling, lyophilization, and nonspecific adsorption of peptides to tube surfaces can often reduce sample recovery in these schemes. In addition, these deleterious effects tend to scale inversely with total input, meaning that the performance for offline systems can be limited for sample quantities below ≈50−100 μg.[45,46] In contrast, we recently demonstrated that electrospray ionization efficiency compensates for chromatographic peak broadening at effluent flow rates below empirical Van Deemter minima,[60] and suggested that a renewed focus on small diameter (≤25 μm i.d.) LC−MS assemblies operated at low nanoliter/minute flow rates would provide significantly improved analytical figures of merit and compliment the continued performance improvements available on state-of-the-art mass spectrometry platforms. Here we successfully coupled our nanoflow 1-D columns with capillary-format high-pH RP and SAX to provide an automated, 3-D separation system for analysis of complex proteomes. The capacity of our RP-SAX-RP platform to provide improved detection and dynamic range is illustrated in Figure 4 in which we observed a systematic improvement in detection of low-abundance proteins as a function of mass spectrometry technology and fractionation depth. Consistent with our experience using 25 μm i.d. analytical columns with integrated electrospray emitters in a 1-D nanoflow configuration[60] or coupled with high-pH RP in an automated RP-RP configuration,[55] our RP-SAX-RP platform has proven to be very robust, providing uninterrupted operation for more than 100 injections spanning several months of analysis time without a major system failure; in fact the majority of data described herein were acquired using one column set over a span of ≈3 months.

CONCLUSIONS

Despite the growing use of mass spectrometry in biomedical research, limited dynamic range continues to hinder the characterization of low-abundance proteins present in complex biological matrixes. Multidimensional peptide fractionation[42,43] improves data quality and facilitates analysis of proteomic results in the context of hypotheses of biological function, although achieving an optimal combination of analytical figures of merit remains a significant challenge. Herein we build upon our recent work in high-performance 1-D[60] and 2-D[55] separations and describe an automated platform for 3-D RP-SAX-RP fractionation of tryptic peptides. Our observations are thematically aligned with other recent efforts[65−71] and provide compelling evidence that capillary-based fractionation formats offer a powerful combination of automation, reproducibility, separation peak capacity, and sample yield.

ASSOCIATED CONTENT

Supporting Information. Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Address: Department of Cancer Biology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Smith 1158A, Boston, MA, 02215-5450. Phone: (617) 632-3150 (office). Fax: (617) 582-7737. E-mail: jarrod_marto@dfci.harvard.edu.

ACKNOWLEDGMENT

The authors thank Eric Smith for assistance with the figures. In addition the authors thank Keith Fadgen, Jim Murphy, and Geoff Gerhardt of Waters Corp. for valuable discussions and technical assistance. Generous support for this work was provided by the Dana-Farber Cancer Institute and the National Institutes of Health, NHGRI (Grant P50HG004233) and NINDS (Grant P01NS047572).

REFERENCES

(1) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal. Chem.* **1996**, *68*, 850–858.
(2) Gingras, A. C.; Gstaiger, M.; Raught, B.; Aebersold, R. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 645–654.
(3) Choudhary, C.; Mann, M. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 427–439.
(4) Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. *Cell* **2006**, *127*, 635–648.
(5) Patwa, T.; Li, C.; Simeone, D. M.; Lubman, D. M. *Mass Spectrom. Rev.* **2010**, *29*, 830–844.
(6) Kim, Y. J.; Freas, A.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 1544–1548.
(7) Madera, M.; Mechref, Y.; Klouckova, I.; Novotny, M. V. *J. Proteome Res.* **2006**, *5*, 2348–2363.
(8) Yang, Z. P.; Harris, L. E.; Palmer-Toy, D. E.; Hancock, W. S. *Clin. Chem.* **2006**, *52*, 1897–1905.

(9) Yu, Y. H.; Hoffhines, A. J.; Moore, K. L.; Leary, J. A. *Nat. Methods* **2007**, *4*, 583–588.
(10) Miller, M. J.; Barrett-Wilt, G. A.; Hua, Z. H.; Vierstra, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 16512–16517.
(11) Blomster, H. A.; Imanishi, S. Y.; Siimes, J.; Kastu, J.; Morrice, N. A.; Eriksson, J. E.; Sistonen, L. *J. Biol. Chem.* **2010**, *285*, 19324–19329.
(12) Xu, P.; Peng, J. M. *Biochim. Biophys. Acta: Proteins Proteomics* **2006**, *1764*, 1940–1947.
(13) Peng, J. M.; Schwartz, D.; Elias, J. E.; Thoreen, C. C.; Cheng, D. M.; Marsischky, G.; Roelofs, J.; Finley, D.; Gygi, S. P. *Nat. Biotechnol.* **2003**, *21*, 921–926.
(14) Schilling, O.; Barre, O.; Huesgen, P. F.; Overall, C. M. *Nat. Methods* **2010**, *7*, 508–U33.
(15) Kleifeld, O.; Doucet, A.; Keller, U. A. D.; Prudova, A.; Schilling, O.; Kainthan, R. K.; Starr, A. E.; Foster, L. J.; Kizhakkedathu, J. N.; Overall, C. M. *Nat. Biotechnol.* **2010**, *28*, 281-U144.
(16) Wiederhold, E.; Veenhoff, L. M.; Poolman, B.; Slotboom, D. J. *Mol. Cell. Proteomics* **2010**, *9*, 431–445.
(17) Andersen, J. S.; Lam, Y. W.; Leung, A. K. L.; Ong, S. E.; Lyon, C. E.; Lamond, A. I.; Mann, M. *Nature* **2005**, *433*, 77–83.
(18) de Godoy, L. M. F.; Olsen, J. V.; Cox, J.; Nielsen, M. L.; Hubner, N. C.; Frohlich, F.; Walther, T. C.; Mann, M. *Nature* **2008**, *455*, 1251-U60.
(19) Swaney, D. L.; Wenger, C. D.; Coon, J. J. *J. Proteome Res.* **2010**, *9*, 1323–1329.
(20) Nevedomskaya, E.; Derks, R.; Deelder, A. M.; Mayboroda, O. A.; Palmblad, M. *Anal. Bioanal. Chem.* **2009**, *395*, 2527–2533.
(21) Wallace, A.; Millar, A.; Langridge, J. *Nat. Methods* **2007**, AN12–AN13.
(22) Shvartsburg, A. A.; Smith, R. D. *Anal. Chem.* **2008**, *80*, 9689–9699.
(23) Giannakopulos, A. E.; Thomas, B.; Colburn, A. W.; Reynolds, D. J.; Raptakis, E. N.; Makarov, A. A.; Derrick, P. J. *Rev. Sci. Instrum.* **2002**, *73*, 2115–2123.
(24) Olsen, J. V.; Nielsen, M. L.; Damoc, N. E.; Griep-Raming, J.; Moehring, T.; Makarov, A.; Schwartz, J.; Horning, S.; Mann, M. *Mol. Cell. Proteomics* **2009**, 2759–2769.
(25) Second, T. P.; Blethrow, J. D.; Schwartz, J. C.; Merrihew, G. E.; MacCoss, M. J.; Swaney, D. L.; Russell, J. D.; Coon, J. J.; Zabrouskov, V. *Anal. Chem.* **2009**, *81*, 7757–7765.
(26) Moehring, T.; Strupat, K.; Delanghe, B.; Muenster, H. *Mol. Cell. Proteomics* **2005**, *4*, S308–S308.
(27) Sterner, J. L.; Johnston, M. V.; Nicol, G. R.; Ridge, D. P. *J. Mass Spectrom.* **2000**, *35*, 385–391.
(28) Nilsson, T.; Mann, M.; Aebersold, R.; Yates, J. R.; Bairoch, A.; Bergeron, J. J. M. *Nat. Methods* **2010**, *7*, 681–685.
(29) Gouw, J. W.; Tops, B. B. J.; Mortensen, P.; Heck, A. J. R.; Krijgsveld, J. *Anal. Chem.* **2008**, *80*, 7796–7803.
(30) Kellmann, M.; Muenster, H.; Zomer, P.; Mol, H. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1464–1476.
(31) Delmotte, N.; Lasaosa, M.; Tholey, A.; Heinzle, E.; van Dorsselaer, A.; Huber, C. G. *J. Sep. Sci.* **2009**, *32*, 1156–1164.
(32) Wang, H.; Chang-Wong, T.; Tang, H. Y.; Speicher, D. W. *J. Proteome Res.* **2010**, *9*, 1032–1040.
(33) Ficarro, S. B.; McCleland, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M. *Nat. Biotechnol.* **2002**, *20*, 301–305.
(34) Zhang, Y.; Wolf-Yadlin, A.; Ross, P. L.; Pappin, D. J.; Rush, J.; Lauffenburger, D. A.; White, F. M. *Mol. Cell. Proteomics* **2005**, *4*, 1240–1250.
(35) Andersen, J. S.; Mann, M. *EMBO Rep.* **2006**, *7*, 874–879.
(36) Burckstummer, T.; Bennett, K. L.; Preradovic, A.; Schutze, G.; Hantschel, O.; Superti-Furga, G.; Bauch, A. *Nat. Methods* **2006**, *3*, 1013–1019.
(37) Ficarro, S. B.; Adelmant, G.; Tomar, M. N.; Zhang, Y.; Cheng, V. J.; Marto, J. A. *Anal. Chem.* **2009**, *81*, 4566–4575.
(38) Wilm, M.; Mann, M. *Anal. Chem.* **1996**, *68*, 1–8.
(39) Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M. *Nature* **1996**, *379*, 466–469.
(40) Piersma, S. R.; Fiedler, U.; Span, S.; Lingnau, A.; Pham, T. V.; Hoffmann, S.; Kubbutat, M. H. G.; Jimenez, C. R. *J. Proteome Res.* **2010**, *9*, 1913–1922.
(41) Fang, Y.; Robinson, D. P.; Foster, L. J. *J. Proteome Res.* **2010**, *9*, 1902–1912.
(42) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R. *Nat. Biotechnol.* **1999**, *17*, 676–682.
(43) Washburn, M. P.; Wolters, D.; Yates, J. R. *Nat. Biotechnol.* **2001**, *19*, 242–247.
(44) Peng, J. M.; Elias, J. E.; Thoreen, C. C.; Licklider, L. J.; Gygi, S. P. *J. Proteome Res.* **2003**, *2*, 43–50.
(45) Dowell, J. A.; Frost, D. C.; Zhang, J.; Li, L. J. *Anal. Chem.* **2008**, *80*, 6715–6723.
(46) Slebos, R. J. C.; Brock, J. W. C.; Winters, N. F.; Stuart, S. R.; Martinez, M. A.; Li, M.; Chambers, M. C.; Zimmerman, L. J.; Ham, A. J.; Tabb, D. L.; Liebler, D. C. *J. Proteome Res.* **2008**, *7*, 5286–5294.
(47) Cargile, B. J.; Talley, D. L.; Stephenson, J. L. *Electrophoresis* **2004**, *25*, 936–945.
(48) Balgley, B. M.; Wang, W. J.; Song, T.; Fang, X. P.; Yang, L.; Lee, C. S. *Electrophoresis* **2008**, *29*, 3047–3054.
(49) Hubner, N. C.; Ren, S.; Mann, M. *Proteomics* **2008**, *8*, 4862–4872.
(50) Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. *J. Sep. Sci.* **2005**, *28*, 1694–1703.
(51) Song, C. X.; Ye, M. L.; Han, G. H.; Jiang, X. N.; Wang, F. J.; Yu, Z. Y.; Chen, R.; Zou, H. F. *Anal. Chem.* **2010**, *82*, 53–56.
(52) Manadas, B.; English, J. A.; Wynne, K. J.; Cotter, D. R.; Dunn, M. J. *Proteomics* **2009**, *9*, 5194–5198.
(53) Phillips, H. L.; Williamson, J. C.; van Elburg, K. A.; Snijders, A. P. L.; Wright, P. C.; Dickman, M. J. *Proteomics* **2010**, *10*, 2950–2960.
(54) Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Yates, J. R. *Nat. Biotechnol.* **2003**, *21*, 532–538.
(55) Zhou, F.; Cardoza, J. D.; Ficarro, S. B.; Adelmant, G. O.; Lazaro, J. B.; Marto, J. A. *J. Proteome Res.* **2010**, *9*, 6242–6255.
(56) Giddings, J. C. *Anal. Chem.* **1984**, *56*, 1258.
(57) Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. *Anal. Chem.* **2005**, *77*, 6426–6434.
(58) van der Heeft, E.; ten Hove, G. J.; Herberts, C. A.; Meiring, H. D.; van Els, C.; de Jong, A. *Anal. Chem.* **1998**, *70*, 3742–3751.
(59) Licklider, L. J.; Thoreen, C. C.; Peng, J. M.; Gygi, S. P. *Anal. Chem.* **2002**, *74*, 3076–3083.
(60) Ficarro, S. B.; Zhang, Y.; Lu, Y.; Moghimi, A. R.; Askenazi, M.; Hyatt, E.; Smith, E. D.; Boyer, L.; Schlaeger, T. M.; Luckey, C. J.; Marto, J. A. *Anal. Chem.* **2009**, *81*, 3440–3447.
(61) Ghaemmaghami, S.; Huh, W.; Bower, K.; Howson, R. W.; Belle, A.; Dephoure, N.; O'Shea, E. K.; Weissman, J. S. *Nature* **2003**, *425*, 737–741.
(62) Picotti, P.; Bodenmiller, B.; Mueller, L. N.; Domon, B.; Aebersold, R. *Cell* **2009**, *138*, 795–806.
(63) Askenazi, M.; Webber, J. T.; Marto, J. A. *Mol. Cell. Proteomics* **2011**, *10*, 7.
(64) Rudnick, P. A.; Clauser, K. R.; Kilpatrick, L. E.; Tchekhovskoi, D. V.; Neta, P.; Blonder, N.; Billheimer, D. D.; Blackman, R. K.; Bunk, D. M.; Cardasis, H. L.; Ham, A. J. L.; Jaffe, J. D.; Kinsinger, C. R.; Mesri, M.; Neubert, T. A.; Schilling, B.; Tabb, D. L.; Tegeler, T. J.; Vega-Montoto, L.; Variyath, A. M.; Wang, M.; Wang, P.; Whiteaker, J. R.; Zimmerman, L. J.; Carr, S. A.; Fisher, S. J.; Gibson, B. W.; Paulovich, A. G.; Regnier, F. E.; Rodriguez, H.; Spiegelman, C.; Tempst, P.; Liebler, D. C.; Stein, S. E. *Mol. Cell. Proteomics* **2010**, *9*, 225–241.
(65) Thakur, S. S.; Geiger, T.; Chatterjee, B.; Bandilla, P.; Froehlich, F.; Cox, J.; Mann, M. *Mol. Cell. Proteomics* **2011**, *10*, M110.003699, 1–9.
(66) Shen, Y. F.; Zhang, R.; Moore, R. J.; Kim, J.; Metz, T. O.; Hixson, K. K.; Zhao, R.; Livesay, E. A.; Udseth, H. R.; Smith, R. D. *Anal. Chem.* **2005**, *77*, 3090–3100.

(67) Di Palma, S.; Stange, D.; van de Wetering, M.; Clevers, H.; Heck, A. J. R.; Mohammed, S. *J. Proteome Res.* **2011**, *10*, 3814–3819.
(68) Di Palma, S.; Boersema, P. J.; Heck, A. J. R.; Mohammed, S. *Anal. Chem.* **2011**, *83*, 3440–3447.
(69) Shen, Y. F.; Jacobs, J. M.; Camp, D. G.; Fang, R. H.; Moore, R. J.; Smith, R. D.; Xiao, W. Z.; Davis, R. W.; Tompkins, R. G. *Anal. Chem.* **2004**, *76*, 1134–1144.
(70) Kong, R. P. W.; Siu, S. O.; Lee, S. S. M.; Lo, C.; Chu, I. K. *J. Chromatogr., A* **2011**, *1218*, 3681–3688.
(71) Luo, Q.; Gu, Y.; Wu, S. L.; Rejtar, T.; Karger, B. L. *Electrophoresis* **2008**, *29*, 1604–11.