RESEARCH ARTICLE

A cost–benefit analysis of multidimensional fractionation of affinity purification-mass spectrometry samples

Wade H. Dunham[1,2], Brett Larsen[1], Stephen Tate[3], Beatriz Gonzalez Badillo[1], Marilyn Goudreault[1], Yasmina Tehami[1], Thomas Kislinger[4,5] and Anne-Claude Gingras[1,2]

[1] Centre for Systems Biology, Samuel Lunenfeld Research Institute, Toronto, Ontario, Canada
[2] Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada
[3] AB SCIEX, Concord, Ontario, Canada
[4] Ontario Cancer Institute, Campbell Family Cancer Research Institute, Toronto, Ontario, Canada
[5] Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada

Affinity purification coupled to mass spectrometry (AP-MS) is gaining widespread use for the identification of protein–protein interactions. It is unclear, however, whether typical AP sample complexity is limiting for the identification of all protein components using standard one-dimensional LC-MS/MS. Multidimensional sample separation is useful for reducing sample complexity prior to MS analysis and increases peptide and protein coverage of complex samples. Here, we monitored the effects of upstream protein or peptide separation techniques on typical mammalian AP-MS samples, generated by FLAG affinity purification of four baits with different biological functions and/or subcellular distribution. As a first separation step, we employed SDS-PAGE, strong cation exchange LC, or reversed-phase LC at basic pH. We also analyzed the benefits of using an instrument with a faster scan rate, the new TripleTOF 5600 mass spectrometer. While all multidimensional approaches yielded a clear increase in spectral counts, the increase in unique peptides and additional protein identification was modest and came at the cost of increased instrument and handling time. The use of a high duty-cycle instrument achieved similar benefits without these drawbacks. An increase in spectral counts is beneficial when data analysis methods relying on spectral counts, including Significance Analysis of INTeractome (SAINT), are used.

Received: September 7, 2010
Revised: March 22, 2011
Accepted: March 28, 2011

Keywords:
Affinity purification coupled to mass spectrometry / Base-stable reversed-phase chromatography / MudPIT / SDS-PAGE fractionation / Systems biology / TripleTOF

1 Introduction

Affinity purification coupled to mass spectrometry (AP-MS) has become a method of choice for the identification of protein interaction partners under near physiological conditions [1]. In most experimental approaches, capturing interaction partners for a protein of interest begins with the expression of an epitope-tagged recombinant protein in a relevant cell line. Capture of the tagged protein with its associated partners can be effected by a single affinity purification step, or via serial purification, as in the case of tandem affinity purification (TAP) [2]. AP sample complexity is largely dictated by the interactions established by the tagged protein, the number and nature of the purification steps and the stringency of the washes. For example, TAP

Correspondence: Anne-Claude Gingras, Centre for Systems Biology, Samuel Lunenfeld Research Institute, 600 University Avenue, Toronto, Ontario, Canada M5G 1X5
E-mail: gingras@lunenfeld.ca
Fax: +1-416-586-8869

Abbreviations: AP, affinity purification; **AP-MS,** affinity purification coupled to mass spectrometry; **SAINT,** Significance Analysis of Interactome; **SCX,** strong cation exchange

Colour online: See the article online to view Figs. 1 and 2 in colour.

purification yields less complex samples than single-step approaches [3]. However, the development of better bioinformatics approaches to identify background contaminants [4–6] has resulted in single-step purifications becoming increasingly used [7–9]. Our laboratory has been utilizing single-step FLAG affinity purification to identify interactors for proteins of interest by AP-MS in mammalian cells [10–13]. It is unclear, however, whether current one-dimensional (1-D) LC-MS/MS approaches are able to effectively identify all of the components in this sample type. Given the increasingly widespread usage of this type of purification strategy, this is an important question to address.

Fractionation of very complex samples (at the organelle, protein or peptide level) prior to mass spectrometric analysis allows for an increased depth of peptide and protein identification [14–17]. Many research groups have employed fractionation approaches for the analysis of mammalian AP-MS samples, including protein-level separation by SDS-PAGE [7, 18] or peptide-level separation via different types of multidimensional liquid chromatography [6, 19]. However, a side-by-side comparison of the benefits of each type of fractionation approach for AP-MS protein identification has not been conducted. In an attempt to better characterize the complexity of our single-step purification AP-MS samples, and to assess the benefits of applying multidimensional separation techniques for the analysis of these samples, we tested three commonly used fractionation techniques: (a) protein-based separation by SDS-PAGE (GeLC) [17], (b) peptide-based separation by ion-exchange chromatography (MudPIT) [19] or (c) an additional reversed-phase separation step (RP-RP; see Section 2). We also assessed the performance of a new generation fast-scanning instrument, the AB SCIEX TripleTOF 5600 in relation to a linear ion trap.

To benchmark the effectiveness of multidimensional separation on an AP-MS sample of known complexity, we selected bait proteins (COPS5, eIF4A2, RAF1, MEPCE) for which multiple binding partners are annotated in the protein interaction database BioGRID [4, 20]. These baits were selected because they have different biological roles and/or subcellular localization. COPS5 is a core component of the COP9 signalosome, a well-defined multifunctional eight-subunit complex involved in the ubiquitin–proteasome system, in part via deneddylation of the cullins with which it physically interacts [4, 21]. RAF1 is a serine–threonine kinase which binds to and is activated by GTP-loaded Ras, leading to the activation of the MAP kinase ERK pathway. Importantly, several chaperones (Hsp90, Cdc37 and immunophilins) and 14-3-3 proteins have been identified as physically interacting with and critical for proper activation of RAF1 [22, 23]; eIF4A2 is a translation initiation factor which is part, together with eIF4E and eIF4G, of the biochemically stable eIF4F complex; through eIF4G, eIF4F associates with all members of the eIF3 multisubunit complex, which in turn binds to the 40S ribosomal subunit [24]. Association with eIF4E and eIF4G also enables association with capped mRNAs, and coincident recruitment of other mRNA binding proteins. Finally, MEPCE, the 7SK snRNA methylphosphate capping enzyme also known as BCDIN3, was recently identified by interaction proteomics screens as a component of an snRNP complex containing both RNA processing and transcription factors [18].

2 Materials and methods

2.1 Generation of stably transfected Flp-In T-REx 293 cell lines

The vector pcDNA5-FRT-FLAG, engineered to inducibly express fusion proteins with a single N-terminal FLAG epitope, was constructed from the parent vector pcDNA5-FRT-TO (Invitrogen) and the vector pcDNA3-FLAG [25] as follows: A *Hin*dIII/*Xho*I cassette from pcDNA3-FLAG (containing the FLAG and the multiple cloning site) was subcloned into the pcDNA5-FRT-TO vector also digested with *Hin*dIII/*Xho*I. An internal *Eco*RI site was subsequently destroyed by mutagenesis, and the modified versions of the resulting vector were sequenced. The coding sequences of COPS5, eIF4A2, RAF1 and MEPCE were amplified by PCR from Mammalian Gene Collection constructs BC001187, BC015842, BC018119 and BC018935, respectively, and cloned into pcDNA5-FRT-FLAG (using *Eco*RI/*Not*I for COPS5 and RAF1, and *Asc*I/*Not*I for eIF4A2 and MEPCE), and the junctions sequenced. The resulting vectors were stably co-transfected with the Flp-recombinase expressing vector pOG44 into Flp-In T-REx 293 cells (Invitrogen). Selection was performed by plating transfected cells at low density in DMEM (high glucose) supplemented with 5% FBS, 5% calf serum, 100 U/mL penicillin/streptomycin and 200 μg/mL hygromycin. Individual hygromycin-resistant clones were picked (after ~2 wk in culture), amplified in selection medium, and recombinant protein expression induced by the addition of tetracycline (200 ng/mL, HyClone) to cell media for 24 h.

2.2 Immunopurification

Cells expressing FLAG-tagged proteins were harvested at a confluence of 90% by scraping with a rubber spatula. Cells were pelleted by centrifugation, washed once with PBS and frozen at −80°C. Cell lysis and FLAG immunoprecipitation (IP) on M2-sepharose were performed essentially as previously described [3], with the following modifications. Briefly, to the frozen cell pellet (from six 150 mm plates), a 1:4 (pellet weight/volume) ratio of lysis buffer was added. Lysis buffer was 50 mM HEPES-KOH (pH 8.0), 100 mM KCl, 2 mM EDTA, 0.1% NP40, 10% glycerol, 1 mM PMSF, 1 mM DTT and protease inhibitor cocktail (Sigma; P8340; 1:500). Cells were lysed on ice, subjected to one freeze–thaw cycle and lysate cleared by centrifugation (20 800 rcf, 20 min,

4°C). Cleared lysate was incubated with 30 μL bed volume of pre-washed FLAG M2 agarose beads (Sigma; A2220), for 2 h at 4°C. FLAG M2-sepharose was washed four times with lysis buffer, followed by two washes with ammonium bicarbonate rinsing buffer (50 mM NH_4HCO_3, pH 8.0, 75 mM KCl), before bound proteins were eluted (3 ×) by incubating the resin for 15 min at 4°C with 150 μL of freshly prepared ammonium hydroxide solution (pH 11–12, prepared by diluting 500 μL NH_4OH in 5 mL HPLC-grade H_2O) per elution. Pooled elutions were evaporated in a speed vac until dryness, after which HPLC-grade water (100 μL) was added to each tube, and lyophilization repeated. Except for GeLC, samples were digested overnight with Sigma Trypsin Singles (Sigma; T7575; 1 μg per sample) in 50 mM NH_4HCO_3, pH 8.0 at 37°C. After overnight digestion, samples were spiked with an additional 0.25 μg trypsin (in 50 mM NH_4HCO_3, pH 8.0) and incubated for 3 h at 37°C before acidification to 1% formic acid and lyophilization. Digested samples were stored at −40°C. COPS5 samples coming from the same batch purification are indicated by superscript numbers in Table 1. All eIF4A2, RAF1 and MEPCE RP and RP-RP analyses were done using samples from the same batch purification.

2.3 1-D LC-MS/MS analysis

For 1-D LC-MS/MS analysis, affinity purified, digested and lyophilized FLAG alone or FLAG-COPS5, eIF4A2, RAF1 or MEPCE samples were re-suspended in 5% formic acid before direct loading onto fused silica capillary columns (0.75 μm id) packed in-house with 10 cm Zorbax C18 (ZorbaxSB, 3.5 μm) and pre-equilibrated with HPLC buffer A. The amount of affinity purified material loaded on column was equivalent to two 90% confluent 150 mm plates. Loaded columns were placed in-line with a LTQ mass spectrometer equipped with an Agilent 1100 capillary HPLC, an LTQ-Orbitrap mass spectrometer equipped with an Eksigent Ultra HPLC or a TripleTOF 5600 equipped with an Eksigent nanoLC. On all platforms, HPLC gradients were delivered at 200 nL/min using a split flow arrangement or nanoflow, respectively. Buffer A was 3% ACN and 0.1% formic acid; buffer B was 80% ACN and 0.1% formic acid. The HPLC gradient program delivered an ACN gradient over 120 min (1–5% buffer B over 4 min, 5–40% buffer B over 100 min, 40–60% buffer B over 5 min, 60–100% buffer B over 5 min, hold buffer B at 100% 3 min and 100–0% B in 2 min). The parameters for data-dependent acquisition on the LTQ mass spectrometer were:

Table 1. Spectral counts, unique peptides, and non-redundant protein identification

Separation method (instrument)	Spectral counts	Fold change	Unique pep	Fold change	Proteins	Fold change
(A) All after background removal						
1-D RP merged (LTQ)[1,2,3,4]	2045 ± 691		576 ± 189		45 ± 19	
1-D RP (LTQ)[1]	1944		626		43	
RP/RP (LTQ)[5]	3234	1.66	723	1.15	51	1.19
GeLC (LTQ)[6]	2712	1.40	675	1.08	47	1.09
1-D RP (TripleTOF 5600)[5]	2892	1.49	657	1.05	49	1.14
1-D RP (ORB)[7]	1365		736		40	
MudPIT (ORB)[7]	2013	1.47	515	0.70	83	2.08
(B) BioGRID annotated interactors						
1-D RP merged (LTQ)[1,2,3,4]	1511 ± 435		433 ± 126		20 ± 2	
1-D RP (LTQ)[1]	1501		483		22	
RP/RP (LTQ)[5]	2649	1.76	541	1.12	22	1.00
GeLC (LTQ)[6]	2233	1.49	540	1.12	24	1.09
1-D RP (TripleTOF 5600)[5]	2327	1.55	514	1.06	23	1.05
1-D RP (ORB)[7]	1066		581		24	
MudPIT (ORB)[7]	1675	1.57	313	0.54	25	1.04
(C) All before background removal						
1-D RP merged (LTQ)[1,2,3,4]	2718 ± 1102		892 ± 381		111 ± 53	
1-D RP (LTQ)[1]	2500		923		111	
RP/RP (LTQ)[5]	4502	1.80	1219	1.32	148	1.33
GeLC (LTQ)[6]	4247	1.70	1223	1.33	123	1.11
1-D RP (TripleTOF 5600)[5]	4687	1.87	1370	1.48	160	1.44
1-D RP (ORB)[7]	2080		1340		143	
MudPIT (ORB)[7]	3390	1.63	1273	0.95	238	1.66

(A) for all proteins identified in COPS5 samples after background contaminant removal (COPS4 and DDB1 spectra were detected in FLAG alone controls, but these proteins were not filtered out due to substantial increase in spectra in FLAG-COPS5 purifications), (B) for the COPS5 interactors reported in BioGRID and detected in our samples or (C) for all proteins prior to background contaminant removal. Immunopurified FLAG-COPS5 was subjected to LC-MS/MS either on an LTQ, an LTQ-Orbitrap (ORB), or a TripleTOF 5600. Samples were fractionated for multidimensional analysis using RP-RP (pH 10/pH 2), GeLC or MudPIT. Only proteins identified with two or more unique peptides and a ProteinProphet probability of at least 0.95 are considered. Sample batch purifications are indicated in superscript.

one centroid MS (mass range 400–2000) followed by MS/MS on the five most abundant ions. On the Orbitrap, one high-resolution MS scan followed by three low-resolution MS/MS scans of the most abundant ions. On the TripleTOF 5600, data-dependent acquisition was done with 1 high-resolution MS scan followed by 20 high-resolution MS/MS scans over a 1.3-s cycle time.

2.4 Multidimensional LC-MS/MS analysis

Because sample fractionation allows for the analysis of more digested material, peptide level fractionation (RP-RP or MudPIT) of FLAG-COPS5 samples was done with the equivalent of four 90% confluent 150 mm plates, and at the protein level (GeLC) with the equivalent of six 90% confluent plates. However, to directly assess whether changes in the number of spectra, peptides and proteins are truly due to the multidimensional separation and not the amount of starting material used, a true parallel comparison for FLAG-eIF4A2, RAF1 and MEPCE was also performed by using the same amount of material for 1-D LC-MS and RP-RP, namely two 90% confluent 150 mm plates.

MudPIT was performed essentially as described [26], pre-columns (150 μm id) and analytical columns (75 μm id) were fused silica packed in-house with Phenomenex strong cation exchange (SCX) resin and Magic C18, respectively. MudPIT samples were desalted by loading onto a capillary column (200 μm id), packed in-house with 5 cm Zorbax C18 (ZorbaxSB, 3.5 μm) and washed with buffer A. Desalted samples were eluted using 50% ACN, 50% buffer A and lyophilized before loading onto SCX resin in buffer A. Loaded samples were bumped sequentially onto the RP resin using 8 μL salt bumps of 100, 200, 300 and 500 mM sodium acetate. The loaded column was placed in-line with an LTQ-Orbitrap XL mass spectrometer equipped with a Proxeon Easy-LC. Buffer A is 0.1% formic acid; buffer B is 99.9% ACN and 0.1% formic acid. The HPLC gradient program delivered an ACN gradient over 120 min at 400 nL/min (0–7% buffer B over 7 min, 7–40% buffer B over 98 min, 40–80% buffer B over 2 min, hold buffer B at 80% for 13 min). Data-dependent acquisition on the mass spectrometer was: one MS scan followed by five data-dependent MS/MS scans.

For RP-RP fractionation, samples were re-suspended in 20 mM ammonium formate (NH_4HCO_2, pH 10) and bomb loaded onto a capillary column (200 μm id), packed in-house with 6 cm X-Bridge base stable C-18 resin (3 μm, Waters), pre-equilibrated with NH_4HCO_2 (pH 10). Sample flow-through was collected and the column was washed using 20 μL of 20 mM NH_4HCO_2 (pH 10). Sample fractions were eluted using serial 20 μL injections of 5, 10, 15, 20 or 50% ACN in 20 mM NH_4HCO_2 (pH 10) and lyophilized. Lyophilized flow-through and sample fractions were reconstituted in 5% formic acid and analyzed using LC-MS/MS as described above.

For GeLC, undigested affinity purified and lyophilized FLAG-COPS5 or FLAG alone samples were resuspended in Laemmli sample buffer, subjected to SDS-PAGE, colloidal blue staining (Thermo Scientific; 24592) and in-gel digestion essentially as described [17, 27]. Sample lanes were divided into 20 proportionally sized fractions and sequential fractions pooled before elution solvent evaporation in a speed vac. Fraction eluates were reconstituted in 10 μL of 5% formic acid, centrifuged at 16 000 rcf to sediment any gel debris and 6 μL analyzed using LC-MS/MS as described above.

2.5 Data analysis

All ThermoFinnigan RAW files were saved in our local interaction proteomics LIMS, ProHits (Liu et al, submitted). mzXML files were generated from ThermoFinnigan RAW files using the ProteoWizard converter implemented within ProHits (–filter 'peakPicking true2'–filter 'msLevel2'). TripleTOF 5600 .wiff files were converted to .mgf format using ProteinPilot software before being saved into ProHits for analysis. The searched database contained the human complement of the RefSeq protein database (version 37; 38 097 entries searched). mzXML files were searched with MASCOT version 2.2 using the following parameters: one missed cleavage site, methionine oxidation and asparagine/glutamine deamidation as variable modifications. The fragment mass tolerance was 0.6 Da (monoisotopic mass) for LTQ and LTQ-Orbitrap data, and the mass window for the precursor was ± 3 Da average mass in the case of LTQ, and ± 12 ppm (monoisotopic) in the case of LTQ-Orbitrap. TripleTOF 5600 fragment mass tolerance was ± 0.2 Da (monoisotopic) and precursor ± 40 ppm. Search results were further analyzed using the PeptideProphet and ProteinProphet TransProteomics Pipeline tools [28, 29]. Peptides that did not meet the minimum PeptideProphet significance *p*-value of 0.05 were filtered out. To combine multiple analyses from the fractionation experiments, pepXML files were combined into a single mzXML using Trans Proteomic Pipeline (TPP). Details on the number of identifications made pre-filtering are presented in Supporting Information Table 1. The increase in the number of spectra, unique peptide and protein identifications made after RP-RP analysis of FLAG-eIF4A2, RAF1 and MEPCE samples was assessed for statistical significance using a paired *t*-test implemented by GraphPad Software (http://www.graphpad.com/quickcalcs/ttest1.cfm). All raw files have been deposited in Tranche, hash: OSV877I4YeXnaQQnBk2BjOxRFOIuyT6qh+Nl8K1mnf413ufsMk2m5kTH30VgJzEutwKXLITEEY3vU2fvbuBgWmyzlagAAAAAAABT4a==.

2.6 SAINT analysis

SAINT calculates, for each prey protein identified in a purification, the probability of true interaction by using spectral counting (semi-supervised clustering, using a number of negative control runs) [5, 30, 31]. SAINT analysis

of eIF4A2 and RAF1 was done using two biological replicates per bait and condition (1-D versus 2-D). BAIT protein samples were analyzed alongside seven negative control runs, consisting of purifications from cells expressing the FLAG tag alone that had been analyzed using either one- or two-dimensional separation. Probability scores, indicating the likelihood for a true protein interaction to exist, were first computed for each prey protein in independent biological replicates before a final probability score for said prey protein was calculated as the average of its probabilities in the two individual replicates (Avg p); final results with Avg $p \geq 0.9$ are reported.

3 Results

To determine whether applying multidimensional separation methods to AP-MS samples increases coverage for this type of sample (both in terms of total number of proteins identified and number of unique peptides or spectral counts observed), FLAG-tagged COPS5, eIF4A2, RAF1, MEPCE and a negative control consisting of the FLAG tag alone were purified by IP from inducible cell lines, as in [3] (Fig. 1A). Samples were analyzed by 1-D LC-MS/MS on an LTQ, an LTQ-Orbitrap (Thermo) or a TripleTOF 5600 instrument (AB SCIEX), and/or subjected to an additional fractionation step prior to mass spectrometric identification (see Section 2 for details; Fig. 1B). Proteins present on a list of common AP-MS contaminants (Supporting Information Table 2A; $n = 232$) or proteins detected in any of the FLAG alone control runs (Supporting Information Table 2B, $n = 126$), were removed from subsequent analysis. Only those proteins identified with at least two unique peptides, and a ProteinProphet [29] probability of 0.95 were used for comparison.

As a first approach, the reproducibility of our standard AP-MS approach was assessed through analysis of four independent biological replicates of FLAG-COPS5 (i.e. APs from four different cell pellets performed at different times) on an LTQ. After subtraction of likely contaminants, an average of 2045 spectra, 576 unique peptides and 45 proteins were identified per biological replicate (Table 1A, first row; unfiltered data are in Table 1C, first row).

Examining the reproducibility of protein identifications in the four different COPS5 biological replicates revealed that 23 proteins were detected in all four samples, 5 in three samples, 11 in two samples and 54 proteins were observed in a single sample (36 of which were associated with COPS5 sample C; Fig. 2A). The database BioGRID [20] (v3.0.67) reports 52 COPS5 interacting partners identified by a variety of different approaches, including 35 which were previously identified by AP-MS [4]. A total of 24 of these proteins were detected in at least one of our four 1-D LC-MS/MS analyses. Importantly, 18 of these previously reported COPS5 interactors were detected across all four samples (Fig. 2B). The proteins observed in only one or two runs yielded low spectral counts (see below). These data indicate that our standard AP-MS method is robust for the identification of known interactors, but that there is some degree of variation in the recovery of a given interacting partner between biological replicates, especially for proteins present in lower abundance in the sample.

Next, we analyzed the effect of applying multidimensional separation to the analysis of AP samples. As expected, additional sample fractionation resulted in an increase in spectral counts for almost all COPS5 interacting proteins detected by MS analysis after background removal; RP-RP, GeLC and MudPIT increased average spectral counts by 1.66-, 1.4- and 1.47-fold, respectively (Table 1A). A similar increase (1.49-fold) was observed when the samples were analyzed on a new generation instrument with a faster scan rate, the AB SCIEX TripleTOF 5600 (of note, this result was obtained with only 50% of the sample volume loaded on the 1-D LTQ, and 25% of the sample volume analyzed by multidimensional separation).

To analyze whether these increases in spectral counts could be reproduced with different AP samples, we performed RP-RP analysis on eIF4A2, RAF1 and MEPCE samples. RP-RP analysis of biological replicate eIF4A2 and

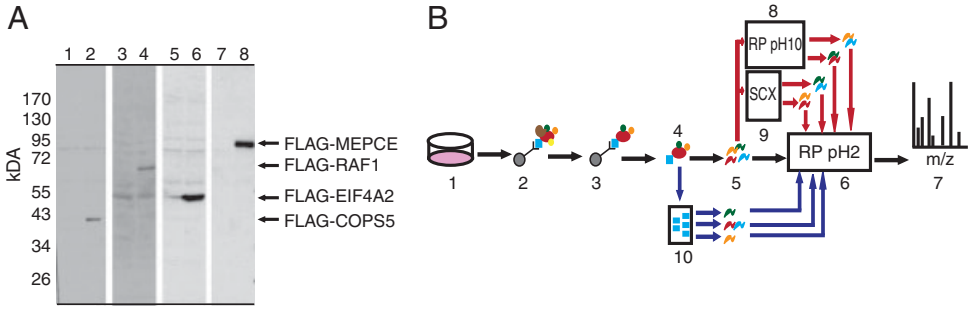

Figure 1. Sample preparation. (A) Western blot showing expression of FLAG-COPS5 (lanes 1 and 2), RAF1 (lanes 3 and 4), eIF4A2 (lanes 5 and 6) and MEPCE (lanes 7 and 8), using an inducible system in 293 Flp-In T-REX cells. Lanes 1, 3, 5 and 7 show uninduced cells, and lanes 2, 4, 6 and 8 induced. (B) Immunopurified FLAG-COPS5 was subjected to LC-MS/MS on either an LTQ, LTQ-Orbitrap or TripleTOF 5600 mass spectrometer. FLAG-eIF4A2, RAF1 and MEPCE were subjected to LC-MS/MS on an LTQ mass spectrometer (B steps 1–7). Samples were fractionated for multidimensional analysis using, reverse-phase liquid chromatography (RP) at pH 10 (B, step 8), strong cation exchange (SCX) liquid chromatography (B, step 9), or SDS-PAGE (B, step 10), prior to RP separation at pH 2 (B, step 6).

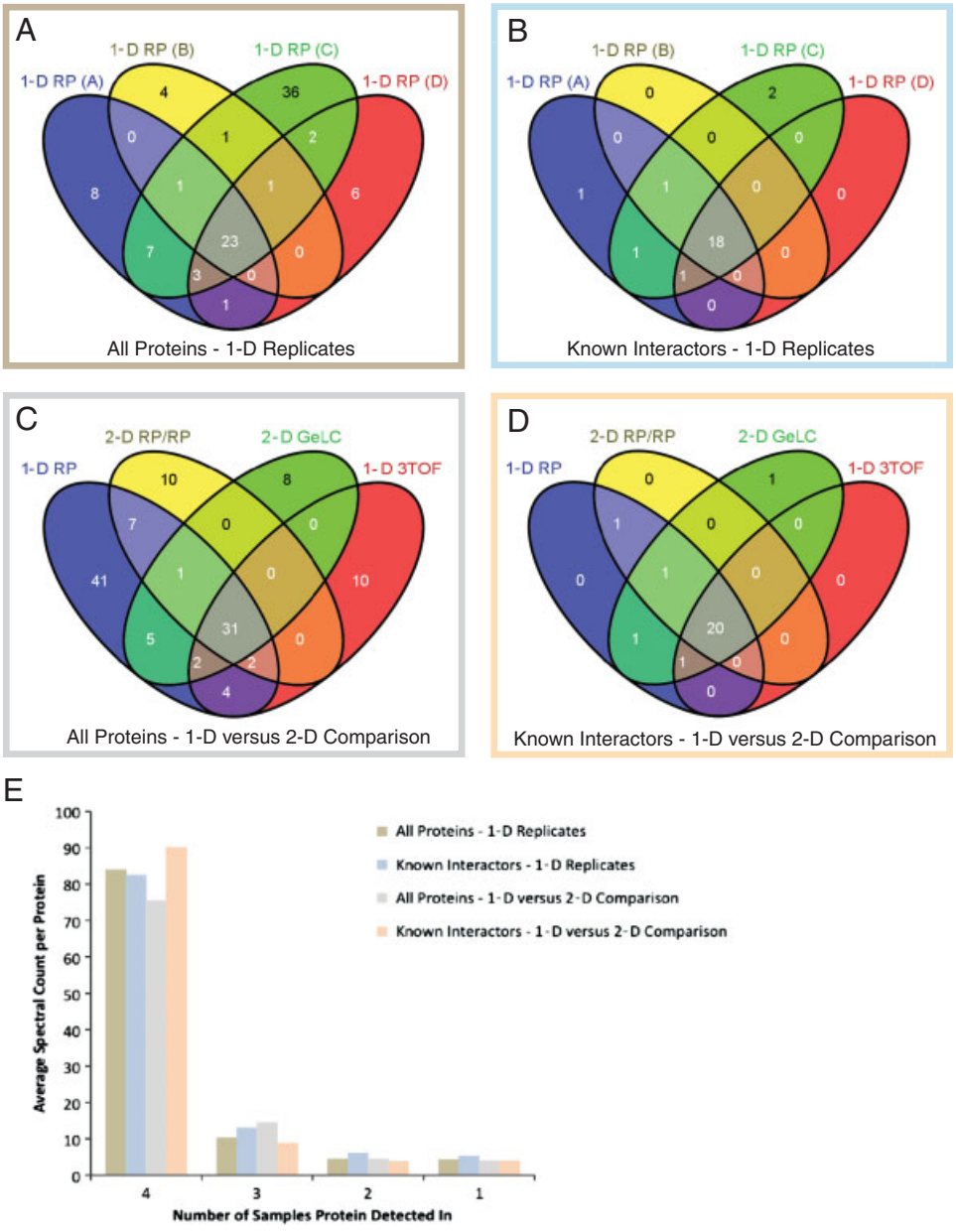

Figure 2. Venn diagrams showing protein identification overlap for COPS5. (A) All proteins identified across four 1-D LC-MS/MS analyses on an LTQ. (B) Protein identification overlap for the samples outlined in (A) when only BioGRID-annotated COPS5 interactors are considered. (C) Protein identification overlap when all four 1-D runs are merged together and compared with samples subjected to multidimensional separation and analysis on an LTQ, or analysis on a TripleTOF 5600. (D) Protein identification overlap for the samples outlined in (C) when only BioGRID-annotated COPS5 interactors were considered. (E) Average spectral count per protein identified in either 4, 3, 2 or 1 samples. Venn diagrams were created using the web application VENNY (An interactive tool for comparing lists with Venn Diagrams. Oliveros, J. C. (2007), http://bioinfogp.cnb.csic.es/tools/venny/index.html).

RAF1 samples increased the average spectral counts observed after background removal by 2.34- and 2.48–fold, respectively (Table 2A). Analysis of one MEPCE sample resulted in an increase in average spectral count of 1.71-fold (Supporting Information Table 3A).

Importantly, however, the increase in spectral counts was accompanied in all cases by a more modest increase in the number of unique peptides detected, an average of 68% of the gain detected for spectral counts across all baits, methods and instruments analyzed, respectively (Tables 1A and 2A and Supporting Information Table 3A). When the same parameters were analyzed for only the previously annotated COPS5, eIF4A2, RAF1 and MEPCE interactors, the same trend was observed: i.e. an overall >2-fold increase in spectral counts in response to an additional degree of separation, and a more modest change in the recovery of known interactors, and in the number of unique peptides assigned to each of these interactors (Tables 1B and 2B and Supporting Information Table 3B). Notably, the increase in unique peptides and proteins observed in the RAF1 RP-RP, MEPCE RP-RP and COPS5 MudPIT analyses after background removal (Tables 1A and 2A and Supporting Information Table 3A) was reduced to levels comparable to those observed across our other samples when only these previously annotated interacting proteins were considered. These observations indicate that conventional 1-D-LC MS/MS may be sufficient for the detection of most interactions.

Table 2. Spectral counts, unique peptides and non-redundant protein identification for two biological replicate analyses of EIF4A2 and RAF1

BAIT/parameter	1-D RP (1)	1-D RP (2)	RP/RP (1)	RP/RP (2)	Ratio (1)	Ratio (2)
(A) All after background removal						
EIF4A2/spectral counts	523	626	1491	1147	2.85	1.83
EIF4A2/unique peptides	363	381	513	460	1.41	1.21
EIF4A2/proteins identified	42	40	62	60	1.48	1.50
RAF1/spectral counts	248	488	747	949	3.01	1.94
RAF1/unique peptides	139	319	272	444	1.96	1.39
RAF1/proteins identified	23	51	63	104	2.74	2.04
(B) BioGRID annotated interactors						
EIF4A2/spectral counts	233	318	689	543	2.96	1.71
EIF4A2/unique peptides	158	189	187	180	1.18	0.95
EIF4A2/proteins identified	10	12	10	11	1.00	0.92
RAF1/spectral counts	80	136	278	301	3.48	2.21
RAF1/unique peptides	45	78	74	76	1.64	0.97
RAF1/proteins identified	7	8	8	9	1.14	1.13
(C) All before background removal						
EIF4A2/spectral counts	932	1070	2612	1855	2.80	1.73
EIF4A2/unique peptides	725	715	1077	892	1.49	1.25
EIF4A2/proteins identified	202	177	278	266	1.38	1.50
RAF1/spectral counts	701	1114	1937	2370	2.76	2.13
RAF1/unique peptides	509	791	862	1331	1.69	1.68
RAF1/proteins identified	180	244	348	582	1.93	2.39

(A) for all proteins after background contaminant removal, (B) for the interaction partners reported in BioGRID or (C) for all protein hits prior to background contaminant removal. Immunopurified FLAG-EIF4A2 or RAF1 were subjected to LC-MS/MS on an LTQ mass spectrometer. Samples were fractionated for multidimensional analysis using RP-RP (pH 10/pH 2). Only proteins identified with two or more unique peptides and a ProteinProphet probability of at least 0.95 are considered.

To investigate whether different fractionation approaches led to the recovery of specific subsets of COPS5 binding partners, we compared the protein identifications made in 1-D-LTQ (by merging non-redundant protein identifications in the four samples) to those made in 2-D RP-RP, GeLC and by the TripleTOF 5600. Similar to what we observed in the four independent 1-D LTQ samples, we saw 31 proteins identified in all methods; 5 additional proteins observed across three methods, 16 proteins across two methods and 69 by a single method (Fig. 2C). Of the 26 known COPS5 interactors that we detected in our studies, the majority (20) were detected by all approaches, 2 proteins by three approaches, 2 by two approaches and only 1 by a single technique (Fig. 2D). Notably, 8 of the 10 proteins detected in our analyses across all methods, but not listed as COPS5 interactors in BioGRID, are annotated interactors of other COP9 signalosome components, indicating that the majority of these proteins are indeed bona fide COPS5-binding partners (Supporting Information Table 4).

Not surprisingly, the trends described above for the entire population of proteins remained in place for individual known COPS5, eIF4A2, RAF1 and MEPCE interactors, with the main benefit of multidimensional fractionation appearing to be an increase in spectral counts, followed by a modest increase in unique peptide identification (Table 3A and B and Supporting Information Tables 5A and B and 6–10).

To further analyze what are the new proteins identified after 2-D separation, we analyzed the spectral count distribution of the proteins detected in each of the COPS5 samples. Analysis of the average spectral count per protein detected across all COPS5 1-D LTQ runs or across all methods revealed an average of 83 spectral counts per protein (Fig. 2E). This number dropped to an average of 12 spectral counts or less per protein once proteins were restricted to those detected in three or fewer samples (and were further reduced for proteins detected in only one or two samples), suggesting that these proteins are low stoichiometry interactors or low-level contaminants that have escaped detection in the FLAG control experiments. It is unclear at this point whether these newly associated proteins are functionally relevant to the COP9 signalosome.

Finally, we were intrigued by the possibility that the increase in spectral counts brought about by 2-D fractionation may be beneficial in conjunction with methods using spectral counts as part of the scoring scheme. SAINT (Significance Analysis of INTeractome) uses spectral count information for a given prey protein across a data set containing negative controls to calculate the probability that a protein is detected in a purification because it is an interactor and not a contaminant [5, 31]. We used seven negative control runs, consisting of purifications from cells expressing the FLAG tag alone that had been analyzed using either one- or two-dimensional separation for the SAINT

Table 3. Fold increase in spectral counts (A) or unique peptides (B) for BioGRID-annotated COPS5 interactors

Interactor	RP/RP (LTQ)	GeLC (LTQ)	1-D RP (TripleTOF)	MudPIT (ORB)	Fold difference
(A)					
COPS1	1.57	2.61	0.87	0.83	>5
COPS2	1.67	1.37	1.16	1.95	4.01–5.00
COPS3	1.90	1.66	3.67	1.71	3.01–4.00
COPS4[a)]	1.60	1.64	2.40	1.80	2.01–3.00
COPS5-bait	1.04	0.91	1.39	0.31	1.01–2.00
COPS6	2.55	3.42	2.68	2.33	1
COPS7A	1.79	1.10	1.17	3.59	0.51–0.99
COPS7B	10.00	4.39	3.17	3.52	0.01–0.50
COPS8	1.28	0.45	0.88	0.26	ND (#)
CUL1	1.78	1.32	2.11	1.08	
CUL2	1.20	1.18	0.61	0.97	
CUL3	2.34	1.13	1.62	1.40	
CUL4A	1.67	1.47	1.09	1.27	
CUL4B	1.27	1.16	0.93	2.38	
DDB1[a)]	2.14	1.72	1.34	1.10	
DDB2	1.77	1.45	1.23	1.35	
VPRBP	3.00	5.00	2.25	1.67	
LRRC14	0.77	3.23	1.31	1.00	
DCAF11	1.27	0.58	1.38	0.83	
BTBD2	3.00	5.67	2.67	4.00	
PPIL5	ND (9)	1.11	0.77	ND (4)	
ERCC8	0.40	ND (5)	ND (5)	4[b)]	
FBX017	1.67	0.67	0.78	0.44	
(B)					
COPS1	1.14	1.59	0.90	0.31	>5
COPS2	0.98	0.72	1.02	0.26	4.01–5.00
COPS3	1.16	0.94	1.10	0.43	3.01–4.00
COPS4[a)]	1.13	0.74	0.94	0.73	2.01–3.00
COPS5-bait	0.87	0.55	0.80	0.24	1.01–2.00
COPS6	0.76	1.18	1.35	0.15	1
COPS7A	1.24	0.56	0.88	0.68	0.51–0.99
COPS7B	2.00	1.75	1.08	1.50	0.01–0.50
COPS8	0.84	0.39	0.74	0.15	ND (#)
CUL1	1.44	1.75	1.56	0.53	
CUL2	1.00	1.33	0.92	0.45	
CUL3	1.12	1.15	1.39	0.59	
CUL4A	1.06	1.39	1.33	0.44	
CUL4B	1.14	1.27	1.04	0.86	
DDB1[a)]	1.18	1.28	0.96	0.45	
DDB2	1.44	1.44	1.33	0.50	
VPRBP	2.67	4.33	2.00	0.93	
LRRC14	0.75	3.25	1.38	0.53	
DCAF11	1.18	0.82	1.27	0.47	
BTBD2	1.33	1.67	1.67	1.50	
PPIL5	ND (4)	2.00	1.00	ND (4)	
ERCC8	0.50	ND (4)	ND (4)	2[b)]	
FBX017	1.00	0.50	0.67	0.29	

RP-RP, GeLC, 1-D RP (Triple TOF) fold increase calculated relative to 1-D LTQ sample A. MudPIT fold increase calculated relative to 1-D ORB.

a) COPS4 and DDB1 spectra were detected in FLAG alone controls, but these proteins were not filtered out due to substantial increase in spectra in FLAG-COPS5 purifications.

b) Number of peptides detected in MudPIT sample, none detected in ORB sample; ND (♯), not detected in 2-D sample (number peptides detected in 1-D RP sample).

semi-supervised modeling, and analyzed the effects of performing RP-RP fractionation to identify interactors for eIF4A2 and RAF1 (we used the AvgP, averaged probability, across both biological replicates). In both cases, SAINT identified a higher number of putative interactors with Avg $p \geq 0.9$, an average of 65 versus 48 and 118 versus 62 when RP-RP fractionation was employed for either eIF4A2 or RAF1, respectively (Supporting Information Fig. 1 and

Supporting Information Table 11). This confirms our hypothesis that the increase in spectral count can be useful, depending on the downstream analysis.

4 Discussion

Multidimensional separation is typically applied to high complexity samples and results in the identification of both more peptides and proteins than standard 1-D LC-MS/MS [15, 16, 32]. However, here we found that when applied to lower complexity AP samples, the benefit of multidimensional separation is not as evident. We observed that multidimensional separation robustly increased the number of spectra observed by MS; however, the effect on unique peptides or proteins observed was more modest, and it was unclear whether the additional identified proteins were true interaction partners or low-level contaminants (Tables 1 and 2).

A paired *t*-test using RP-RP LTQ data for bait proteins which had paired samples separated in 1-D (eIF4A2, RAF1, MEPCE) revealed that interactors annotated in BioGRID had a statistically significant ~2.2-fold spectral enrichment ($p = 0.0178$) while unique peptides and number of unique proteins were only mildly enriched (1.15- and 1.05-fold, respectively; not statistically significant; Supporting Information Table 12).

Most notably, the increase in the number of spectra detected by multidimensional separation may prove useful if spectral counts are used either for quantification [30] or statistical noise filtering, as in the SAINT analysis discussed above [6]. In these cases, additional spectral observations may increase the confidence that an identified protein is a biologically relevant interactor, and/or provide a better estimate of its relative abundance in a given sample. Note, however, that there is a significant level of variation in the spectral counts (before any filtering is applied) across our biological replicates, even in 1-D LC-MS/MS (Table 1C, first row, up to ~40% variation; the standard deviation between technical replicates is lower, at ~16%, Supporting Information Table 7). This variability will need to be considered with approaches based on spectral counting. In the SAINT example shown above, we have solely focused, for example, on those proteins that were statistically enriched in both of the eIF4A2 or RAF1 biological replicates.

Unfortunately, the increased sensitivity brought by multidimensional fractionation comes at the cost of increased handling and instrument time. For example, to obtain the average 2.17-fold increase in spectral counts yielded by RP-RP across all replicates of all four baits analyzed, mass spectrometer analysis time was increased from 2 to 12 h per sample, and the handling time was increased from 4 to 7 h. For GeLC, the 1.40-fold increase in sensitivity for COPS5 was attained with an increase in instrument time from 2 to 20 h, and an increase in sample preparation time from 4 to 8 h. Additionally, the loading of multiple sample fractions required for some multidimensional separation techniques (i.e. RP-RP, GeLC) increases the probability of sample loss due to the increased complexity of the platform and column clogging, column leakage or loss of spray. Finally, we note that in some cases, fractionation can induce selective loss of some of the proteins/peptides.

In this study, we have employed four baits with different cellular functions and/or intracellular localization and for which several interaction partners were already annotated in the BioGRID interaction database (to enable benchmarking). These also tend to lean toward the highest level of complexity for AP samples, with fairly large protein complexes being co-precipitated [4, 18, 20]. While the trends in spectral counts and unique peptide increases were observed for each of these baits, it is still possible that some baits bringing down an even more complex set of associating proteins may benefit to a larger degree.

Notably, we observed that employing an instrument with a faster scan rate enabled us to achieve results similar to those obtained by multidimensional separation on older mass spectrometers. Additionally, the reduced sample volume required to generate these results allows for smaller scale experiments to be performed at reduced cost, or additional technical replicates to be conducted in the same amount of time as a single multidimensional analysis. With the increased availability of fast-scanning mass spectrometers, it is likely that the advantages of multidimensional fractionation for AP-MS samples will be further decreased.

We thank Mariana Gomez for modifying the pcDNA5-FRT-FLAG vector, Michael Mullin for providing samples for early analyses, Frank Liu for help with data analysis and Brian Raught for critical review of the manuscript. Work in the Gingras Lab was supported by a grant from the Canadian Institutes of Health Research (MOP-84314). W.H.D. is supported by a grant from the Canadian Cancer Society Research Institute (CCSRI). A.C.G. is a Canada Research Chair in Functional Proteomics and the Lea Reichmann Chair in Cancer Proteomics. Work in the Kislinger Lab was supported by a grant from the Canadian Institutes of Health Research (MOP-93772). T.K. is a Canada Research Chair in Proteomics in Cancer Research.

The authors have declared no conflict of interest.

5 References

[1] Gingras, A. C., Gstaiger, M., Raught, B., Aebersold, R., Analysis of protein complexes using mass spectrometry. *Nat. Rev. Mol. Cell. Biol.* 2007, *8*, 645–654.

[2] Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M. et al., A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 1999, *17*, 1030–1032.

[3] Chen, G. I., Gingras, A. C., Affinity-purification mass spectrometry (AP-MS) of serine/threonine phosphatases. *Methods* 2007, *42*, 298–305.

[4] Sowa, M. E., Bennett, E. J., Gygi, S. P., Harper, J. W., Defining the human deubiquitinating enzyme interaction landscape. *Cell* 2009, *138*, 389–403.

[5] Breitkreutz, A., Choi, H., Sharom, J. R., Boucher, L. et al., A global protein kinase and phosphatase interaction network in yeast. *Science* 2010, *328*, 1043–1046.

[6] Sardiu, M. E., Cai, Y., Jin, J., Swanson, S. K. et al., Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. *Proc. Natl. Acad. Sci. USA* 2008, *105*, 1454–1459.

[7] Ewing, R. M., Chu, P., Elisma, F., Li, H. et al., Large-scale mapping of human protein–protein interactions by mass spectrometry. *Mol. Syst. Biol.* 2007, *3*, 89.

[8] Behrends, C., Sowa, M. E., Gygi, S. P., Harper, J. W., Network organization of the human autophagy system. *Nature* 2010, *466*, 68–76.

[9] Mak, A. B., Ni, Z., Hewel, J. A., Chen, G. I. et al., A lentiviral functional proteomics approach identifies chromatin remodeling complexes important for the induction of pluripotency. *Mol. Cell. Proteomics* 2010, *9*, 811–823.

[10] Lawo, S., Bashkurov, M., Mullin, M., Ferreria, M. G. et al., HAUS, the 8-subunit human Augmin complex, regulates centrosome and spindle integrity. *Curr. Biol.* 2009, *19*, 816–826.

[11] Goudreault, M., D'Ambrosio, L. M., Kean, M. J., Mullin, M. J. et al., A PP2A phosphatase high density interaction network identifies a novel striatin-interacting phosphatase and kinase complex linked to the cerebral cavernous malformation 3 (CCM3) protein. *Mol. Cell. Proteomics* 2009, *8*, 157–171.

[12] Chen, G. I., Tisayakorn, S., Jorgensen, C., D'Ambrosio, L. M. et al., PP4R4/KIAA1622 forms a novel stable cytosolic complex with phosphoprotein phosphatase 4. *J. Biol. Chem.* 2008, *283*, 29273–29284.

[13] Nakada, S., Tai, I., Panier, S., Al-Hakim, A. et al., Non-canonical inhibition of DNA damage-dependent ubiquitination by OTUB1. *Nature* 2010, *466*, 941–946.

[14] Horvatovich, P., Hoekman, B., Govorukhina, N., Bischoff, R., Multidimensional chromatography coupled to mass spectrometry in analysing complex proteomics samples. *J. Sep. Sci.* 2010, *33*, 1421–1437.

[15] Gilar, M., Olivova, P., Chakraborty, A. B., Jaworski, A. et al., Comparison of 1-D and 2-D LC MS/MS methods for proteomic analysis of human serum. *Electrophoresis* 2009, *30*, 1157–1167.

[16] Elschenbroich, S., Ignatchenko, V., Sharma, P., Schmitt-Ulms, G. et al., Peptide separations by on-line MudPIT compared to isoelectric focusing in an off-gel format: application to a membrane-enriched fraction from C2C12 mouse skeletal muscle cells. *J. Proteome Res.* 2009, *8*, 4860–4869.

[17] Schirle, M., Heurtier, M. A., Kuster, B., Profiling core proteomes of human cell lines by one-dimensional PAGE and liquid chromatography-tandem mass spectrometry. *Mol. Cell. Proteomics* 2003, *2*, 1297–1305.

[18] Jeronimo, C., Forget, D., Bouchard, A., Li, Q. et al., Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Mol. Cell.* 2007, *27*, 262–274.

[19] Washburn, M. P., Wolters, D., Yates, J. R., 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* 2001, *19*, 242–247.

[20] Breitkreutz, B. J., Stark, C., Reguly, T., Boucher, L. et al., The BioGRID Interaction Database: 2008 update. *Nucleic Acids Res.* 2008, *36*, D637–D640.

[21] Wei, N., Serino, G., Deng, X. W., The COP9 signalosome: more than a protease. *Trends Biochem. Sci.* 2008, *33*, 592–600.

[22] Tzivion, G., Luo, Z., Avruch, J., A dimeric 14-3-3 protein is an essential cofactor for Raf kinase activity. *Nature* 1998, *394*, 88–92.

[23] Stancato, L. F., Chow, Y. H., Owens-Grillo, J. K., Yem, A. W. et al., The native v-Raf.hsp90.p50 heterocomplex contains a novel immunophilin of the FK506 binding class. *J. Biol. Chem.* 1994, *269*, 22157–22161.

[24] Gingras, A. C., Raught, B., Sonenberg, N., eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu. Rev. Biochem.* 1999, *68*, 913–963.

[25] Gingras, A. C., Caballero, M., Zarske, M., Sanchez, A. et al., A novel, evolutionarily conserved protein phosphatase complex involved in cisplatin sensitivity. *Mol. Cell. Proteomics* 2005, *4*, 1725–1740.

[26] Taylor, P., Nielsen, P. A., Trelle, M. B., Horning, O. B. et al., Automated 2D peptide separation on a 1D nano-LC-MS system. *J. Proteome Res.* 2009, *8*, 1610–1616.

[27] Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., Mann, M., In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* 2006, *1*, 2856–2860.

[28] Keller, A., Nesvizhskii, A. I., Kolker, E., Aebersold, R., Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 2002, *74*, 5383–5392.

[29] Nesvizhskii, A. I., Keller, A., Kolker, E., Aebersold, R., A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 2003, *75*, 4646–4658.

[30] Choi, H., Larsen, B., Lin, Z. Y., Breitkreutz, A. et al., SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nat. Methods.* 2011, *8*, 70–73.

[31] Skarra, D. V., Goudreault, M., Choi, H., Mullin, M. et al., Label-free quantitative proteomics and SAINT analysis enable interactome mapping for the human Ser/Thr protein phosphatase 5. *Proteomics* 2011, *11*, 1508–1516.

[32] Gilar, M., Olivova, P., Daly, A. E., Gebler, J. C., Orthogonality of separation in two-dimensional liquid chromatography. *Anal. Chem.* 2005, *77*, 6426–6434.

