Characterization of Biases in Phosphopeptide Enrichment by Ti 4+ -Immobilized Metal Affinity Chromatography and TiO 2 Using a Massive Synthetic Library and Human Cell Digests

ARTICLE *in* ANALYTICAL CHEMISTRY · JULY 2014

Impact Factor: 5.64 · DOI: 10.1021/ac501803z · Source: PubMed

4 AUTHORS, INCLUDING:

Henk W P van den Toorn
Utrecht University
32 PUBLICATIONS **482** CITATIONS
SEE PROFILE

Albert J R Heck
Utrecht University
674 PUBLICATIONS **21,819** CITATIONS
SEE PROFILE

Shabaz Mohammed
University of Oxford
157 PUBLICATIONS **7,074** CITATIONS
SEE PROFILE

analytical chemistry

Article

pubs.acs.org/ac

Characterization of Biases in Phosphopeptide Enrichment by Ti^{4+}-Immobilized Metal Affinity Chromatography and TiO_2 Using a Massive Synthetic Library and Human Cell Digests

Lucrece Matheron,[†] Henk van den Toorn,[†] Albert J. R. Heck,[†] and Shabaz Mohammed*[,†,‡,§]

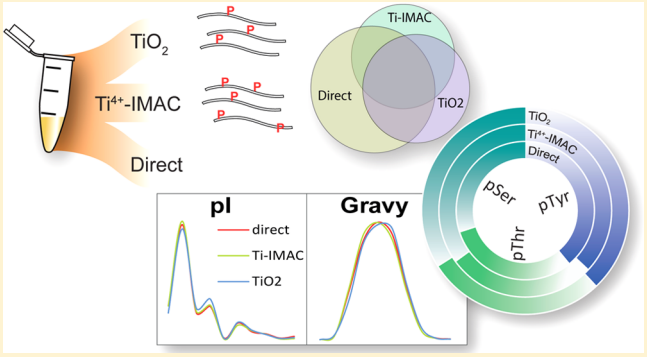
[†]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute of Pharmaceutical Sciences and The Netherlands Proteomics Centre, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands

[‡]Chemistry Research Laboratory, Department of Chemistry, University of Oxford, Mansfield Road, OX1 3TA, Oxford, United Kingdom

[§]Department of Biochemistry, University of Oxford, South Parks Road, OX1 3QU, Oxford, United Kingdom

Supporting Information

ABSTRACT: Outcomes of comparative evaluations of enrichment methods for phosphopeptides depend highly on the experimental protocols used, the operator, the source of the affinity matrix, and the samples analyzed. Here, we attempt such a comparative study exploring a very large synthetic library containing thousands of serine, threonine, and tyrosine phosphorylated peptides, being present in roughly equal abundance, along with their nonphosphorylated counterparts, and use an optimized protocol for enrichment by TiO_2 and Ti^{4+}-immobilized metal affinity chromatography (IMAC) by a single operator. Surprisingly, our data reveal that there are minimal differences between enrichment of phosphopeptides by TiO_2 and Ti^{4+}-IMAC when considering biochemical and biophysical parameters such as peptide length, sequence surrounding the site, hydrophobicity, and nature of the amino acid phosphorylated. Similar results were obtained when evaluating a tryptic digest of a cellular lysate, representing a more natural source of phosphopeptides. All the data presented are available via ProteomeXchange with the identifier PXD000759.



In addition to the identification of proteins, one of the crucial current challenges in proteomics is to identify and localize the wide range of post-translational modifications (PTMs) present. Protein phosphorylation is among the most important and best explored PTMs due to its key involvement in the regulation of many biological processes. Because of the low in vivo abundance of phosphorylation events, a dedicated enrichment step is necessary. The main enrichment methods are based on targeted affinity chromatography, such as immobilized metal affinity chromatography (IMAC)[1,2] with Fe(III),[3] Ga(III),[4] or Ti(IV)[5,6] and metal oxide affinity chromatography (MOAC) with mostly TiO_2.[7,8] Alternatively, specific antibodies targeting phosphorylated tyrosine residues have been successfully developed,[9,10] but immuno-affinity strategies for serine and threonine phosphorylation currently still require motif-specific antibodies.[11,12]

Affinity chromatography and immuno-affinity enrichment methods, often in conjunction with extensive peptide fractionation, have in the past decade been successfully used in high-throughput studies and allowed the identification of tens of thousands of phosphorylation sites.[6,13−15] Still no phosphoproteome has been mapped to completion. Many factors render the full detection of the phosphoproteome challenging, among which are its high complexity and its highly dynamic regulation. Another issue could be the existence of experimental biases in enrichment methods or use of proteases that would only target specific subsets of the phosphoproteome.[16] One of the main difficulties to determine these potential biases is that phosphopeptides are difficult to detect without enrichment.

Many recent studies have attempted to detect specificities by evaluating the characteristics of the enriched peptides or by comparing them across several procedures. Although data are accumulating, there is still debate on this subject. It is commonly hypothesized that each enrichment method targets specific sequences, since relatively small overlaps have been found between Fe^{3+}-IMAC and TiO_2[17] or Ti^{4+}-IMAC and TiO_2.[18] These materials are believed to have different affinities for phosphopeptides, and the current consensus ranking is $TiO_2 \approx Ti^{4+}$-IMAC[19] > Fe^{3+}-IMAC[5,17,20,21] > Ga^{3+}-IMAC.[22] In terms of the phosphorylated amino acid, TiO_2 showed an unbiased level of enrichment toward serine, threonine, and

Received: May 14, 2014
Accepted: July 28, 2014
Published: July 28, 2014

   dx.doi.org/10.1021/ac501803z | Anal. Chem. 2014, 86, 8312−8320

tyrosine residues.[23] In terms of the enriched sequences, an over-representation of acidic[17,18,24,25] and hydrophobic motifs[18] was observed when compared to the human genome.

Here, we report on our evaluation of the performance of one of the most well-established phosphopeptide enrichment protocols, TiO_2, and of Ti^{4+}-IMAC. We used a, recently described, large synthetic library[26] containing tens of thousands of peptides equally phosphorylated on serine, threonine, and tyrosine residues. Although synthetic, the sequences are based on naturally occurring phosphopeptides and offer a unique resource to test enrichment performance. We also used human cell digests. In essence, we find that phosphopeptides enriched by TiO_2 or Ti^{4+}-IMAC do not show any substantial sequence bias toward the characteristics studied, although some were only enriched by one method.

MATERIALS AND METHODS

Synthetic Phosphopeptide Library. We used the large set of 96 reference synthetic phosphopeptide libraries described recently[26] that were graciously provided by the Kuster laboratory (Technische Universität München).

Cell Culture, Lysis, and Digestion. HeLa cells were grown and digested as previously described.[6] Briefly, cells were lyzed by gentle sonication on ice. Proteins were reduced with 4 mM dithiothreitol at 56 °C for 30 min and alkylated with 8 mM iodoacetamide at room temperature for 30 min in the dark, and 4 mM DTT was added again. A first digestion was carried out with Lys-C at an enzyme-to-protein ratio of 1:75 (w/w) at 37 °C for 4 h. The samples were diluted 4 times and further digested with trypsin at an enzyme-to protein-ratio of 1:100 (w/w) at 37 °C overnight.

Phosphopeptide Enrichment by Ti^{4+}-IMAC and TiO_2. Enrichment with Ti^{4+}-IMAC and TiO_2 was performed following the protocol described in detail for Ti^{4+}-IMAC.[19] The synthetic libraries were resuspended in 40 μL of 10% FA (1 nmol/μL). A slurry of 10 mg/mL of the Ti^{4+}-IMAC or TiO_2 beads (Sachtopore, Sachtleben Chemie, Germany) was prepared in 30% acetonitrile (ACN)/0.1% trifluoroacetic acid (TFA) (v/v). 40 μL of the slurry was packed by centrifugation (100*g*) in a GELoader using a C8 plug. The microcolumns were conditioned with 50 μL of loading buffer (80% ACN, 6% TFA) by centrifugation (150*g*). The samples were resuspended in loading buffer (8 nmol for the synthetic libraries, 100 μg for the HeLa digest) and loaded on the microcolumns by centrifugation (100*g*). The microcolumns were washed with (i) 50 μL of 50% ACN/0.5% TFA/200 mM NaCl and (ii) 50 μL of 50% ACN/0.1% TFA by centrifugation (150*g*). Phosphopeptides were eluted into 25 μL of 10% FA with (i) 20 μL of 10% NH_4OH and (ii) 3 μL of 80% ACN/2% FA by centrifugation (100*g*). 3 μL of 100% FA was added for further acidification. The libraries were dried down in vacuo and resuspended in 10% FA; the HeLa digests were used as is.

LC-MS/MS Analysis. Samples (2 nmol for the direct analysis of the synthetic libraries, 50% of the enriched libraries and 30% of the enriched HeLa digests) were analyzed by an LTQ-Orbitrap Elite (Thermo Fisher Scientific, Bremen) equipped with an electron transfer dissociation (ETD) source (Thermo Fisher Scientific) connected to an ultra HPLC (UHPLC) Proxeon EASY-nLC 1000 (Thermo Scientific). Peptides were trapped on a double-fritted trap column (Dr. Maisch, Reprosil C18; 3 μm, 2 cm × 100 μm) and separated on an analytical column (Agilent, Poroshell 120 EC-C18; 2.7 μm, 50 cm × 50 μm). Solvent A was 0.1% FA, and solvent B was 0.1% FA in ACN (Biosolve). Samples were loaded at a pressure of 800 bar with 100% solvent A. Peptides were separated at a flow rate of 100 nL/min by the following gradient: 7% to 30% solvent B in 91 min; 30% to 100% solvent B in 3 min; 100% solvent B for 5 min; 100% to 7% solvent B in 1 min; 7% solvent B for 20 min. The LTQ-Orbitrap Elite was operated in positive ion and data dependent acquisition mode. FT full-scan spectra were acquired at 60 000 resolution. The 20 most intense precursors were selected for CID or ETD fragmentation under control of an in-house developed data dependent decision tree (CID fragmentation for 2+ precursors, 3+ and 4+ precursors with m/z > 800, 5+ precursors with m/z > 950, ETD fragmentation otherwise; isolation width 1.5 Da, AGC target 5000, ETD reagent ion AGC target 1×10^5, ETD reaction time 50 ms).[27] Dynamic exclusion was enabled for a 40 s repeat duration and a 40 s exclusion duration with a repeat count of 1.

Data Analysis. Raw data corresponding to the synthetic libraries and HeLa digests were analyzed with Proteome Discoverer (version 1.3, Thermo). Mascot (Matrix Science, version 2.4) was used to search the MS/MS data against the human SwissProt database (release 2013_07, 20 272 entries) (for the libraries, the database was supplemented with 96 entries, each containing concatenations of all theoretically possible peptides within a synthetic library). Parameters included: trypsin cleavage (5 missed cleavages for library and 2 for HeLa); precursor mass tolerance ±50 ppm; product ion tolerance 0.6 Da; oxidation of methionines and phosphorylation (S, T, Y) as variable modifications; the phosphoRS node (version 3.1)[28] was used for site localization. Peptide-to-spectrum match (PSM) areas were calculated. For the library, the PSMs were filtered sequentially according to the following criteria, in this order: (i) search engine rank 1; (ii) 0 < Δscore < 1; (iii) sequence in agreement with the corresponding library sequence, with 0 or 1 phosphorylation site and the phosphorylation site in the right position; (iv) Mascot score ≥40; (v) pRS site probability ≥99% for the phospho PSMs. PSMs were grouped into peptide sequences. The area was calculated as the median of the PSM areas; the score was the one of the highest scoring PSM.

For HeLa digests, a decoy database search was enabled. The PSMs were filtered according to the following criteria: (i) false discovery rate (FDR) of less than 1% (Percolator-based algorithm[29,30]); (ii) Mascot score ≥20; (iii) search engine rank 1; (iv) pRS site probability ≥99%. PSMs were grouped into peptide sequences. Peptide areas were calculated as for the library peptides.

The package Bio.SeqUtils.ProtParam from Biopython 1.63[31] on Python 2.7.5 32-bit for Windows was used to calculate theoretical pI and Gravy hydropathy indexes. IceLogo[20] was used to determine potential over- or under-represented sequence motifs. IceLogo compares the experimental data set with a chosen background data set in order to determine, with a set confidence interval (*p*-value of 0.05 here), if the amino acid compositions differ. For the library experiments, the experimental set was generated by extracting a 3 amino acid sequence with the phosphorylation site in the middle. The background was artificially generated as 3 amino acids sequences with a serine, threonine, or tyrosine in the middle and all possible amino acids combinations in positions −1 and +1. For the Hela experiments, the negative and positive sets were generated by extracting an 11 amino acid sequence with the site in position 6. Linear correlation parameters were calculated with the least-squares method.

dx.doi.org/10.1021/ac501803z | *Anal. Chem.* 2014, 86, 8312−8320

RESULTS AND DISCUSSION

In order to evaluate phosphopeptide enrichment by Ti^{4+}-IMAC and TiO_2, we primarily used a set of 96 reference synthetic libraries.[26] Their sequences are based on 96 singly phosphorylated peptides unambiguously identified in the human proteome, covering a wide range of characteristics such as length (9 to 27 amino acids), relative site position, and hydrophobicity. From each of these 96 seed sequences, a library was generated by combinatorial chemistry. The phosphorylated site was permuted to a serine, threonine, or tyrosine, phosphorylated or not, and the 2 amino acids around it (position −1 and +1) were permuted for all 20 amino acids. These libraries theoretically contain more than 200 000 known sequences. Half of these are phosphorylated, equally on serine, threonine, and tyrosine residues. In addition, the cross-evaluation of Ti^{4+}-IMAC and TiO_2 was also applied to a HeLa cell digest, closer to the biological reality of phosphoproteomics experiments.

Evaluation of Ti^{4+}-IMAC and TiO_2 Enrichment in the Synthetic Libraries. *Experimental Design and Data Processing.* Each library was enriched in parallel by Ti^{4+}-IMAC and TiO_2 by a protocol originally optimized for Ti^{4+}-IMAC[19] but also providing optimal results for TiO_2. The libraries were analyzed directly and in parallel after the two enrichment procedures. All the comparisons were therefore made between these three conditions. Due to the high similarity of the peptide sequences, the resulting peptide-to-spectrum matches (PSMs) were manually filtered. The goal of the study was not to identify as many phosphopeptides as possible, but rather to study trends based on unambiguous identification and site localization (see the Materials and Methods for details of the stringent filtering). The PSMs were grouped into peptide sequences; a peptide area was calculated as the median of all PSMs areas.

Global Evaluation of the Analyses. A detailed overview of the results for each library is given in Table S1, Supporting Information. Combining the results across all libraries, the direct analysis allowed the quantification of 15 990 phosphopeptides and the analysis after Ti^{4+}-IMAC of 10 956 and the one after TiO_2 of 11 795 phosphopeptides. The overlap between the analyses in the 3 conditions is shown in Figure S1, Supporting Information. 5006 phosphopeptides are common between the 3 conditions, which represents approximately 1/3rd of the peptides quantified directly and almost half of those quantified after enrichment. With the chosen search criteria, no library was fully identified, and the overlap is relatively poor, which has to be related to (i) the synthesis itself, as was shown in the initial study presenting the libraries,[26] (ii) random sampling in the MS process, and (iii) the stringent filters chosen. However, it might also be possible that each condition (direct analysis, Ti^{4+}-IMAC, and TiO_2) favors a distinct subset of the library phosphopeptides, and this is the hypothesis being tested by this study. We thus determined if the phosphopeptides in the 3 conditions could be distinguished by the properties commonly identified in the literature as having a potential influence on the enrichment.

Nature of the Phosphorylated Amino Acid. A crucial point when dealing with phosphopeptide enrichment is whether the procedure discriminates against one of the three commonly phosphorylated amino acids. From several large-scale phosphoproteomics experiments on human cell lines, the percentages of phosphorylated serine, threonine, and tyrosine residues are typically of 90%, 10%, and 0.1%, respectively.[32] The question remains if these proportions are truly reflecting the state of the sample, as they are mostly calculated after enrichment protocols. The libraries are especially suited to answer this, because the three phosphorylated amino acids are present in similar abundances, and their ratio can be determined before and after enrichment.

Figure 1 shows the proportion of quantified phosphorylated serine, threonine, and tyrosine residues, in terms of number of peptides identified (Figure 1A) and of total peptide abundance (Figure 1B). In both measurement methods, the proportion for each of the residues in the 3 protocols (direct, after Ti^{4+}-IMAC, and after TiO_2) is very similar, within experimental variability. This shows that the enrichment methods do not favor one type of phosphorylated residue in this synthetic library-based experiment.

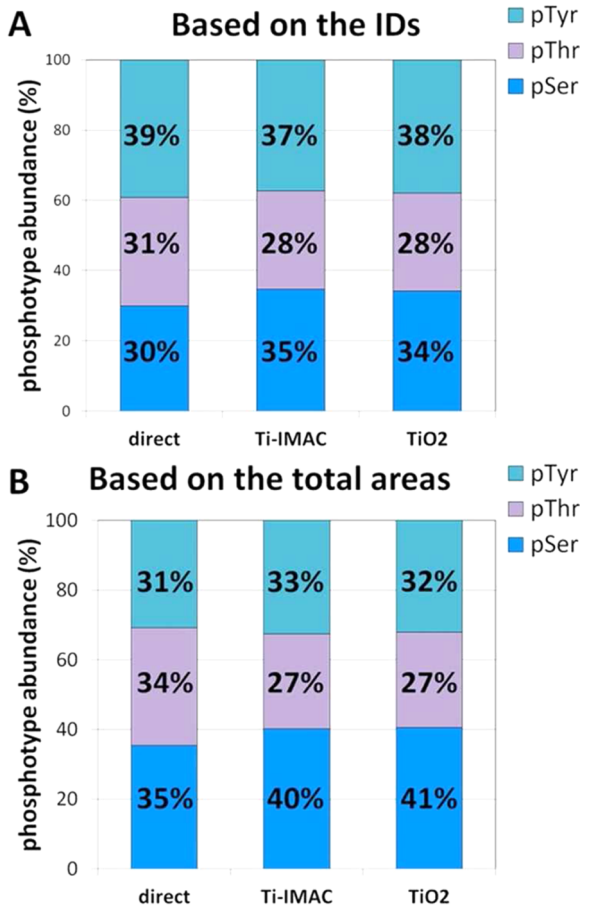


Figure 1. Relative distribution of phosphorylated amino acids within the identified library peptides in terms of (A) number of identifications and (B) total areas extracted from ion chromatograms.

In terms of number of identifications, peptides bearing a phosphorylated tyrosine are over-represented, similarly in all 3 protocols. In mass spectrometry, phosphotyrosine residues tend to be identified with higher scores and better localized than phosphorylated serines and threonines,[26] because of the less abundant neutral losses during CID or HCD fragmentation.[33,34] By applying stringent Mascot and localization score thresholds, we would favor the identification of peptides bearing phosphorylated tyrosine residues. This over-representation is not seen when comparing the areas, many of the

dx.doi.org/10.1021/ac501803z | Anal. Chem. 2014, 86, 8312−8320

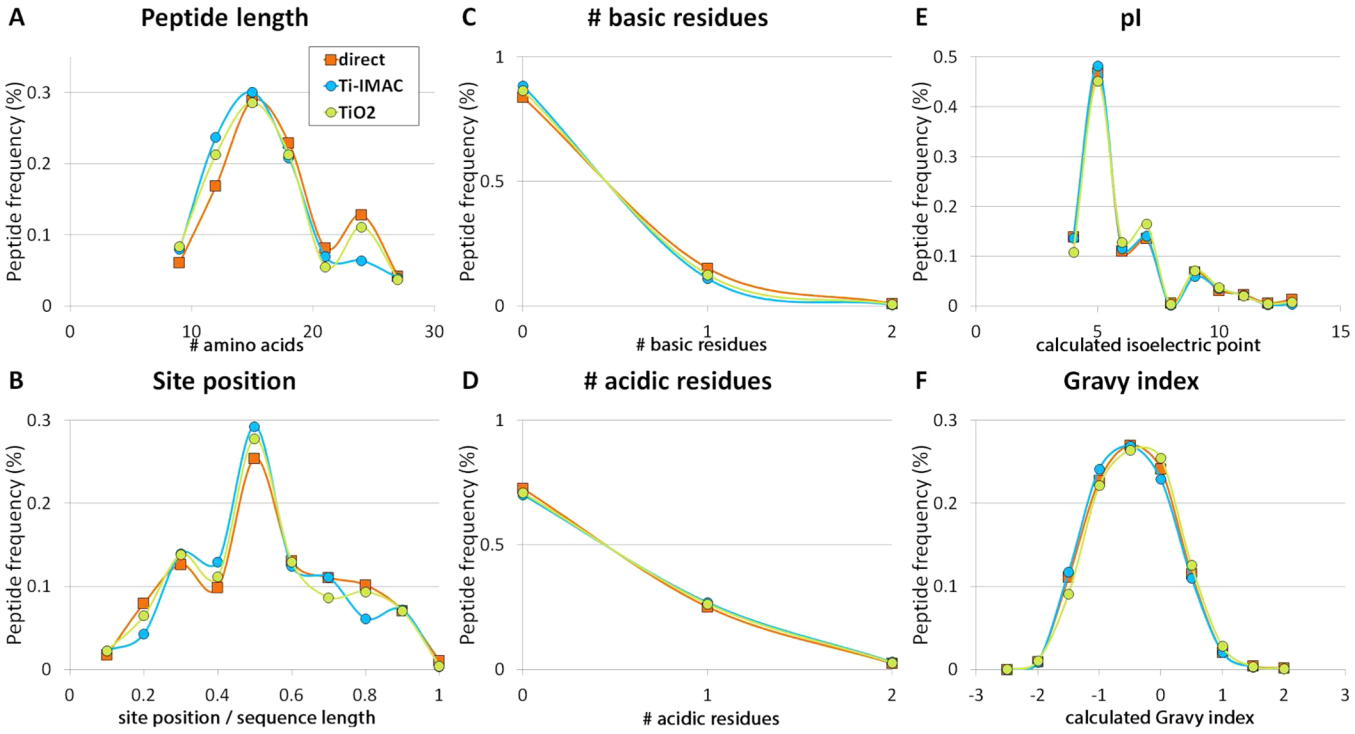


Figure 2. Frequency plots representing profiles of physicochemical characteristics of the quantified library phosphopeptides, in the direct analysis (orange squares), the analysis after Ti^{4+}-IMAC (blue circles), and the analysis after TiO_2 (green circles). The *y*-axis shows the peptide frequency in percentage. The properties shown are (A) peptide length, (B) relative position of the phosphorylation site (ratio site position/peptide length), (C) number of basic residues in the 3 permuted amino acids (phosphorylation site, positions −1 and +1), (D) number of acidic residues in the 3 permuted amino acids, (E) calculated pI of the peptides, and (F) calculated Gravy hydropathy index of the peptides.

peptides with a phosphorylated tyrosine residue having a relatively low intensity (Figure S2, Supporting Information). At the area level, the peptides bearing phosphorylated serine residues are the most abundant.

Characteristics of the Peptides Identified. The peptides identified in the direct analysis and after enrichment were compared in terms of their global biochemical and biophysical properties. We focused on the peptide length, the relative position of the phosphorylation site, the peptide pI, and the Gravy hydropathy index. In addition, the nature of the 2 permuted amino acids in positions −1 and +1 was also considered. In Figure 2, frequency plots for these characteristics are shown for the peptides observed in the direct analysis and for those enriched by Ti^{4+}-IMAC or TiO_2. Most of the trends identified here for the direct analysis are in good agreement with what was already published on these synthetic phosphopeptides.[26]

The distribution of the peptide length in the 3 experiments is shown in Figure 2A, with very similar trends. The libraries most successfully analyzed ranged between 12 and 18 amino acids, in which 70−75% of the peptides quantified were concentrated. This can not only be attributed to the library design, according to which that range should contain about 50% of the peptides (Figure S3A, Supporting Information). The libraries around 20 amino acids were less successful, in accordance to the generally less good performance of the long synthetic libraries,[26] causing a dip in the curve of Figure 2A.

The distribution of the relative position of the phosphorylation site is shown in Figure 2B. In the initial design of the libraries, if the site was in the first position, only the amino acid in position +1 could be permuted, and the theoretical number of phosphopeptides in that library is 60 instead of 1200. Similarly, if the site was in the one-before-last position, only the amino acid in position −1 could be permuted, to still mimic tryptic peptides. These two types of libraries are thus much smaller, hence their low frequency in Figure 2B. This figure shows that both enrichment protocols and direct analysis result in alike distributions. There is a preference for a phosphorylation site in the middle of the sequence, originating largely from the library design (Figure S3B, Supporting Information).

Figure 2C,D shows the proportion of basic (arginine and lysine here) and acidic (aspartic and glutamic acids) amino acids in the permuted positions −1 and +1. The curves observed with the three parallel protocols can essentially be superimposed, showing that the enrichment procedures do not discriminate on the basis of acido-basic balance. However, in the samples themselves, acidic residues are favored. Indeed, if the amino acid distribution was fully random, the probability of having one of two specific amino acids (i) in both positions −1 and +1 amounts to 1%, (ii) in either position −1 or +1 amounts to 18%, and (iii) in none of these two positions amounts to 81% (Figure S3C, Supporting Information). Consequently, the proportion of basic amino acids in the 2 permuted positions corresponds well to this random distribution (Figure 2C), whereas acidic residues are over-represented (Figure 2D). Since this is already observed in the direct analysis, it can be related to a more efficient synthesis or more likely to a facilitated LC-MS/MS identification of acidic phosphorylated peptides, amide bonds adjacent to acidic residues being prone to fragmentation.[35]

Figure 2E,F shows the distribution of the calculated isoelectric point, i.e., pI, and the Gravy hydropathy index of the phosphopeptides. Once again, the curves corresponding to the 3 protocols are highly similar. Figure 2E shows that acidic

dx.doi.org/10.1021/ac501803z | *Anal. Chem.* 2014, 86, 8312−8320

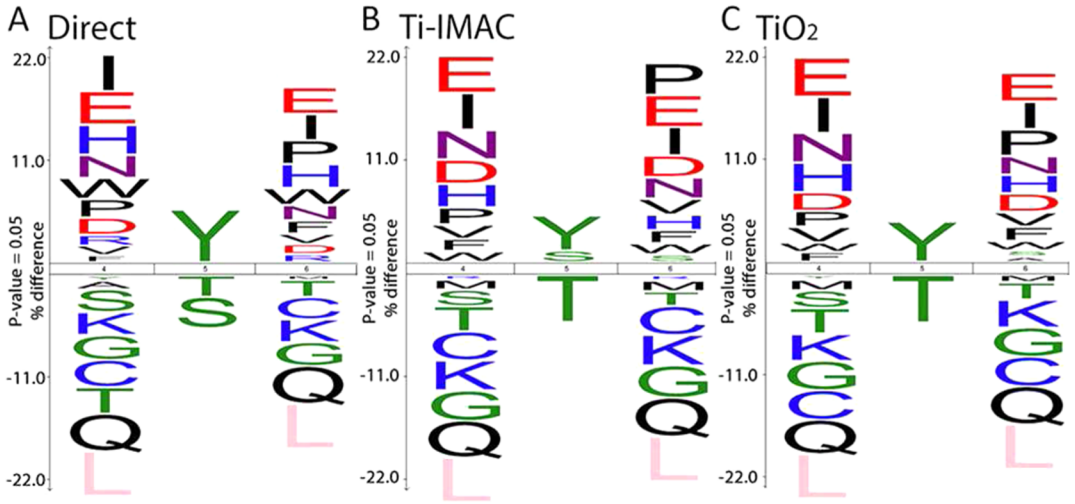


Figure 3. IceLogo profiles for the library phosphopeptides: motif analysis of the three permuted amino acids (phosphorylation site, positions −1 and +1) identified in (A) the direct analysis, (B) the analysis after Ti^{4+}-IMAC, and (C) the analysis after TiO_2. An artificial background was used consisting of three amino acids sequences where the medium position is Ser, Thr, or Tyr and the −1 and +1 positions are all possible combinations of two amino acids. The amino acid enrichment was filtered to a *p* value of <0.05. Over-represented amino acids are displayed on the top and under-represented ones on the bottom, and the *y*-axis displays their relative frequencies.

peptides tend to be overrepresented, as already exemplified by Figure 2C. On the other hand, the hydrophobicity is more balanced, with a slight tendency toward the hydrophilic side as shown by the apex of the Gravy index curve at −0.5. Altogether, these analyses suggest that the putative difference between the direct analysis and the enrichment methods, as well as the apparent complementarity between Ti^{4+}-IMAC and TiO_2, cannot be explained by a bias toward the peptide length, the site position, the acidic and basic neighboring residues, or the peptide isoelectric point and hydrophobicity

Favored Motifs for the Library Phosphopeptides. In order to have a global view of a potential sequence bias surrounding the phosphorylation site in the enrichment procedures, we performed a motif analysis of the 3 permuted amino acids (phosphorylation site, position −1 and +1) against an artificial 3-amino acids background (see Materials and Methods). Figure 3 shows these computed motifs. The first striking point is that the 3 different protocols are nearly indistinguishable. Any sequence bias is already present in the direct analysis. Whether this initial bias is due to the efficiency of the synthesis or the MS/MS identification is hard to determine. Although there are some sequence preferences, the actual proportion by which these amino acids are favored or disfavored is low. Also interesting is the high similarity between the positions −1 and +1. A certain type of amino acid can be preferred, but its actual position is not important. To further refine this analysis, we also plotted the same motifs for the sites specific to one of the three methods (not identified with the others) (Figure S4, Supporting Information).

Some already mentioned features are verified: phosphorylated tyrosine residues are over-represented, particularly in the sites specific to the direct analysis, as are acidic residues in positions −1 and +1. Lysine residues are disfavored, but the direct analysis slightly favors histidine and arginine residues which is in contrast to the sites identified after enrichment. There is also a small preference for proline residues around the phosphorylation site. Hydrophobic amino acids are slightly favored. The presence of an additional serine or threonine residue in position −1 or +1 is slightly disfavored, probably because the confident localization of a single phosphate group is harder with several potential sites in close proximity.

Altogether, none of the properties studied allowed us to distinguish between the phosphopeptides in the direct analysis or after enrichment by Ti^{4+}-IMAC or TiO_2. Importantly, we have shown that there is no preferential enrichment of phosphorylated serine, threonine, or tyrosine residues, although identification alone with stringent filtering criteria leads to a positive bias in favor of phosphotyrosine. Our analysis still reveals that some phosphopeptides are more efficiently enriched by either Ti^{4+}-IMAC or TiO_2, although we find no clear rules to predict.

Evaluation of Ti^{4+}-IMAC and TiO_2 Enrichment with HeLa Cell Digests. *Experimental Design and Data Processing.* Although we argue that, by design, these synthetic libraries are favorable for comparative enrichment studies, they do not represent a "real-life" sample. Therefore, we also investigated if our results were transferable to a typical phosphoproteomics experiment on a tryptic digest of a human cell lysate. The library experiments were performed in batches, and each time, a Ti^{4+}-IMAC and TiO_2 enrichment was also performed in parallel on 100 μg of a digest of HeLa cells. This resulted in 10 replicates for each of the enrichment methods, spanning over a 10 month period. These samples were analyzed and processed in a similar manner to the libraries, although in this case a target/decoy approach could and was used (see Materials and Methods).

Global Evaluation of the Enrichment. An overview of the results of the identification is given in Table S2, Supporting Information. The enrichment efficiencies for each of the replicates, calculated as the percentage of identified unique phosphopeptides, are shown in Figure S4, Supporting Information. On average, the enrichment by Ti^{4+}-IMAC allowed the identification of ≈1500 unique phosphopeptides, with a 90% specificity, whereas the TiO_2 procedure generated ≈1700 unique phosphopeptides with an 83% specificity. The Ti^{4+}-IMAC procedure is consistently more specific, while the TiO_2 is more reproducible. Moreover, by combining the 20 experiments, 4833 unique phosphopeptides and 5499 unique

dx.doi.org/10.1021/ac501803z | *Anal. Chem.* 2014, 86, 8312−8320

phosphorylation sites were identified. 60% are common to Ti^{4+}-IMAC and TiO_2 (Figure S5, Supporting Information), and each method selectively identifies approximately 20% of the sites. The complementarity of the two methods is best exemplified by Figure 4: For each enrichment method, the ten replicates were

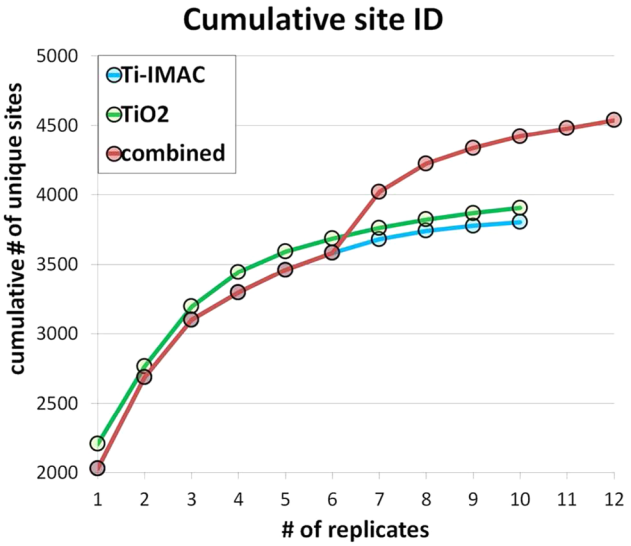


Figure 4. For the HeLa analyses, cumulative identification of unique phosphorylated peptides (i) for the 10 Ti^{4+}-IMAC replicates (blue), ranked in the order of new unique phosphopeptides they provided, (ii) for the 10 TiO_2 replicates (green), ranked similarly, and (iii) for the first 6 Ti^{4+}-IMAC replicates followed by the 6 first TiO_2 replicates (red), ranked similarly.

ranked by the number of new phosphopeptides they provided, and the cumulated number of identification was plotted. For both Ti^{4+}-IMAC (in blue) and TiO_2 (in green), six replicates were necessary to reach 95% of the identifications. We then took these six most informative Ti^{4+}-IMAC and TiO_2 replicates and calculated the cumulative identifications with these 12 experiments (in red). This lead to an increase in the number of identifications that could never be reached by using only one method.

We also examined how often, in the 10 replicates, each phosphopeptide was identified. For all the singly phosphorylated peptides identified in either Ti^{4+}-IMAC or TiO_2 (Figure S6A, Supporting Information), 25% of the peptides are only observed in 1 replicate, but there is still a decent number with a higher frequency of observation (approximately 400 peptides seen in 2 replicates, and 200 between 3 and 10 replicates). However, if we repeat this analysis specifically for the peptides observed only in Ti^{4+}-IMAC or TiO_2 (Figure S6B, Supporting Information), 63% of the TiO_2-specific peptides and 80% of the Ti^{4+}-IMAC-specific peptides are seen in only 1 of the 10 replicates. These numbers indicate that these sites might just be hard to enrich and detect. We also evaluated the run-to-run variability with the areas obtained across the 10 replicates (Figure S7, Supporting Information). The median coefficient of variation is 63% for Ti^{4+}-IMAC and 50% for TiO_2, while the average coefficient of variation is 99% for Ti^{4+}-IMAC and 90% for TiO_2.

Characteristics of the HeLa Peptides Identified. Figure S7, Supporting Information, provides an overview of the peptides identified with multiple phosphorylation sites. The average proportion of singly phosphorylated peptides was 89% for TiO_2 and 82% for Ti^{4+}-IMAC, which is a small albeit significant difference (p = 0.00087, two-tailed t test with unequal variances). This can partly explain the lower identifications with Ti^{4+}-IMAC: confident site localization is harder with multiple phosphorylation sites, and these peptides would more likely be filtered out by the 99% phosphoRS threshold applied. In Figure S6C,D, Supporting Information, the frequency of observation of multiply phosphorylated peptides *per* replicates is plotted. Comparing these figures with the corresponding graphs for singly phosphorylated peptides (Figure S6A,B, Supporting Information) shows that the proportions between Ti^{4+}-IMAC and TiO_2 are indeed reversed. This is the first clear bias that we detect between Ti^{4+}-IMAC and TiO_2.

We next evaluated the peptides' physicochemical characteristics after Ti^{4+}-IMAC and TiO_2 (Figure S8, Supporting Information). As in the library experiments, they are very comparable: thus, these 2 enrichment procedures have very similar performance and specificity.

As previously noted, the preferred peptide length is ~15 amino acids (Figure S8A, Supporting Information). The peptides tend to be longer in the analyses after enrichment. From the known content of the libraries, we already established that it is not because of the enrichment procedures themselves. It can in fact be explained by the higher number of missed cleavages in phosphopeptides than in normal peptides.[36,37] This is also illustrated by Figure S8C, Supporting Information, showing that the peptides after enrichment have a higher proportion of basic residues. The relative position of the phosphorylation site is more homogeneous than in the library experiments (Figure S8B, Supporting Information). There are more acidic residues in the phosphopeptides identified than in the peptides in the direct analysis (Figure S8D, Supporting Information). Finally, the curves showing the distribution of the isoelectric point (Figure S8E, Supporting Information) and of the Gravy hydropathy index (Figure S8F, Supporting Information) show the same trends as the library experiments.

Nature of the Phosphorylated Amino Acid. We, also, determined the proportion of phosphorylated serine, threonine, and tyrosine residues quantified in the HeLa cell digests after enrichment by Ti^{4+}-IMAC or TiO_2 (Table S2, Supporting Information). In addition, in Figure S9, Supporting Information, we plotted the proportion of sites occurring on a phosphorylated serine, threonine, and tyrosine residue obtained by combining the 10 replicates, in terms of identification or of total areas. Phosposerines dominate and the identified proportion is around 5−7% for phosphorylated threonine residues and 0.3% for phosphorylated tyrosine residues. These proportions are the same for Ti^{4+}-IMAC and TiO_2 and reproducible across the ten replicates. This is similar to previous findings for single-runs enrichments,[6,32] although other studies performed with several dimensions of separation reported proportions around 14−18% for phosphothreonines and 2−3% for phosphotyrosines.[24,36]

Favored Motifs Detected in the HeLa Phosphopeptides. We performed a motif analysis with IceLogo to compare the sites that were identified exclusively with only one of the protocols. The Ti^{4+}-IMAC sites are plotted on the top part of the graph, and the TiO_2 sites are on the bottom part (Figure S10A, Supporting Information). When all the specific sites are compared, regardless of how many phosphates the peptides they originate from were bearing, the main feature of the Ti^{4+}-IMAC-specific sites is a higher proportion of serine residues. In truth, the average number of serine *per* sequence in the Ti^{4+}-

dx.doi.org/10.1021/ac501803z | *Anal. Chem.* 2014, 86, 8312−8320

IMAC-specific sites (2.33) is only slightly higher as for the TiO_2-specific sites (2.17). However, Ti^{4+}-IMAC allowed the identification of more multiple phosphorylated peptides, and multiple phosphorylation sites on a peptide are close to each other. By aligning all the sequences with the phosphorylation site in position 6, we are artificially repeating them, shifting the position of the other serine residues a few residues apart. To avoid this artifact, the same motif comparison was performed for only the sites extracted from singly phosphorylated peptides (Figure S10B, Supporting Information). Note that this could skew the results for the Ti^{4+}-IMAC sites, because this method, more sensitive toward multiply phosphorylated peptides, leads to the identification of less singly phosphorylated ones. TiO_2-specific sites show a slight over-representation of both basic and acidic amino acids in the neighborhood of the phosphorylation site and of proline residues before the phosphate, whereas Ti^{4+}-IMAC rather favors prolines after the site. Whether these small differences are caused by the less complete view of the phosphoproteome for Ti^{4+}-IMAC is hard to determine. Nevertheless, a more complete picture could be offered by combining the two enrichment methods.

Correlation of the Peptide Areas after Ti^{4+}-IMAC and TiO_2. Finally, we also studied the correlation of the peptide abundances after the two enrichment procedures (Figure 5).

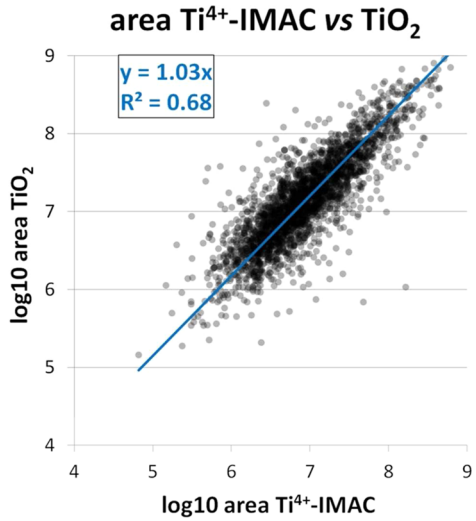


Figure 5. Correlation plot between the areas (logarithmic scale) extracted from ion chromatograms of phosphopeptides from the HeLa cells digest observed after enrichment by Ti^{4+}-IMAC or by TiO_2 and parameters of the linear regression in blue. The areas plotted are the median of the individual areas observed across the ten replicates.

The plot generates an R^2 of 0.68. Considering that this is a label free quantification across 10 replicates (over a 10 month period), this correlation is very satisfactory. Moreover, the slope of the linear correlation is 1.03, showing that, in a biologically relevant sample, the Ti^{4+}-IMAC and TiO_2 procedures are enriching phosphopeptides with comparable performance and lead to similar quantification. In addition, the phosphopeptides identified only by one method (which then have a null area for the other method) have an area spread over the whole range (Figure S11, Supporting Information), which shows that the abundance of phosphopeptides does not bias their enrichment by Ti^{4+}-IMAC or TiO_2 comparatively. Moreover, the multiply phosphorylated peptides (in red in Figure S11, Supporting Information), although more identified by Ti^{4+}-IMAC, show a trend similar to the singly phosphorylated ones.

DISCUSSION

The goal of this study was to determine if the phosphopeptide enrichment methods using Ti^{4+}-IMAC and TiO_2 were biased for a specific type of peptide. Establishing the performance of such widely used methods will not only give confidence to a laboratory using them but also aid interlaboratory comparison and reproducibility.

Comparative performance of enrichment methods is still a debate in the community. TiO_2 has been reported to outperform Fe^{3+}-IMAC[8,22] or to lead to a similar level of phosphopeptide identifications.[38] Due to its stronger affinity for phosphopeptides, TiO_2 is generally recognized as being more specific than Fe^{3+}-IMAC,[38,39] although this does not hold true for Ti^{4+}-IMAC.[40] On the basis of the experiments presented here, our main conclusion is that, although a significant number of phosphorylated peptides is only detected in one of our three protocols (direct analysis, enrichment by Ti^{4+}-IMAC and TiO_2), they do not differ significantly by any of the biochemical properties that we have studied.

Sequence specificities were observed in the detected phosphorylated peptides, predating the enrichment. In our data set, there is a preference for acidic, hydrophobic, and proline residues in the positions −1 and +1 with regards to the phosphorylation site. These motifs are commonly found in large-scale, unbiased phosphoproteomics studies, overwhelmed with the proline-directed S/TP motif[17,21,41] and with acidophilic kinases patterns,[17,18,24,25] while hydrophobic residues are slightly favored[18] and basic phosphopeptides are usually under-represented.[40,42] It is of interest that such motifs, although primarily coming from the biological activity of their corresponding kinases, are also slightly favored by the LC-MS/MS analysis. Importantly, we did not detect any preference for one type of phosphorylated amino acid, although relying on identification only with stringent filtering criteria leads to an over-representation of phosphorylated tyrosine residues. Thanks to the use of the synthetic libraries, we are able to state that, for these phosphopeptides and in single-run (2 h) LC-MS/MS analysis, the properties we have studied do not seem to induce a bias in phosphopeptides observed using the two enrichment procedures.

Although the libraries gave us very valuable information, their major drawback is that the phosphopeptide abundance is much higher than that observed in a biological sample. We thus applied the same enrichment procedures to HeLa cell digests. The main finding here, that could not be achieved with the synthetic libraries, is that Ti^{4+}-IMAC enriches multiply phosphorylated peptides better. Previous studies have already shown that Fe^{3+}-IMAC was better than TiO_2 for the enrichment of multiply phosphorylated peptides[17,43] and Ga^{3+}-IMAC better than Fe^{3+}-IMAC.[44] The affinities of these materials for phosphopeptides rank in the opposite order (TiO_2 > Fe^{3+}-IMAC > Ga^{3+}-IMAC), and multiply phosphorylated peptides might interact too strongly with TiO_2 to be eluted efficiently. The use of stronger elution conditions, such as sequential solution with 5% ammonium hydroxide, 5% piperidine, and 5% pyrrolidine[45] might improve the recovery of multiply phosphorylated peptides from TiO_2.

Contrarily to the library experiments, there is now significant sequence bias in the phosphopeptides quantified after enrichment, related to the previously cited kinase motifs. In these

dx.doi.org/10.1021/ac501803z | Anal. Chem. 2014, 86, 8312−8320

conditions, Ti^{4+}-IMAC and TiO_2 behave very similarly. However, the combination of both enrichment methods still allows an increase in the number of phosphorylated peptides observed, that could not be achieved by using a single protocol and performing more replicates. Despite our best effort presented in this study, we could not find any biochemical difference in these enrichment-specific sites. However, in the current study, the HeLa cell digests are analyzed in single runs. Differences between the two enrichment methods might become more distinct when digging deeper into the phosphoproteome through the use of fractionation but that may introduce new biases that would need to be addressed.

To summarize, analyzing 23 000 synthetic phosphopeptides and a large number of HeLa cells tryptic digests, we find that there are no clear differences between the pools of phosphopeptides observed using Ti^{4+}-IMAC and TiO_2 when considering general biochemical properties such as peptide length, site position, isoelectric point, hydrophobicity, and phosphorylation motif. The largest difference detected is the preferential enrichment of multiply phosphorylated peptides by Ti^{4+}-IMAC. We also observed that the abundance of the phosphopeptides enriched is rather similar when using these techniques. We expect that these findings will provide the community the reassurance that observations made in a phosphoproteomics study using TiO_2 or Ti^{4+}-IMAC can actually be related to the initial state of the sample and not to an artifact originating from a particular enrichment method.

■ 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

*E-mail: shabaz.mohammed@chem.ox.ac.uk.

Notes

The authors declare no competing financial interest.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository[46] with the data set identifier PXD000759.

■ ACKNOWLEDGMENTS

This work has been supported by the PRIME-XS project, grant agreement number 262067, funded by the European Union 7th Framework Programme, and by The Netherlands Organization for Scientific Research (NWO) with the VIDI grant for S.M. (700.10.429) and the large scale proteomics facility *Proteins@Work* (project 184.032.201) headed by A.J.R.H. The authors would like to thank Prof Dr. Bernhard Kuster and Dr. Simone Lemeer for the synthetic libraries and for many fruitful discussions, Harm Post for experimental assistance, and the PRIDE Team for their help in depositing the data.

■ REFERENCES

(1) Gruhler, A.; Olsen, J. V.; Mohammed, S.; Mortensen, P.; Faergeman, N. J.; Mann, M.; Jensen, O. N. *Mol. Cell Proteomics* **2005**, *4*, 310−327.

(2) Villen, J.; Gygi, S. P. *Nat. Protoc.* **2008**, *3*, 1630−1638.

(3) Andersson, L.; Porath, J. *Anal. Biochem.* **1986**, *154*, 250−254.

(4) Posewitz, M. C.; Tempst, P. *Anal. Chem.* **1999**, *71*, 2883−2892.

(5) Zhou, H.; Ye, M.; Dong, J.; Han, G.; Jiang, X.; Wu, R.; Zou, H. *J. Proteome Res.* **2008**, *7*, 3957−3967.

(6) Zhou, H.; Di Palma, S.; Preisinger, C.; Peng, M.; Polat, A. N.; Heck, A. J.; Mohammed, S. *J. Proteome Res.* **2013**, *12*, 260−271.

(7) Pinkse, M. W.; Uitto, P. M.; Hilhorst, M. J.; Ooms, B.; Heck, A. J. *Anal. Chem.* **2004**, *76*, 3935−3943.

(8) Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jorgensen, T. J. D. *Mol. Cell. Proteomics* **2005**, *4*, 873−886.

(9) Bergstrom Lind, S.; Molin, M.; Savitski, M. M.; Emilsson, L.; Astrom, J.; Hedberg, L.; Adams, C.; Nielsen, M. L.; Engstrom, A.; Elfineh, L.; Andersson, E.; Zubarev, R. A.; Pettersson, U. *J. Proteome Res.* **2008**, *7*, 2897−2910.

(10) Rikova, K.; Guo, A.; Zeng, Q.; Possemato, A.; Yu, J.; Haack, H.; Nardone, J.; Lee, K.; Reeves, C.; Li, Y.; Hu, Y.; Tan, Z.; Stokes, M.; Sullivan, L.; Mitchell, J.; Wetzel, R.; Macneill, J.; Ren, J. M.; Yuan, J.; Bakalarski, C. E.; Villen, J.; Kornhauser, J. M.; Smith, B.; Li, D.; Zhou, X.; Gygi, S. P.; Gu, T. L.; Polakiewicz, R. D.; Rush, J.; Comb, M. J. *Cell* **2007**, *131*, 1190−1203.

(11) White, C. D.; Toker, A. *Curr. Protoc. Mol. Biol.* **2013**, Chapter 18, Unit 18 20; DOI: 10.1002/0471142727.mb1820s101.

(12) Giansanti, P.; Stokes, M. P.; Silva, J. C.; Scholten, A.; Heck, A. J. *Mol. Cell. Proteomics* **2013**, *12*, 3350−3359.

(13) Zanivan, S.; Meves, A.; Behrendt, K.; Schoof, E. M.; Neilson, L. J.; Cox, J.; Tang, H. R.; Kalna, G.; van Ree, J. H.; van Deursen, J. M.; Trempus, C. S.; Machesky, L. M.; Linding, R.; Wickstrom, S. A.; Fassler, R.; Mann, M. *Cell Rep.* **2013**, *3*, 552−566.

(14) Bodenmiller, B.; Aebersold, R. *Methods Enzymol.* **2010**, *470*, 317−334.

(15) Jedrychowski, M. P.; Huttlin, E. L.; Haas, W.; Sowa, M. E.; Rad, R.; Gygi, S. P. *Mol. Cell. Proteomics* **2011**, *10*, M111 009910.

(16) Gauci, S.; Helbig, A. O.; Slijper, M.; Krijgsveld, J.; Heck, A. J.; Mohammed, S. *Anal. Chem.* **2009**, *81*, 4493−4501.

(17) Bodenmiller, B.; Mueller, L. N.; Mueller, M.; Domon, B.; Aebersold, R. *Nat. Methods* **2007**, *4*, 231−237.

(18) Lai, A. C.; Tsai, C. F.; Hsu, C. C.; Sun, Y. N.; Chen, Y. J. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 2186−2194.

(19) Zhou, H.; Ye, M.; Dong, J.; Corradini, E.; Cristobal, A.; Heck, A. J.; Zou, H.; Mohammed, S. *Nat. Protoc.* **2013**, *8*, 461−480.

(20) Colaert, N.; Helsens, K.; Martens, L.; Vandekerckhove, J.; Gevaert, K. *Nat. Methods* **2009**, *6*, 786−787.

(21) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 12130−12135.

(22) Cantin, G. T.; Shock, T. R.; Park, S. K.; Madhani, H. D.; Yates, J. R., 3rd *Anal. Chem.* **2007**, *79*, 4666−4673.

(23) Li, Q.; Shen, F.; Zhang, X.; Hu, Y.; Zhang, Q.; Xu, L.; Ren, X. *Anal. Chim. Acta* **2013**, *795*, 82−87.

(24) Di Palma, S.; Zoumaro-Djayoon, A.; Peng, M.; Post, H.; Preisinger, C.; Munoz, J.; Heck, A. J. *J. Proteomics* **2013**, *91*, 331−337.

(25) Yue, X. S.; Hummon, A. B. *J. Proteome Res.* **2013**, *12*, 4176−4186.

(26) Marx, H.; Lemeer, S.; Schliep, J. E.; Matheron, L.; Mohammed, S.; Cox, J.; Mann, M.; Heck, A. J.; Kuster, B. *Nat. Biotechnol.* **2013**, *31*, 557−564.

(27) Frese, C. K.; Altelaar, A. F.; Hennrich, M. L.; Nolting, D.; Zeller, M.; Griep-Raming, J.; Heck, A. J.; Mohammed, S. *J. Proteome Res.* **2011**, *10*, 2377−2388.

(28) Taus, T.; Kocher, T.; Pichler, P.; Paschke, C.; Schmidt, A.; Henrich, C.; Mechtler, K. *J. Proteome Res.* **2011**, *10*, 5354−5362.

(29) Kall, L.; Storey, J. D.; Noble, W. S. *Bioinformatics* **2008**, *24*, i42−48.

(30) Spivak, M.; Weston, J.; Bottou, L.; Kall, L.; Noble, W. S. *J. Proteome Res.* **2009**, *8*, 3737−3745.

(31) Cock, P. J.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; de Hoon, M. J. *Bioinformatics* **2009**, *25*, 1422−1423.

(32) Rainer, M.; Sonderegger, H.; Bakry, R.; Huck, C. W.; Morandell, S.; Huber, L. A.; Gjerde, D. T.; Bonn, G. K. *Proteomics* **2008**, *8*, 4593−4602.

(33) Boersema, P. J.; Mohammed, S.; Heck, A. J. *J. Mass Spectrom.* **2009**, *44*, 861−878.
(34) DeGnore, J. P.; Qin, J. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 1175−1188.
(35) Paizs, B.; Suhai, S. *Mass Spectrom Rev.* **2005**, *24*, 508−548.
(36) Molina, H.; Horn, D. M.; Tang, N.; Mathivanan, S.; Pandey, A. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 2199−2204.
(37) Gershon, P. D. *J. Proteome Res.* **2014**, *13*, 702−709.
(38) Wilson-Grady, J. T.; Villen, J.; Gygi, S. P. *J. Proteome Res.* **2008**, *7*, 1088−1097.
(39) Paradela, A.; Albar, J. P. *J. Proteome Res.* **2008**, *7*, 1809−1818.
(40) Zhou, H.; Low, T. Y.; Hennrich, M. L.; van der Toorn, H.; Schwend, T.; Zou, H.; Mohammed, S.; Heck, A. J. *Mol. Cell. Proteomics* **2011**, *10*, M110 006452.
(41) Schwartz, D.; Gygi, S. P. *Nat. Biotechnol.* **2005**, *23*, 1391−1398.
(42) Matheron, L.; Sachon, E.; Burlina, F.; Sagan, S.; Lequin, O.; Bolbach, G. *Anal. Chem.* **2011**, *83*, 3003−3010.
(43) Jensen, S. S.; Larsen, M. R. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 3635−3645.
(44) Tsai, C. F.; Hsu, C. C.; Hung, J. N.; Wang, Y. T.; Choong, W. K.; Zeng, M. Y.; Lin, P. Y.; Hong, R. W.; Sung, T. Y.; Chen, Y. J. *Anal. Chem.* **2014**, *86*, 685−693.
(45) Kyono, Y.; Sugiyama, N.; Imami, K.; Tomita, M.; Ishihama, Y. *J. Proteome Res.* **2008**, *7*, 4585−4593.
(46) Vizcaino, J. A.; Cote, R. G.; Csordas, A.; Dianes, J. A.; Fabregat, A.; Foster, J. M.; Griss, J.; Alpi, E.; Birim, M.; Contell, J.; O'Kelly, G.; Schoenegger, A.; Ovelleiro, D.; Perez-Riverol, Y.; Reisinger, F.; Rios, D.; Wang, R.; Hermjakob, H. *Nucleic Acids Res.* **2013**, *41*, D1063−1069.