Protein Phosporylation Hot Paper

DOI: 10.1002/anie.201409440

Screening for Protein Phosphorylation Using Nanoscale Reactions on Microdroplet Arrays**

Simon K. Küster, Martin Pabst, Renato Zenobi, and Petra S. Dittrich**

Abstract: *We present a novel and straightforward screening method to detect protein phosphorylations in complex protein mixtures. A proteolytic digest is separated by a conventional nanoscale liquid chromatography (nano-LC) separation and the eluate is immediately compartmentalized into microdroplets, which are spotted on a microarray MALDI plate. Subsequently, the enzyme alkaline phosphatase is applied to every second microarray spot to remove the phosphate groups from phosphorylated peptides, which results in a mass shift of $n \times -80$ Da. The MALDI-MS scan of the microarray is then evaluated by a software algorithm to automatically identify the phosphorylated peptides by exploiting the characteristic chromatographic peak profile induced by the phosphatase treatment. This screening method does not require extensive MS/MS experiments or peak list evaluation and can be easily extended to other enzymatic or chemical reactions.*

The investigation of post-translational protein modifications is pivotal for gaining a deeper understanding of protein functions and their roles in cellular mechanisms.[1] Amongst the large family of post-translational modifications (PTMs),[2] phosphorylation is found to be one of the most frequent.[3] It is estimated that approximately one third of all proteins in eukaryotic cells are phosphorylated at one point in their lifetime[4] and that approximately two percent of the human genome encodes kinases and phosphatases.[5] Despite the great regulatory importance of protein phosphorylation, their proteomic study is still far from being routine.

[*] S. K. Küster,[+] Dr. M. Pabst,[+] Prof. R. Zenobi, Prof. P. S. Dittrich[$]
Department of Chemistry and Applied Biosciences, ETH Zurich
Vladimir-Prelog-Weg 3, 8093 Zurich (Switzerland)
E-mail: martin.pabst@alumni.ethz.ch
petra.dittrich@bsse.ethz.ch

[$] Current address: Department of Biosystems Science and Engineering, ETH Zurich (Switzerland)

[+] These authors contributed equally to this work and share first authorship.

[**] This work was supported by the European Research Council (ERC Starting Grant nμLIPIDS; Grant No. 203428) and the Swiss KTI (Kommission für Technologie und Innovation; Grant No. 13123.1 PFNM-NM). Simon K. Küster acknowledges financial support from the scholarship fund of the Swiss Chemical Industry (SSCI). Furthermore, we thank Tobias Schibli from AB Sciex and Xiangyang Zhang from the MS-service of our institute. Moreover, we would like to thank Konstantins Jefimovs from EMPA Dübendorf and Fabian Wahl from Sigma–Aldrich for their support in microarray development.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201409440.

State-of-the-art methods to study protein phosphorylation commonly rely on bottom-up proteomic workflows using liquid chromatography (LC) coupled online to electrospray ionization (ESI) tandem mass spectrometry (MS/MS).[6] However, those and other common approaches suffer from various limitations and are far from delivering a comprehensive picture. Usually, in a single data-dependent analysis run, a substantial number of (mainly low intensity) peaks are not selected for fragmentation and consequently remain unnoticed. This is particularly problematic for the detection of phosphopeptides as they are known to be of low abundance and reported to exhibit poor ionization efficiencies[7] and poor fragmentation behavior.[8] These limitations often result in a large number of false positives as well as in a low probability score in automated MS/MS search algorithms.[9] False positives have also been reported as being introduced by sample preparation artifacts in combination with improper precursor selection,[10] or by insufficient mass accuracy and precision, which, for example, makes it impossible to discriminate between phosphorylation and sulfation.[11] Furthermore, depending on the scoring and reporting procedure, different tandem MS search algorithms can generate significantly different results.[12]

A common fragmentation-free approach employs enzymatic dephosphorylation prior to LC-MS analysis,[4] so that the removal of n phosphate groups results in an indicative mass shift of $n \times -80$ Da. However, this approach requires at least two consecutive LC-MS runs to detect the mass shift by peak lists comparison.[6e,g,13] This can be challenging as the removal of the phosphate group(s) influences the elution time of the peptides and thus, large elution areas from separate runs have to be correlated.[14]

Enzyme-based phosphopeptide identification has also been demonstrated using a workflow combining LC and matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) with intermediate on-plate digestion.[6a,7c] However, the commercially available fractionation devices deteriorate the separation by pooling eluate for dozens of seconds on one spot. This results in competitive ionization, which prevents the identification of low-abundant signals. Furthermore, an on-plate digest after MALDI matrix addition is hardly feasible.

Here we describe a novel and straightforward screening method to detect protein phosphorylation in complex protein mixtures. The method is based on the integration of on-plate nanoliter phosphatase reactions into a nano-LC-MALDI-MS workflow. This is achieved by using droplet microfluidics which has proven to be a powerful technology for (bio)chemical syntheses[15] and (bio)analytical applications.[16] Building on our recently introduced interface between droplet micro-

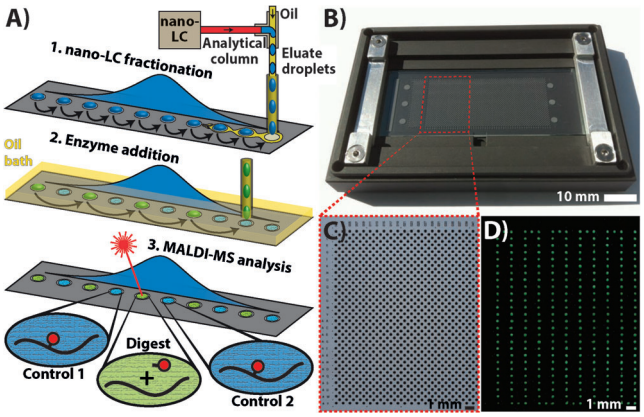

Figure 1. A) Workflow for phosphopeptide screening: 1) The nano-LC eluate is compartmentalized into picoliter droplets using droplet microfluidics and collected on a microarray plate with high temporal resolution (final volume of each fraction on the plate is 5 nL). 2) The microarray plate is placed under a protecting film of perfluorinated oil and every second fraction is digested by adding droplets which contain alkaline phosphatase. 3) MALDI matrix is added to each spot (not shown) and each spot is analyzed by MALDI-MS. B) Photograph of a microarray plate with 2780 hydrophilic spots (spot diameter: 300 μm) inside a sample holder. C) Micrograph showing the left part of the microarray after eluate spotting. D) Fluorescence micrograph showing the same field of view after enzyme addition. The enzyme droplets, which also contain low amounts of fluorescein (used to visualize the droplets), are only spotted on every second eluate fraction.

fluidics and MALDI-MS,[17] we used a modified platform that facilitates spotting of nano-LC eluate fractions on a high-density microarray with subsequent on-plate dephosphorylation (Figure 1). This is enabled by a new system, in which the microarray is immersed in a perfluorinated oil bath that prevents evaporation during the addition and incubation of the enzyme-containing droplets. The combination of the on-plate nanoliter enzymatic reactions and eluate fractionation with high temporal resolution enables the simultaneous detection of phosphorylated and enzymatically dephosphorylated fractions using the same nano-LC run. As a result, our method expands the standard analytical dimensions of LC-MS analysis, that is, retention time and mass, by adding a time-wise correlated enzymatic step—in this case the specific dephosphorylation—as a third analytical dimension. This enables a robust and straightforward phosphopeptide screening for both low- and high-abundant signals. Phosphorylated peptides are identified without a priori knowledge about the sample, without any sample enrichment steps, and without fragmentation experiments.

In the experimental workflow, the proteolytic digest of a protein mixture (containing α-casein S1 and S2, β-casein, and BSA) was injected into a reversed-phase nano-LC system. Subsequently, the outflow was fractionated at 1 Hz rates onto a microarray plate using droplet microfluidics.[17b] The eluate compartmentalization into picoliter droplets preserves the chromatographic separation and enables the fast and reliable deposition of eluate fractions (final volume of 5 nL; Figure 1 A1). Therefore, even narrow peaks of only a few seconds width were divided into multiple fractions and stored over several adjacent microarray spots containing virtually identical compositions.

After archiving the nano-LC run on the microarray plate, droplets containing alkaline phosphatase were deposited on every other spot on the microarray (Figure 1 A2 and S1 in the Supporting Information, SI). During this procedure, the microarray plate and the spotting capillary were immersed in a liquid bath of perfluorodecalin, which enabled elongated incubation times. After incubation, most of the oil was removed by decantation and the remaining traces of oil evaporated residue-free. Finally, MALDI matrix was spotted on every spot using the same procedure and the entire chip was analyzed by MALDI-MS.

The phosphatase treatment turns a highly resolved continuous peak of an eluting phosphopeptide into a characteristic zigzag pattern. At the same time, a peak originating from the dephosphorylated peptide with a mass shift of $n \times -80$ Da relative to the phosphopeptide arises solely on the treated spots generating a second zigzag pattern. The resulting characteristic interlocked zigzag pattern is demonstrated in Figure 2 for the phosphorylated peptide KTVDME(p)-STEVFTK (α-casein S2). The observed mass shift also shows the number of phosphorylations (here: one). Non-phosphorylated peptides or peptides with modifications of different chemical nature (e.g., sulfated peptides) are not affected by the enzymatic reaction (Figure 2 B).

Peaks appearing exclusively in control spots or exclusively in adjacent digest spots are therefore directly indicative of phosphorylated peptides and their dephosphorylated counterpart, respectively. This procedure significantly reduced the duration and complexity of data evaluation as compared to conventional approaches in which the dephosphorylation of a part of the sample is performed before separation.[6c,e,g] Furthermore, our data evaluation approach based on the detection of characteristic peak patterns turned out to be more specific than the more commonly used approach based on the comparison of large peak list tables. The reduced complexity of our method therefore paved the way for automated data evaluation.

We were able to develop a Matlab script to systematically screen the entire data set (containing 1430 MS spectra for 25 min LC run time) for the characteristic zigzag peak profiles. Figure 3 (row 1) shows details of three extracted ion chromatograms (XICs) representing a pair of a phosphorylated (A1) and its corresponding dephosphorylated (B1) peptide as well as a non-phosphorylated peptide (C1). To discriminate between the MS signals originating from these different peptide classes, the base-peak-intensity ratio of each spot (n) to the next spot ($n+1$) was calculated (row 2). Subsequently, the intensity ratio values were separated into two groups (rows 3 and 4). The intensity ratio of each control spot to its subsequent digest spot is only high for phosphorylated peaks, because the intensity was typically reduced to the noise level on digest spots due to the enzymatic dephosphorylation (A3). Similarly, high-intensity ratios of each digest spot to its subsequent control spot appear solely for dephosphorylated peptides, because they do not appear on control spots (B4). Peaks from non-phosphorylated peptides do not show up after this procedure (C3, C4).

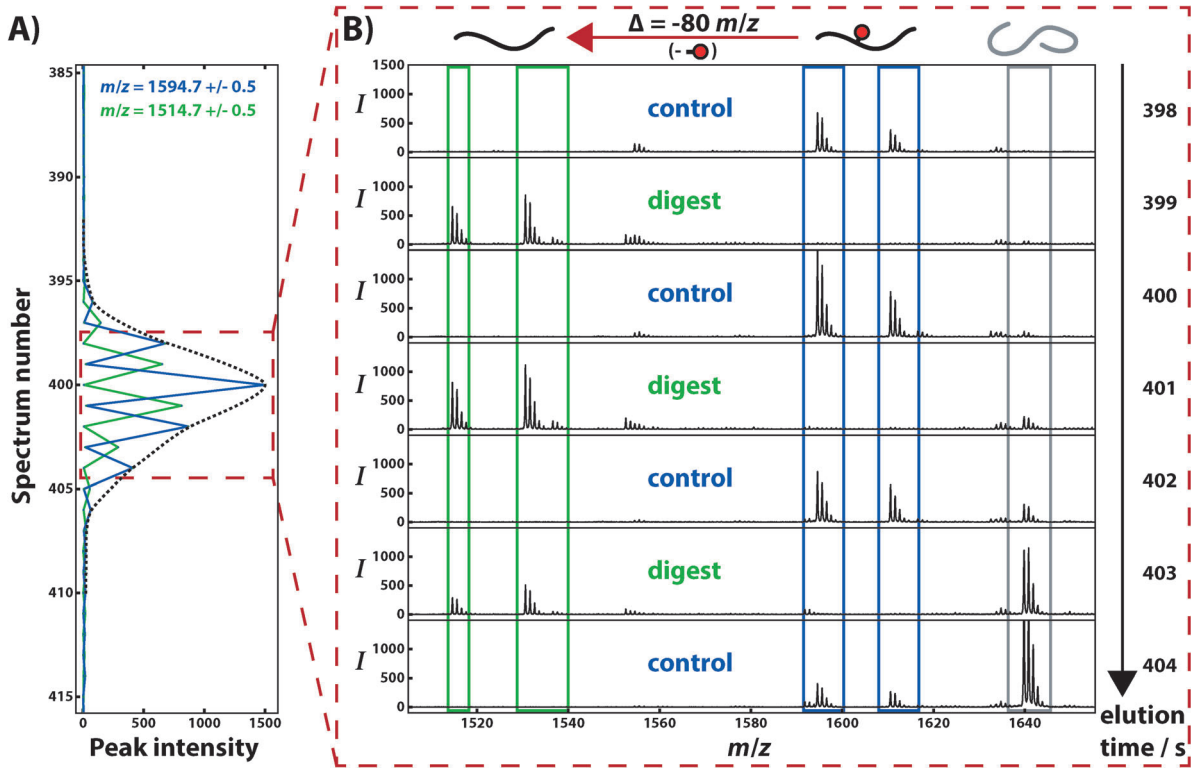

Figure 2. A) Extracted ion chromatograms for $m/z = 1594.7 \pm 0.5$ (green) and $m/z = 1514.7 \pm 0.5$ (blue) illustrating the characteristic interlocked zigzag pattern corresponding to the phosphorylated and dephosphorylated version of KTVDME(p)STEVFTK (α-casein S2). The dashed line indicates the hypothetical shape of the untreated phosphopeptide peak. B) Relevant m/z region of mass spectra from seven subsequent one-second eluate fractions. Peaks with a molecular ion mass of 1594.7 Da and 1610.7 Da (+16 Da; methionine oxidation) appear only in control spots. Other peaks with a molecular ion mass of 1514.7 Da and 1530.7 Da (+16 Da; methionine oxidation) appear only in digest spots. The mass shift of −80 Da shows that the peptide is singly phosphorylated. Peaks corresponding to non-phosphorylated peptides do not show the characteristic intensity fluctuations (example at 1639.8 Da is marked in gray).

Finally, a threshold analysis is used to classify a signal as a potential hit or as background. Using this approach, even low-abundant phosphorylated or dephosphorylated peptides can be detected (see Figure 3, column A).

The classification of a peak as a "hit", a "potential hit", or as "negative" was performed on the basis of two criteria: 1) The appearance of a zigzag peak profile for any peptide due to phosphatase treatment, and 2) the appearance of a second zigzag peak profile interlocked with the first peak profile at a mass shift of $n \times 80$ Da. Peptides fulfilling both criteria were rated as a "hit". Peptides showing a characteristic zigzag peak profile but without an automatically detected interlocked counterpart were rated as a "potential hit". In this case, we applied a "count-up" or "count-down" method. This method implies the overlaid plotting of the elution trace of a "potential hit" peptide together with elution traces of the

Table 1: Summary of all transitions, which were detected in our validation sample.[a]

Entry	α-casein S1 (#P)	α-casein S2 (#P)	β-casein (#P)	Not assigned to model proteins (Part 1) (#P)	Not assigned to model proteins (Part 2) (#P)
1	1660.7 → 1580.8 (1)	1466.5 → 1386.6 (1)	2061.7 → 1981.8 (1)	3184.4 → 3104.4 (1)	1933.9 → 1853.9 (1)
2	1951.9 → 1871.9 (1)	1594.6 → 1514.6 (1)	2431.9 → 2352.0 (1)	2794.4 → 2714.4 (1)	1991.9 → 1912.0 (1)
3	2079.9 → 2000.0 (1)	1738.7 → 1658.7 (1)	2556.0 → 2476.1 (1)	2755.2 → 2675.2 (1)	1989.8 → 1909.8 (1)
4	2548.2 → 2468.2 (1)	1418.5 → 1338.5 (1)	2984.2 → 2904.2 (1)	2445.0 → 2365.1 (1)	1682.7 → 1602.8 (1)
5	2720.9 → 2321.0 (5)	1546.5 → 1466.5 (1)	3122.2 → 2802.3 (4)	2424.1 → 2344.1 (1)	1550.5 → 1390.5 (2)
6	1927.6 → 1767.7 (2)	1539.5 → 1379.5 (2)		2109.7 → 2029.8 (1)	1482.5 → 1402.5 (1)
7	1943.6 → 1783.7 (2)	2716.1 → 2556.2 (2)		2083.7 → 2003.7 (1)	1453.6 → 1373.6 (1)
8	1879.5 → 1719.6 (2)	3008.1 → 2688.2 (4)		2050.0 → 1970.0 (1)	
9	1832.8 → 1752.8 (1)	2093.8 → 2013.8 (1)			
10		3132.0 → 2812.2 (4)			

[a] The first, second, and third column show entries for α-casein S1, α-casein S2, and β-casein, respectively. The last two columns show unassigned transitions (split in two parts). The number of phosphorylations (#P) inducing each transition is shown in brackets.

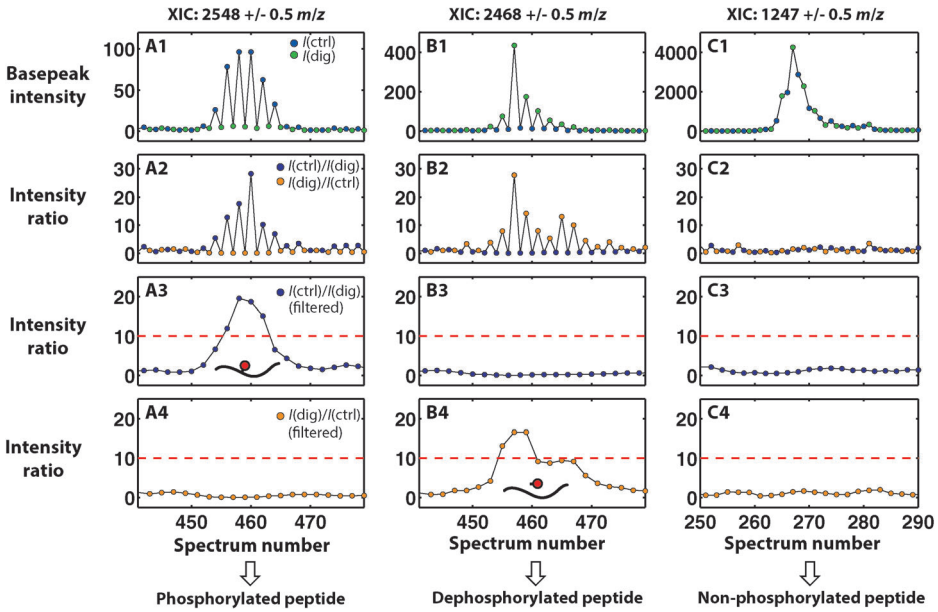

Figure 3. Automated peak classification procedure for phosphorylated and dephosphorylated peptides illustrated on the basis of three exemplarily selected mass windows (A to C) with 1 *m*/*z* width (out of a total number of 2900 mass windows): A1 shows the extracted ion chromatogram (XIC) containing the signal from a phosphorylated peptide (from α-casein S1), and B1 the XIC of its dephosphorylated counterpart. C1 contains the XIC of a non-phosphorylated peptide (from α-casein S2). The intensity ratios of each spectrum to the next are calculated for each XIC (A2, B2, C2) and are displayed separately for I(ctrl)/I(dig) and I(dig)/I(ctrl) (A3, B3, C3 and A4, B4, C4, respectively). A simple threshold analysis (red dashed line) is used to discard non-phosphorylated peaks (C3, C4) and to differentiate between phosphorylated (A3) and dephosphorylated (B4) peptides.

same peptide with a consecutively increasing or decreasing number of phosphates (e.g., see Figure S6C, SI). In this way, even very low-abundant counterparts, which were not detected with the automatic data evaluation, could be identified.

The method presented here was validated through the analysis of the tryptically digested model protein mixture (α-casein S1 and S2, β-casein, and BSA). Table 1 gives a condensed summary of the evaluation results (see also Table S1; SI). Almost 40 hits for pairs of peptides with interlocked zigzag peak profiles and clear transitions of $n \times -80$ Da were found (an overlaid XIC for each hit can be found in the SI). The majority could be assigned to the predicted phosphorylation sites of the model proteins of which most sites resulted in multiple hits due to additional variable modifications. Furthermore, a substantial number of hits could not be assigned to any of the model proteins, most of them in the lower intensity range (Table 1). Additionally, we discovered an even larger number of low-intensity potential hits for peptides appearing exclusively after alkaline phosphatase treatment. Since it was not within the scope of this work to investigate the background proteins of our standards, the nature and origin of these peptides were not analyzed any further.

In conclusion, the method presented here describes a novel screening strategy for protein phosphorylation in complex protein mixtures. By integrating a nanoscale phosphatase digest step into a standard nano-LC-MALDI-MS workflow, our system enables an MS/MS-free detection of phosphopeptides based on the appearance of a characteristic zigzag peak profile. This screening approach enables the detection of low-abundant phosphopeptides, which are typically missed in MS/MS experiments. Moreover, it is still optionally possible to fragment both peptide forms (phosphorylated and dephosphorylated) if sequencing or phosphosite localization is desired. Finally, a method for automatic data analysis is presented, which is indispensable for complex phosphoproteomic analyses.

The nanoliter reaction platform described here will certainly be of interest beyond the field of protein phosphorylation, because it can be advantageously integrated into most analytical LC-MALDI-MS workflows. Finally, coupling our platform to a more complex microfluidic chip for droplet generation and advanced droplet handling, e.g., droplet splitting[16f] or sorting,[16d] has great future potential.

Received: September 24, 2014
Published online: December 12, 2014

Keywords: droplet microfluidics · mass spectrometry · microarrays · protein modifications · protein phosphorylation

[1] a) P. Minguez, I. Letunic, L. Parca, P. Bork, *Nucleic Acids Res.* **2013**, *41*, D306–311; b) R. Aebersold, B. F. Cravatt, *Trends Biotechnol.* **2002**, *20*, S1–S2.
[2] a) S. Prabakaran, G. Lippens, H. Steen, J. Gunawardena, *Wiley Interdiscip Rev. Syst. Biol. Med.* **2012**, *4*, 565–583; b) C. T. Walsh, S. Garneau-Tsodikova, G. J. Gatto, *Angew. Chem. Int. Ed.* **2005**, *44*, 7342–7372; *Angew. Chem.* **2005**, *117*, 7508–7539.
[3] G. A. Khoury, R. C. Baliban, C. A. Floudas, *Sci. Rep.* **2011**, *1*, DOI: 10.1038/srep00090.
[4] M. Mann, S. E. Ong, M. Gronborg, H. Steen, O. N. Jensen, A. Pandey, *Trends Biotechnol.* **2002**, *20*, 261–268.
[5] J. C. Venter, *Science* **2001**, *291*, 1304–1351.
[6] a) M. P. Torres, R. Thapar, W. F. Marzluff, C. H. Borchers, *J. Proteome Res.* **2005**, *4*, 1628–1635; b) H. Steen, B. Küster, M. Fernandez, A. Pandey, M. Mann, *Anal. Chem.* **2001**, *73*, 1440–1448; c) H. Wang, J. Duan, L. Zhang, Z. Liang, W. Zhang, Y. Zhang, *J. Sep. Sci.* **2008**, *31*, 480–487; d) R. D. Unwin, J. R. Griffiths, M. K. Leverentz, A. Grallert, I. M. Hagan, A. D. Whetton, *Mol. Cell. Proteomics* **2005**, *4*, 1134–1144; e) H.-Y. Wu, V. S.-M. Tseng, P.-C. Liao, *J. Proteome Res.* **2007**, *6*, 1812–1821; f) A. Wolf-Yadlin, S. Hautaniemi, D. A. Lauffenburger, F. M. White, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 5860–5865; g) H.-Y. Wu, V. S.-M. Tseng, L.-C. Chen, Y.-C. Chang, P. Ping, C.-C. Liao, Y.-G. Tsay, J.-S. Yu, P.-C. Liao, *Anal. Chem.* **2009**, *81*,

7778 – 7787; h) R. D. Unwin, J. R. Griffiths, A. D. Whetton, *Nat. Protoc.* **2009**, *4*, 870 – 877; i) P. Picotti, R. Aebersold, *Nat. Methods* **2012**, *9*, 555 – 566; j) N. St-Denis, A.-C. Gingras, *Prog. Mol. Biol. Transl. Sci.* **2012**, 106, 3 – 32; k) A. Leitner, M. Sturm, W. Lindner, *Anal. Chim. Acta* **2011**, *703*, 19 – 30.

[7] a) D. S. Jones, W. Heerma, P. D. van Wassenaar, J. Haverkamp, *Rapid Commun. Mass Spectrom.* **1991**, *5*, 192 – 195; b) K. M. Loyet, *Mol. Cell. Proteomics* **2005**, *4*, 235 – 245; c) P. C. Liao, J. Leykam, P. C. Andrews, D. A. Gage, J. Allison, *Anal. Biochem.* **1994**, *219*, 9 – 20; d) T. Miliotis, P. O. Ericsson, G. Marko-Varga, R. Svensson, J. Nilsson, T. Laurell, R. Bischoff, *J. Chromatogr. B* **2001**, *752*, 323 – 334.

[8] a) A. Leitner, A. Foettinger, W. Lindner, *J. Mass Spectrom.* **2007**, *42*, 950 – 959; b) M. M. Savitski, M. L. Nielsen, R. A. Zubarev, *Mol. Cell. Proteomics* **2005**, *4*, 1180 – 1188.

[9] E. Kapp, F. Schütz, *Curr. Protoc. Protein Sci.* **2007**, Chapter 25, Unit25.2.

[10] R. Krüger, C.-W. Hung, M. Edelson-Averbukh, W. D. Lehmann, *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1709 – 1716.

[11] a) R. E. Bossio, A. G. Marshall, *Anal. Chem.* **2002**, *74*, 1674 – 1679; b) B. W. Gibson, P. Cohen, *Methods Enzymol.* **1990**, *193*, 480 – 501.

[12] B. M. Balgley, T. Laudeman, L. Yang, T. Song, C. S. Lee, *Mol. Cell. Proteomics* **2007**, *6*, 1599 – 1608.

[13] H. Yamaguchi, M. Miyazaki, H. Kawazumi, H. Maeda, *Anal. Biochem.* **2010**, *407*, 12 – 18.

[14] a) T. Kawakami, K. Tateishi, Y. Yamano, T. Ishikawa, K. Kuroki, T. Nishimura, *Proteomics* **2005**, *5*, 856 – 864; b) H. Steen, J. A. Jebanathirajah, J. Rush, N. Morrice, M. W. Kirschner, *Mol. Cell. Proteomics* **2006**, *5*, 172 – 181; c) M. E. Boehm, J. Seidler, B. Hahn, W. D. Lehmann, *Proteomics* **2012**, *12*, 2167 – 2178.

[15] a) C. J. Gerdts, D. E. Sharoyan, R. F. Ismagilov, *J. Am. Chem. Soc.* **2004**, *126*, 6327 – 6331; b) L. Frenz, A. El Harrak, M. Pauly, S. Bégin-Colin, A. D. Griffiths, J.-C. Baret, *Angew. Chem. Int. Ed.* **2008**, *47*, 6817 – 6820; *Angew. Chem.* **2008**, *120*, 6923 – 6926; c) A. B. Theberge, E. Mayot, A. El Harrak, F. Kleinschmidt, W. T. S. Huck, A. D. Griffiths, *Lab Chip* **2012**, *12*, 1320 – 1326; d) R. Ameloot, F. Vermoortele, W. Vanhove, M. B. J. Roeffaers, B. F. Sels, D. E. De Vos, *Nat. Chem.* **2011**, *3*, 382 – 387.

[16] a) P. S. Dittrich, M. Jahnz, P. Schwille, *ChemBioChem* **2005**, *6*, 811 – 814; b) J. S. Edgar, G. Milne, Y. Zhao, C. P. Pabbati, D. S. W. Lim, D. T. Chiu, *Angew. Chem. Int. Ed.* **2009**, *48*, 2719 – 2722; *Angew. Chem.* **2009**, *121*, 2757 – 2760; c) Q. Li, J. Pei, P. Song, R. T. Kennedy, *Anal. Chem.* **2010**, *82*, 5260 – 5267; d) J. J. Agresti, E. Antipov, A. R. Abate, K. Ahn, A. C. Rowat, J. C. Baret, M. Marquez, A. M. Klibanov, A. D. Griffiths, D. A. Weitz, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4004 – 4009; e) A. B. Theberge, G. Whyte, W. T. S. Huck, *Anal. Chem.* **2010**, *82*, 3449 – 3453; f) D. Lombardi, P. S. Dittrich, *Anal. Bioanal. Chem.* **2011**, *399*, 347 – 352; g) J.-Y. Kim, S.-W. Cho, D.-K. Kang, J. B. Edel, S.-I. Chang, A. J. deMello, D. O'Hare, *Chem. Commun.* **2012**, *48*, 9144 – 9146.

[17] a) S. K. Küster, S. R. Fagerer, P. E. Verboket, K. Eyer, K. Jefimovs, R. Zenobi, P. S. Dittrich, *Anal. Chem.* **2013**, *85*, 1285 – 1289; b) S. K. Küster, M. Pabst, K. Jefimovs, R. Zenobi, P. S. Dittrich, *Anal. Chem.* **2014**, *86*, 4848 – 4855.

