ANALYTICAL SCIENCES SEPTEMBER 2018, VOL. 34



# Retention Order Reversal of Phosphorylated and Unphosphorylated Peptides in Reversed-Phase LC/MS

**Kosuke OGATA,* Oleg V. KROKHIN,** and Yasushi ISHIHAMA*†**

**Graduate School of Pharmaceutical Sciences, Kyoto University, 46-29 Yoshidashimoadachicho, Sakyo, Kyoto 606–8501, Japan*
***Manitoba Centre for Proteomics and Systems Biology and Department of Internal Medicine, University of Manitoba, 799 JBRC, 715 McDermot Avenue, Winnipeg, R3E 3P4, Canada*

Protein phosphorylation is one of the most ubiquitous post-translational modifications in humans, and trypsin-digested phosphorylated peptides have been analyzed by reversed phase LC/MS using C18-silica columns under acidic conditions to profile human phosphoproteomes. Here, we report that phosphopeptides generally exhibit stronger retention than their unphosphorylated counterparts when C18-silica columns are used with acetic acid or formic acid as an ion-pairing reagent, whereas the retention order is reversed when less hydrophobic stationary phases such as C4-silica columns are employed. Similarly the retention reversal is observed when more hydrophobic ion-pairing reagents such as trifluoroacetic acid are used with C18-silica columns. These phenomena could be explained by the smaller *S*-values of phosphopeptides in linear solvation strength theory, based on the reduced net charge caused by intramolecular interaction between phosphate and basic groups.

**Keywords** Phosphopeptides, retention order reversal, linear solvation strength theory, reversed-phase LC, ion-pairing

**(Received July 10, 2018; Accepted July 19, 2018; Advance Publication Released Online by J-STAGE July 27, 2018)**

## Introduction

In LC/MS-based shotgun proteomics, peptide retention time in reversed-phase LC (RPLC) is a meaningful attribute for peptide identification, helping to increase the confidence of sequence assignments.[1] To maximize its utility, further improvement in predicting peptide retention times is needed. The accuracy of peptide retention time prediction models has been improved dramatically in recent years due to the availability of large datasets obtained by LC/MS/MS.[2] However, retention time prediction for post-translationally modified peptides still represents a major challenge because the retention mechanisms of the modified peptides in RPLC are poorly understood.

Protein phosphorylation is well known as the key factor of intracellular signal transduction.[3] In-depth analyses of the entire phosphoproteomes of interest using LC/MS/MS have significantly contributed to profiling the complex cellular signaling network at the molecular level and to understanding disease and the mechanisms of the action of drugs.[4] Phosphopeptides are believed to elute earlier than their unphosphorylated counterparts in RPLC because of the hydrophilic nature of the phosphate group. Indeed, it has been reported that the retention times increase after dephosphorylation when neutral[5] or acidic conditions with perfluoroalkanoic acids (including trifluoroacetic acid (TFA))[6] as ion-pairing modifier have been applied. However, it is known that TFA and other perfluoroalkanoic acids are not suitable for proteomic LC/MS/MS because of the signal suppression.[7] Under acidic conditions with MS-compatible ion-pairing reagents such as acetic acid (AA) or formic acid (FA), phosphopeptides exhibit stronger retention than their corresponding unmodified counterparts.[8–10] A recent large-scale analysis using synthetic phosphopeptides also supports this observation.[11] So far no retention model has been proposed to explain this retention behavior quantitatively,[12] as there is still the need to investigate the underlying mechanism responsible for such dramatic changes.

In this paper, we employed a non-porous C18-silica column and wide-pore C4-silica columns with ion-pairing reagents such as AA and TFA to elucidate the phosphopeptide retention behavior in RPLC. We systematically profiled the retention times of phosphopeptides and their corresponding unmodified counterparts using an offline RPLC fractionation approach followed by nanoLC/MS/MS. We also directly connected the C4 or C18 columns to MS to measure the retention behavior of phosphorylated and unphosphorylated peptides to confirm our findings obtained by offline 2D-LC/MS/MS.

## Experimental

*Materials*

An Eprogen NPS-ODS-I column (4.6 mm i.d., 33 mm length, 1.5 μm non-porous C18-silica) was purchased through Tokyo Chemical Industry (Tokyo, Japan). A Protein-RP column (2.0 mm i.d., 150 mm length, 5 μm C4-silica (USP L26), 20 nm pore) was purchased from YMC (Kyoto, Japan). An Inertsil WP300 C4 column (2.1 mm i.d., 150 mm length, 5 μm C4-silica, 30 nm pore) was purchased from GL Sciences (Tokyo,

† To whom correspondence should be addressed.
E-mail: yishiham@pharm.kyoto-u.ac.jp.

Japan). A Luna C18(2) column (1.0 mm i.d., 100 mm length, 5 µm silica, Phenomenex, Torrance, CA) was used for isocratic measurements. SGVVVGDGFGGR and its phosphorylated analog pSGVVVGDGFGGR were synthesized by JPT (Berlin, Germany). AA, TFA, and acetonitrile (ACN) were purchased from Fujifilm Wako Pure Chemical (Osaka, Japan). Porous titanium dioxide beads (TitansphereTiO, 10 µm) were obtained from GL Sciences. Water was purified by a Millipore Milli-Q system (Bedford, MA, USA).

*Sample preparation and fractionation*

Hela cell lysates were digested by phase transfer surfactant (PTS)-aided trypsin digestion protocol as described previously.[13] After digestion, the sample was desalted using SDB-XC StageTips[14] and lyophilized. Digests were dissolved in 4% ACN with 0.5% TFA and injected onto the Eprogen NPS-ODS-I column, the YMC Protein-RP column or the GL Sciences Inertsil WP300 C4 column using an ACQUITY UPLC H-Class Bio system (Waters, Milford, MA, USA). The mobile phases consisted of (A) water with 0.5% AA or 0.1% TFA and (B) 80% ACN with 0.5% AA or 0.1% TFA. Peptides were separated with a two-step linear gradient of 5 – 60% in 30 min, 60 – 99% B in 1 min and 99% B for 4 min at a flow rate of 0.8 mL/min (NPS-ODS-I) or 0.2 mL/min (YMC Protein-RP and GL Sciences Inertsil WP300 C4). Fractions were collected every 2 min. One tenth of each fraction was directly analyzed by nanoLC/MS/MS. Remaining samples were subjected to the subsequent phosphopeptide enrichment step.

*Phosphopeptide enrichment*

Metal oxide chromatographic (MOC) tips were prepared as described previously.[15] Briefly, C8 StageTips packed with $TiO_2$ beads (0.5 mg/tip) were equilibrated with 80% ACN with 0.1% trifluoroacetic acid (TFA) and 300 mg/mL lactic acid as a selectivity enhancer (solution A). The fractionated samples were diluted with an equal amount of solution A and loaded onto the MOC tips. After washes with solution A and 80% ACN with 0.1% TFA, phosphopeptides were eluted with 0.5% piperidine. The eluate was acidified by 10% TFA and desalted using SDB-XC StageTips.

*LC/MS/MS analysis*

NanoLC/MS/MS analyses were performed on a TripleTOF 5600+ system (SCIEX, Framingham, MA, USA) or a Q Exactive (Thermo Fisher Scientific, Waltham, MA, USA), which were connected to an Ultimate 3000 pump (Thermo Fisher Scientific) and a HTC-PAL autosampler (CTC Analytics, Zwingen, Switzerland). Peptides were separated by self-pulled needle columns (150 mm length, 100 µm i.d., 6 µm needle opening) packed with Reprosil-Pur 120 C18-AQ 3 µm reversed phase material (Dr. Maisch, Ammerbuch, Germany). For nanoLC/MS/MS analysis with YMC-C4 capillary column, peptides were separated by self-pulled needle columns (150 mm length, 100 µm i.d., 6 µm needle opening) packed with YMC Protein-RP 5 µm reversed phase material. The injection volume was 5 µL, and the flow rate was 500 nL/min. The mobile phases consisted of (A) 0.5% acetic acid and (B) 0.5% acetic acid and 80% ACN. A two-step linear gradient of 5 – 40% B in 20 min, 40 – 99% B in 1 min and 99% B for 4 min was employed.

*Database searching*

Peptides and proteins were identified by means of automated database searching using Mascot v2.6 (Matrix Science, London) against UniprotKB/Swiss-Prot release 2017/04) with a precursor mass tolerance of 20 ppm (TripleTOF 5600+) or 5 ppm (Q Exactive), a fragment ion mass tolerance of 0.1 Da (TripleTOF 5600+) or 20 ppm (Q Exactive), and strict trypsin/P specificity allowing for up to two missed cleavages. Cysteine carbamidomethylation was set as a fixed modification. Methionine oxidation was allowed as a variable modification. For phosphopeptide identification, phosphorylation of serine, threonine, and tyrosine were additionally allowed as variable modifications.

*Isocratic RPLC measurements*

An Agilent 1100 series HPLC system with UV detector (214 nm) with 10 µL injection loop, 1 × 100 mm Luna C18(2) column and 150 µL/min flow rate was used for separations. Acetonitrile concentration (0.5% AA as ion-pairing modifier) was varied in a range of 9 – 17% to measure retention coefficients for the SGVVVGDGFGGR/pSGVVVGDGFGGR pair (~2 µg of each peptide per injection).

## Results and Discussion

We first investigated the C18-silica column with AA or TFA as an ion-pairing reagent to fractionate the tryptic phosphorylated and unmodified peptides from HeLa proteins. In order to avoid the influence of the pore size of the packing material on the retention behavior of peptides, we selected the non-porous particle (Eprogen NPS-ODS-I). Fractions were collected every 2 min and each fraction was analyzed (with and without phosphopeptide enrichment) by nanoLC/MS/MS, resulting in identification of 166 and 111 pairs of phospho/unphosphopeptides for AA and TFA conditions, respectively. Average retention time difference expressed in ACN concentration for each fraction (average %ACN of fractions of phosphopeptides – average %ACN of fractions of unmodified peptides) was found to be +0.87%ACN with AA, and –0.48%ACN with TFA (Fig. 1(A)).

The retention order of phospho- and unphosphopeptides with AA on the non-porous C18-silica column was reversed by using TFA instead of AA, which is in agreement with the previous reports where 12 and 20 nm pore C18-silica gels were used,[11] indicating that the particle pore size is not the major factor to control the retention order of phosphopeptides and their unmodified counterparts.

To investigate the mechanism of the phenomena further, we employed two different C4-silica columns with two different ion-pairing reagents for the fractionation followed by nanoLC/MS/MS using the same samples. As a result, 213 and 293 pairs (GL Sciences Inertsil WP300 C4, GLS-C4) and 89 and 124 pairs (YMC Protein-RP, YMC-C4) of modified/unmodified peptides were identified from AA and TFA fractions, respectively. In the case of AA, the retention difference between phosphorylated and unmodified peptides on C4 was closer to zero (GLS-C4) or even negative (YMC-C4), indicating that relative retention of unmodified peptides became stronger, compared with that on C18 (Figs. 1(B) and 1(C)). The situation was identical in the case of TFA, where the relatively weak retention of phosphopeptides was additionally emphasized. These results indicate that phosphopeptides tend to be weakly retained on C4 compared to C18 columns both with AA and TFA. Additionally, for both C4 columns, the retention of unmodified peptides became stronger when TFA was used instead of AA, as was the case with the C18 column, although this effect was not so significant for the YMC-C4 column. Note that we did not observe any difference between phosphoserine-, phosphothreonine- and phosphotyrosine-containing peptides in their retention behaviors, although the content of phosphotyrosine

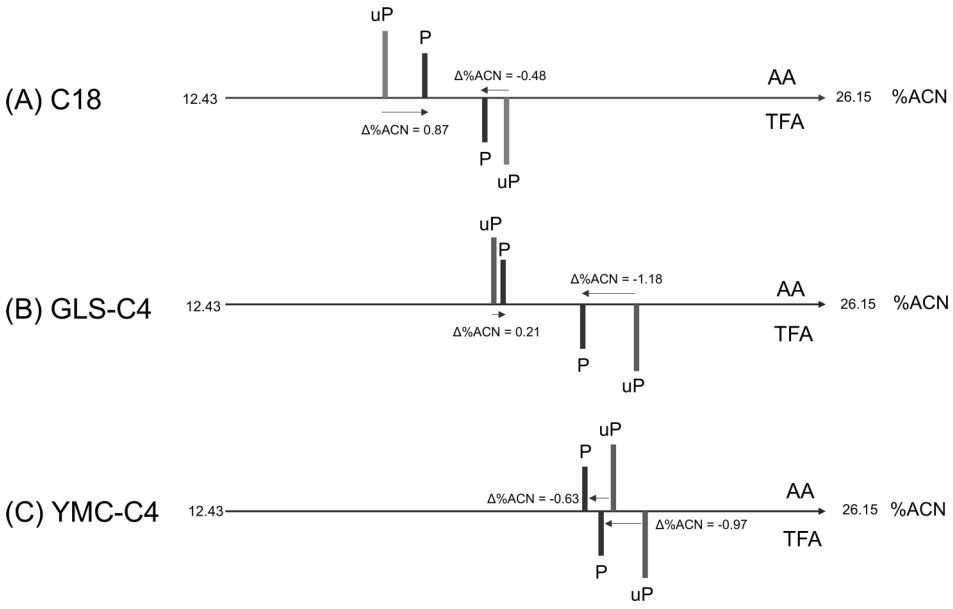


Fig. 1 Average retention time shift between phosphorylated and unmodified peptides using various columns with different ion-pairing reagents. (A) Non-porous C18-silica column with AA (above the line) and TFA (below the line), (B) C4-silica column from GL Sciences (GLS-C4) with AA and TFA, and (C) C4-silica column from YMC (YMC-C4) with AA and TFA. Retention time is described as %ACN from 12.43 to 26.15% in the gradient elution for fractionation after the gradient delay correction. P: Phosphopeptides, uP: unmodified peptides.

Table 1 Retention times of five phosphopeptides and their corresponding unmodified peptides in nanoLC/MS/MS using capillary columns packed with YMC C4-silica and Reprosil-Pur C18-AQ with AA as ion-pairing modifier

| Sequence | YMC C4 | | | Reprosil C18 | | |
|---|---|---|---|---|---|---|
| | RT_uP[a] | RT_P[b] | Delta RT[c] | RT_uP | RT_P | Delta RT |
| QVPDpSAATATAYLcmCGVK | 33.78 | 33.41 | –0.37 | 44.57 | 47.53 | 2.96 |
| DELHIVEAEAMNYEGpSPIK | 50.04 | 45.02 | –5.02 | 54.82 | 56.97 | 2.15 |
| EGRPSGEAFVELEpSEDEVK | 32.95 | 28.89 | –4.06 | 40.57 | 42.35 | 1.79 |
| SLYASpSPGGVYATR | 23.78 | 22.16 | –1.62 | 32.86 | 37.66 | 4.81 |
| KQPPVpSPGTALVGSQK | 20.04 | 20.91 | 0.87 | 26.69 | 30.04 | 3.35 |

a. Retention time (min) of unmodified peptide. b. Retention time (min) of phosphopeptide. c. Delta RT = RT_P – RT_uP.

peptides was only 1 - 2% in this study.

We further confirmed the retention order reversal for C4 and C18 columns with AA by using online nanoLC/MS/MS with in-house packed YMC-C4 and Reprosil-C18 capillary columns (0.1 mm i.d.). Because the column efficiency of the capillary C4 column was not so high, we were able to identify only five pairs of phospho- and unmodified peptides as listed in Table 1. The retention order reversal was observed for four pairs, which is consistent with the off-line fractionation experiment, whereas the retention order of KQPPVSPGTALVGSQK and KQPPVpSPGTALVGSQK on C18 was not reversed on C4 although smaller delta RT was found for C4.

So far, the effect of TFA (compared to AA or FA) on the retention behavior of tryptic peptides on C18 columns has been described by the ion-pairing formation, resulting in the increase of the retention times. However, this effect is considered to be less significant for phosphopeptides because the phosphate group hinders the ion-pairing formation, resulting in weaker retention of phosphopeptides than unmodified peptides on C18 with TFA.[6,16] In the case of AA or FA, the hydrophobicity of the ion-pair is not so high as that of the TFA. However, intramolecular ion-pairing between the basic group and phosphate group increases the retention due to the reduced net charge, resulting in stronger retention of phosphopeptides than unphosphopeptides in most cases.[8-10] Note that there are always some exceptions to this rule,[11,12] which could be attributed to the secondary structure effects.[12] The decrease in retention of phosphopeptides compared to the unmodified peptides on C4 columns (opposite to C18) both with TFA and AA represents a new finding, which requires elucidating on its mechanism.

It is well known that the reversal of peptide retention order in RPLC is often observed when the different gradient slopes are applied, and it can be quantitatively explained by linear solvation strength (LSS) theory, in which the retention factor $k$ can be described as

$$\log k = \log k_0 - S\varphi \quad (1)$$

where $\varphi$ is the volume fraction of the less polar component in the water–organic mobile phase, $k_0$ is the value of $k$ for the

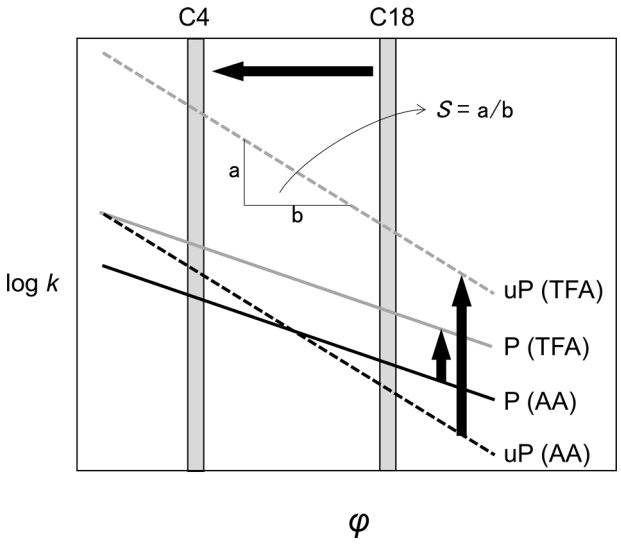


Fig. 2 A proposed LSS-based model to describe retention order reversal of phosphorylated and unmodified peptides. A representative phosphopeptide P (solid line) and its unmodified peptide uP (dot line) are analyzed under the isocratic conditions with various ACN concentration $\varphi$ with AA (black) and TFA (gray). The retention order in C18 at a certain $\varphi$ and that in C4 at the corresponding $\varphi$ are indicated in the gray square labeled as C18 and C4, respectively.

solute at the start of the gradient in the initial mobile phase ($\varphi = 0$) and $S$ is a constant characteristic for a given analyte and chromatographic system.[17] When $k_0$ is similar but $S$-values for two peptides are sufficiently different, it is likely that the retention order reversal happens between two different gradient elution conditions. We hypothesized that phosphopeptides have smaller $S$-values than the corresponding unmodified molecules. The major factor to contribute positively to $S$-values of peptides is molecular weight[17,18] and more recently Shinoda *et al.*,[19] and Spicer *et al.*,[20] independently reported that the number of hydrophobic amino acids negatively contribute to $S$, whereas the numbers of basic amino acids and hydrophilic amino acids positively contribute to $S$. Based on these data, the NZHI model (N: peptide length, Z: net charge, HI: hydrophobicity index) for $S$ prediction was proposed.[20] Since the phosphate group reduces the net charge, smaller $S$-values are expected for phosphopeptides in general. To confirm this, we measured experimental $S$-values of SGVVVGDGFGGR and pSGVVVGDGFGGR under the isocratic conditions with different ACN concentrations. As a result, we obtained log $k_0$ and $S$ as follows:

$$\log k = 3.32 - 24.0\varphi \quad \text{(for SGVVVGDGFGGR, } r^2 = 0.9993) \quad (2)$$

$$\log k = 3.53 - 21.4\varphi \quad \text{(for pSGVVVGDGFGGR, } r^2 = 0.9984) \quad (3)$$

As expected from the NZHI model, the $S$-value of phosphopeptide was smaller than that of unmodified peptide, although the retention order reversal was not observed in the range of ACN concentrations employed in this case.

Taken together, we proposed the LSS-based model of retention order reversal found in this study, as shown in Fig. 2. In both cases of C4 and C18 columns, TFA as an ion-pairing reagent increases the retention of both phosphorylated and unmodified peptides compared to AA, and this effect is more pronounced for the latter due to larger net charges. On the other hand, in both cases of TFA and AA, the relative retention of unmodified peptides to the phosphopeptides is stronger on C4 than on C18 because of the smaller $S$-value of phosphopeptides. C4 columns have weaker hydrophobic interaction than C18, thus the contribution of peptide hydrophobicity to the retention time becomes smaller, resulting in stronger retention of unmodified peptides. Note that Fig. 2 assumes that the C4 column has exactly the same properties as the C18 column except the hydrophobicity, *i.e.*, the C4 column gives smaller log $k_0$ with the same $S$ for each condition, meaning that the retention order on C4 is the same as that on C18 at lower $\varphi$. In reality, however, for the C4 and C18 columns employed in this study, the column length, column diameter, particle diameter, pore size, alkyl silylation chemistry and phase ratio are different. This is why the observed retention (%ACN) of peptides on the non-porous C18 column was generally less than that on the C4 columns shown in Fig. 1.

Since the retention order of phosphorylated and unmodified peptides depends on the gradient conditions according to the proposed LSS model, collection of larger experimental data is needed for the development of an accurate prediction model using MS compatible conditions (AA or FA as ion-pairing modifiers). This model should take into account sequence-specific effects of phosphorylation on both, log $k_0$ and $S$, as these values determine peptides' behavior in RPLC systems.

In conclusion, we observed the retention order reversal of phosphorylated and unmodified peptide pairs between C4 and C18 columns with AA as an ion-pairing reagent. The same phenomenon was observed when we used C18 columns with AA and TFA. To understand the retention behavior of these peptides, we proposed a retention model based on LSS theory. According to a previously reported model, smaller $S$-values were expected for phosphopeptides. This conclusion was confirmed *via* independent measurements under isocratic RPLC conditions. We believe that this proposed model provides a key finding for the development of a comprehensive phosphopeptide retention time prediction tool.

## Acknowledgements

This work was supported by the Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research No. 17H03605 and No. 17H05667, Japan Science and Technology Agency A-STEP program No. AS2915165U (Y. I.) and by grant from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2016-05963; O. V. K.).

## Supporting Information

The MS raw data and analysis files have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) *via* the jPOST partner repository (http://jpostdb.org)[21] with the data set identifier PXD010327.

## References

1. L. Pasa-Tolić, C. Masselon, R. C. Barry, Y. Shen, and R. D. Smith, *Biotechniques*, **2004**, *37*, 621.
2. V. Krokhin, S. Ying, J. P. Cortens, D. Ghosh, V. Spicer, W. Ens, K. G. Standing, R. C. Beavis, and J. A. Wilkins, *Anal. Chem.*, **2006**, *78*, 6265.
3. T. Pawson and J. D. Scott, *Science*, **1997**, *278*, 2075.

4. C. Choudhary and M. Mann, *Nat. Rev. Mol. Cell. Biol.*, **2010**, *11*, 427.
5. J. Lucas and A. Henschen, *J. Chromatogr.*, **1986**, *369*, 357.
6. J. Kim, K. Petritis, Y. Shen, D. G. Camp, R. J. Moore, and R. D. Smith, *J. Chromatogr. A*, **2007**, *1172*, 9.
7. Y. Ishihama, *J. Chromatogr. A*, **2005**, *1067*, 73.
8. Y. Ishihama, F. Y. Wei, K. Aoshima, T. Sato, J. Kuromitsu, and Y. Oda, *J. Proteome Res.*, **2007**, *6*, 1139.
9. T. Kawakami, K. Tateishi, Y. Yamano, T. Ishikawa, K. Kuroki, and T. Nishimura, *Proteomics*, **2005**, *5*, 856.
10. H. Steen, J. A. Jebanathirajah, J. Rush, N. Morrice, and M. W. Kirschner, *Mol. Cell. Proteomics*, **2006**, *5*, 172.
11. H. Marx, S. Lemeer, J. E. Schliep, L. Matheron, S. Mohammed, J. Cox, M. Mann, A. J. Heck, and B. Kuster, *Nat. Biotechnol.*, **2013**, *31*, 557.
12. T. Y. Perlova, A. A. Goloborodko, Y. Margolin, M. L. Pridatchenko, I. A. Tarasova, A. V. Gorshkov, E. Moskovets, A. R. Ivanov, and M. V. Gorshkov, *Proteomics*, **2010**, *10*, 3458.
13. T. Masuda, M. Tomita, and Y. Ishihama, *J. Proteome Res.*, **2008**, *7*, 731.
14. J. Rappsilber, M. Mann, and Y. Ishihama, *Nat. Protoc.*, **2007**, *2*, 1896.
15. N. Sugiyama, T. Masuda, K. Shinoda, A. Nakamura, M. Tomita, and Y. Ishihama, *Mol. Cell. Proteomics*, **2007**, *6*, 1103.
16. C. Dass, P. Mahalakshmi, and D. Grandberry, *J. Chromatogr. A*, **1994**, *678*, 249.
17. M. A. Stadalius, H. S. Gold, and L. R. Snyder, *J. Chromatogr. A*, **1985**, *327*, 27.
18. Y. Sakamoto, N. Kawakami, and T. Sasagawa, *J. Chromatogr.*, **1988**, *442*, 69.
19. K. Shinoda, M. Tomita, and Y. Ishihama, *Bioinformatics*, **2008**, *24*, 1590.
20. V. Spicer, M. Grigoryan, A. Gotfrid, K. G. Standing, and O. V. Krokhin, *Anal. Chem.*, **2010**, *82*, 9678.
21. S. Okuda, Y. Watanabe, Y. Moriya, S. Kawano, T. Yamamoto, M. Matsumoto, T. Takami, D. Kobayashi, N. Araki, A. C. Yoshizawa, T. Tabata, N. Sugiyama, S. Goto, and Y. Ishihama, *Nucleic Acids Res.*, **2017**, *45*, D1107.