DOI 10.1002/pmic.201400569 *Proteomics* 2015, *15*, 2966–2971

TECHNICAL BRIEF

Pitfalls in histone propionylation during bottom-up mass spectrometry analysis

Paulien Meert, Elisabeth Govaert, Ellen Scheerlinck, Maarten Dhaenens and Dieter Deforce**

Laboratory of Pharmaceutical Biotechnology, Ghent University, Belgium

Received: December 4, 2014
Revised: April 13, 2015
Accepted: May 19, 2015

Despite their important role in regulating gene expression, posttranslational histone modifications remain technically challenging to analyze. For identification by bottom-up MS, propionylation is required prior to and following trypsin digestion. Hereby, more hydrophobic peptides are generated enabling RP HPLC separation. When histone dynamics are studied in a quantitative manner, specificity, and efficiency of this chemical derivatization are crucial. Therefore we examined eight different protocols, including two different propionylation reagents. This revealed amidation (up to 70%) and methylation (up to 9%) of carboxyl groups as a side reaction. Moreover, incomplete (up to 85%) as well as a specific propionylation (up to 63%) can occur, depending on the protocol. These results highlight the possible pitfalls and implications for data analysis when doing bottom-up MS on histones.

Keywords:
Cell biology / Histone / Method optimization / MS–LC-MS / Propionylation

Additional supporting information may be found in the online version of this article at the publisher's web-site

Histones are subjected to a diverse array of PTMs, thereby regulating the accessibility of the underlying DNA, effecting both physiology and disease [1, 2]. MS has become a powerful tool to simultaneously identify and quantify these PTMs. Nevertheless, sample preparation for bottom-up MS strategies is complicated by the requirement of chemical derivatization such as propionylation prior to and following trypsin digestion (Fig. 1A) [3]. This step modifies all free primary amine groups (the N-termini and the ε-amino group of unmodified and monomethylated lysine (K)), hence changing the tryptic into Arg-C specificity and resulting in larger (6–20 amino acids), more hydrophobic and readily identifiable peptides. However, in order to optimize identification and study histone dynamics in a quantitative way, this propionylation reaction has to be specific as well as efficient. Only then, accuracy and reproducibility can be guaranteed. Propionylation specificity of histones can be hampered by the high abundance of hydroxyl containing residues (serine (S), threonine (T), and tyrosine (Y)) that can be a specifically propionylated in addition to the primary amines, hereafter referred to as "overpropionylation." On the other hand, an efficient reaction implies that all free primary amine groups should react, leaving no peptides "underpropionylated." Finally, unanticipated side-reactions can hinder the peptide annotation rate and bias quantification.

Based on literature we developed four different propionylation methods (A–D) comprising two different types of propionylation agents, several buffer types, and incubation temperatures (Fig. 1.B) [3–7]. The methods using propionic anhydride were performed with a single as well as a double round of propionylation pre- and postdigestion, to monitor the advantage of an extra round of propionylation. For similar reasons Method D was carried out using two different concentrations of NHS-propionate. Each protocol was performed in triplicate on 10 μg bovine histones and subsequently 1 μg of each sample was analyzed by MS, using a label-free information-dependent acquisition strategy on a TripleTOF

Correspondence: Professor Dieter Deforce, Laboratory of Pharmaceutical Biotechnology, Ghent University, Ottergemsesteenweg 460, 9000 Ghent, Belgium
E-mail: Dieter.Deforce@UGent.be
Fax: +32-0-9-220-66-88

Abbreviations: **QC**, quality control; **XIC**, extracted ion chromatogram

*These authors equally contributed to this work.

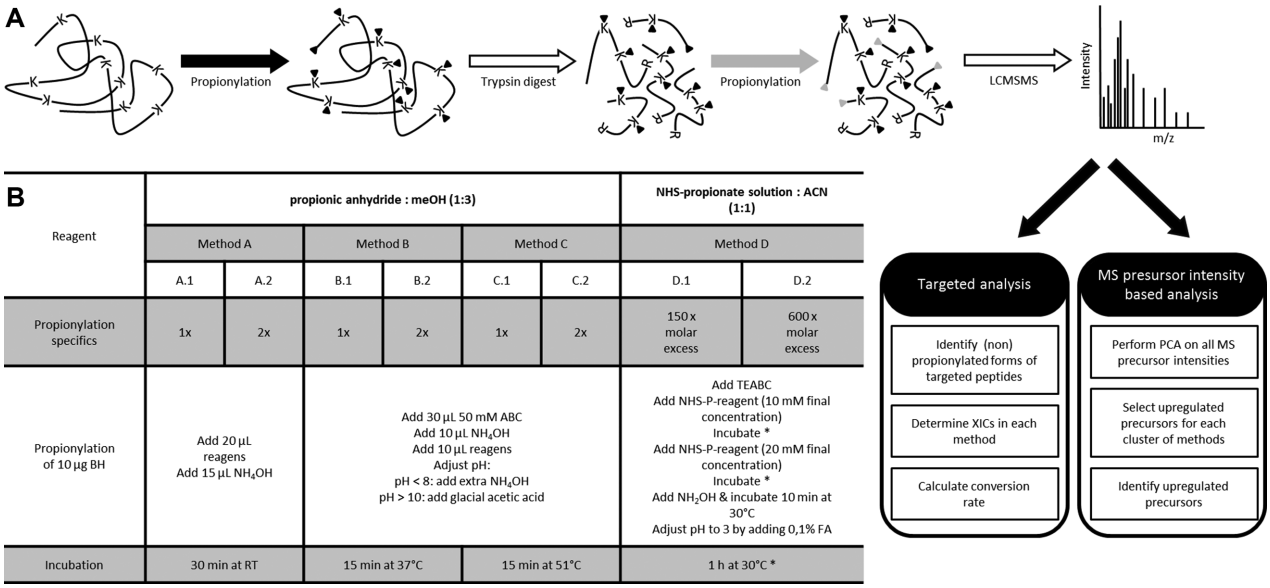

Figure 1. Propionylation workflow and overview of the different protocols. (A) Propionylation is carried out prior to (▼) as well as post-digestion (▽), followed by LCMSMS analysis. During the first propionylation reaction *K*, monomethylated *K* and the protein N-terminus become derivatized. After digestion the newly generated peptide N-termini get propionylated as well. The generated LCMSMS data were analyzed using two approaches: (i) left box: a targeted approach, defining the conversion rate based on identified peptides (ii) right box: an untargeted approach, based on differential MS precursor intensities in between methods. The first strategy can be used to determine efficiency and specificity of the protocol, the latter is used to monitor for unexpected side reactions that differ in between protocols. (B) Table representing the differences between the propionylation methods. The four different methods vary in propionylation reagent, buffer, and reaction temperature. Methods A to C are performed with a single (1×) as well as a double (2×) round of propionylation before and after digestion, marked as Method X1 and X2, respectively. Method D is carried out with a 150 times molar excess (Method D1) and a 600 × molar excess (Method D2) of NHS-propionate to the bovine histones.

5600 (AB Sciex) (Supporting Information protocols for experimental details). In a first targeted data analysis step on this dataset, eight different manually validated peptides were monitored throughout all eight methods to gain a first insight into the conversion rate, specificity, and efficiency of each propionylation protocol. In the subsequent untargeted evaluation strategy, a PCA analysis (Progenesis QI, NonLinear Dynamics, Waters) on all MS precursor intensities present, 11.247 in total, was performed in order to verify clustering (and thus reproducibility) of the experimental conditions and to check for outliers without prior knowledge of these peptides' identity. For each cluster of methods MS precursors were selected that were significantly (p-value ≤ 0.0001) most abundant (highest mean) within this cluster. Each group of extracted MS precursors was then subjected to consecutive rounds of searches to define the occurrence of any unanticipated side-reactions (Fig. 1A).

First we compared the average propionylation conversion rate of eight different peptides in the triplicate runs of each protocol. The chosen peptides are both nonmodified as well as biologically modified and originate from histone H3 and H4. Since acetylation on peptide KQLATKAAR results in two isobaric coeluting forms (H3K18Ac and H3K23Ac) that cannot be distinguished on MS level, we here refer to these isobaric species as KQLATKAAR+Ac. To determine this conversion rate for a specific peptide, the peak area of the extracted ion chromatogram (XIC) for the desired product was divided by the sum of the peak areas representing the total pool of this peptide: the desired form, overreacted products (overpropionylated) and incomplete products (underpropionylated) (Fig. 2A andB). Method A2 performs best for all eight peptides with an average conversion rate between 93 and 100%. Methods B and C on the other hand have average conversion rates lower than 70% for 7 out of 8 peptides (Fig. 2C). The one outlier with a conversion rate of 99% in these protocols coordinately introduces the notion of sequence-dependent propionylation efficiency. This phenomenon can also be seen in the methods using NHS chemistry, where the conversion rate is over 80%, except for the peptides DAVTYTEHAKR and K(Me)SAPATGGVKKPHR where it stays below 65%.

The low conversion rates of Methods B to D can either be due to a specific overpropionylation, inefficient propionylation or a combination of both. Hence, we calculated the average contribution of overpropionylation as well as underpropionyation for all peptides in each protocol based on XICs. When using NHS-chemistry, overpropionylation is the main reason for a low conversion rate, with a peak of up to 60% overpropionylation of peptides DAVTYTEHAKR and K(Me)SAPATGGVKKPHR (Supporting Information Fig.1A).

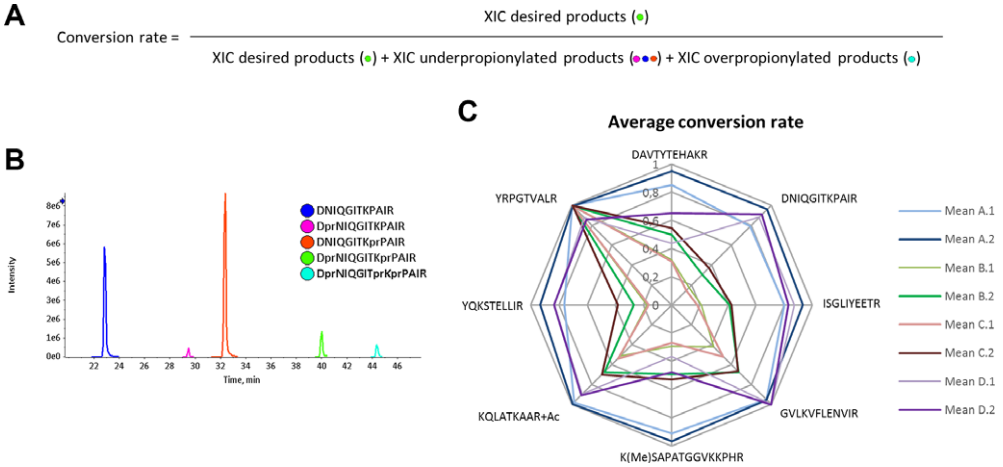

Figure 2. Targeted data analysis. (A) Formula to determine the conversion rate of a peptide, based on XICs of identified forms. (B) Composite representation of the XICs of five different forms of peptide DNIQGITKPAIR, generated after the propionylation (pr) workflow: underpropionylated products (●●●), desired products (●), overpropionylated products (●). This clearly illustrates the increasing retention that is induced by propionylation. (C) Radar chart representing the average conversion rate for eight targeted peptides. Each peptide is located on one angle of the radar chart and each method is represented by another color. The conversion rate for each peptide using the different methods is shown on the radius, whereby a conversion rate of 0 is located in the center, increasing outwards.

Methods B and C on the other hand mainly suffer from underpropionylation due to incomplete reaction, thereby hampering conversion (Supporting Information Fig. 1B). Method A2 is both specific and efficient for all peptides, which explains the high conversion rate mentioned before.

Next, we focused on the effect of a second round of propionylation prior to as well as post digestion on the conversion rate. Therefore we calculated the average increase in conversion rate for all six peptides when using Method X2 instead of Method X1 (Supporting Information Fig. 1C). These results indicate that a second round of propionylation increases the conversion rate by lowering underpropionylation without any increase in overpropionylation. This effect is most notable for method B and C with an increase of the conversion rate by a factor of 1.55 for the protocol at 37°C and up to a factor 1.82 for the protocol at 51°C. When using method A an increase by only a factor of 1.10 was observed. This can be explained by the high conversion rate that was already found when performing one propionylation round. Of interest, technical variation of propionylation in between triplicates as well as sequence-dependent propionylation in between the peptides subjected to the same protocol, lowered considerably when performing a second round of propionylation (Supporting Information Table 1). Especially Method A2 thus shows little sequence-dependent propionylation at all, with average conversion rate in between 93 and 99% for all six peptides as opposed to 79–99% when only one round of propionylation was applied (Method A1). Thus, a second round of propionylation increases the conversion rate and thereby lowers technical variation as well as sequence-dependent propionylation whereby Method A2 emerges as the best candidate protocol for subsequent quantitative MS analysis.

While of obvious value in finding the best possible sample preparation protocol, these analyses are all based on preceding peptide identification, and are therefore targeted. Nevertheless, it is possible that unanticipated side reactions occur, thereby generating peptides that remain unidentified using standard search parameters. In order to search for these unanticipated side reactions a "quantify-then-identify" strategy was applied under the form of an "MS precursor intensity based" PCA (Progenesis QI, Nonlinear Dynamics, Waters) on the total of 11.247 MS precursors, (Fig. 3). Four quality control (QC) samples were included, next to the triplicate method samples, resulting in a total of 144.784 MSMS spectra generated over all different runs. QC samples are identical and contain equal amounts of each sample and 1 μg in total. Because they contain all possible precursors within one sample, they can be used as a precursor alignment template for the Progenesis QI software, which is very important here because considerable differences were induced by the different protocols. Since this PCA is carried out using MS precursors instead of identified peptides, no prior knowledge concerning annotation parameters is required. The aggregation of the different methods in this PCA analysis (87% of the MS precursors has a power > 0.8) shows three different clusters: Method A, Method B and C and Method D with PC1 explaining 44% and PC2 19% of the variation. MSMS spectra from precursors that are significantly most abundant for a cluster of methods were extracted into a separate *.mgf file (ANOVA p-value < 0.0001; q-value < 2 e-6), generating three clusters of MS precursors: most abundant in Method A, most abundant in Method B and C and most abundant in Method D. These three separate *.mgf files (comprising 21.406, 37.736, and 36.141 MSMS spectra, respectively) were then each subjected to an error tolerant search with both N-terminal and K-specific

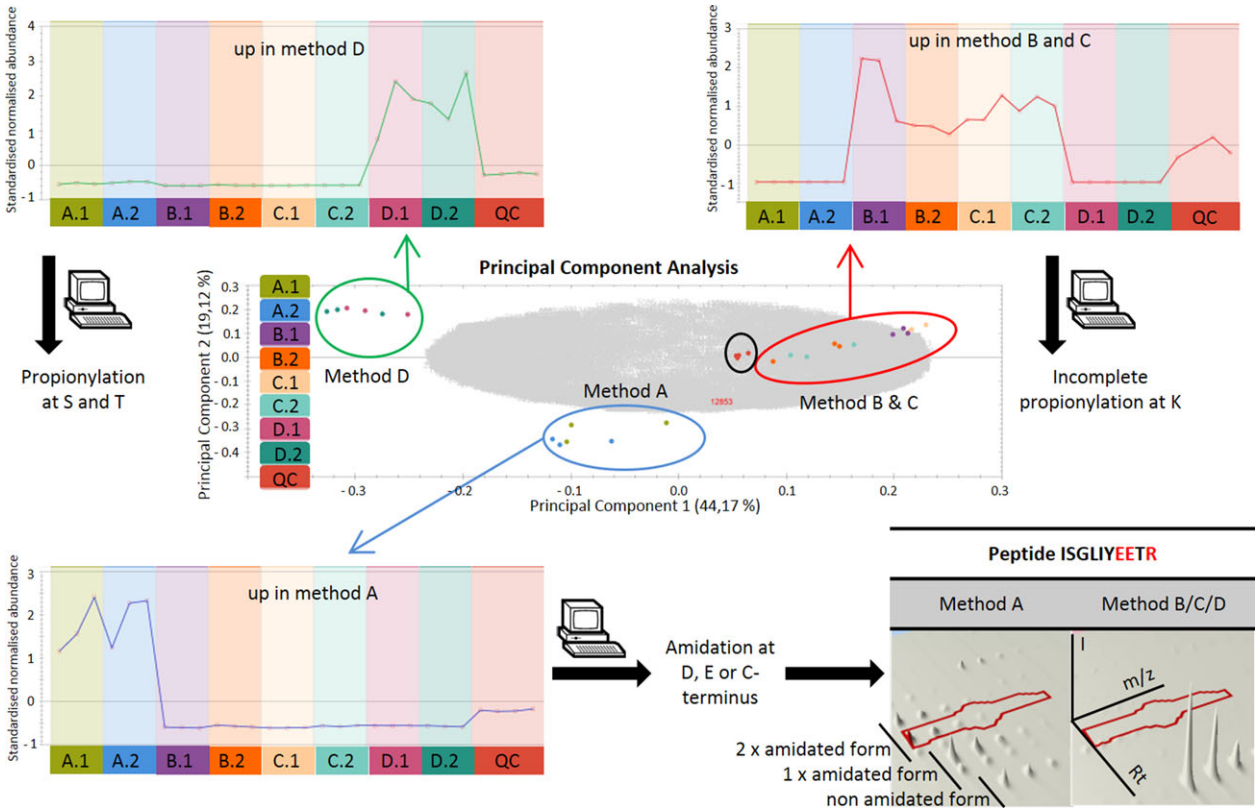

Figure 3. Untargeted, precursor-based data analysis. PCA was performed on MS precursor intensities from a label-free IDA analysis of triplicate experiments and four QC samples resulting in the clustering of Method A, Method D, and Methods B and C. Precursors with an ANOVA *p*-value ≤ 0.0001 and significantly most abundant for one cluster of methods were filtered. A representative abundance profile of a differential precursor is shown for each cluster (Method A: blue; Method B and C: red; Method D: green). Identification of the MSMS spectra linked to these exported precursors confirmed that there is an enrichment of overpropionylated peptides in Method D, and that underpropionylation is mainly found in Methods B and C. A new modification was revealed to be enriched in Method A: amidation of D, E, and the C-terminus. This can result in a dispersion of precursor signal intensity over the generated peptide forms, illustrated by peptide ISGLIYEETR. The amino acids susceptible for amidation are highlighted in red and the different peptide forms are marked.

propionylation set as variable modification (Mascot 2.5, Matrix Science). The most occurring error tolerant modification was identified for each cluster, using the modification statistics (Supporting Information Fig. 2). Indeed, in vitro induced modification significantly outnumbered biologically relevant PTMs such as acetylation. Earlier findings of the targeted analysis were confirmed herein: (i) precursors that were most abundant in method D were mainly identified as overpropionylated (propionylation on S and T as the most identified error tolerant modification), (ii) precursors that were most abundant in Method B and C mainly suffer from underpropionylation. The latter group thus also comprised "semi-ArgC" peptides, in which the protein was "nonspecifically" cleaved at a K because this amino acid was not sufficiently propionylated in the reaction prior to the trypsin digest. Surprisingly, the third group of differential precursors (iii) that were most abundant in method A, also shared a common PTM: amidation (–0.9840 Da) on aspartic (D) and glutamic acid (E), as well as on the C-terminus. In Supporting Information, these identifications (MSMS spectra of both the amidated and nonamidated peptide form) are shown in Fig. 3. Remarkably, the reactivity of carboxyl groups in method A is not only limited to amidation. Also methylation of D, E, and C-terminus can occur when mixing propionic anhydride with methanol. Yet, this side reaction was not identified for method B and C even though methanol was used as well. In order to estimate the impact of these side reactions (amidation and methylation of COOH-groups) on quantification we calculated the relative abundance of both the amidated and carboxy methylated peptide form compared to its non-reacted counterpart. All eight peptides used for targeted analysis were investigated, yet two peptides were not included in Supporting Information Fig. 4. Especially for peptide KQLATKAAR+Ac, accurate quantitation was impaired by the fact that the amidated form elutes first and its naturally occurring first isotope coelutes with the nonamidated peptide precursor mass, hindering XIC-based quantification (Supporting Information Fig. 5). For the other peptide, K(Me)SAPATGGVKKPHR, amidation nor methylation was identified. As shown in Supporting Information Fig. 4,

Table 1. Table summarizing the pitfalls for each method and suggestions for data analysis

Method	Pitfall	Suggestions for data analysis				
		Peptide identification				Quantification of histone modification
		Amino acid	Modification	Variable	Fixed	
A	Amidation at COOH	C-terminus/D/E	Amidation	x		Use relative quantification within the same run
	Methylation at COOH	C-terminus/D/E	Methylation	x		
		N-terminus	Propionylation		x	
B	Underpropionylation	N-terminus/K/Kme	Propionylation	x		Use relative quantification within the same run
		K	Nonspecific cleavage	x		
C	Underpropionylation	N-terminus/K/Kme	Propionylation	x		Use relative quantification within the same run
		K	Non specific cleavage	x		
D	Overpropionylation	S/T/Y	Propionylation	x		Use relative quantification within the same run
		N-terminus	Propionylation		x	

D, aspartic acid; **E,** glutamic acid; **K,** lysine; **S,** serine; **T,** threonine; **Y,**tyrosine.

amidation at COOH occurs far more frequently and intense than methylation in this protocol. The latter was only detected for peptide ISGLIYEETR and no more than 9% was affected. Amidation on the other hand can convert up to 70% of a peptide in an amidated form, depending on the presence of D and E in the sequence.

As shown here, each method has its own shortcomings. However, the most often used metric for studying biology is the relative abundance of PTMs in which the intensity of modified peptides are expressed relative to all peptides sharing that same sequence. Thus, we quantified the relative abundance of one such PTM, H3K23ac, within the bovine histone sample in all different protocols. Hereby we assumed the researcher to be blind to an unexpected side reaction, underpropionylation, or overpropionylation. This was estimated as the percentage of peptide KQLATKAAR+Ac by all peptides sharing that same sequence (side reactions, underpropionylation, or overpropionylation not taken into account). As shown in Supporting Information Table 2, estimations only vary slightly (between 28 and 35%). While this is reassuring for the conclusions on biology reported to date, it emphasizes the importance of calculating the relative abundance in each run before comparing the samples, as opposed to directly comparing precursor intensities in between separate runs. Also, the standard deviation of the relative abundance in protocols using only a single round of propionylation cautions for the limited accuracy of estimation for these methods. Equally important however, abundant side reactions such as amidation not only reduce the signal of the in vivo relevant precursor ion, they also generate large amounts of uninformative new precursors that are being selected for MSMS during the DDA acquisition, as these new forms have a different retention time (Fig. 3). It is therefore advisable to check for side reactions (amidation / methylation at COOH), underpropionylation or overpropionylation. These peptides can then be taken into account when reporting on the total amount of features detected and identified.

In conclusion, this dataset pointed out that a second round of propionylation increases the conversion rate as well as the reproducibility of precursor quantification and is therefore strongly recommended. Nevertheless, several pitfalls in propionylating histones for bottom-up MS were disclosed: incomplete derivatization, aspecific propionylation, and side reactions on carboxyl groups. Each of these events has its own implications during data analysis (Table 1). When focusing on identification of histones the following should be taken into account: using Method A, amidation and methylation of COOH-groups should be added as a variable modification; Method B and C will benefit from allowing nonspecific cleavage at K (or using trypsin as enzyme in the search parameters with a high number of missed cleavages) and setting N-terminal propionylation as variable modification instead of fixed; adding propionylation on S, T, and Y will increase identifications when using Method D. For accurate quantification, we strongly recommend to use relative abundances, as these appear reproducible between different protocols and can thus be considered the most robust option. Amidation was only found as a side reaction thanks to an untargeted evaluation strategy based on a PCA on MS precursor intensities and identification of differential MS precursor abundance in between protocols. Therefore, we would like to stress the importance of using such an approach and recommend including it when comparing or evaluating other protocols.

This research was mainly funded by a PhD grant from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen), awarded to Paulien Meert. This research was (partially) financed by GOA- project

no. 01G01112 (Ellen Scheerlinck) and BOF Mandate 01D23313 (Elisabeth Govaert). This work was supported by a grant from the Fund of Scientific Research Flanders (FWO, G073112N).

The authors have declared no conflict of interest.

References

[1] Jenuwein, T., Allis, C. D., Translating the histone code. *Science* 2001, *293*, 1074–1080.

[2] Strahl, B. D., Allis, C. D., The language of covalent histone modifications. *Nature* 2000, *403*, 41–45.

[3] Garcia, B. A., Mollah, S., Ueberheide, B. M., Busby, S. A. et al., Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat. Protoc.* 2007, *2*, 933–938.

[4] Glibert, P., Van Steendam, K., Dhaenens, M., Deforce, D., iTRAQ as a method for optimization: enhancing peptide recovery after gel fractionation. *Proteomics* 2014, *14*, 680–684.

[5] Lin, S., Garcia, B. A., Examining histone posttranslational modification patterns by high-resolution mass spectrometry. *Methods Enzymol.* 2012, *512*, 3–28.

[6] Xu, G., Shin, S. B., Jaffrey, S. R., Global profiling of protease cleavage sites by chemoselective labeling of protein N-termini. *Proc. Natl. Acad. Sci. USA* 2009, *106*, 19310–19315.

[7] Zhang, C., Liu, Y., Andrews, P. C., Quantification of histone modifications using N metabolic labeling. *Methods* 2013, *106*, 46.

