Natural isotope correction improves analysis of protein modification dynamics

Jörn Dietze,[†] Alienke van Pijkeren,[‡,¶] Mathias Ziegler,[§] Marcel Kwiatkowski,[*,‖] and Ines Heiland[*,†,⊥]

†*Department of Arctic and Marine Biology, UiT The Arctic University of Norway*
‡*Lab for Metabolic Signaling, Department of Biochemistry, University of Innsbruck, Austria*
¶*Department of Analytical Biochemistry, Groningen Research Institute of Pharmacy, The Netherlands*
§*Department of Biomedicine, University of Bergen, Norway*
‖*Functional Proteo-Metabolomics, Department of Biochemistry, University of Innsbruck, Austria*
⊥*Department of Clinical Medicine, University of Bergen, Norway*

E-mail: Marcel.Kwiatkowski@uibk.ac.at; ines.heiland@uit.no

Abstract

Stable isotope labelling in combination with high resolution mass spectrometry approaches are increasingly used to analyse both metabolite and protein modification dynamics. To enable correct estimation of the resulting dynamics it is critical to correct the measured values for naturally occurring stable isotopes, a process commonly called isotopologue correction or deconvolution. While the importance of isotopologue correction is well recognized in metabolomics, it has received far less attention in proteomics approaches. Although several tools exist that enable isotopologue correction of mass spectrometry data, none of them is universally applicable for all potential experimental approaches. We here present PICor which has been streamlined for multiple isotope labelling isotopologue correction in proteomics or metabolomics approaches. We demonstrate the importance for accurate measurement of the dynamics of protein modifications, such as histone acetylation.

Introduction

Protein modifications such as phosphorylation, methylation or acetylation have an important regulatory function in biology and are the basis of environmental adaptation and phenotypic variations. While it is well known that protein phosphorylations are highly dynamic and result from a complex interplay between protein kinases and phosphatases, the dynamics of other protein modifications such as methylation and acetylation has been less well investigated. The availability of stable isotope labelled compounds and high resolution mass spectrometry has, however, more recently paved the way to analyse the dynamics of protein methylation, acetylation, phosphorylation and glycosylation[1,2].

As basically all protein modifications are effectuated by small metabolic compounds, one can use partially or fully labelled metabolic precursors to monitor changes in concentrations of both metabolic intermediates and protein modifications over time. Figure 1 depicts a typical experimental workflow for stable isotope labelling experiments. Following incubation of cells with labelled precursors, metabolites and proteins are extracted at various time points, separated by

gas chromatography (GC) or liquid chromatography (LC) and analyzed by mass spectrometry (MS).

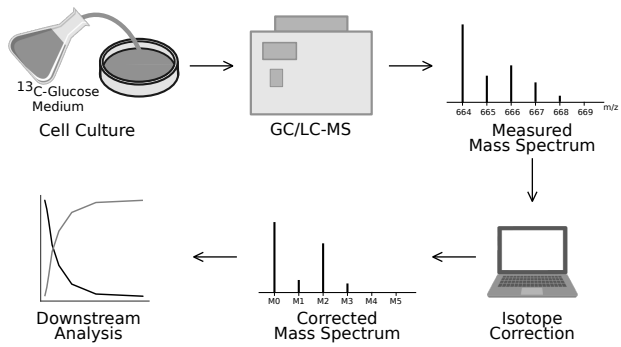


Figure 1: Typical workflow of ^{13}C labelling experiment. Adapted from Midani et al.[3]

Molecules with the same composition of isotopes, so-called isotopomers, are chemically identical. Their intact precusor ions are isobaric and therefore indistinguishable by MS analysis. An additional fragmentation by tandem mass spectrometry (MS/MS or MS2) using e.g. collision induced dissociation (CID), allows to localize the isotopes based on their presence in the individual fragments. The obtained data need to be appropriately processed, in order to correctly calculate metabolite or protein modification dynamics. A further complication in this analysis is the necessity to correct the obtained values for natural isotope abundance, called isotop(ologu)e correction or deconvolution. This is critical, because naturally occurring compounds already contain 1.1% of the stable heavy isotope ^{13}C. Thus, integration of ^{13}C-labeled precursors requires the correction for this naturally occuring isotope and potentially also for other stable natural isotopes such as deuterium, ^{18}O, or ^{15}N, depending on whether or not these isotopes can be resolved by the detection method used[4]. Especially the labeling of large metabolites as well as peptides and proteins could otherwise lead to large errors in the estimation of the label integration dynamics.

In recent years, a number of tools have been made available that enable isotopologue correction for metabolites. But very little attention has been paid to natural isotope correction in proteomics or for the analyses of protein modification turnover. The theoretical basis and application of available tools for metabolites varies. Correction can for example be done based on spectra from naturally occurring compounds. This approach is, however, not feasible for proteomics applications, as it would require to have all peptides of interest available as standards, let alone inclusion of protein modifications. Similarly unfeasible for proteomics related measurements is the application of tools that require the reporting of all expected isotopologues as is the case for example for IsoCor[5]. The approach we choose here, corrects measured values based on the calculated theoretical isotope distribution. The latter can be described using probability matrices. Commonly, two general approaches exist ([3,6,7]): In the so-called *classical* approach, the same spectrum or distribution is used to correct all isotopologues. This is also the basis for correction methods based on measured spectra from naturally occurring substances. In molecules, however, that have all light isotopes replaced by their heavy counterparts, no heavier isotopologues can exist. These methods therefore lead to an estimation bias towards heavy isotopologues especially for large molecules with high label incorporation. We therefore used a so-called *skewed* matrix approach, that corrects each isotoplogue with its own set of theoretical calculated correction factors (schematic representation see figure 2a). This method is for example implemented in the R-based toolbox IsoCorrectoR[8] and in AccuCor[4].

For the identification and quantification of isobaric modified histone species, Yuan et al.[9] developed an alternative method for isotope correction using a deconvolution approach based on MS/MS measurements. The corresponding toolbox EpiProfile has been primarily developed to analyse isobaric modified histone peptides irrespective of whether they arise from label integration or are caused by similar chemical composition. As this method allows to identify which part of the molecule is labelled with heavy isotopes it does enable some correction for natural isotope abundance. But, beside the disadvantage of requiring MS/MS measurements, it does not provide a full isotopologue correction as visualized in figure 2b. The resulting estimation

error is increasing with the number of possible isotopologues and the number of atoms labelled.

We have developed PICor (`https://github.com/MolecularBioinformatics/PICor`), a python based isotopologue correction tool that is applicable for the analysis of both labelled metabolites as well as peptides and protein modifications. It enables the correction for incorporation of multiple different isotopes for example both ^{13}C and ^{15}N and uses the skewed matrix approach. Importantly, our approach is based on the theoretical distribution of isotopologues. Therefore, only the most abundant or important isotopologues need to be measured and thus incomplete datasets can be used. Using time course measurements of acetylated histone peptides we demonstrate that appropriate correction for natural isotope distribution considerably improves the accuracy of the analyses.

Results

To highlight the importance of isotope correction and to compare different tools available we used the large metabolite NAD (figure 3a) as example. Using an imaginary NAD labelling spectrum makes visualization more comprehensible as the number of potential isotopologues is much lower as for any peptide and it allows the comparison with other tools, as some of them cannot be used for the analysis of peptides. Still NAD is sufficiently large to visualize effects occurring in larger compounds. As shown in figure 3a the corrected spectra derived with PICor match those of other tools, that use the same correction approach, like IsoCorrectoR and AccuCor. The results do, however, clearly differ from those correction methods using the classical matrix correction, such as IsoCor. It should be noted that AccuCor cannot be used to do isotopologue correction in experiments using label combinations such as for example ^{13}C and ^{15}N, even though correction for all naturally occurring isotopes present in the molecule is done. Analyses of labelled compounds with more than one incorporated isotope can to our knowledge only be done with IsoCorrectoR and IsoCor, with the latter having the disadvantage

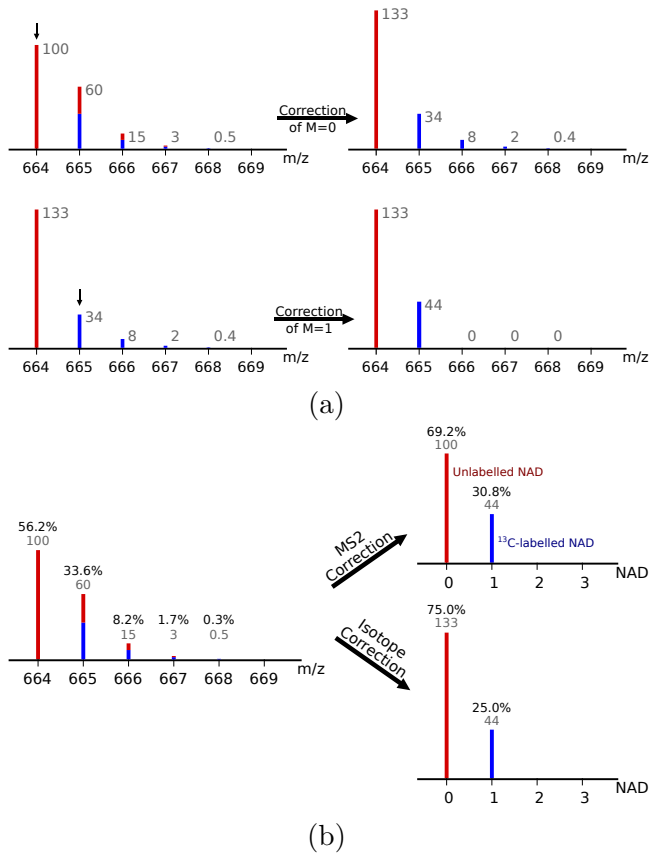


Figure 2: Schematic representation of isotopologue correction methods using an artificial spectrum of ^{13}C-labelled NAD containing 75% unlabelled NAD and 25% single ^{13}C-labelled NAD. Red depicts the contributions of unlabelled NAD and blue the ones of the single labelled NAD. (a) Visualisation of the skewed matrix approach. M+0 correction requires the addition of all naturally labelled heavier isotopologues. The lower panel demonstrate the sequential process for M+1 correction. This correction process needs to be done for each isotopologue. The lower right panel shows the fully corrected spectrum. (b) Comparison between the isotope deconvolution done by EpiProfile and the full isotopologue correction done by our approach.

of using the classical correction.

When correcting for natural isotopes it is important to consider the mass resolution of the experiments[4]. Nowadays, it is common to use a measurement resolution that can separate ^{13}C and ^{15}N. But, resolving the mass difference between ^{13}C and deuterium would require a resolution that is unfeasible in practice, due to required measurement time. PICor, therefore, enables isotopologue correction for all label combinations according to the mass resolution used in the experiments, making it especially relevant for labelling experiments using combinations of ^{13}C and deuterium. To make PICor applicable for proteomics approaches one only needs to provide a table with measurements for the isotopologues of interest and not a full list for all possible isotopologues as for example required for the analysis with IsoCor. Table 1 summarizes the potential application areas as well as the correction method used by the aforementioned isotope correction tools.

Histone Acetylation

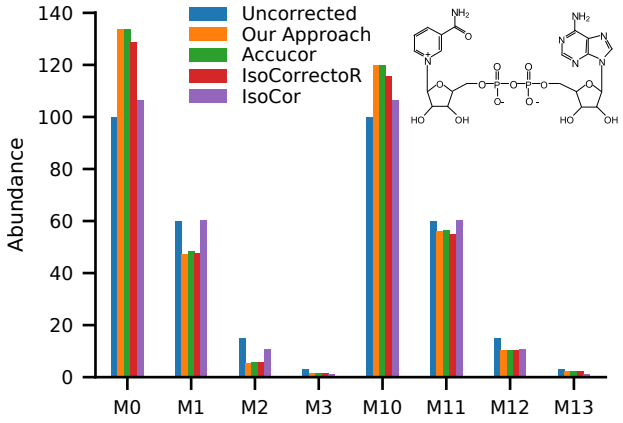
To demonstrate the effect of natural isotope correction on labelling experiments in proteomics we analysed experimental time course data from ^{13}C-labelled glucose supplementation on histone acetylations. Histone acetylations are important for transcriptional regulation. The acetylation of lysine residues in histones neutralizes the charge of the side chain and thus changes the ability of histones to bind DNA: this in turn leads to changes in the 3 dimensional structure and accessibility of the DNA.

Figure 3b shows an experimental time course for ^{13}C-glucose labelling. In such a metabolic labelling experiment, the naturally occuring ^{12}C-glucose is replaced with ^{13}C-glucose during cell culture. Labelled glucose is consequently converted into ^{13}C-acetyl-Co-enzyme A. This is, among others, the substrate for histone acetyl transferases that catalyze the transfer of the acetyl group onto a nitrogen in the lysine side chain of a protein, resulting in a M+2 isotopologue. As an example we here measured the histone 3 peptide H3(18-26) that can be modified at two lysine residues (K18, K23). Upon complete modificationa M+4 isotopologue of the peptide was observed. Figure 3b shows the dynamics of the different labels over a 24 hour time course, both for the corrected and the uncorrected measurements. The slope of the time courses for the corrected and the uncorrected data are considerably different due to the overrepresentation of labelled compounds at time point zero that stems from naturally occuring ^{13}C.

SILAC – Quantitive proteomics

Besides metabolic labelling of proteins and protein modifications, stable isotopes are routinely applied in quantitative proteomics to investigate differentially expressed proteins. For example in SILAC (**S**table **I**sotope **L**abeling with **A**mino acids in **C**ell culture) experiments, cells are cultured in media containing either unlabelled lysine or arginine, or ^{13}C or ^{13}C and ^{15}N labelled amino acids. SILAC is not only used for a quantitative comparison of protein abundance under different experimental conditions, but has also been used to investigate protein turnover. After cell lysis and protein extraction, differentially labelled samples are mixed and subjected to quantitative LC-MS based proteome analysis. The sample origin is determined based on the mass shift. To date it is not common to correct for natural isotopes in SILAC based quantitative proteomics. This might be due to the fact that usually only fold changes above 2 were considered. However, with increasing accuracy and requests for higher precision and the goal to identify more fine grain changes between different experimental conditions, isotope correction will also become more relevant in these experiments. Figure 4 shows the potential effect of isotope correction in triple SILAC experiments. PICor has been used to calculate the theoretical difference between corrected and uncorrected values for the sample peptide SLDIFTVGGSGDGVVTNSAWETFQR. Even though the errors are much smaller than for metabolic labelling, it would be expected to contribute a systematical error if SILAC is for example used in time course experiments to estimate synthesis and degradation rates.

(a) Metabolite: NAD



(b) Peptide: KQLATKAAR

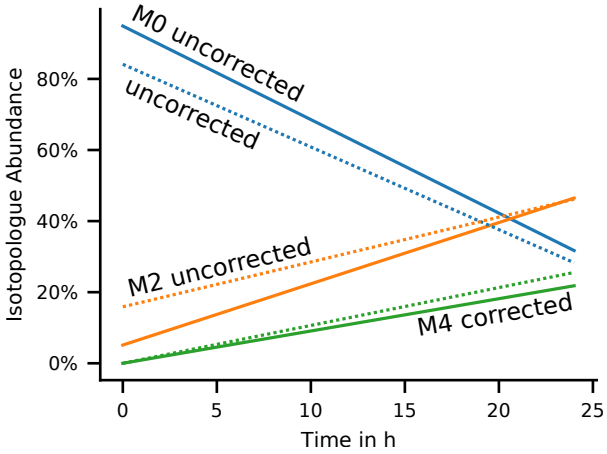


Figure 3: (a) Comparison of different isotope correction tools using an artificial spectrum of ^{13}C labelled NAD as example. (b) MS time course of histone 3 peptide H3(18-26) acetylated at none (M0), one (M2) or two (M4) lysine residues (K18, K23) solid line – isotope corrected data; dashed line – uncorrected data

Table 1: Advantages and application area of different tools for isotopologue correction. (Adj.: high resolution mode with resolution dependent correction)

	PICor	IsoCorrectoR	AccuCoR	IsoCor
Programming language	Python	R	R	Python
Multiple isotope labels	Yes	Yes	No	No
Resolution	Adj.	Low & High	Adj.	Low & Adj.
Correction matrix	Skewed	Skewed	Skewed	Classic

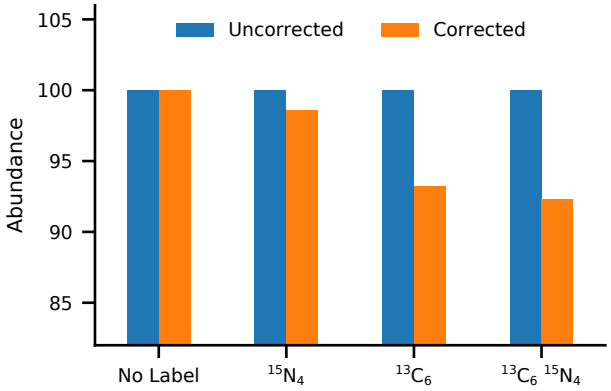


Figure 4: Comparison between corrected and uncorrected values in a triple SILAC approach. Value are given relative to the unlabelled sample. The artificial dataset analysed had equal amounts of uncorrected values and the following peptide sequence was used: SLDIFTVGGSGDGVVTNSAWETFQR

Outline

New high-resolution mass spectrometry approaches now enable the analysis of metabolite and protein modification dynamics using metabolic stable isotope labelling techniques. Due to the high number of atoms in a peptide the amount of naturally labelled peptides is relatively large compared to small metabolic compound. Therefore, correct and complete isotopologue correction is essential to correctly estimate protein modification dynamics. PICor is a tool developed to enable isotopologue correction for both metabolites and proteins applicable to complex labelling experiments.

Acknowledgement JD, MZ and IH have been supported by the Norwegian Research Foundation. MK has received funding from the Promotion program for Young Scientists at

University of Innsbruck (project no: 316826).

References

(1) Zheng, Y.; Sweet, S. M. M.; Popovic, R.; Martinez-Garcia, E.; Tipton, J. D.; Thomas, P. M.; Licht, J. D.; Kelleher, N. L. Total kinetic analysis reveals how combinatorial methylation patterns are established on lysines 27 and 36 of histone H3. *Proceedings of the National Academy of Sciences* **2012**, *109*, 13549–13554.

(2) van Pijkeren, A.; Bischoff, R.; Kwiatkowski, M. Mass spectrometric analysis of PTM dynamics using stable isotope labeled metabolic precursors in cell culture. *Analyst* **2019**, *144*, 6812–6833.

(3) Midani, F. S.; Wynn, M. L.; Schnell, S. The importance of accurately correcting for the natural abundance of stable isotopes. *Analytical Biochemistry* **2017**, *520*, 27–43.

(4) Su, X.; Lu, W.; Rabinowitz, J. D. Metabolite Spectral Accuracy on Orbitraps. *Analytical Chemistry* **2017**, *89*, 5940–5948.

(5) Millard, P.; Delépine, B.; Guionnet, M.; Heuillet, M.; Bellvert, F.; Létisse, F. IsoCor: isotope correction for high-resolution MS labeling experiments. *Bioinformatics* **2019**, 1–4.

(6) Rosenblatt, J.; Chinkes, D.; Wolfe, M.; Wolfe, R. R. Stable isotope tracer analysis by GC-MS, including quantification of isotopomer effects. *American Journal of Physiology-Endocrinology and Metabolism* **1992**, *263*, E584–E596.

(7) Fernandez, C. A.; Rosiers, C. D.; Previs, S. F.; David, F.; Brunengraber, H. Correction of13C Mass Isotopomer Distributions for Natural Stable Isotope Abundance. *Journal of Mass Spectrometry* **1996**, *31*, 255–262.

(8) Heinrich, P.; Kohler, C.; Ellmann, L.; Kuerner, P.; Spang, R.; Oefner, P. J.; Dettmer, K. Correcting for natural isotope abundance and tracer impurity in MS-, MS/MS- and high-resolution-multiple-tracer-data from stable isotope labeling experiments with IsoCorrectoR. *Scientific Reports* **2018**, *8*, 1–10.

(9) Yuan, Z. F.; Sidoli, S.; Marchione, D. M.; Simithy, J.; Janssen, K. A.; Szurgot, M. R.; Garcia, B. A. EpiProfile 2.0: A Computational Platform for Processing Epi-Proteomics Mass Spectrometry Data. *J Proteome Res* **2018**, *17*, 2533–2541.

Experimental

Chemicals

Water, acetonitrile (ACN), formic acid (FA, all HPLC-grade) were obtained from Thermo Fisher Scientific (Dreieich, Germany). All other chemicals were purchased from Sigma-Aldrich (Munich, Germany) unless otherwise noted.

Metabolic labeling with uniformely labeled 13C-glucose and histone extraction

The murine macrophage-like cell line RAW264.7 (American Type Culture Collection, Wesel, Germany) was cultured in 6-well plates at 37°C and 5% CO2 atmosphere in DMEM with 4.5 g/L glucose. Media was supplemented with 2 mM extra L-Glutamine and 10% (v/v) heat-inactivated FBS. For the labeling experiment, RAW264.7 cells were first cultured in media containing uniformly 12C-glucose ([U–^{12}C]-Glc) for 38 hours. Subsequently, the culture media was replaced by [U–^{13}C]-Glc (Euroisotope, Saarbrücken, Germany) containing media and the cells were cultured for 24 hours. The cells were harvested by scraping before the exchange of culture medium (t= 0h) or after 24 hours incubation in [U–^{13}C]-Glc containing media (t= 24h).

For histone extraction, the cells were washed twice with a PBS solution (T= 37°C) and lysed by adding 1mL of ice-cold lysis buffer (PBS,

0.5% Triton X-100 (v/v), 1 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM sodium butyrate, 10 μM suberanilohydroxamic acid (SAHA), 10 mM nicotinamide). The cells were subsequently sonificated on ice (3 tunes, 10 sec. 50%) and centrifuged for 10 min at 10,000 rcf at 4°C. After aspirating the supernatant, the pellets were resuspended in 0.5 mL ice-cold resuspension buffer (13 mM EDTA, 10 mM 2-Amino-2-(hydroxymethyl)-1,3-propanediol (Tris base), pH 7.4). After centrifugation (10 min at 10,000 rcf, 4°C) and aspiration of the supernatant, pellets were resuspended in 200 μL ice-cold HPLC-H_2O. Sulfuric acid (H_2SO_4) was added to the samples to a final concentration of 0.4M. Samples were incubated on ice for 1 hour on a shaker using 400 rpm. After centrifugation (10 min at 10,000 rcf, 4°C), the supernatant containing the histone proteins, was added to 1.5 mL acetone and left at -20°C overnight to precipitate the proteins. After centrifugation, acetone was removed and the pellet was dried at room temperature. The pellet was subsequently redissolved in 50 μL HPLC-H_2O. Protein concentration was determined using the microplate BCA protein assay kit (Thermo Fisher Scientific, Dreieich, Germany) following the manufacturer's instructions.

Tryptic in-solution digestion

For tryptic digestion, 6 μg of the histone extract were diluted in HPLC-H_2O to a final volume of 50 μL. For reduction, the samples were incubated with 10 mM dithiothreitol (DTT, dissolved in 100 mM ammonium bicarbonate (ABC), pH 8.3) for 10 min at 57°C on a shaker (600 rpm). Afterwards, the samples were alkylated with 20 mM iodacetamide (IAA, dissolved in 100 mM ABC, pH 8.3) and incubated for 30 min at room temperature in the dark. Subsequently, free IAA was quenched by adding 10 mM of DTT (dissolved in 100 mM ABC, pH 8.3). For tryptic digestion, the samples were incubated with 1 μL of a trypsin solution (c = 0.2 μg/μL sequencing grade modified trypsin, dissolved in trypsin resuspension buffer, Promega, Walldorf, Germany) for 16 hours at 37°C. Afterwards, the samples were acidified by adding formic acid (FA) to final concentration of 0.1% FA.

LC-MS/MS analysis and preprocessing

For LC-MS/MS analysis, 100 ng of the tryptic peptide digests were injected on a nano-ultra pressure liquid chromatography system (Dionex UltiMate 3000 RSLCnano pro flow, Thermo Scientific, Bremen, Germany) coupled via electrospray-ionization (ESI) source to a tribrid orbitrap mass spectrometer (Orbitrap Fusion Lumos, Thermo Scientific, San Jose, CA, USA). The samples were loaded (15 μL/min) on a trapping column (nanoE MZ Sym C18, 5μm, 180umx20mm, Waters, Germany, buffer A: 0.1% FA in HPLC-H_2O; buffer B: 80% ACN, 0.1% FA in HPLC-$H_2$2O) with 5% buffer B. After sample loading, the trapping column was washed for 2 min with 5% buffer B (15 μL/min) and the peptides were eluted (250 nL/min) onto the separation column (nanoE MZ PST CSH, 130 A, C18 1.7μ, 75μ mx250mm, Waters, Germany) and separated with a gradient of 5–30% B in 25 min, followed by 30-50% in 5 min. The spray was generated from a steel emitter (Fisher Scientific, Dreiech, Germany) at a capillary voltage of 1900 V. For MS/MS measurements, an inclusion list was used for the two times acetylated isotopologue species of the H3 peptide KQLATKAAR. The fragment spectra were acquired in the orbitrap mass analyzer using a resolution of 15,000 (at m/z 200), a normalized HCD collision energy of 30, an AGC target of 5e4 and a maximum injection time of 120 ms. The full MS spectra were acquired in the orbitrap mass analyzer with a resolution of 60,000 (at m/z 200), an AGC target of 4e5, a maximum injection time of 50 ms and over a m/z-range of 350-1650 using the ETD source for internal mass calibration.

Data analysis was performed in FreeStyle 1.6 (Version 1.6.75.20, Thermo Scientific, Bremen, Germany). For the three two-times acetylated KQLATKAAR isotopologues (K(ac12C)QLATK(ac12C)AAR, m/z: 535,8195; K(ac12C)QLATK(ac13C)AAR, m/z: 536,8228; K(ac13C)QLATK(ac13C)AAR, m/z: 537,8262), extracted ion chromatograms (EIC) of all isotopes of the isotopic distributions of the different

isotopologues were generated with a mass tolerance of +/- 3 ppm. For peak detection, the Genesis algorithm was used with the following parameters: percent of highest peak: 10, minimum peak height (signal/noise): 5, signal-to-noise threshold: 3, tailing factor: 3. The peak area and the peak height was exported for each isotope of the isotopic distributions of the three different isotopologues to a tab-delimited text file.

Natural isotope correction

We used the skewed matrix approach for correction of high resolution natural isotope abundance as described Heinrich et al. 2018[8]. PICor is a Python based implementation of the isotopologue correction tool and is freely available on GitHub `https://github.com/MolecularBioinformatics/PICor` and can be used for any organic compound. The user needs to specify the chemical composition and the type of labeling experiment and supply a data file with the peak integration values from MS measurements for the isotopologues of interest. It is not required to provide measurements for all possible isotopologues, which makes it applicable for proteomics data. Full documentation can be found in the GitHub repository.

Software for isotopologue correction

The following software versions were used for the comparison of isotopologue corrections:

- AccuCor: v0.2.3.9000
- IsoCor: v2.2.0
- IsoCorrectoR: v1.6.2
- PICor (our tool): v1.0.0