analytical chemistry

pubs.acs.org/ac Article

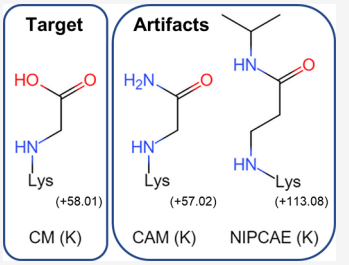
Pitfalls and Solutions in Mass Spectrometry-Based Identification of Protein Glycation

Wendong Ma, Irene Ling Ang, Kate M.K. Lei, Melody Man Ting Lam, Pengwei Zhang, and Terence Chuen Wai Poon*

Cite This: *Anal. Chem.* 2023, 95, 1829−1837 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Emerging evidence suggests that advanced glycation end-products (AGEs) such as N^{ϵ}-(carboxymethyl)lysine (CML) and N^{ϵ}-(carboxymethyl)lysine (CEL) may play important roles in certain human diseases. Reliable analytical methods are needed for their characterizations and measurements. Pitfalls have been reported for applications of LC−MS/MS to identify various types of post-translational modifications, but not yet for the case of AGEs. Here, we showed that in the absence of manual inspection, cysteine alkylation with 2-iodoacetamide (IAA) can result in false-positive/ambiguous identifications of CML >20%. They were attributed to offsite alkylation together with incorrect monoisotopic peak assignment (pitfall 1) or together with deamidation (pitfall 2). For pitfall 1, false-positive identifications can be alleviated using a peptide mass error tolerance ≤5 ppm during the database search. Pitfall 2 results in ambiguous modification assignments, which may be overcome by using other alkylation reagents. According to calculations of theoretical mass shifts, the use of other common alkylation reagents (iodoacetic acid, 2-chloroacetamide, and acrylamide) should face similar pitfalls. The use of acrylamide can result in false-positive identifications of CEL instead of CML. Subsequently, we showed that compared to IAA, the use of *N*-isopropylacrylamide (NIPAM) as an alkylation reagent achieved similar levels of proteome coverage, while reducing the offsite alkylation reactions at lysine by more than five times. Furthermore, false-positive/ambiguous identifications of CML due to the two types of pitfalls were absent when using NIPAM. NIPAM alkylation results in a unique mass shift that allows reliable identifications of CML and most likely other AGEs, such as CEL.



Protein glycation is a type of post-translational modifications (PTMs) involving complex non-enzymatic reactions of reducing sugars or reactive dicarbonyls with proteins, generating a heterogeneous group of advanced glycation end products (AGEs).[1] It predominantly occurs at free amino groups of proteins. Glyoxal (GO) and methylglyoxal (MGO) react with a lysine residue to form two common kinds of AGEs: N^{ϵ}-(carboxymethyl)lysine (CML) and N^{ϵ}-(carboxymethyl)lysine (CEL),[2] respectively. Emerging evidence suggests that AGEs on proteins contribute to development of chronic diseases/disorders such as cardiovascular diseases. AGEs play important roles in inflammation and apoptosis through interaction with receptors for AGEs (RAGEs).[1−3] Depending on their locations on proteins, AGEs can alter protein structure and function.[4] Therefore, it is important to have reliable methods for identification and characterization of the major advanced glycation sites.

To date, mass spectrometry (MS) is the most efficient technology for proteome-wide characterization of PTMs including protein glycation, revealing their identities and modification sites. Pitfalls in MS-based bottom-up proteomics approaches were reported for various types of PTMs, e.g., deamidation, phosphorylation, and ubiquitination. False PTM identifications can be resulted from incorrect identity assignment, misidentification of co-eluting isobaric peptides, incomplete database search,[5] unexpected mass shift caused by side reactions,[6] the presence of naturally occurring or artificially labeled isotopes,[7−9] etc. Whether MS-based global identification of protein glycation suffers from similar problems has not been investigated.

In bottom-up proteomics, 2-iodoacetamide (IAA) is the most commonly used reagent for cysteine alkylation. IAA reacts with the sulfhydryl group of a cysteine residue, i.e., carbamidomethylation (CAM). However, offsite alkylation can occur at the protein *N*-terminus and other amino acid residues such as lysine residue.[10] In our recent attempts to examine the global protein glycation patterns, we found that some of the identified CMLs can be false positives associated with the use of IAA. This led to the present study. Here, for the first time, we provided evidence that cysteine alkylation with IAA can result in false-positive identifications of CML. Similar pitfall might happen when identifying CEL in peptides alkylated with acrylamide (AA). Lastly, we showed that the use of *N*-isopropylacrylamide (NIPAM) for alkylation can avoid the

Received: March 22, 2022
Accepted: December 30, 2022
Published: January 11, 2023

https://doi.org/10.1021/acs.analchem.2c01261
Anal. Chem. 2023, 95, 1829−1837

false-positive identifications of CML and most likely other AGEs as well, such as CEL.

EXPERIMENTAL SECTION

Prediction of False Identifications Associated with IAA. False identifications of PTM on peptides are commonly attributed to the same or highly close mass shifts caused by the expected modification(s) and the actual modification(s). The latter is incorrectly identified as the former. The actual mass shift can be caused by a single modification, or a combination of co-occurring modifications. An additional mass shift may have resulted from an incorrect monoisotopic peak assignment. IAA is the most commonly used alkylation reagent in bottom-up proteomics. The mass shift caused by IAA-derived CAM (+57.021464 Da) is 0.984016 Da less than the mass shift caused by GO-derived carboxymethylation(CM) (+58.005479 Da). Both modifications can happen at lysine. To balance the mass shift difference, there are two possible ways, pitfall 1 and pitfall 2, as follows.

Pitfall 1: Correct monoisotopic peak assignment is an important step for both data-dependent acquisition and protein sequence database search. It is not uncommon that the second isotope of a peptide has been mistakenly assigned as the monoisotope, resulting in a peptide mass of 1.003355 Da higher. Incorrect monoisotopic peak assignment is a common cause of false-positive identification of deamidated peptides, even when a high-resolution mass spectrometer is used for acquisition of LC−MS/MS data.[11] An incorrect monoisotopic peak assignment together with an instance of CAM can result in a total mass shift of 58.024819 Da, which is only slightly higher than the mass shift caused by an instance of GO-derived CM (mass shift difference = 0.019339 Da). When the second isotopic peak of a carbamidomethylated peptide (e.g., containing one ^{13}C) is selected for the protein sequence database search, it can be incorrectly assigned as a carboxymethylated peptide.

Pitfall 2: Deamidation of asparagine is an artifact known to be frequently introduced during preparation of proteomic samples. It also occurs on glutamine, but at a lower rate. Deamidation has been related to misidentification of *N*-linked glycosylation sites.[6,8,9] We speculate that the mass shift difference between IAA-derived CAM and GO-derived CM can be balanced by a single instance of deamidation (+0.984016 Da) co-occurring with CAM. An ambiguous identification can be caused by an absence of *b*-ions or *y*-ions generated by fragmentating between a modified lysine and an (deamidated) asparagine.

Datasets for Protein Glycation and Artifact Analysis. Two sets of raw LC−MS/MS data, both generated by using orbitrap-based platforms and higher-energy collisional dissociation (HCD), were analyzed in the present study. The first data set was generated and published by Keilhauer et al. for examining protein glycation in HeLa cell lysates, accessible via ProteomeXchange with identifier PXD004182.[12] In their experiments, samples were analyzed using a Q Exactive HF mass spectrometer (Thermo Fisher Scientific, Massachusetts, USA). Full scans (m/z 300−1650) were acquired with a resolution (FWHM at 200 m/z) of 60,000, and fragmentation scans were acquired with 16,000 resolution. The second set of data were generated in-house for examination of protein glycation in *Escherichia coli* (*E. coli*) cell lysates. For evaluation of performance of NIPAM compared with IAA as the alkylation reagent, in-house data generated from the *E. coli* proteome and the HEK293 cell line proteome were analyzed. All in-house samples were analyzed using a Q Exactive mass spectrometer (Thermo Fisher Scientific). Full scans (m/z 300−2000) were acquired with 70,000 resolution, and fragmentation scans were acquired with 17,500 resolution.

Preparation of Lysate Digests. *E. coli* lysate proteins were obtained from Bio-Rad (California, USA). HEK293 cells (ATCC, CRL-1573) were harvested with a lysis buffer (0.1 M Tris−HCl, pH 7.5, 2% sodium dodecyl sulfate, 0.1 M dithiothreitol), and lysate proteins were obtained by sonication. Tryptic digests were produced according to the standard filter-aided sample preparation (FASP) protocol.[13] Lysate proteins were reduced with 0.1 M dithiothreitol at 56 °C for 30 min. After washing with U8 solution (8 M urea in 0.1 M Tris/HCl, pH 8.5), the proteins were alkylated with 50 mM IAA or NIPAM (purity ≥99%, Sigma-Aldrich, Missouri, USA) in U8 solution at room temperature in the dark for 20 min.

LC−MS/MS Measurement. The tryptic digests were cleaned up with C18 ZipTip (Millipore) according to the manufacturer's instructions. After SpeedVac drying, the cleaned peptides were reconstituted in 5% acetonitrile (ACN) containing 0.1% formic acid. The peptides were separated on an EASY-Spray LC Column (75 μm × 50 cm, 2 μm 100 Å particles, Thermo Fisher Scientific) with an ACN gradient from 5 to 35% in 0.1% formic acid over 300 min at a constant flow of 250 nL/min. Peptides eluting from the column were analyzed on the Q Exactive mass spectrometer operated in a data-dependent acquisition mode. Survey full-scan MS spectra (300−2000 m/z) were acquired at 70,000 resolution. Top 20 most intense ions were sequentially isolated and fragmented by HCD at normalized collision energy of 25. MS/MS spectra were acquired at 17,500 resolution.

Data Analysis. All LC−MS/MS data were searched against the human or *E. coli* protein sequence database using PEAKS Studio Xpro[14] (PEAKS for short, Bioinformatics Solutions Inc., Ontario, Canada) and MaxQuant[15,16] (version 2.0.1.0, Max Planck Institute of Biochemistry, Germany). A reviewed *E. coli* K12 reference proteome (2033 sequences, August 2021) and a reviewed human reference proteome (20,371 sequences, August 2021) from UniProt were used. For all searches, oxidation of methionine (+15.994915 Da), deamidation of asparagine (+0.984016 Da), and acetylation of protein *N*-termini (+42.010565 Da) were set as variable modifications. When using IAA as the alkylation reagent, CAM of cysteine (+57.021464 Da) was set as a fixed modification. When identifying protein glycation, CM of lysine (i.e., CML, +58.005479 Da) was set as an additional variable modification. This is referred to as "CML search" in the following sections. When searching for false-positive identifications of protein glycation caused by the use of IAA as the alkylation reagent, CAM of lysine (+57.021464 Da) was set as an additional variable modification. This is referred to as "pitfall search". When using PEAKS, the peptide mass error tolerance was 5 or 10 ppm and the fragment mass error tolerance was 0.05 Da. When using MaxQuant, the default mass error tolerance values were used for the peptides and the fragments, i.e., first search peptide mass tolerance of 20 ppm, main search peptide mass tolerance of 4.5 ppm, and MS/MS match tolerance of 20 ppm. Default cleavage site specificity for trypsin was used. Two missed cleavages were allowed. For PEAKS, the proteins were identified at PSM FDR of 1%, peptide FDR of 1%, and protein FDR of 1%. For MaxQuant, the default settings were used, i.e., 1% PSM FDR and 1% protein FDR. A schematic flowchart

https://doi.org/10.1021/acs.analchem.2c01261
Anal. Chem. 2023, 95, 1829−1837

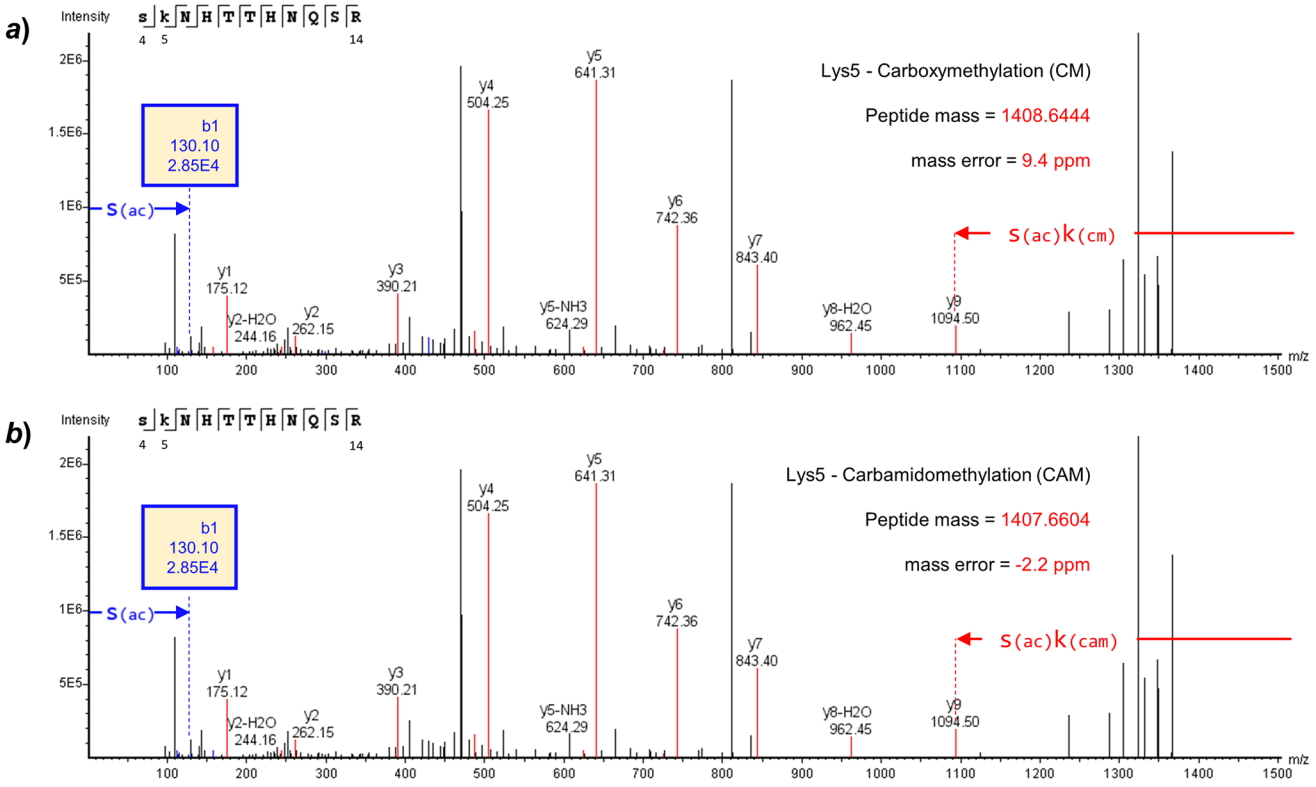


Figure 1. A representative case of the false-positive identification of CML on a peptide due to pitfall 1. MS/MS spectra of the $_4$SKNHTTHNQSR$_{14}$ peptide identified in the HeLa digests: (a) mapped to the peptide sequence containing an instance of CML (mass = 1408.6444, m/z = 470.5598, charge = +3, error = 9.4 ppm, RT = 10.65 min) and (b) mapped to the same peptide sequence containing an instance of CAM (mass = 1407.6604, m/z = 470.2264, charge = +3, error = −2.2 ppm, RT = 10.65 min). RT: retention time; blue boxes and arrows highlight the key fragment peaks and *b*-ions, red arrows highlight the key fragment *y*-ions in identification; ac: acetylation; cm: carboxymethylation; cam: carbamidomethylation (modified by IAA).

(Figure S1) is provided to illustrate how various criteria were used to identify false-positive cases due to pitfall 1 or ambiguous cases due to pitfall 2.

When using NIPAM for alkylation, the data analysis method was the same as abovementioned, except that *N*-isopropylcarbamidoethylation (NIPCAE, +113.084064 Da) was used to replace CAM. For comparing the alkylation performance between NIPAM and IAA, degrees of offsite modification at lysine residues were examined through calculating the percentage of unique proteins/peptides modified with NIPAM or IAA at lysine, i.e., number of unique proteins (or peptides) with offsite alkylation at one or more lysine residues ÷ total number of unique proteins (or peptides) identified × 100%. Two-tailed t test was used to compare the results between the experiments using NIPAM (n = 5) and those using IAA (n = 5).

RESULTS AND DISCUSSION

Pitfall 1—Incorrect Monoisotopic Peak Assignment. For a peptide having CAM on lysine but being wrongly identified to have CML due to an incorrect assignment of its second isotopic form as the monoisotopic form, it has the following six characteristics. First, being the monoisotopic form, its amino acid sequence should be the same as that of the second isotopic form, i.e., the peptide with a false-positive identification of CML. Second, its monoisotopic form should be 1.0 Da lighter than the second isotopic form. Third, observed retention times (RTs) of the isotopic forms of a peptide should be almost identical. Fourth, the two isotopic forms should co-occur in one or more consecutive MS scans. Fifth, they should have the same charge state. Sixth, an incorrect monoisotopic peak assignment together with an instance of CAM should result in a total mass shift of 58.024819 Da, which is only slightly higher than the mass shift caused by an instance of CM (+58.005479 Da, mass shift difference = 0.019339 Da). In both Keilhauer et al.'s study[12] and the present study, high-resolution orbitrap mass spectrometers were used to acquire the LC−MS/MS data. When the second isotope of a tryptic peptide containing an instance of CM was mistakenly identified to be a peptide having a CML, the peptide mass error should be usually more than 5 ppm. In contrast, for a correct identification of CM on a lysine residue, the peptide mass error should be usually less than 5 ppm. Theoretically, using a peptide mass error tolerance of 5 ppm in the databases search, the number of false-positive identifications due to pitfall 1 should be greatly reduced.

Keilhauer et al.'s LC−MS/MS data for the HeLa cell lysates was first subjected to the CML search using PEAKS at a peptide mass error tolerance of 10 ppm. Five tentative false positives that were containing CAM on lysine and fulfilling all the aforementioned characteristics were found.

Their corresponding peptides with a false-positive identification of CML were 1.0 Da heavier and had peptide mass errors greater than 5 ppm. All of them were successfully confirmed by manual inspection of the LC−MS/MS spectra (Tables S1 and S2, Figure 1, Figures S2−S5). After changing the mass error tolerance of the peptides from 10 to 5 ppm in

https://doi.org/10.1021/acs.analchem.2c01261
Anal. Chem. 2023, 95, 1829−1837

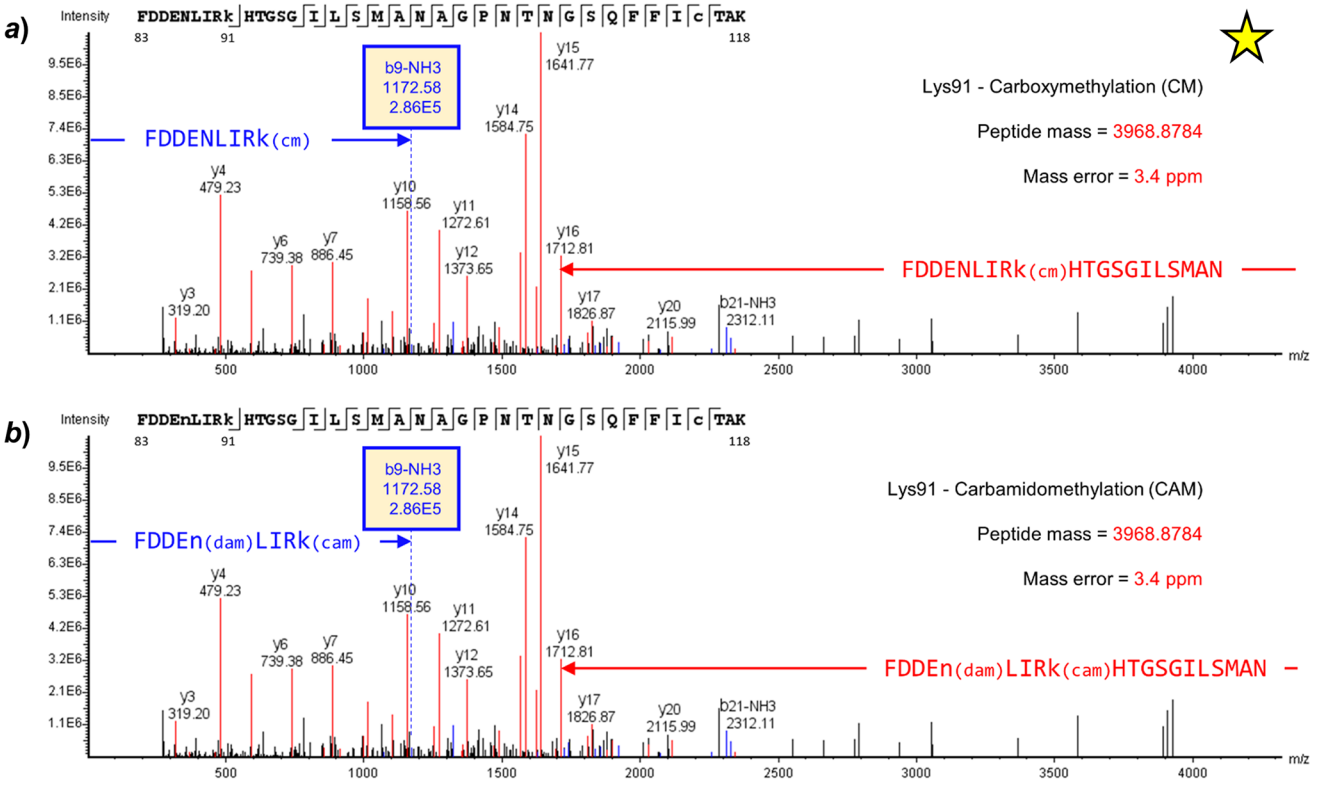


Figure 2. A representative case of the ambiguous identification of CML on a peptide due to pitfall 2. MS/MS spectra of the $_{83}$FDDENLIRKHTGSGILSMANAGPNTNGSQFFICTAK$_{118}$ peptide identified in the HeLa digests: (a) mapped to a peptide sequence containing an instance of CML (mass = 3968.8784, m/z = 1323.9712, charge = +3, error = 3.4 ppm, RT = 105.04 min) and (b) mapped to the same peptide sequence containing an instance of CAM and an instance of deamidation at asparagine (mass = 3968.8784, m/z = 1323.9712, charge = +3, error = 3.4 ppm, RT = 105.04 min) under a peptide mass error tolerance of 5 ppm. RT: retention time; blue boxes and arrows highlight the key fragment peaks and *b*-ions, red arrows highlight the key fragment *y*-ions in identification; dam: deamidation; cm: carboxymethylation; cam: carbamidomethylation (modified by IAA); star: Lys91 was recognized as a site for CML in Keilhauer et al.'s article.[12]

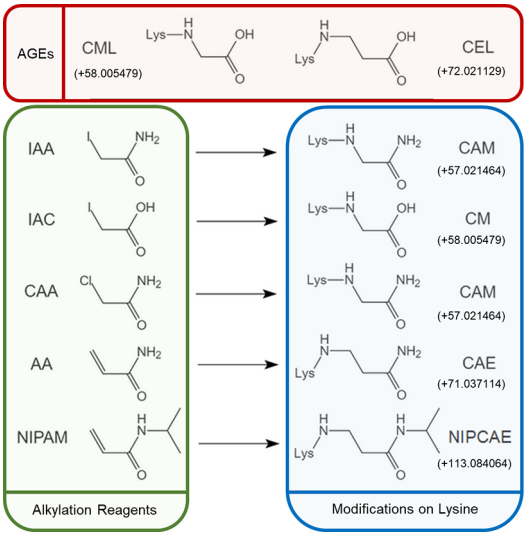


Figure 3. Structures of typical AGEs (red) and modifications (blue) at lysine caused by four common alkylation reagents and NIPAM (green). The resulting mass shifts are provided in the brackets.

the CML search, only one tentative false positive was found. However, it was subsequently rejected by manual inspection.

A representative false-positive case is provided in Figure 1. From the CML search, K5 of peptide $_{4}$SKNHTTHNQSR$_{14}$ of 60S ribosomal protein L29 was identified as a CM site under the peptide mass error tolerance of 10 ppm (Figure 1a). The observed peptide mass error was >5 ppm. Through the pitfall search and careful manual inspection on the MS/MS spectra, charge state (+3), RT (10.65 min), and mass difference between the two peptides (0.33 Da for a peptide having a charge state of +3), we revealed that the identification of K5 as a CM site from the CML search under a peptide mass error tolerance of 10 ppm should have been a false-positive identification attributed to pitfall 1. A correct modification assignment (Figure 1b) resulted in a much lower peptide mass error (−2.2 ppm, compared with 9.4 ppm).

The presence of pitfall 1 was verified with the in-house LC−MS/MS data of the *E. coli* lysates. With a peptide mass error tolerance of 10 ppm, seven tentative false positives were found and two were subsequently confirmed by manual inspection (Tables S1 and S2, Figure 5a,b, Figure S6). With a peptide mass error tolerance of 5 ppm and manual inspection, no false positive due to pitfall 1 was found.

Next, we examined whether the use of MaxQuant with the recommended peptide mass error tolerance (i.e., 20 ppm in the first search and 4.5 ppm in the main search) can avoid false-positive identifications of CML. After manual inspection, no false-positive due to pitfall 1 was found in both Keilhauer et al.'s LC−MS/MS data of the HeLa cell lysates and the in-house LC−MS/MS data of the *E. coli* lysates.

In the proteomic analyses based on using the high-resolution Orbitrap HCD-MS/MS, a peptide mass error tolerance at 10 ppm is still widely used during database search. While for other

https://doi.org/10.1021/acs.analchem.2c01261
Anal. Chem. 2023, 95, 1829−1837

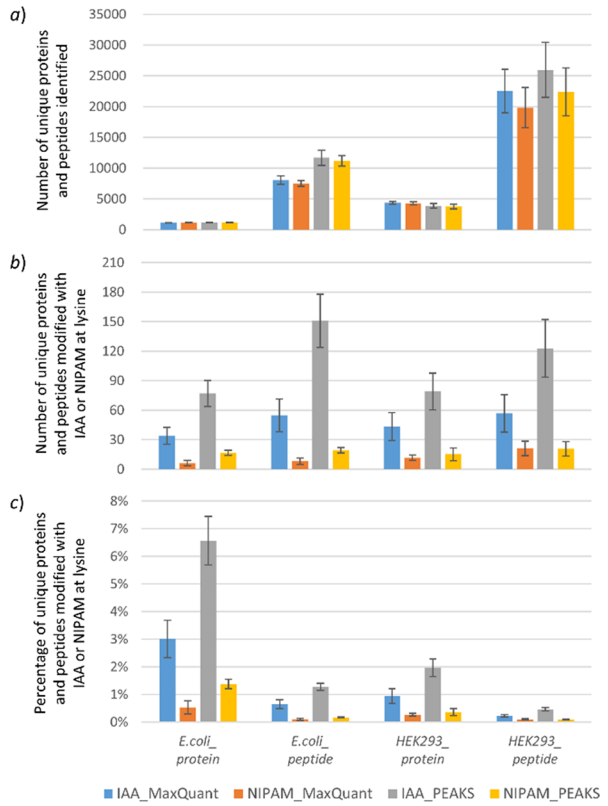


Figure 4. Comparison of bottom-up proteomic profiling results for lysate digests of *E. coli* and HEK293 cells with IAA and NIPAM as alkylation reagents. The proteome database search was performed using PEAKS or MaxQuant with the peptide mass error tolerance of 5 or 4.5 ppm, respectively. (a) The number of unique proteins and unique peptides identified. (b) The number of unique proteins and unique peptides modified with IAA or NIPAM at lysine. (c) The percentage of unique proteins and unique peptides modified with IAA or NIPAM at lysine. The average values obtained from five independent experiments are provided. The error bars indicate the standard error of the means (SEMs).

LC−MS instruments, for instance, triple-TOF MS/MS, a higher peptide mass error tolerance such as 15 ppm is commonly recommended. This can possibly aggravate the mistakes due to pitfall 1. To avoid pitfall 1, a peptide mass error tolerance at 5 ppm or a lower value should be suggested.

Pitfall 2—Co-Occurrence of Deamidation. For a peptide having CAM on lysine but being wrongly assigned to have a CML due to the co-occurrence of deamidation, both modification assignments should be originated from the monoisotopic form of the same peptide. We searched for any peptides that were assigned to have CM at lysine in the CML search and also assigned to have CAM at lysine and deamidation at asparagine in the pitfall search (Figure S1). For the HeLa lysates, 17 and 18 tentative ambiguous cases were found when using PEAKS with the peptide mass error tolerance of 10 and 5 ppm, respectively; 12 tentative ambiguous cases were found when using MaxQuant at the default settings. After manual inspection, two and nine peptides with pitfall 2-associated ambiguous assignments were found using PEAKS and MaxQuant with the peptide mass error tolerance of 5 and 4.5 ppm, respectively (Tables S1 and S2, Figure 2, Figures S7 and S11−S19).

One representative case with an ambiguous identification of CML was found using both PEAKS (Figure 2) and MaxQuant (Figure S11). As shown in Figure 2, according to the CML search results, K91 of peptide $_{83}$FDDENLIRKHTGSGILSMA-NAGPNTNGSQFFICTAK$_{118}$ from peptidyl-prolyl *cis-trans* isomerase A-like 4C was identified as a site of CM (Figure 2a) under the peptide error mass tolerance of 5 ppm. Through the pitfall search under the same peptide error mass tolerance and careful manual inspection on the MS spectra, MS/MS spectra, charge state (+3), RT (105.04 min), and observed *m*/*z* values (expecting to be the same, i.e., 1323.9712 at a charge state of +3), the same peptide was found to have CAM at K91 and deamidation at N87 (Figure 2b). With the co-existence of deamidation at N87, yielding a mass increase of 0.984016 Da, the carbamidomethylated fragment $_{83}$FDDENLIRKHTGSG$_{96}$ with a total mass shift of +58.005479 Da can probably be misidentified as the fragment containing a single instance of CM (+58.005479 Da) at lysine. Due to the identical mass shift in both cases, this ambiguity cannot be distinguishable. Of note, this case shown in Figure 2 (and Figure S11) and two other ambiguous cases (Figures S12 and S15) had been recognized as CML peptides by Keilhauer et al.,[12] accounting for 27% of their 11 reported CML peptides.

We subsequently attempted to verify the presence of pitfall 2 using the in-house LC−MS/MS data of the *E. coli* lysates. After manual inspection, four ambiguous cases due to pitfall 2 were found using PEAKS and four were found using MaxQuant, including one overlapping case (Tables S1 and S2, Figure 6a,b, Figures S8−S10 and S20−S23). These ambiguous cases confirmed that pitfall 2 can commonly happen in protein glycation analysis when using IAA for alkylation.

Summary of False-Positive/Ambiguous Cases. The number of false-positive/ambiguous cases associated with pitfall 1 or pitfall 2 is summarized in Table S2. In the absence of manual inspection of the MS/MS spectra, cysteine alkylation with IAA can result in tentative false-positive/ambiguous identifications of CML >20% (23% for HeLa lysates, 28% for *E. coli* lysates, PEAKS with the peptide mass error tolerance of 5 ppm). Majority of the tentative false-positive/ambiguous cases were associated with pitfall 2. It appeared that PEAKS identified more tentative false-positive/ambiguous cases compared with MaxQuant. After manual inspection of the MS/MS spectra for the results obtained with the peptide mass error tolerance of 5 ppm (PEAKS) or 4.5 ppm (MaxQuant), the total number of confirmed false-positive/ambiguous cases was fewer than 10 (ranging from 2 to 9) and all were due to pitfall 2. The number of confirmed false-positive/ambiguous cases was similar for the two protein sequence search engines.

False-positive identifications of CML due to pitfall 1 can be alleviated by using a peptide mass error tolerance ≤5 ppm. However, pitfall 2 leads to an undistinguishable ambiguity. It is not possible to tell whether it is a true CML or the co-occurrence of CAM at lysine and deamidation. To solve this issue, other strategies need to be developed, such as using another alkylation reagent to replace IAA.

Searching for an Alternative to IAA. When choosing an alternative for protein alkylation, it should not only avoid false-positive identifications of CML but also avoid false-positive identifications of other AGEs, such as CEL. Assuming for all alkylation reagents, offsite alkylation reactions can occur at other amino acids, including lysine, we examined the possibility of similar pitfalls by calculating the total mass shift values for the second isotopic form and for the deamidated form of a peptide with an instance of alkylation at lysine and

https://doi.org/10.1021/acs.analchem.2c01261
Anal. Chem. 2023, 95, 1829−1837

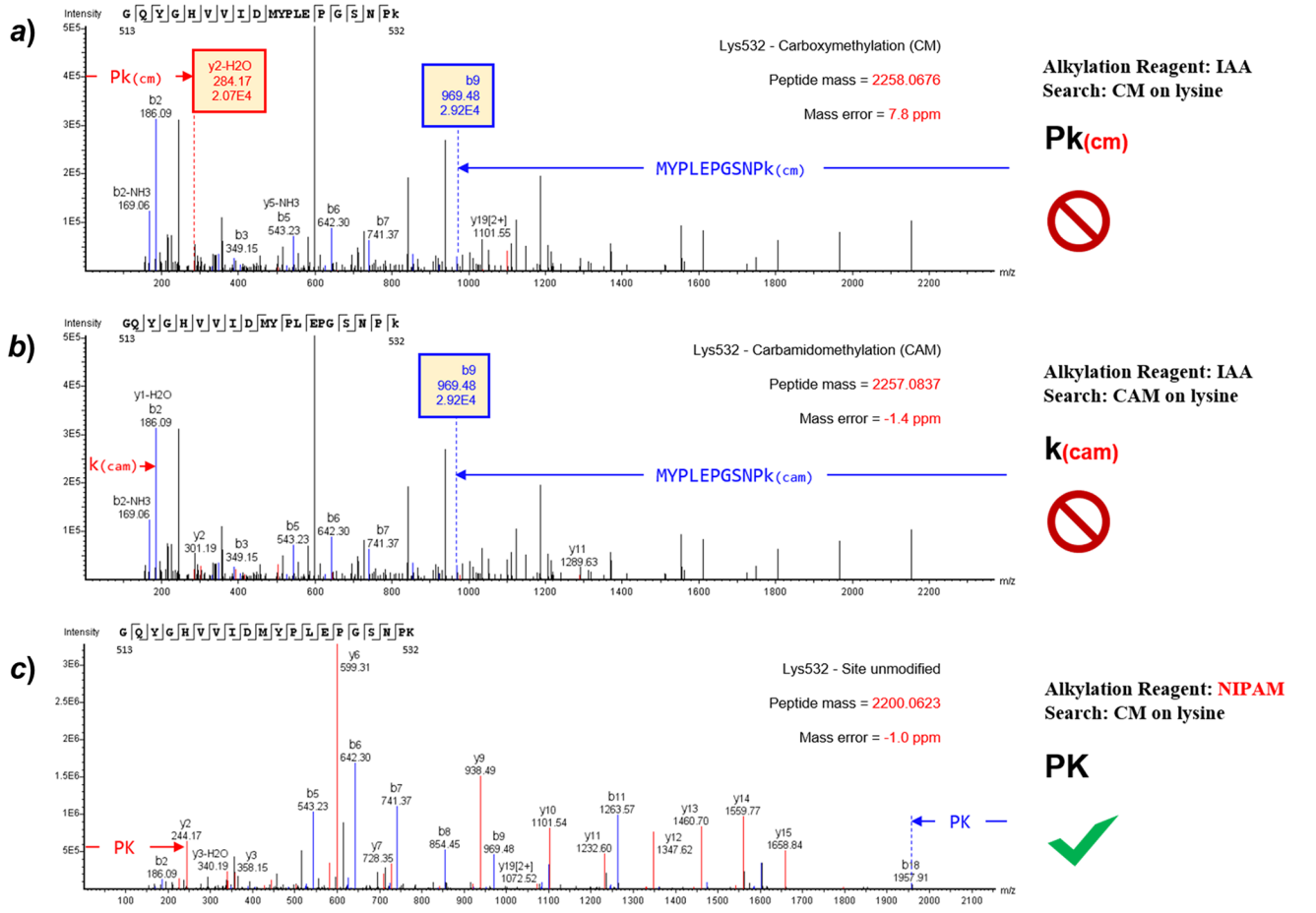


Figure 5. A representative case of using NIPAM to alleviate the false-positive identification of CML on a peptide due to pitfall 1. MS/MS spectra of $_{513}$GQYGHVVIDMYPLEPGSNPK$_{532}$ peptide in the *E. coli* digests: (a) mapped to a peptide sequence containing an instance of CML (alkylation reagent = IAA, mass = 2258.0676, m/z = 753.7024, charge = +3, error = 7.8 ppm, RT = 96.96 min), (b) mapped to the same peptide sequence containing an instance of CAM (alkylation reagent = IAA, mass = 2257.0837, m/z = 753.3674, charge = +3, error = −1.4 ppm, RT = 96.96 min), and (c) mapped to the same peptide sequence not containing any modification (alkylation reagent = NIPAM, mass = 2200.0623, m/z = 1101.0374, charge = +2, error = −1.0 ppm, RT = 153.22 min). RT: retention time; blue boxes and arrows highlight the key fragment peaks and *b*-ions, red boxes and arrows highlight the key fragment peaks and *y*-ions in identification; cm: carboxymethylation; cam: carbamidomethylation (modified by IAA).

compared with the mass shift values caused by a CML (+58.005479 Da) or a CEL (+72.021129 Da). Figure 3 and Table S3 summarize the modifications and mass shift values for four commonly used alkylation reagents, namely, IAA, iodoacetic acid (IAC), 2-chloroacetamide (CAA), and acrylamide (AA).[10] The use of IAC can directly result in CM at lysine (+58.005479 Da), which cannot be differentiated from a CML. The use of CAA can result in a mass shift (+57.021464 Da) equal to that of IAA (+57.021464 Da). The use of AA leads to carbamidomethylation (CAE, +71.037114 Da). This can result in false-positive identifications of CEL instead of CML. Therefore, none of the four commonly used alkylation reagents can avoid false-positive identifications of CML and CEL.

Compared to iodine-containing alkylation reagents, AA was suggested to be a better alkylation reagent for minimizing offsite reactions in shotgun proteomics.[10] Therefore, we attempted to search for a stable derivative of AA, of which the mass is slightly higher and pure preparation is commercially available at low cost. We finally found *N*-isopropylacrylamide (NIPAM). Alkylation with NIPAM will lead to a formation of NIPCAE with a mass shift of +113.084064 Da, which is obviously different from those of CML and CEL (Figure 3), as well as other common PTMs.

Evaluation of NIPAM as an Alkylation Reagent. Lysate digests of *E. coli* cells and HEK293 cells were prepared with the alkylation reagent IAA or NIPAM and subjected to proteome profiling and identification of offsite alkylation reactions at lysine. When NIPAM was used, the numbers of unique proteins and peptides obtained were not significantly different from those obtained for IAA (P values >0.500). However, the numbers and percentages of offsite alkylation reactions at lysine can be more than five times lower when using NIPAM (P values ≤0.050 for MaxQuant, P values ≤0.002 for PEAKS). This is consistent with the reported advantage of AA in terms of reducing offsite alkylation reactions.[10] These results are summarized in Figure 4 and Table S4. It is worth noting that in general, compared with MaxQuant, PEAKS identified more unique proteins/peptides modified with IAA or NIPAM at lysine. This can be the underlying cause for more tentative false-positive/ambiguous CML cases when IAA and PEAKS were used in the analysis (Table S2).

Alleviation of False-Positive Identifications of CML Using NIPAM. We found that the replacement of IAA with

https://doi.org/10.1021/acs.analchem.2c01261
Anal. Chem. 2023, 95, 1829−1837

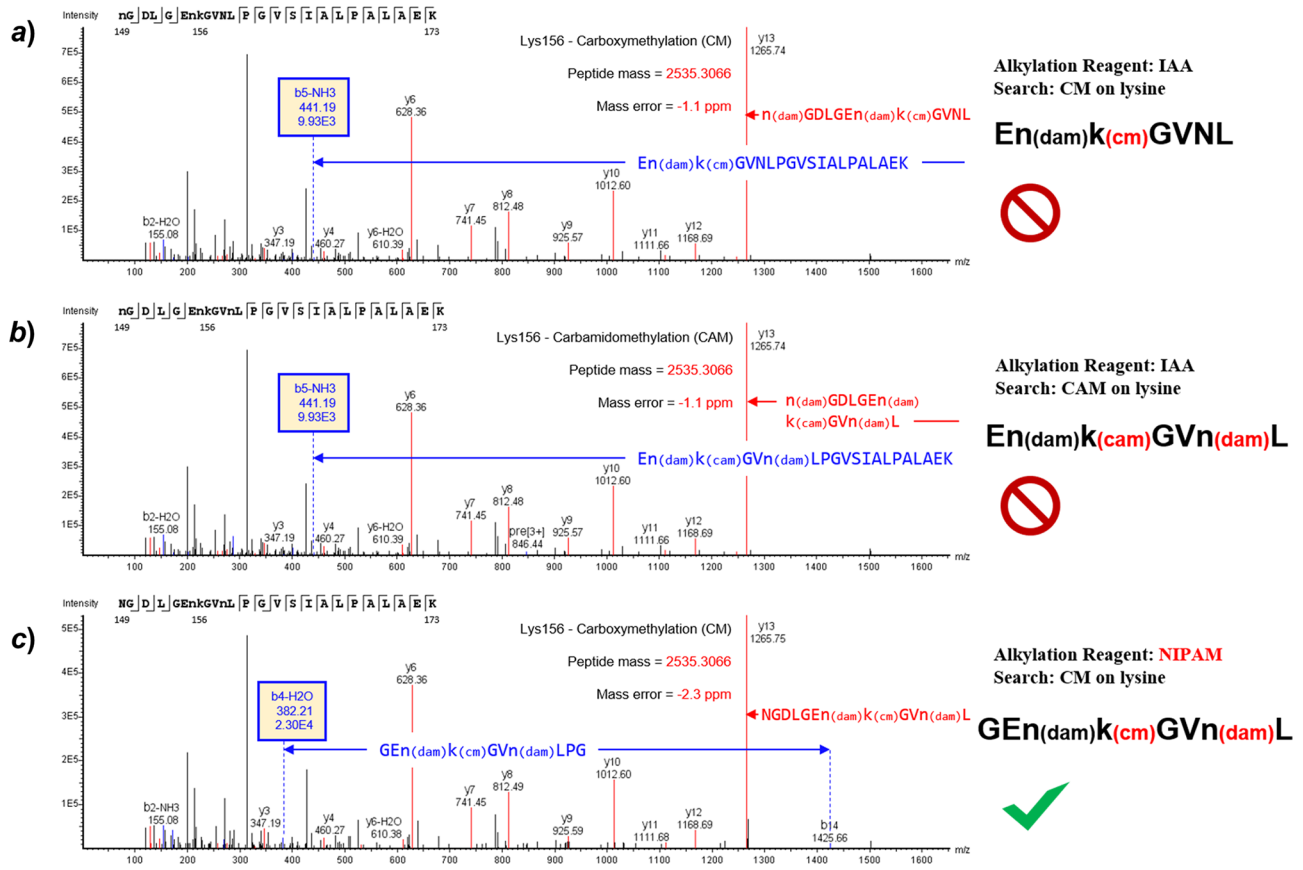


Figure 6. A representative case of using NIPAM to alleviate the ambiguous identification of CML on a peptide due to pitfall 2. MS/MS spectra of the $_{149}$NGDLGENKGVNLPGVSIALPALAEK$_{173}$ peptide in the *E. coli* digests: (a) mapped to a peptide sequence containing an instance of CML and an instance of deamidation at asparagine (alkylation reagent = IAA, mass = 2535.3066, m/z = 846.1085, error = −1.1 ppm, charge = +3, RT = 272.95 min), (b) mapped to the same peptide sequence containing an instance of CAM and two instances of deamidation at asparagine (alkylation reagent = IAA, mass = 2535.3066, m/z = 846.1085, error = −1.1 ppm, charge = +3, RT = 272.95 min), and (c) mapped to the same peptide sequence containing an instance of CML and an instance of deamidation at asparagine (alkylation reagent = NIPAM, mass = 2535.3066, m/z = 846.1075, error = −2.3 ppm, charge = +3, RT = 274.59 min). RT: retention time; blue boxes and arrows highlight the key fragment peaks and *b*-ions, red arrows highlight the key fragment *y*-ions in identification; dam: deamidation; cm: carboxymethylation; cam: carbamidomethylation (modified by IAA).

NIPAM can alleviate the false-positive or ambiguous identifications of CML in the *E. coli* lysate. For pitfall 1, Figure 5 illustrates a representative false-positive case aforementioned (Table S1). When using IAA as the alkylation reagent, we suspected that the K532 of peptide $_{513}$GQYGHVVIDMYPLEPGSNPK$_{532}$ of elongation factor G was a CAM site (peptide mass error = −1.4 ppm, Figure 5b) and the second isotope of this IAA-modified peptide was wrongly assigned to have a CML (peptide mass error = 7.8 ppm, Figure 5a), resulting in a peptide mass error >5 ppm. When using NIPAM as the alkylation reagent, this CML-containing peptide cannot be found anymore. However, a peptide with the same amino acid sequence but not carrying any modifications remained to be detectable (peptide mass error = −1.0 ppm, Figure 5c). Our results strongly supported that the identification of a CML on peptide $_{513}$GQYGHVVIDMYPLEPGSNPK$_{532}$ was a false positive. This modified peptide did not exist in the *E. coli* digests.

For pitfall 2, Figure 6 illustrates a representative ambiguous case aforementioned (Table S1). When using IAA as the alkylation reagent, peptide $_{149}$NGDLGENKGVNLPGVSIALPALAEK$_{173}$ of pyruvate kinase I was identified to have CM at K156 in the CML search (peptide mass error = −1.1 ppm, Figure 6a). However, the same peptide was identified to have CAM at K156 in the pitfall search (peptide mass error = −1.1 ppm, Figure 6b). When using NIPAM as the alkylation reagent, the same peptide was observed (similar RT, similar m/z, same charge state, similar MS/MS spectrum) and identified to have CM at K156 (peptide mass error = −2.3 ppm, Figure 6c). This confirmed that the identification of a CML on peptide $_{149}$NGDLGENKGVNLPGV-SIALPALAEK$_{173}$ was a true-positive identification.

The numbers of false-positive/ambiguous cases identified by using NIPAM are summarized in Table S2. When using NIPAM instead of IAA, the number of unique peptides identified to have a CML dropped by 30 to 37% in the *E. coli* lysates. In summary, applying NIPAM avoided both pitfalls.

CONCLUSIONS

We revealed, for the first time, the presence of two pitfalls when applying the existing MS-based strategy for proteome-wide identification of CMLs. IAA is the most commonly used reagent for alkylation of cysteine during the sample preparation. False-positive identifications of CML were attributed to the offsite IAA alkylation together with incorrect monoisotopic peak assignment (pitfall 1) or together with

https://doi.org/10.1021/acs.analchem.2c01261
Anal. Chem. 2023, 95, 1829−1837

deamidation (pitfall 2). For pitfall 1, false-positive identifications of CML can be alleviated using a peptide mass error tolerance ≤5 ppm during the database search. For pitfall 2, ambiguous modification assignments resulted. To overcome this ambiguity, IAA should be replaced by another alkylation reagent.

Offsite alkylation is a common phenomenon for the four commonly used alkylation reagents (IAA, IAC, CAA, and AA). We here only provide solid evidence for the presence of false identifications of CML in lysate digests alkylated with IAA. However, according to the prediction of mass shifts upon alkylation, we speculate that use of IAC, CAA, or AA also suffers from similar pitfalls. The use of AA will not result in false-positive identifications of CML but can result in false-positive identifications of CEL instead.

Last but not least, we have shown, for the first time, that NIPAM was a suitable alkylation reagent for preparing bottom-up proteomic samples. Compared with IAA, NIPAM can result in fewer offsite reactions at lysine, as inherited from AA. NIPAM alkylation results in a mass shift substantially greater than those caused by protein glycation. This allows reliable identifications of CML and most likely other AGEs as well, such as CEL. Considering that NIPAM is chemically stable and commercially available at low cost, the advantage of using NIPAM as an alkylation reagent in proteomics is worth further exploration.

■ ASSOCIATED CONTENT

Data Availability Statement

The MS proteomics data and MaxQuant output files have been deposited to the ProteomeXchange[17] Consortium via the PRIDE[18] partner repository with the dataset identifier PXD032235.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.2c01261.

Table S1. Summary of -peptides found to have CML that was either false-positive or ambiguous. Table S2. Comparison of the numbers of CML-peptides and their respective false-positive/ambiguous cases identified in the bottom-up proteomic profiling experiments under different conditions. Table S3. Comparison of the mass shifts between an AGE (CML or CEL) and a modification at lysine caused by an offsite alkylation with IAA, IAC, CAA, AA, or NIPAM. Table S4. Comparison of bottom-up proteomic profiling results for lysate digests of *E. coli* and HEK293 cells with IAA or NIPAM as the alkylation reagent. Figure S1. A schematic flowchart illustrating the major steps for finding peptides with a false-positive/ambiguous identification of CML. Figures S2−S23. Cases of the false-positive identifications of CML on a peptide due to pitfall 1 or 2 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Terence Chuen Wai Poon − *Pilot Laboratory, MOE Frontier Science Centre for Precision Oncology, Centre for Precision Medicine Research and Training, Institute of Translational Medicine, Faculty of Health Sciences, University of Macau, Macau 999078, China*; Phone: +853 8822-4501; Email: tcwpoon@um.edu.mo

Authors

Wendong Ma − *Pilot Laboratory, MOE Frontier Science Centre for Precision Oncology, Centre for Precision Medicine Research and Training, Institute of Translational Medicine, Faculty of Health Sciences, University of Macau, Macau 999078, China*; orcid.org/0000-0001-7805-7910

Irene Ling Ang − *Pilot Laboratory, MOE Frontier Science Centre for Precision Oncology, Centre for Precision Medicine Research and Training, Institute of Translational Medicine, Faculty of Health Sciences, University of Macau, Macau 999078, China*

Kate M.K. Lei − *Pilot Laboratory, MOE Frontier Science Centre for Precision Oncology, Centre for Precision Medicine Research and Training, Institute of Translational Medicine, Faculty of Health Sciences, University of Macau, Macau 999078, China*

Melody Man Ting Lam − *Proteomics Core, Faculty of Health Sciences, University of Macau, Macau 999078, China*; orcid.org/0000-0002-7136-5029

Pengwei Zhang − *Department of Laboratory Medicine, The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, Guangdong 510120, China*; orcid.org/0000-0002-7692-8468

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.2c01261

Author Contributions

The manuscript was written through contributions of all authors.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This project was supported by the Macau Science and Technology Development Fund (FDCT) Grant 0011/2019/AKP.

■ REFERENCES

(1) Singh, R.; Barden, A.; Mori, T.; Beilin, L. *Diabetologia* **2001**, *44*, 129−146.

(2) Vistoli, G.; De Maddis, D.; Cipak, A.; Zarkovic, N.; Carini, M.; Aldini, G. *Free Radical Res.* **2013**, *47*, 3−27.

(3) Stirban, A.; Gawlowski, T.; Roden, M. *Mol. Metab.* **2014**, *3*, 94−108.

(4) Zheng, Q.; Omans, N. D.; Leicher, R.; Osunsade, A.; Agustinus, A. S.; Finkin-Groner, E.; D'Ambrosio, H.; Liu, B.; Chandarlapaty, S.; Liu, S.; David, Y. *Nat. Commun.* **2019**, *10*, 1289.

(5) Kim, M. S.; Zhong, J.; Pandey, A. *Proteomics* **2016**, *16*, 700−714.

(6) Palmisano, G.; Melo-Braga, M. N.; Engholm-Keller, K.; Parker, B. L.; Larsen, M. R. *J. Proteome Res.* **2012**, *11*, 1949−1957.

(7) Wang, X.; Swensen, A. C.; Zhang, T.; Piehowski, P. D.; Gaffrey, M. J.; Monroe, M. E.; Zhu, Y.; Dong, H.; Qian, W. J. *J. Proteome Res.* **2020**, *19*, 1863−1872.

(8) Gonzalez, J.; Takao, T.; Hori, H.; Besada, V.; Rodriguez, R.; Padron, G.; Shimonishi, Y. *Anal. Biochem.* **1992**, *205*, 151−158.

(9) Angel, P. M.; Lim, J. M.; Wells, L.; Bergmann, C.; Orlando, R. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 674−682.

(10) Müller, T.; Winter, D. *Mol. Cell. Proteomics* **2017**, *16*, 1173−1187.

(11) Hao, P.; Ren, Y.; Alpert, A. J.; Sze, S. K. *Mol. Cell. Proteomics* **2011**, *10*, O111.009381.

(12) Keilhauer, E. C.; Geyer, P. E.; Mann, M. *J. Proteome Res.* **2016**, *15*, 2881−2890.

(13) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. *Nat. Methods* **2009**, *6*, 359−362.

(14) Zhang, J.; Xin, L.; Shan, B.; Chen, W.; Xie, M.; Yuen, D.; Zhang, W.; Zhang, Z.; Lajoie, G. A.; Ma, B. *Mol. Cell. Proteomics* **2012**, *11*, M111.010587.
(15) Cox, J.; Mann, M. *Nat. Biotechnol.* **2008**, *26*, 1367−1372.
(16) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. *J. Proteome Res.* **2011**, *10*, 1794−1805.
(17) Vizcaíno, J. A.; Deutsch, E. W.; Wang, R.; Csordas, A.; Reisinger, F.; Ríos, D.; Dianes, J. A.; Sun, Z.; Farrah, T.; Bandeira, N.; Binz, P. A.; Xenarios, I.; Eisenacher, M.; Mayer, G.; Gatto, L.; Campos, A.; Chalkley, R. J.; Kraus, H. J.; Albar, J. P.; Martinez-Bartolomé, S.; et al. *Nat. Biotechnol.* **2014**, *32*, 223−226.
(18) Perez-Riverol, Y.; Bai, J.; Bandla, C.; García-Seisdedos, D.; Hewapathirana, S.; Kamatchinathan, S.; Kundu, D. J.; Prakash, A.; Frericks-Zipper, A.; Eisenacher, M.; Walzer, M.; Wang, S.; Brazma, A.; Vizcaíno, J. A. *Nucleic Acids Res.* **2022**, *50*, D543−D552.

https://doi.org/10.1021/acs.analchem.2c01261
Anal. Chem. 2023, 95, 1829−1837