Mass spectrometry-based degradomics analysis of toxoid vaccines

Michiels, T.J.M.

4

NOVEL FORMALDEHYDE-INDUCED MODIFICATIONS OF LYSINE RESIDUE PAIRS IN PEPTIDES AND PROTEIN: IDENTIFICATION AND RELEVANCE TO VACCINE DEVELOPMENT

Thomas J.M. Michiels [1,2], Christian Schöneich [3], Martin R.J. Hamzink [2], Hugo D. Meiring [2], Gideon F.A. Kersten [1,2], Wim Jiskoot [1], Bernard Metz [2]

[1] Division of BioTherapeutics, Leiden Academic Centre for Drug Research (LACDR), Leiden University, 2333 CC, Leiden, The Netherlands

[2] Intravacc, Institute for Translational Vaccinology, 3721 MA, Bilthoven, The Netherlands

[3] Department of Pharmaceutical Chemistry, The University of Kansas, Lawrence, KS 66047, USA

Mol Pharm **2020**, *17* (11), 4375-4385.

Abstract

Formaldehyde-inactivated toxoid vaccines have been in use for almost a century. Despite formaldehyde's deceptively simple structure, its reactions with proteins are complex. Treatment of immunogenic proteins with aqueous formaldehyde results in heterogenous mixtures due to a variety of adducts and crosslinks. In this study we aimed to further elucidate the reaction products of formaldehyde reaction with proteins, and report unique modifications in formaldehyde-treated cytochrome c and corresponding synthetic peptides. Synthetic peptides (Ac-GDVEKGAK and Ac-GDVEKGKK) were treated with isotopically labeled formaldehyde ($^{13}CH_2O$ or CD_2O), followed by purification of the two main reaction products. This allowed for their structural elucidation by (2D)-NMR and nanoscale LC-MS analysis. We observed modifications resulting from: formaldehyde-induced deamination and formation of α,β-unsaturated aldehydes, and methylation on two adjacent lysine residues; and formaldehyde-induced methylation and formylation of two adjacent lysine residues. These products react further to form intramolecular cross-links between the two lysine residues. At higher peptide concentrations, these two main reaction products were also found to subsequently cross-link to lysine residues in other peptides, forming dimers and trimers. The accurate identification and quantification of formaldehyde-induced modifications improves our knowledge of formaldehyde-inactivated vaccine products, potentially aiding the development and registration of new vaccines.

Keywords

formaldehyde; vaccines; antigens; NMR; mass spectrometry; structural elucidation; protein modification

Introduction

Formaldehyde is involved in a wide range of applications and processes. It is an important precursor in the synthesis of many chemicals, such as polymers and resins [1]. Furthermore, the chemical is a potent disinfectant and sterilant. It is either obtained as a 37 wt.% aqueous solution (known as formalin, usually stabilized with 10-15 wt.% methanol) or vaporized from paraformaldehyde. As a disinfectant it is effective against a wide variety of bacteria, fungi and viruses [2]. Besides disinfecting equipment, formaldehyde is also used in other medical applications, for instance in dentistry [3]. In histology and pathology, it is used as fixation agent [4]. In pharmaceutical applications, formaldehyde-inactivation of antigens remains an important method for chemical inactivation of pathogens (reviewed elsewhere [5]) in the production of vaccines almost a century after its discovery [6]. Marketed vaccines that are inactivated this way range from toxins (*e.g.*, diphtheria toxin, tetanus toxin) to viruses (such as the poliovirus) and bacteria (such as whole-cell *Bordetella pertussis* vaccines) [6-8]. Formaldehyde-inactivation also has potential for the development of new vaccines, such as enteroviruses (*i.e.*, enterovirus 71 [9] and coxsackieviruses [10]) or coronaviruses (*i.e.*, SARS-CoV-1 [11]). Nevertheless, some formaldehyde-inactivated vaccine concepts have failed dramatically, such as the formaldehyde-inactivated respiratory syncytial virus (RSV) vaccine, which enhanced the severity of RSV [5]. Recently, this has been linked with the usage of suboptimal concentrations of formaldehyde, resulting in misfolding of the RSV fusion protein [12, 13]. This failure underlines the importance of product characterization and a general understanding of the mechanisms involved. The use of formaldehyde itself in vaccine production is sometimes a source of criticism because of its toxicity [14]; however, formaldehyde is also an endogenous product in various metabolic processes *in vivo* and present at higher concentrations in the human body than in vaccines [15]. It has been shown that endogenous formaldehyde induces immunogenic adducts; increased immunogenicity of formaldehyde-treated proteins is also observed in some vaccine products [16, 17]. The mechanism of formaldehyde-mediated inactivation and other formaldehyde reactions with proteins and amino acids have been studied by several groups [18-26]. As the formaldehyde is mixed with a solution containing proteins, imine and hydroxymethyl adducts are formed on amines, amides and thiols. These then react further with other amino acid residues in the mixture such as tyrosine and arginine residues. This results in intermolecular (with other proteins or amino acids in the solution) and intramolecular crosslinks [19]. An overview of the most common formaldehyde-

induced modifications is depicted in Table S1. Although a lot of progress has been made in the identification of formaldehyde-induced modifications, the complexity and heterogeneity of the reaction products still hinder a complete understanding of the processes involved. Thorough vaccine product characterization and understanding is key to monitoring batch-to-batch consistency. If the various chemical modifications and the degree of these modification are known, batch release could be based on *in vitro* tests in a so-called consistency approach, instead of traditional *in vivo* release tests, measuring immunogenicity and (absence of) residual toxicity [27-30]. Moreover, better understanding of formaldehyde-inactivated vaccines can aid the development and registration of new vaccines.

In previous work, we have analyzed the influence of formaldehyde modifications on the kinetics of proteolytic digestion of various model proteins [31]. To identify formaldehyde-induced modifications in diphtheria toxoid or model proteins, such as cytochrome c, proteins were treated with aqueous solutions of CH_2O or CD_2O [20]. After incubation, the resulting mixtures were pooled in a 1:1 ratio. Subsequent protease digestion was used to obtain peptides that were analyzed by nanoscale LC-MS. Classic formaldehyde modifications yield mass spectral doublet peaks with a 2 Da mass difference or a multiple of 2 Da [18, 20] with equal intensities. The structure of these formaldehyde modifications in cytochrome c have been assigned [31]. However, several atypical spectral doublet peaks were observed that could not be addressed to these classic formaldehyde modifications.

In this study we aim to further elucidate the reaction products of formaldehyde with proteins, and report new modifications in formaldehyde-treated cytochrome c and corresponding synthetic peptides. Synthetic peptides were treated with isotopically labeled formaldehyde ($^{13}CH_2O$ or CD_2O) followed by purification of the two main reaction products. This allowed their structural elucidation by NMR and nanoscale liquid chromatography-mass spectrometry (LC-MS) analysis. These modifications involved: formaldehyde-induced deamination and formation of vinylic aldehydes, and methylation on two adjacent lysine residues; formaldehyde-induced methylation and formylation of two adjacent lysine residues. At higher peptide concentrations, these two main reaction products were found to subsequently cross-link to lysine residues in other peptides, forming dimers and trimers.

Materials & Methods

Synthetic peptides were purchased from Pepscan with >95% purity as TFA salt. In a typical reaction the peptides were dissolved in water (LC-MS grade, Biosolve) and added to a 100 mM phosphate buffer (pH 7.4, obtained as a 1 M solution from Sigma Aldrich) containing 120 mM formaldehyde (Sigma Aldrich). The reaction mixture was then placed at 40 °C, typically for two days. To stop the reaction and allow for nanoscale LC-MS analysis, 1 µL aliquots were diluted in 1 mL 0.1 vol% formic acid (Merck) or 10-µL aliquots added to 90 µL 1 vol% formic acid for conventional LC-MS. To stop the reaction and allow for purification, the pH of the mixture would be adjusted to ~2 by addition of 10 vol% TFA (Sigma Aldrich).

Reductive methylation

Solutions of synthetic Ac-GDVEKGKK and Ac-GDVEKGKKIFVQ (1 mM) were treated with 123 mM formaldehyde (CH_2O) in phosphate buffered saline (Gibco) with a total volume of 100 µL. After four days of incubation at 37 °C, an aliquot was taken for nanoscale LC-MS analysis (diluted 1:100 in 0.1 vol% FA). Subsequently, 10 µL 1.23M $NaBH_3CN$ and another 10 µL 1.23 M formaldehyde were added and the samples incubated for 30 minutes at room temperature after which another sample was taken.

Purification

Purification was achieved by reducing the total reaction volume tenfold by vacuum concentration in an Eppendorf vacuum centrifuge. The reaction products were separated by subsequent 20-µL injections on an XSelect Peptide CSH C18, 130 Å, 5 µm 4.6 mm x 250 mm semi-preparative column (Waters) in an Agilent 1290 Infinity II HPLC system with UV detection at 215 nm. Eluent A consisted of water with 0.02 vol% TFA and eluent B consisted of 80 vol% acetonitrile (LC-MS grade, Biosolve) with 20 vol% water and 0.02 vol% TFA. To separate the products a 10% to 17% eluent B gradient in 6 minutes was used, followed by 17% to 50% B in 1 minute. Fractions were collected manually and subsequently concentrated in the vacuum centrifuge to approximately 50 µL. For NMR analysis 600 µL D_2O (Sigma Aldrich) and 10 µL D_2O containing 0.75 wt% TMSP (Sigma Aldrich) were added. Products were stored at room temperature in NMR tubes at a pH <2.5 due to residual TFA.

Stability of product 2a

90 µL of purified 2a (prepared from Ac-GDVEKGKK) was taken from the NMR tube (15 µg) and mixed with 10 µL 1M phosphate buffer (pH 7.4). The pH was verified using pH indicator paper. Immediately afterwards, a 1-µL aliquot was diluted into 1000 µL 0.1 vol% formic acid to quench the reaction. The remaining mixture was incubated at 40 °C for 18 hours and sampled subsequently.

Acetaldehyde reaction

Acetaldehyde (Sigma Aldrich) was dissolved in 100 mM phosphate buffer containing 0.1 mg/mL Ac-GDVEKGAK and the mixture was placed at room temperature for 1 day. Subsequently a 1-µL aliquot was diluted into 1000 µL 0.1vol% formic acid, and 1 µL of an LC-MS sample of purified 8a (from Ac-GDVEKGAK treated with $^{13}CH_2O$) was added as a ^{13}C labeled reference for nanoscale LC-MS analysis.

Nanoscale LC-MS

The peptides were analyzed by reversed phase nanoscale LC-MS using a vented column system as described by Meiring et al [32]. An Agilent 1290 Infinity HPLC system was used in conjunction with a 100-µm inner diameter (I.D.) trapping column packed to a bed length (L) of 20mm with Reprosil-Pur C18-AQ 5 µm particles and a 31 cm, L x 50 µm I.D. analytical column packed with Reprosil-Pur C18-AQ 3 µm particles, coupled to a gold-coated nano electrospray ionization spray tip with a 3.5-µm tip diameter (all prepared in-house). 10-µL of the sample was injected onto the trapping column. The trapping column was then washed with 0.1 vol% formic acid in water at a 5 µL/min flow rate for 10 minutes. The gradient consisted of water with 0.1 vol% formic acid (eluent A) and acetonitrile with 0.1 vol% formic acid (eluent B). The gradient started at 4% B to 34% in 15 minutes followed by washing steps and re-equilibration. A flow restrictor was used to ensure a flow rate of approximately 125 nL/min through the analytical column. The LC system was coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher). Analysis was done in Data Dependent Acquisition mode, with MS1 scans at 120,000 FWHM resolution from 300-1500 m/z. Collision Induced Fragmentation (CID) MS2 scans were measured in either iontrap or orbitrap mode, where the orbitrap resolution would be decreased to 7,500 FWHM resolution for increased acquisition speed. All MS1 orbitrap m/z readouts were corrected using fluoranthene as Internal Mass Calibrant. Typical mass errors expected on MS1 were <1 ppm and <20 ppm on

MS2 (in orbitrap measurements). *Note*: for ease of reading the number of digits behind the decimal separator in some figures is lower than those still considered accurate.

Conventional LC-MS

Reaction kinetics of Ac-GDVEKGKK, Ac-GDVEKGAK, Ac-GDVEAGKK and Ac-GDVEKGKA were compared by using conventional LC-MS. To this end, we used an Agilent Poroshell 120 EC C18 column (2.1 mm I.D. x 50mm L) packed with 1.9 µm particles with an Agilent 1200SL HPLC system coupled to an LTQ Orbitrap XL mass spectrometer (Thermo Fisher) with electrospray ionization. The same eluent system as with nanoscale LC was used with a 1% to 30% B in a 4-minute gradient. 5-µL of the sample at a concentration of 0.01 mg/mL peptide was injected. Measurements were performed in the ion trap at normal scan rate, from 300 to 900 m/z.

NMR

NMR experiments were carried out on a JEOL JNM-ECZ400S/L1 400 MHz NMR in D_2O with trimethylsilylpropanoic acid as internal reference. Default experimental settings as provided by JEOL were used. ^{1}H-NMR spectra were recorded using Robust5 water suppression (1064 scans). Standard ^{13}C spectra, DEPT-135 and DEPT-90 spectra were recorded until signal/noise was sufficient (typically, >30,000 scans). For HSQC and HMBC the HSQC_wet and HMBC_wet experiments were used to suppress the water signal.

Results

Detection of unusual formaldehyde modifications in cytochrome c

In a previous study, chemical modifications were identified in cytochrome c after formaldehyde treatment [31]. However, some unexpected modifications were identified by the PEAKS Studio PTM module, which could not be explained. These observations triggered further investigation. In the previous study, proteins were treated with CH_2O and CD_2O separately and after incubation, the samples were mixed 1:1 and digested with the protease cathepsin S. Instead of the usual 1:1 ratio of the light and heavy peptide, the isotope pattern was skewed towards the light peptide (Figure S1). This observation triggered further investigation. Moreover, the mass increase after formaldehyde treatment was +42.0106 Da, which commonly corresponds to acetylation ($+C_2H_2O$, on top of the standard acetylation

of the cytochrome c *N*-terminus), but acetylation by formaldehyde is not obvious. As the mass increase between the light peptide and the heavy peptide was 4 Da, 2 formaldehyde molecules (each containing 2 deuterons) must have been incorporated in the modification. Peptides derived from formaldehyde-treated cytochrome c that included this mass increase (+42.0106 Da) were Ac-GDVEKGKK, Ac-GDVEKGKKIFVQ, KGKKHKTGPNL and AYLKK. All these peptides contain at least 2 lysine residues, making the involvement of two lysine residues in one modification or crosslink very likely.

Formaldehyde modifications in synthetic Ac-GDVEKGKK

To study this modification in detail, a synthetic peptide with the sequence Ac-GDVEKGKK was subjected to 120 mM formaldehyde (CH_2O or CD_2O) in 100 mM phosphate buffer (pH 7.4). Treatment of the native peptide resulted in the same modification with ΔM = +42.0106 Da, as observed for cytochrome c, along with a variety of other modifications. The most important of these other modifications were ΔM = +24.9840 Da (+[C_2O]-[NH]) and ΔM = +73.0051 Da (+[$C_3H_3O_3$]-[N]), where the latter is suggested to be an analogue where an additional molecule has reacted with the former. A representative chromatogram is shown in Figure 1. Because of the complicated nature and the many reaction products formed, we decided to focus on the most important modifications, which form the basis for the other reaction products: ΔM = +24.9840 Da (**1a-b**) and ΔM = +42.0106 Da (**2a-b**). The proposed structures for these modifications are depicted in Scheme 1. Experimental evidence for these structures is described below.

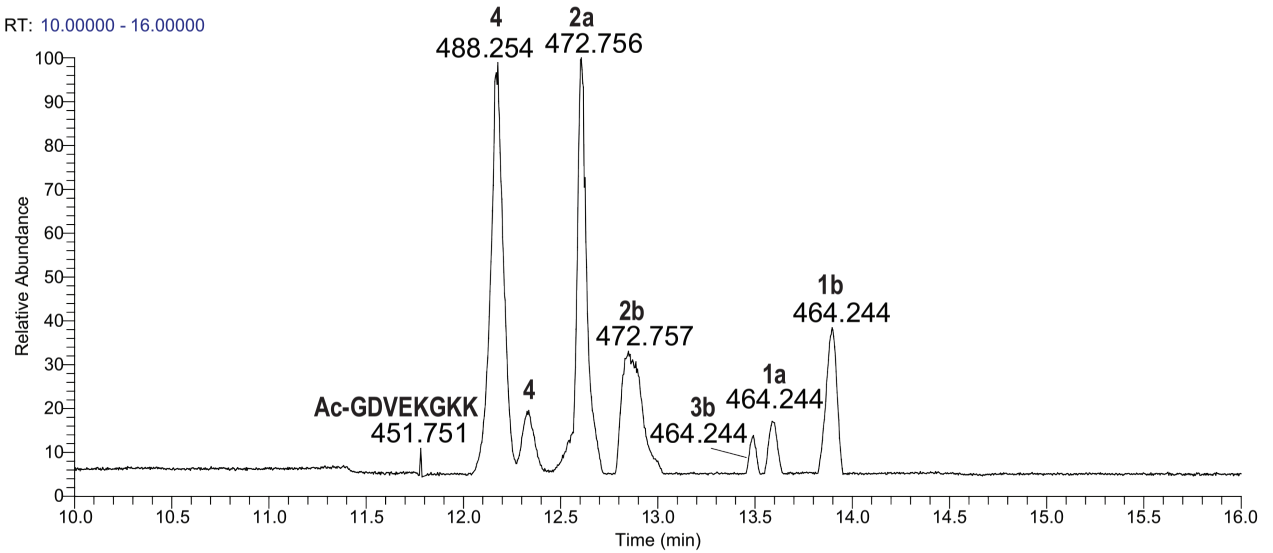
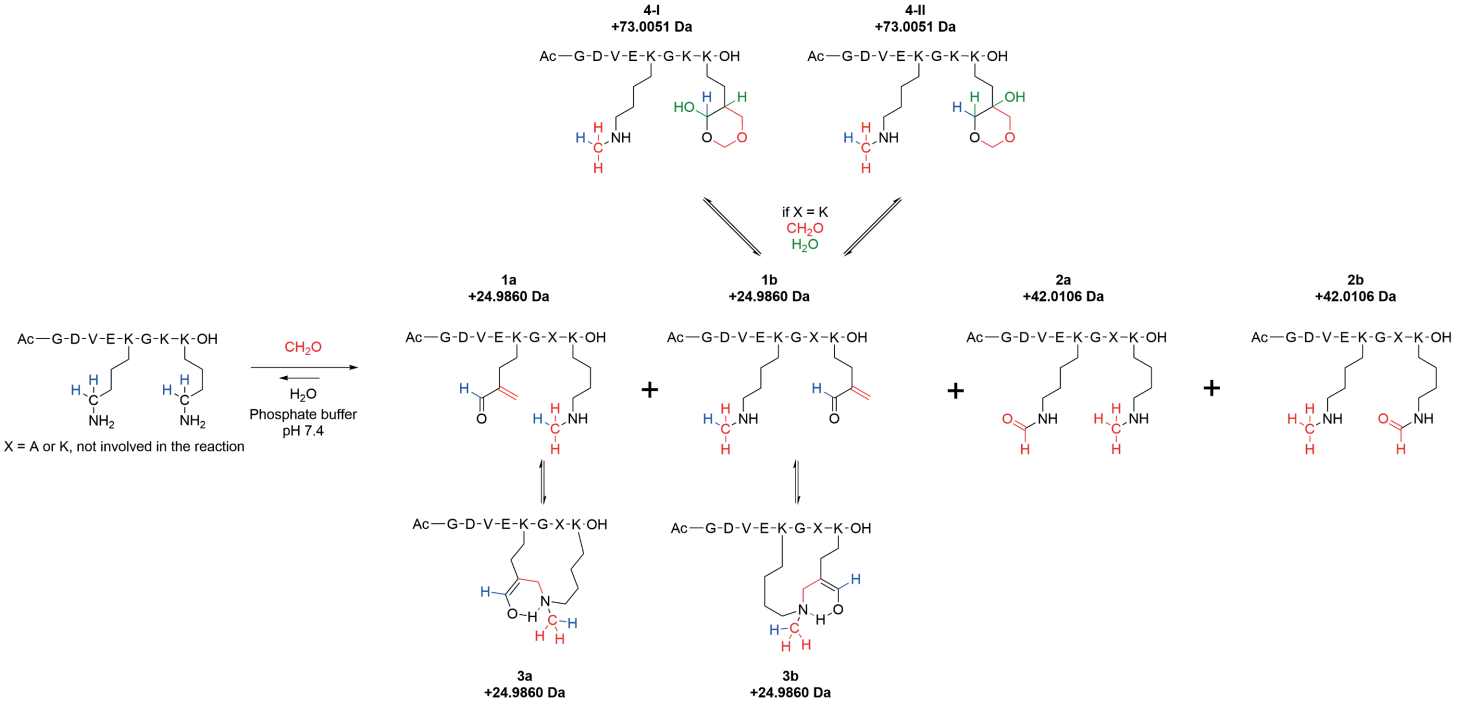


Figure 1. Representative nanoscale LC-MS base peak (m/z 400-500) chromatogram of synthetic Ac-GDVEKGKK treated with aqueous CH_2O for 46 hours (z=2 for all peaks). The exact regioisomers of the two peaks corresponding to product **4** could not be assigned.



Scheme 1. Ac-GDVEKGXK (X=K or A) reacts with formaldehyde in aqueous buffer to form **2a**-**b** (**a-b** represent two regioisomers), **1a-b** and tautomers **3a-b**. If X=K **4-I** and **4-II** are formed as well. Reaction conditions: 100 mM formaldehyde, 100 mM phosphate buffer (pH 7.4) in water at 40 °C for 24 hours.

Reaction kinetics

To investigate which lysine residues were involved in these formaldehyde modifications (+24.9860 and +42.0106 Da), peptides with the general sequence Ac-GDVEXGXX were made, where X = K or A in all possible permutations. Treatment of the peptides that contained only 1 lysine did not result in the formation of significant additional peaks, with the exception of a small amount of the classic hydroxymethyl and imine adducts (data not shown). The other peptides showed different degrees of modification, where the extent of modification decreased in the following order: Ac-GDVEKGKK ≡ Ac-GDVEKGAK > Ac-GDVEKGKA >> Ac-GDVEAGKK (Figure 2). The formation of **2** was diminished by the use of deuterated formaldehyde (72% reduction for Ac-GDVEKGKK), while the effect of deuterated formaldehyde on the formation of **1** was less substantial. Ac-GDVEKGKK and Ac-GDVEKGAK form similar amounts of **2**, with similar kinetics, but **1** was formed faster from Ac-GDVEKGAK. The use of D_2O as a solvent instead of H_2O did not affect the structure of the reaction products, but slightly decreased the speed of the reaction (Figure S2).

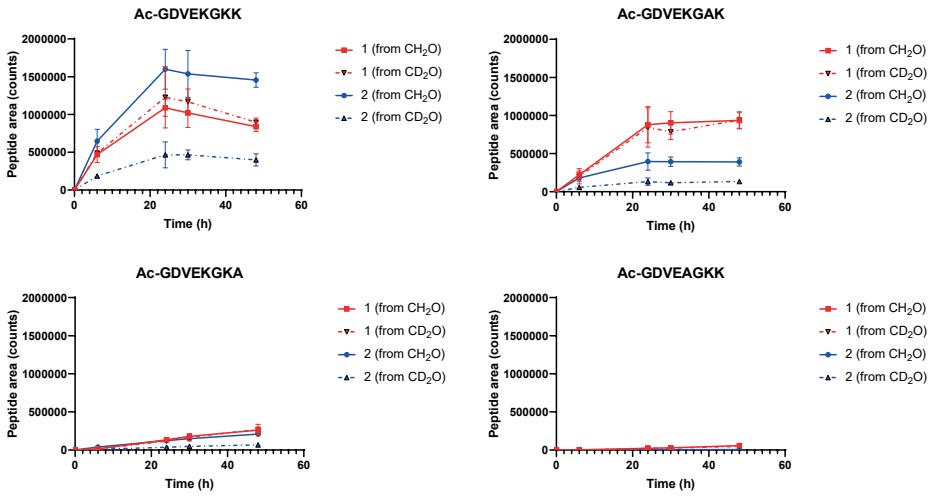


Figure 2. Comparison of the reaction between synthetic analogues of peptide Ac-GDVEKGKK (where one of the various lysine residues is replaced with an alanine residue) and the use of CH_2O or CD_2O on the formation of products **1** and **2**. The area of the regioisomers has been summed.

Reductive methylation of the reaction products

To elucidate the structures of reaction products **1** and **2**, the reaction mixture of Ac-GDVEKGKK with formaldehyde was subjected to reductive methylation by the addition of more formaldehyde and $NaCNBH_3$. Unreacted Ac-GDVEKGKK was methylated 6 times, twice on each lysine residue. Although a variety of reaction products were formed, some of the main peaks observed were those corresponding to products **1** and **2** with three additional methyl groups. Thus, the original formaldehyde-induced adducts on **1** and **2** were both stable under mild reducing conditions, and three positions on the amine groups could still form imines and the corresponding methyl groups (Scheme S1). These results are consistent with the proposed structures **1** and **2**.

Purification and stability of products 1 and 2

To enable NMR analysis of products **1** and **2**, the reaction mixture of $^{13}CH_2O$ and Ac-GDVEKGAK was purified by reversed phase, preparative reversed-phase HPLC. Ac-GDVEKGAK instead of Ac-GDVEKGKK (found in cytochrome c) was used as the reactions were cleaner, which simplifies the purification. The products were reasonably stable in aqueous solution at low pH (<2.5) but not stable when drying, and thus no dry products **1** and **2** could be obtained. Instead of completely removing the solvent, the solutions were concentrated with almost complete removal of the acetonitrile. The pH was typically <3 due to residual trifluoroacetic acid (TFA). Typical yields (NMR) were around 10%.

Despite product **2a** being baseline separated from the chromatographically succeeding products **2b**, **1a** and **1b**, the concentrated solution always contained some **1a** (Figure S3 A). Presumably, **1a** was formed during concentration or other steps of the workup, after collection of **2a**. Likewise, the purification of **1a** always resulted in some **3a** being present, despite this chromatographic peak eluting at the same time as the removed product **2a** (Figure S3 B). The presence of these impurities indicates that **1a** can be formed from **2a** and that **3a** can be formed from **1a** without the presence of additional formaldehyde.

To further investigate the stability of the reaction products, purified **2a** (from Ac-GDVEKGKK) was placed in phosphate buffer at pH 7.4 and stored at 37 °C for 1 day. Despite **2a** being stable for weeks at low pH, exposure to pH 7.4 resulted in the formation of **1a** and **1b,** although **1a** was the more abundant degradation product (Figure S4).

NMR of purified 1b and 2a

By using $^{13}CH_2O$, sufficient material could be obtained to perform ^{13}C NMR experiments on two purified fractions containing **1b** or **2a** derived from Ac-GDVEKGAK. However, only carbons originating from the formaldehyde modifications could be detected this way. To provide further structural information, ^{1}H (Robust-5 water suppression), DEPT and HSQC were combined with HMBC (Figure 3 and Figure 4, described in detail in supporting information) and confirmed the structures of **1b** and **2a** respectively. Despite purification, nanoscaleLC-MS showed more than one peak, although the major product was **1b.** Indeed, in the NMR analysis, two ^{13}C signals were observed other than the peaks assigned to **1b**. These signals correspond to structure **3** (a tautomer of **1**), and free formaldehyde.

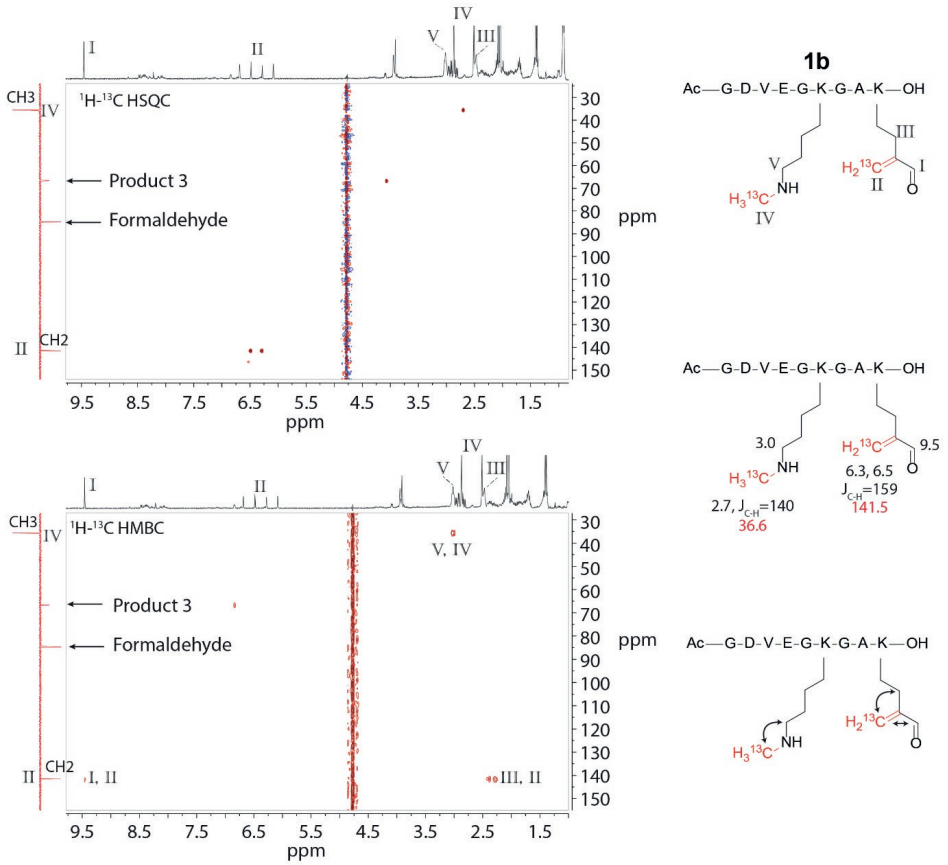


Figure 3. ^{1}H-^{13}C HSQC and ^{1}H-^{13}C HMBC of purified **1b** obtained from a reaction with Ac-GDVEKGAK with ^{13}C labeled formaldehyde (adducts containing ^{13}C are marked in red, ^{13}C NMR shifts are also marked in red). F1: DEPT-135 and F2: Robust5-^{1}H (for water suppression) spectra.

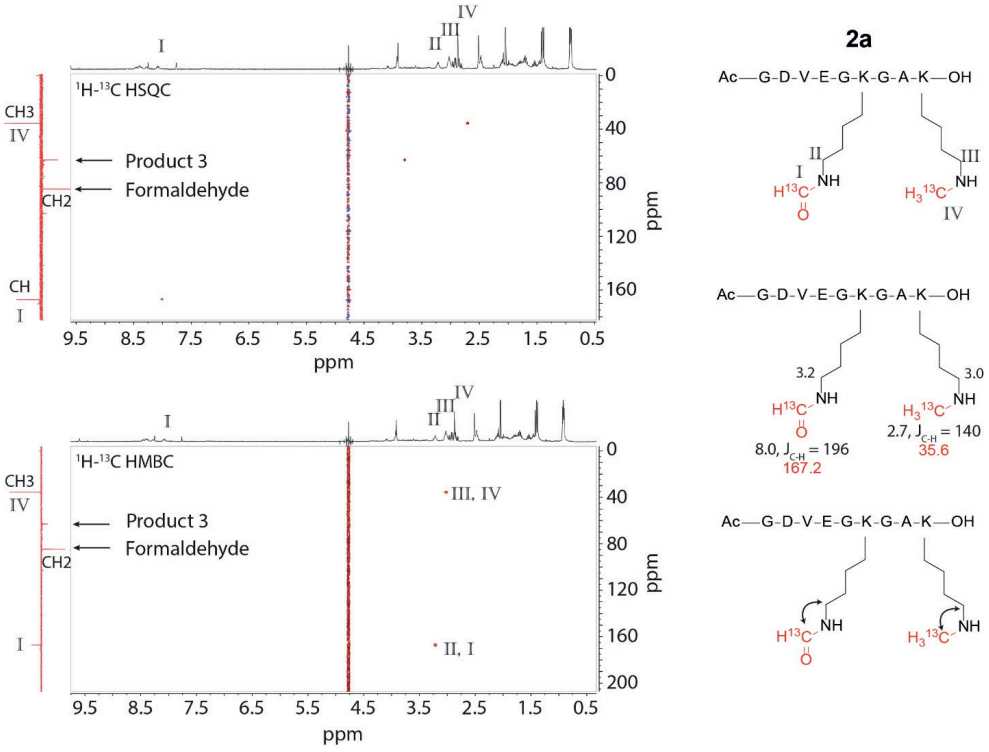


Figure 4. ^{1}H-^{13}C HSQC and ^{1}H-^{13}C HMBC of purified **2a** obtained from a reaction with Ac-GDVEKGAK with ^{13}C labeled formaldehyde (adducts containing ^{13}C are marked in red, ^{13}C NMR shifts are also marked in red). F1: DEPT-135 and F2: Robust5-^{1}H (for water suppression) spectra.

MS2 analysis of products 1, 2 and 3

Subjecting Ac-GDVEKGKK, exposed to various isotopically enriched formaldehyde solutions (CH_2O, $^{13}CH_2O$ and CD_2O), to MS2 analysis supported the NMR findings. Structures **1** and **2** were both found to exist as regioisomers, with mainly lysine-5 and lysine-8 being modified. Either lysine-8 contained the methyl group (structures **1a** and **2a**), with the other modification on lysine-5, or *vice versa* (structures **1b** and **2b**) (Figure S5).

Fragmentation of **2a** revealed that the y1 ion contains an additional +14.02 Da, consistent with methylation of the C-terminal lysine (Figure 5 A). ^{13}C-labeled formaldehyde resulted in the y1, y2 and y3 ions to be 1.00 Da heavier than the corresponding y ions resulting from

regular formaldehyde treatment (Figure 5 B). The ^{13}C-labeled formaldehyde resulted in a y4 ion with ΔM = +2.01 Da, which is in compliance with the proposed structure and the incorporation of two formaldehyde molecules on the peptide. Fragmentation of CD_2O-treated Ac-GDVEKGKK resulted in the y1 ion being 3.02 Da heavier than that obtained after CH_2O treatment, indicating that a deuteride was transferred from the modification on lysine-5 to the C-terminal lysine. Indeed, the mass difference between the y3 and y4 ions after CD_2O treatment was 1.01 Da (ΔM = K+29.00) compared to the y3 and y4 ions obtained after CH_2O treatment (ΔM = K+27.99).

Fragmentation of CH_2O-treated **1a** revealed methylation of lysine-8 and a ΔM = +10.97 Da (-[NH_3]+[CO]) on lysine-5 (Figure 6 A). Contrary to **2a**, CD_2O-treated **1a** showed a methylation of the C-terminal lysine with just 2 deuterium atoms (Figure 6 C). The y4 ions of **1a** show that the C-terminal lysine contains a modification which is 2 Da heavier when using CD_2O, corresponding to incorporation of both deuterium atoms on this residue.

MS2 analysis of the chromatographic peaks corresponding to structure **3b** showed specific fragmentation resulting in only the y4-H_2O ion (data not shown). The lack of backbone fragmentation between lysine-5 and lysine-8 supports a cyclic structure. Ionization of **3a** already resulted in significant fragmentation and detection of the b4 ion in the MS1.

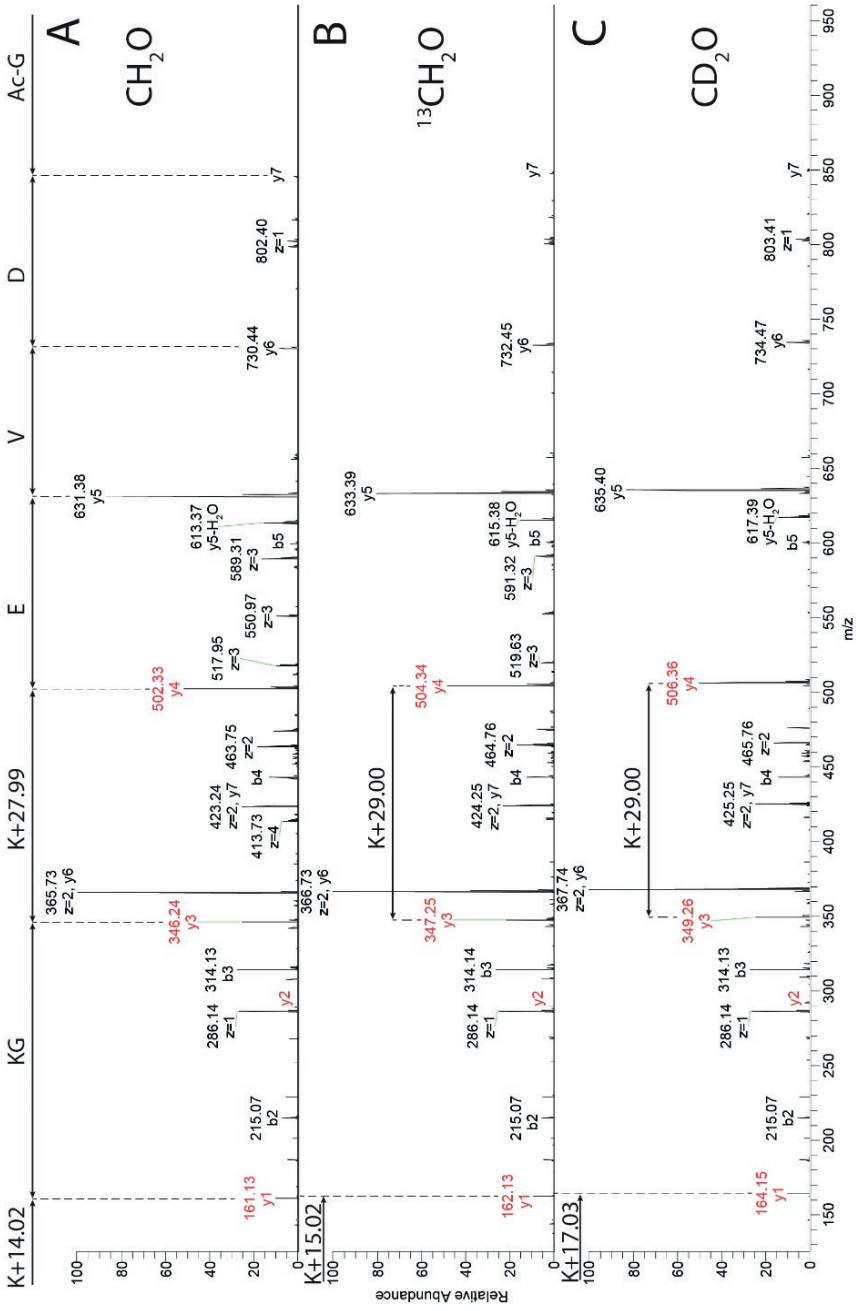


Figure 5. High resolution MS2 of **2a**. MS2 spectrum of Ac-GDVEKGKK treated with A) CH_2O, B) $^{13}CH_2O$, C) CD_2O.

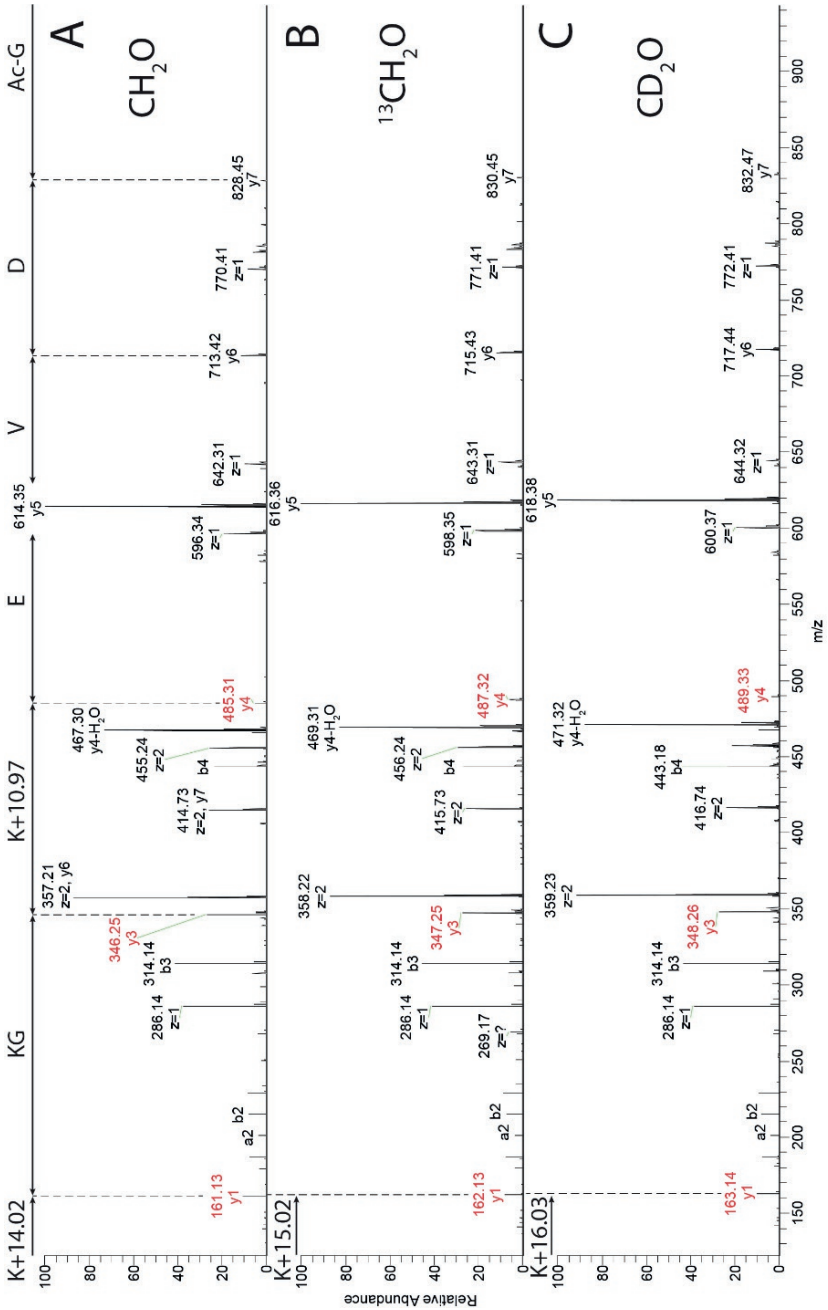


Figure 6. High resolution MS2 of **1a**. MS2 spectrum of Ac-GDVEKGKK treated with A) CH_2O, B) $^{13}CH_2O$, C) CD_2O.

MS2 of product 4

Product **4** was only observed for the peptide containing three lysine residues. Collision induced dissociation fragmentation (CID) (ion trap detection) showed identical ions for the two isobaric chromatographic peaks but at different relative intensities (Figure S6). This would fit having both the Markovnikov and anti-Markovnikov product present, but with different intensities. From this data the two isomers could not be assigned to a specific chromatographic peak.

Dimerization and oligomerization

In order to obtain as much purified reaction product as possible, the reaction of formaldehyde was performed with various concentrations of synthetic Ac-GDVEKGAK. At the lowest peptide concentration (0.25 mg/mL), nanoscale LC-MS analysis revealed a few reaction products, mainly **1** and **2** (Figure S7). When the peptide concentrations were increased to as high as 5.75 mg/mL, reaction products other than **1** and **2** started to form. Some of the most intense peaks may belong to cross-linked structures, such as the trimer **7** (Figure S7) and the dimers **8** and **9**. These proposed structures are based on their total mass (MS1 mass error < 1 ppm). Both **1** and **2** seem to be involved in oligomerization, where **1** can form multiple reaction products, such as hydrogen bond stabilized enolates (**7**) or hetero-Diels-Alder products (**8**). A variety of reaction products with different numbers of methyl groups were observed.

Reactions with acetaldehyde

The formation of isobaric reaction products under the influence of acetaldehyde is described in the supporting information.

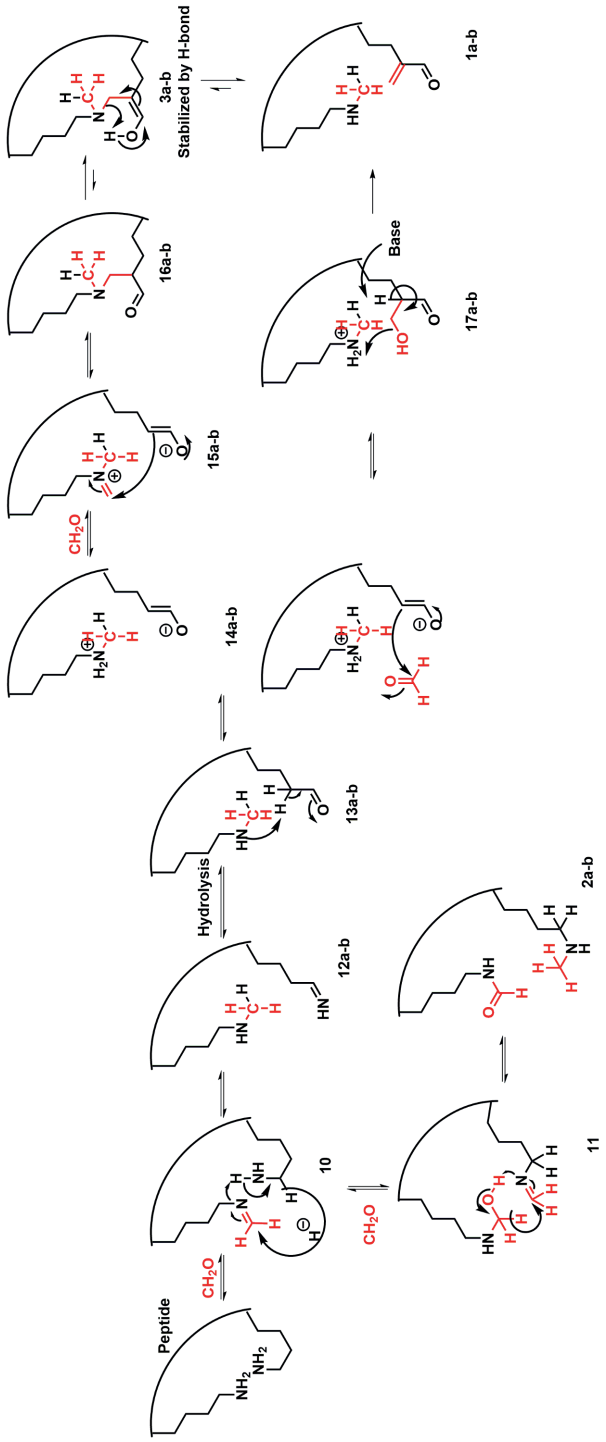
Discussion

Two new groups of formaldehyde-induced modifications on peptide sequences containing (at least) two lysine residues were identified. The first modification involved methylation and formylation of the lysine pair, resulting in ΔM = +42.01 Da (product **2**). In the second modification, one lysine of the lysine pair was converted to a α,β-unsaturated aldehyde while the other lysine residue was methylated (product **1**). Structure **1** could also form a tautomer where both lysine residues are crosslinked (product **3**). While lysine-lysine cross-links in the form of aminals have been suggested [5], to our knowledge no NMR evidence for formaldehyde-induced lysine-lysine cross-links have been reported yet.

Formaldehyde modification (product **2**) (ΔM = +42.01 Da) was not found in our previous studies [20], where we used proteins treated with either CH_2O or CD_2O in a 1:1 mixture. Hydroxymethyl adducts exchange their oxygen atom with the oxygen in water through the equilibrium with the corresponding imine. However, deuterons of the common formaldehyde-induced protein modifications were not expected to be exchanged [19]. This assumption implies that formaldehyde-induced modifications are always present at an almost perfect 1:1 ratio. In our current study, we identified a new modification where slower reaction kinetics with deuterated formaldehyde resulted in a skewed isotope pattern. These reaction kinetics point to the transfer of a hydride (or deuteride) from one part of the molecule to another. This transfer would be the rate-limiting step in the formation of **2**. Our suggested mechanism for the formation of **2** (Scheme 2) involves the formation of an imine on the ε-NH_2 group of one of two lysine residues and the corresponding hydroxymethyl on the other lysine residue. We propose that the nitrogen of the imine acts as a base, deprotonating the hydroxymethyl OH; and subsequent transfer of the hydroxymethyl α-hydride to the adjacent imine form product **2**. This last step would explain the observed kinetic isotope effect, which is further supported by MS2 analysis of the deuterated **2a**, showing a methylated y1 ion 3 Da heavier than its non-deuterated counterpart. NMR evidence for this structure is strong with ^{1}H, DEPT-90, DEPT-135 and ^{1}H-^{13}C HMBC supporting this structure. Formylation in combination with methylation by formaldehyde has been observed before in solutions of the amino acid ornithine, but not on lysine [21], indicating that the orientation of the amino groups is pivotal for this reaction. Indeed, the spacing between the lysine residues was important for this reaction to occur; the fastest reaction rates were observed when two spacing amino acids were located between the two lysine residues.

The modification with ΔM = +24.98 Da (**1**) was less affected by the use of deuterated formaldehyde. Our suggested mechanism for the formation of **1** (Scheme 2) again starts with formation of an imine on one of the lysine residues. However, instead of involvement of another formaldehyde molecule, the side chain of the nearby lysine residue is oxidized to an imine. The hydrogen of the lysine side chain is transferred to the imine adduct originating from formaldehyde to form intermediate **12**. Subsequent hydrolysis of imine **12** results in aldehyde **13**. Similar to the formation of **2**, this results in methylation of one lysine residue; however, when deuterated formaldehyde was used, not 3 but 2 deuterons were present on the methyl group, the other hydrogen originates from the lysine side

chain. Intermediate **13** has been observed by LC-MS. Tautomerization of **13** leads to enolate **14**, which can then perform a nucleophilic attack on a formaldehyde molecule to give **17**. Subsequent β-elimination gives product **1**. Product **1** could tautomerize to form enol **3**. This enol could be stabilized by a hydrogen bond between the lysine nitrogen and the enol proton. Hydrogen bond-stabilized enols are well described [33]. The limited fragmentation of **3** during MS2 analysis further supports a cyclic structure, with a bridge between the two lysine residues which hinders complete fragmentation of the y4 ion. The high amount of reactive intermediates would suggest that multiple reaction products could be formed through various pathways. Indeed, the NMR spectra of purified reaction products always contained some impurities, likely formed due to side reactions after purification. The main impurity in purified **2a** is **3**, a tautomer of **1**, which happens to co-elute with **2a**. Furthermore, formaldehyde is present in the purified samples. While it is possible that this is due to incomplete removal of formaldehyde through the preparative-HPLC, it is more likely that the formaldehyde is released from the reaction products through equilibria, as the formaldehyde is present in a similar amount. Storage at low pH (<3) seemed to keep both products **1** and **2** stable for weeks, probably because protonation of the amines decreases nucleophilicity and prevents them from acting as a Brønsted base which could catalyze further reactions. Increasing the pH (to 7.4) resulted in relatively fast (days) conversion of **2** to **1**. Due to the reactivity of α,β-unsaturated aldehydes we expect these reaction products to react further when exposed to suitable nucleophiles in biological matrices. Indeed, the modification with ΔM = +24.98 Da (**1**) was observed only in reactions with synthetic peptides, presumably because either the modification is not formed in cytochrome c due to folding or because of the modified lysine residues reacting further with other amino acids to form other reaction products. As a large number of low-intensity side products were observed by nanoscale LC-MS after formaldehyde treatment of a simple purified peptide, high reactivity of the reaction products combined with a more complicated protein would be a likely explanation for the difference we observed between peptide reactivity and the presence of these modifications in cytochrome c.



Scheme 2. Proposed mechanism for the formation of **1** and **2**.

Product **4** is likely formed from **1b** (Ac-GDVEKGKK), as the modification of lysine-8 involves loss of a nitrogen, similar to product **1b**. Presence of the b7 ion supports a methylation on either lysine-5 or lysine-7 and not lysine-8, while y1 related ions confirm modification of lysine-8 by ΔM = +58.9895 Da. Overall, product **4** is thought to be one of many reaction products that are eventually formed from **1** (or **2**). Our suggested mechanism involves the formation of an hydroxymethylated **1b** as an intermediate (Scheme S2). Purification and structural elucidation of all reaction products present in this peptide-formaldehyde reaction mixture is beyond the scope of the present study; thus, no further purification and subsequent NMR analysis of **4** or other side products was attempted.

Several dimers and trimers were also observed. The most probable structures are suggested based on their accurate mass (ΔM <1 ppm), as MS2 fragmentation was not informative and purification and subsequent NMR analysis was deemed unfeasible. Because of the relatively simple reaction components being the peptide and formaldehyde, the possible combinations that lead to the same exact mass are limited. The bulk of the mass has to be made up by a number of peptides, where the residual mass is likely a combination of the main reaction products **1** and **2** along with other formaldehyde adducts. Trimer **7** is most likely a ring of three peptides, one peptide is one of the regioisomers of **1**, the other two are unmodified peptides. Although a linear structure with one aminal and a hetero-Diels-Alder product similar to dimer **8** is possible, a ring would be more thermodynamically favorable. Hydrogen bonding between the enol hydrogen and the amine could further stabilize this product. Aminals are not very stable in water, especially at lower pH, but in a ring system held together in a hetero-Diels-Alder product, this reaction could reverse, making the cyclic structure more favorable than a linear one. It is interesting that we managed to observe these aminals, as prior to analysis, aliquots of the reaction mixture were diluted in 0.1 vol% formic acid, which would usually result in rapid hydrolysis of the aminal [34]. Two species of dimers were observed, one consisting of reaction product of **2** with an unmodified peptide (**9**), and one consisting of a reaction product of **1** with an unmodified peptide (**8**). To get to the mass of product **9,** two extra formaldehyde molecules and three extra ring double bond equivalents (RDBe) are required. Thus, an aminal formed from the methylated lysine and lysine with an imine adduct, and a new ring involving the formylated lysine and a lysine with an imine adduct is the most logical explanation. Product **8** most likely consists of the product of a hetero-Diels-Alder reaction between an imine adduct on a lysine

and the vinyl aldehyde of product **1**, and an aminal connecting the other lysine residues. The imine corresponding to the aminal of **8** would have the same total mass, but it would be entropically favorable that the aminal forms first, so that the intramolecular Diels-Alder can happen afterwards as these reactions require decent activation energy [35, 36]. In reality, the reaction mixture probably contains a mixture of the cyclic reaction product and the imine, which are in equilibrium.

Conclusion

Overall, we identified two new groups of formaldehyde-induced modifications of lysine residue pairs in peptides. Both the formylation/methylation modification pair and the α,β-unsaturated aldehyde/methylation modification pair were involved in further crosslinking reactions, which should be addressed in future studies. These reaction products should be taken into consideration when analyzing biological samples exposed to exogenous or endogenous formaldehyde. Especially biologicals, such as toxoid vaccines, are notoriously heterogeneous because of their formaldehyde modifications [20, 37]. These newly discovered modifications could help us to further understand the nature of these vaccine products. As formaldehyde-induced modifications are linked to alterations in immunogenicity [16, 17], their accurate identification and quantification may improve our understanding of formaldehyde-inactivated vaccine products, potentially aiding the development and registration of future formaldehyde-inactivated vaccines.

Supporting Information (available online)

Contains Table S1, Figure S1-S9, Scheme S1-S3, a detailed description of NMR spectra, and a description of similar reactions with acetaldehyde instead of formaldehyde.

Acknowledgements

The authors thank Joost Uittenbogaard and Maarten Danial for their advice, Marjolein Zohlandt for critical review of the manuscript, and Adolfo Botana of JEOL for assistance with the NMR analysis. Funding: This work was supported, in part, by the Ministry of Agriculture, Nature and Food Quality, the Netherlands.

References

1. Christjanson, P.; Pehk, T.; Siimer, K. Hydroxymethylation and polycondensation reactions in urea–formaldehyde resin synthesis. *Journal of Applied Polymer Science* **2006,** *100*, (2), 1673-1680.
2. Rutala, W.; Weber, D. HICPAC Guideline for Disinfection and Sterilization in Healthcare Facilities. https://www.cdc.gov/infectioncontrol/guidelines/disinfection/ (31-Jul-2020).
3. Chandrashekhar, S.; Shashidhar, J. Formocresol, still a controversial material for pulpotomy: A critical literature review. *Journal of Restorative Dentistry* **2014,** *2*, (3), 114.
4. Buesa, R. J. Histology without formalin? *Annals of Diagnostic Pathology* **2008,** *12*, (6), 387-396.
5. Delrue, I.; Verzele, D.; Madder, A.; Nauwynck, H. J. Inactivated virus vaccines from chemistry to prophylaxis: merits, risks and challenges. *Expert Rev Vaccines* **2012,** *11*, (6), 695-719.
6. Glenny, A. T.; Hopkins, B. E. Diphtheria Toxoid as an Immunising Agent. *Br J Exp Pathol* **1923,** *4*, (5), 283-288.
7. Gupta, R. K.; Sharma, S. B.; Ahuja, S.; Saxena, S. N. The effects of different inactivating agents on the potency, toxicity and stability of pertussis vaccine. *J Biol Stand* **1987,** *15*, (1), 87-98.
8. WHO Production and Control of Tetanus Vaccine: Module III: Principles of Tetanus Vaccine Production. (WHO/VSQ/94.4). (26th of March).
9. Reed, Z.; Cardosa, M. J. Status of research and development of vaccines for enterovirus 71. *Vaccine* **2016,** *34*, (26), 2967-2970.
10. Stone, V. M.; Hankaniemi, M. M.; Laitinen, O. H.; Sioofy-Khojine, A. B.; Lin, A.; Diaz Lozano, I. M.; Mazur, M. A.; Marjomaki, V.; Lore, K.; Hyoty, H.; Hytonen, V. P.; Flodstrom-Tullberg, M. A hexavalent Coxsackievirus B vaccine is highly immunogenic and has a strong protective capacity in mice and nonhuman primates. *Sci Adv* **2020,** *6*, (19), eaaz2433.
11. Darnell, M. E.; Plant, E. P.; Watanabe, H.; Byrum, R.; St Claire, M.; Ward, J. M.; Taylor, D. R. Severe acute respiratory syndrome coronavirus infection in vaccinated ferrets. *J Infect Dis* **2007,** *196*, (9), 1329-38.
12. Killikelly, A. M.; Kanekiyo, M.; Graham, B. S. Pre-fusion F is absent on the surface of formalin-inactivated respiratory syncytial virus. *Sci Rep* **2016,** *6*, 34108.
13. Zhang, W.; Zhang, L. J.; Zhan, L. T.; Zhao, M.; Wu, G. H.; Si, J. Y.; Chen, L.; Lin, X.; Sun, Y. P.; Lin, M.; Yu, C.; Fang, M. J.; Wang, Y. B.; Zheng, Z. Z.; Xia, N. S. The Optimal Concentration of Formaldehyde is Key to Stabilizing the Pre-Fusion Conformation of Respiratory Syncytial Virus Fusion Protein. *Viruses* **2019,** *11*, (7).
14. Kata, A. A postmodern Pandora's box: Anti-vaccination misinformation on the Internet. *Vaccine* **2010,** *28*, (7), 1709-1716.
15. Mitkus, R. J.; Hess, M. A.; Schwartz, S. L. Pharmacokinetic modeling as an approach to assessing the safety of residual formaldehyde in infant vaccines. *Vaccine* **2013,** *31*, (25), 2738-43.
16. Nakamura, J.; Shimomoto, T.; Collins, L. B.; Holley, D. W.; Zhang, Z.; Barbee, J. M.; Sharma, V.; Tian, X.; Kondo, T.; Uchida, K.; Yi, X.; Perkins, D. O.; Willis, M. S.; Gold, A.; Bultman, S. J. Evidence that endogenous formaldehyde produces immunogenic and atherogenic adduct epitopes. *Sci Rep* **2017,** *7*, (1), 10787.
17. Metz, B.; Jiskoot, W.; Hennink, W. E.; Crommelin, D. J.; Kersten, G. F. Physicochemical and immunochemical techniques predict the quality of diphtheria toxoid vaccines. *Vaccine* **2003,** *22*, (2), 156-67.
18. Metz, B.; Kersten, G. F.; Baart, G. J.; de Jong, A.; Meiring, H.; ten Hove, J.; van Steenbergen, M. J.; Hennink, W. E.; Crommelin, D. J.; Jiskoot, W. Identification of formaldehyde-induced modifications in proteins: reactions with insulin. *Bioconjug Chem* **2006,** *17*, (3), 815-22.

19. Metz, B.; Kersten, G. F.; Hoogerhout, P.; Brugghe, H. F.; Timmermans, H. A.; de Jong, A.; Meiring, H.; ten Hove, J.; Hennink, W. E.; Crommelin, D. J.; Jiskoot, W. Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. *J Biol Chem* **2004,** *279*, (8), 6235-43.

20. Metz, B.; Michiels, T.; Uittenbogaard, J.; Danial, M.; Tilstra, W.; Meiring, H. D.; Hennink, W. E.; Crommelin, D. J. A.; Kersten, G. F. A.; Jiskoot, W. Identification of formaldehyde-induced modifications in diphtheria toxin. *J Pharm Sci* **2019**.

21. Kamps, J. J. A. G.; Hopkinson, R. J.; Schofield, C. J.; Claridge, T. D. W. How formaldehyde reacts with amino acids. *Communications Chemistry* **2019,** *2*, (1), 126.

22. Trezl, L.; Rusznak, I.; Tyihak, E.; Szarvas, T.; Szende, B. Spontaneous N epsilon-methylation and N epsilon-formylation reactions between L-lysine and formaldehyde inhibited by L-ascorbic acid. *Biochem J* **1983,** *214*, (2), 289-92.

23. Yamada, M.; Funaki, S.; Miki, S. Formaldehyde interacts with RNA rather than DNA: Accumulation of formaldehyde by the RNA-inorganic hybrid material. *Int J Biol Macromol* **2019,** *122*, 168-173.

24. Gold, T. B.; Smith, S. L.; Digenis, G. A. Studies on the influence of pH and pancreatin on 13C-formaldehyde-induced gelatin cross-links using nuclear magnetic resonance. *Pharm Dev Technol* **1996,** *1*, (1), 21-6.

25. Toews, J.; Rogalski, J. C.; Clark, T. J.; Kast, J. Mass spectrometric identification of formaldehyde-induced peptide modifications under *in vivo* protein cross-linking conditions. *Anal Chim Acta* **2008,** *618*, (2), 168-83.

26. Toews, J.; Rogalski, J. C.; Kast, J. Accessibility governs the relative reactivity of basic residues in formaldehyde-induced protein modifications. *Anal Chim Acta* **2010,** *676*, (1-2), 60-7.

27. Halder, M.; Depraetere, H.; Delannois, F.; Akkermans, A.; Behr-Gross, M. E.; Bruysters, M.; Dierick, J. F.; Jungback, C.; Kross, I.; Metz, B.; Pennings, J.; Rigsby, P.; Riou, P.; Balks, E.; Dobly, A.; Leroy, O.; Stirling, C. Recommendations of the VAC2VAC workshop on the design of multi-centre validation studies. *Biologicals* **2018,** *52*, 78-82.

28. De Mattia, F.; Chapsal, J. M.; Descamps, J.; Halder, M.; Jarrett, N.; Kross, I.; Mortiaux, F.; Ponsar, C.; Redhead, K.; McKelvie, J.; Hendriksen, C. The consistency approach for quality control of vaccines- a strategy to improve quality control and implement 3Rs. *Biologicals* **2011,** *39*, (1), 59-65.

29. De Mattia, F.; Hendriksen, C.; Buchheit, K. H.; Chapsal, J. M.; Halder, M.; Lambrigts, D.; Redhead, K.; Rommel, E.; Scharton-Kersten, T.; Sesardic, T.; Viviani, L.; Ragan, I. The vaccines consistency approach project: an EPAA initiative. *Pharmeur Bio Sci Notes* **2015,** *2015*, 30-56.

30. Metz, B.; Jiskoot, W.; Mekkes, D.; Kingma, R.; Hennink, W. E.; Crommelin, D. J.; Kersten, G. F. Quality control of routine, experimental and real-time aged diphtheria toxoids by *in vitro* analytical techniques. *Vaccine* **2007,** *25*, (39-40), 6863-71.

31. Michiels, T. J. M.; Meiring, H. D.; Jiskoot, W.; Kersten, G. F. A.; Metz, B. Formaldehyde treatment of proteins enhances proteolytic degradation by the endo-lysosomal protease cathepsin S. *Scientific Reports* **2020,** *10*, (1), 11535.

32. Meiring, H. D.; van der Heeft, E.; ten Hove, G. J.; de Jong, A. P. J. M. Nanoscale LC–MS(n): technical design and applications to peptide and protein analysis. *Journal of Separation Science* **2002,** *25*, (9), 557-568.

33. Imashiro, F.; Maeda, S.; Takegoshi, K.; Terao, T.; Saika, A. Structure-dependent intermolecular hydrogen-bond effects on carbon-13 NMR chemical shifts in enol forms of 1,3-diketones. *Journal of the American Chemical Society* **1987,** *109*, (17), 5213-5216.

34. Sawatzky, E.; Drakopoulos, A.; Rölz, M.; Sotriffer, C.; Engels, B.; Decker, M. Experimental and theoretical investigations into the stability of cyclic aminals. *Beilstein Journal of Organic Chemistry* **2016,** *12*, 2280-2292.

35. Carneiro de Oliveira, J.; Laborie, M. P.; Roucoules, V. Thermodynamic and Kinetic Study of Diels-Alder Reac-

tion between Furfuryl Alcohol and N-Hydroxymaleimides-An Assessment for Materials Application. *Molecules* **2020,** *25*, (2).

36. Otto, S.; Blokzijl, W.; Engberts, J. B. F. N. Diels-Alder Reactions in Water. Effects of Hydrophobicity and Hydrogen Bonding. *The Journal of Organic Chemistry* **1994,** *59*, (18), 5372-5376.

37. Bayart, C.; Peronin, S.; Jean, E.; Paladino, J.; Talaga, P.; Borgne, M. L. The combined use of analytical tools for exploring tetanus toxin and tetanus toxoid structures. *J Chromatogr B Analyt Technol Biomed Life Sci* **2017,** *1054*, 80-92.