THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 8, Issue of February 20, pp. 6235–6243, 2004

Printed in U.S.A.

Identification of Formaldehyde-induced Modifications in Proteins

REACTIONS WITH MODEL PEPTIDES*

Received for publication, September 29, 2003, and in revised form, November 18, 2003
Published, JBC Papers in Press, November 24, 2003, DOI 10.1074/jbc.M310752200

Bernard Metz‡§, Gideon F. A. Kersten‡, Peter Hoogerhout‡, Humphrey F. Brugghe‡, Hans A. M. Timmermans‡, Ad de Jong¶, Hugo Meiring¶, Jan ten Hove¶, Wim E. Hennink§, Daan J. A. Crommelin§, and Wim Jiskoot§‖

From the ‡Unit Research and Development, The Netherlands Vaccine Institute, 3720 AL Bilthoven, The Netherlands, §Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences, Faculty of Pharmaceutical Sciences, Utrecht University, 3508 TB Utrecht, The Netherlands, and ¶Laboratory for Analytical Chemistry, National Institute of Public Health and the Environment, 3720 BA Bilthoven, The Netherlands

Formaldehyde is a well known cross-linking agent that can inactivate, stabilize, or immobilize proteins. The purpose of this study was to map the chemical modifications occurring on each natural amino acid residue caused by formaldehyde. Therefore, model peptides were treated with excess formaldehyde, and the reaction products were analyzed by liquid chromatography-mass spectrometry. Formaldehyde was shown to react with the amino group of the N-terminal amino acid residue and the side-chains of arginine, cysteine, histidine, and lysine residues. Depending on the peptide sequence, methylol groups, Schiff-bases, and methylene bridges were formed. To study intermolecular cross-linking in more detail, cyanoborohydride or glycine was added to the reaction solution. The use of cyanoborohydride could easily distinguish between peptides containing a Schiff-base or a methylene bridge. Formaldehyde and glycine formed a Schiff-base adduct, which was rapidly attached to primary N-terminal amino groups, arginine and tyrosine residues, and, to a lesser degree, asparagine, glutamine, histidine, and tryptophan residues. Unexpected modifications were found in peptides containing a free N-terminal amino group or an arginine residue. Formaldehyde-glycine adducts reacted with the N terminus by means of two steps: the N terminus formed an imidazolidinone, and then the glycine was attached via a methylene bridge. Two covalent modifications occurred on an arginine-containing peptide: (*i*) the attachment of one glycine molecule to the arginine residue via two methylene bridges, and (*ii*) the coupling of two glycine molecules via four methylene bridges. Remarkably, formaldehyde did not generate intermolecular cross-links between two primary amino groups. In conclusion, the use of model peptides enabled us to determine the reactivity of each particular cross-link reaction as a function of the reaction conditions and to identify new reaction products after incubation with formaldehyde.

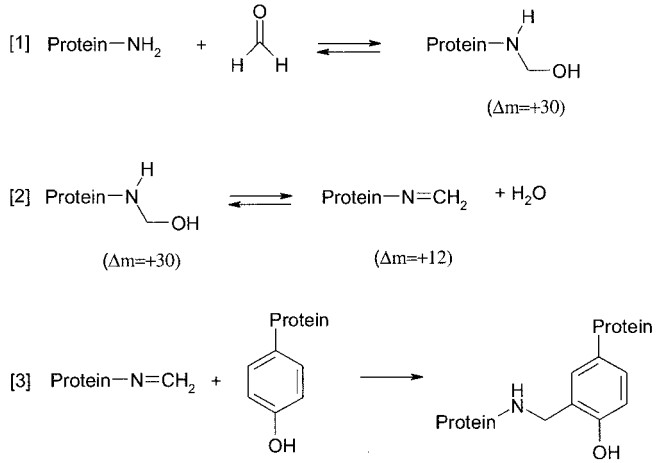
Aldehydes, such as formaldehyde and glutaraldehyde are widely employed reagents in the biochemical, biomedical, and pharmaceutical fields. Formaldehyde, for example, is applied to inactivate toxins and viruses for the production of vaccines, such as diphtheria, tetanus toxoid, hepatitis A, anthrax, and inactivated polio vaccine, and to stabilize recombinant pertussis toxin (1–4). The vaccine quality depends to a considerable extent upon the chemical modifications caused by the formaldehyde treatment (1, 5, 6). Formaldehyde is also used for isotope-labeling of proteins (7–9), for studying protein-protein interactions, *e.g.* histone organization in nucleosomes (10–12), and for fixation of cells and tissues (13). Glutaraldehyde is utilized for the preparation of bioprostheses such as heart valves and vascular grafts (14–16) and for conjugation of enzymes to carrier systems (17). These examples demonstrate the wide range of roles of aldehydes in the biomedical field. Besides the use of aldehydes in diverse applications, they can also destroy important sites of proteins, such as crucial epitopes or active sites in enzymes.

Several decades ago, extensive model studies were performed on reactions of formaldehyde with mixtures of amino acids and derivatives to determine which amino acids can cross-link (18–21). It was demonstrated that formaldehyde reacts first with the amino and thiol groups of amino acids and forms methylol derivatives. In the case of primary amino groups, the methylol groups partially undergo condensation to an imine, also called a Schiff-base (Scheme 1). Subsequently, the imine can cross-link with glutamine, asparagine, tryptophan, histidine, arginine, cysteine, and tyrosine residues. Some of the chemical structures of the proposed adducts have been elucidated by NMR (22). This knowledge, however, is not sufficient to predict all possible modifications in proteins that are induced by formaldehyde. Moreover, the formation of modifications is influenced by various factors, such as the rate of a particular cross-link reaction, the position and local environment of each reactive amino acid in the protein, the pH, the components present in the reaction solution, and the reactant concentrations. Importantly, the nature of all possible chemical modifications in proteins caused by formaldehyde has not yet been fully elucidated, in part because of the low resolution and sensitivity of the analytical methods available at the time the above studies were performed (18–21). However, the current availability of tandem high-performance liquid chromatography-mass spectrometry provides more detailed insight into the chemistry of protein-formaldehyde reactions.

The purpose of this study was to elucidate the chemical nature of the reactions between formaldehyde and proteins. Therefore, a set of model peptides was prepared and used to

* This study was supported, in part, by Grant 9802.086.0/3170.0039 from the "Platform Alternatieven voor Dierproeven" (the Dutch platform on alternatives to animal experiments). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‖ To whom correspondence should be addressed: Dept. of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences, Faculty of Pharmaceutical Sciences, Sorbonnelaan 16, P. O. Box 80082, 3508 TB Utrecht, The Netherlands. Tel.: 31-30-253-6970; Fax: 31-30-251-7839; E-mail: W.Jiskoot@pharm.uu.nl.



SCHEME 1. **The reaction of formaldehyde with proteins starts with the formation of methylol adducts on amino groups [1].** The methylol adducts of primary amino groups are partially dehydrated, yielding labile Schiff-bases [2], which can form cross-links with several amino acid residues, *e.g.* with tyrosine [3].

map systematically the different chemical modifications induced by formaldehyde treatment. The selected peptides can be divided into two groups (see Table I): the first group had the amino acid sequence Ac-VELXVLL, in which one amino acid residue (X) varies and the remaining amino acid residues are non-reactive with formaldehyde. The second group was synthesized for studying the possible formation of intramolecular, formaldehyde-mediated cross-links between two reactive residues and contained peptides with the following sequence: Ac-LOENXLLZF-NH_2, where O, X, and Z are either a (non-reactive) alanine residue or an arginine, lysine, and histidine residue in different permutations (see Table I). The reaction conditions were largely based on the detoxification process of diphtheria toxin for vaccine production (1). Because glycine is used as a reagent during the inactivation of diphtheria toxin by formaldehyde for the preparation of diphtheria toxoid vaccines (6), it was especially chosen to study in detail the cross-link reaction with peptides. The conversion of peptides was monitored by tandem reversed-phase liquid chromatography, electrospray ionization mass spectrometry (LC/MS).[1] In this paper, we present an overview of the major conversion products resulting from reactions between model peptides and formaldehyde (in the absence and presence of glycine), several of which have not been identified before. Our data can be used for the prediction and identification of reactive sites in proteins after exposure to formaldehyde.

EXPERIMENTAL PROCEDURES

Chemicals—Formaldehyde (37%), formic acid (99%), glycine, potassium dihydrogen phosphate ($KH_2PO_4 \cdot 3H_2O$), and dipotassium hydrogen phosphate ($K_2HPO_4 \cdot 3H_2O$) were purchased from Merck (Amsterdam, The Netherlands). Sodium cyanoborohydride ($NaCNBH_3$) was obtained from Sigma (Zwijndrecht, The Netherlands). N^{α}-Acetylarginine methyl ester (Ac-Arg-OMe) was obtained from Bachem Ag (Bubendorf, Switzerland). Dimethyl sulfoxide (Me_2SO) ultra-grade was acquired from Acros Organics ('s-Hertogenbosch, The Netherlands). Endoproteinase Glu-C was bought from Roche Applied Science (Almere, The Netherlands).

Peptides—Peptides (Table I) were synthesized on a 30-mmol scale by using an automated multiple peptide synthesizer equipped with a 96-column reaction block (SYRO II, Fa. MultiSynTech Gmbh, Witten, Germany). Couplings were performed with *N*-(9-fluorenyl)methoxycarbonyl (Fmoc)-amino acid (90 mmol), benzotriazolyloxy-*tris*-[*N*-pyrrolidino]phosphonium hexafluorophosphate (90 mmol), and *N*-methylmorpholine (180 mmol). Single couplings were performed in cycles 1–9 and double couplings from cycle 10. The Fmoc group was cleaved with piperidine/*N*,*N*-dimethylacetamide, 2/8 (v/v). Side-chain deprotection and cleavage from the solid support was effected with trifluoroacetic acid (TFA)/water (95/5, v/v), except for cysteine-, methionine-, and tryptophan-containing peptides, which were treated with TFA/ethanethiol (95/5, v/v). The peptides were purified by reversed-phase (C8 column) high performance liquid chromatography and their identity was confirmed by LC/MS. Before use, peptides were dissolved in water or Me_2SO/water (50/50, v/v) to a final concentration of 10 mM.

Standard Reactions with Peptides—For the reaction of peptides with formaldehyde, 10 μl of a 10 mM peptide solution, 10 μl of 1 M potassium phosphate, pH 7.2, and 5 μl of a second agent (1.0 M glycine, 1.0 M $NaCNBH_3$, 1.0 M Ac-Arg-OME, or water) were added to 70 μl of water. The reaction was started by adding 5 μl of an aqueous solution of 1.0 M formaldehyde. After mixing, the solution was incubated for 48 h at 35 °C. Samples were stored at −20 °C before analysis.

Variations in Reaction Conditions—The effect of different reaction conditions was investigated by varying the reaction time, pH, reagent concentrations, and the moment of addition of $NaCNBH_3$. The reaction of peptides with formaldehyde and glycine was monitored for 6 weeks. Aliquots (10 μl) were taken after 2, 6, and 24 h, 2, 6 and 24 days, and 6 weeks, and stored at −20 °C before analysis. To investigate the effect of pH, reactions were performed in potassium phosphate buffer at pH 5.2, 7.2, and 9.2. The influence of the concentration of the reagents on adduct formation was studied by varying the formaldehyde or the glycine concentration to final concentrations of 5, 50, and 500 mM. To determine internal cross-links in peptides, $NaCNBH_3$ was added 48 h after formaldehyde addition.

Removal of Excess Formaldehyde—Removal of formaldehyde was performed on a high-performance liquid chromatography system equipped with a 10 cm long × 200-μm inner diameter column filled with Poros 10 R2 (5 μm; PerSeptive Biosystems). The sample was diluted with water to a peptide concentration of 100 μM, and 10 μl of the diluted sample was trapped on the column. The column was rinsed for 10 min with solvent A (0.075% TFA in water) at a flow rate of 3 μl/min to remove formaldehyde. The peptide was eluted by a linear gradient from 0–60% solvent B (0.075% TFA in acetonitrile) in 25 min. The fraction containing the peptide was dried in a vacuum centrifuge (Concentrator 5301, Eppendorf) and dissolved in 100 μl of water. Sample was stored at −20 °C before analysis.

Formaldehyde Treatment of Ac-Arg-OME—Formaldehyde, glycine, and an arginine derivative, Ac-Arg-OME, were dissolved or diluted in D_2O to final concentrations of 1.0 M. A reaction mixture was prepared by successively adding 400 μl of glycine solution, 100 μl of Ac-Arg-OME solution, and 200 μl of formaldehyde solution to 300 μl of D_2O. After each addition, the solution was homogenized by gentle mixing. The preparation was incubated for 48 h at 35 °C. Sample was stored at −20 °C before analysis.

Digestion by Endoproteinase Glu-C—Peptides were digested by mixing 5 μl of 1 mM peptide solution, 5 μl of 1.0 M potassium phosphate buffer, pH 9.0, 1.0 μl of 1 μg/μl endoproteinase Glu-C solution, and 39 μl of water, followed by incubation for 24 h at 37 °C. Subsequently, samples were stored at −20 °C before analysis.

Nano-electrospray MS—Analytes were diluted to a concentration of 10 μM in water containing 5% (v/v) Me_2SO and 5% (v/v) formic acid. A gold-coated nano-electrospray needle with an orifice of 1–2 μm inner diameter was loaded with 10 μl of the sample. A stable spray was obtained by an overpressure of 0.5 bar onto the needles and adjusting the electrospray voltage to 0.75 kV. The capillary was heated to 150 °C. MS-spectra were acquired from *m/z* 50–2000, followed by successive stages of collision-induced dissociation (up to MS^4 measurements). The collision energies were optimized for each individual collision-induced dissociation mass analysis (between 30–35%).

LC/MS—Peptide samples were analyzed by nano-scale reversed phase-liquid chromatography (HP 1100 Series LC system, Hewlett Packard Gmbh, Waldbronn, Germany) coupled to electrospray mass spectrometry (LCQ™ Classic quadrupole ion trap), essentially as described previously by Meiring *et al.* (23). Briefly, each peptide sample was diluted to a concentration of 0.1 μM in water containing 5% (v/v) Me_2SO and 5% (v/v) formic acid. An injection volume of 10 μl was used for analysis. To desalt the samples for MS analysis, analytes were trapped on a 15 mm long × 100 μm inner diameter trapping column with Aqua C18 (5 μm; Phenomenex) at a flow rate of 3 μl/min and by using 100% solvent A (0.1 M acetic acid in water) as eluent for 10 min. Then, analytes were separated by reversed-phase chromatography by

[1] The abbreviations used are: LC/MS, tandem liquid chromatography-electrospray ionization mass spectrometry; Fmoc, *N*-(9-fluorenyl) methoxycarbonyl; Me_2SO, dimethyl sulfoxide; TFA, trifluoroacetic acid; Ac-Arg-OME, N^{α}-acetylarginine methyl ester; MS, mass spectrometry.

TABLE I
Peptides involved in this study and their mass increments after formaldehyde treatment under standard conditions[a]

Peptide	Peptide sequence	Formaldehyde, Δm	Formaldehyde/ $NaCNBH_3$, Δm	Formaldehyde/glycine, Δm	Formaldehyde/ Ac-Arg-OMe, Δm
		Da	*Da*	*Da*	*Da*
1	Ac-VEL**A**VLL-OH	0	-[b]	0	-
2	Ac-VEL**C**VLL-OH	30	-	30	242
3	Ac-VEL**D**VLL-OH	0	-	0	-
4	Ac-VEL**F**VLL-OH	0	-	0	-
5	Ac-VEL**H**VLL-OH	0	-	87	0
6	Ac-VEL**K**VLL-OH	0	28	0	254
7	Ac-VEL**M**VEL-OH[c]	0	-	0	-
8	Ac-VEL**N**VLL-OH	0	-	87	0
9	Ac-VEL**P**VLL-OH	0	-	0	-
10	Ac-VEL**Q**VLL-OH	0	-	87	0
11	Ac-VEL**R**VLL-OH	30	-	99/198	0
12	Ac-VEL**S**VLL-OH	0	-	0	-
13	Ac-VEL**T**VLL-OH	0	-	0	-
14	Ac-VEL**W**VLL-OH	12/30	-	12/87	0
15	Ac-VEL**Y**VEL-OH[c]	0	-	87/174	-
16	**L**AENALLAF-NH_2	12	28/26[d]	99	-
17	Ac-LAENALLAF-NH_2	0	-	0	-
18	Ac-LAENALL**H**F-NH_2	30	-	87	-
19	Ac-LAEN**K**LLAF-NH_2	12/30	28	12/30	-
20	Ac-L**R**ENALLAF-NH_2	30	-	99/198	-
21	Ac-L**R**ENALL**H**F-NH_2	30	-	99/186/198/285	-
22	Ac-LAEN**K**LL**H**F-NH_2	12/30	28/26[d]	12/30/87	-
23	Ac-L**R**EN**K**LLAF-NH_2	24	28/24[d]	24/99/111/123/198/210	-
24	Ac-L**R**EN**K**LL**H**F-NH_2	24	28/24[d]	24/99/111/123/153/ 186/198/210/285/297	-

[a] For details, see "Experimental Procedures."
[b] Experiment was not performed.
[c] The peptides Ac-VELMVLL-OH and Ac-VELYVLL-OH could not be obtained in acceptable purity. Therefore, the peptides Ac-VELMVEL-OH and Ac-VELYVEL-OH were synthesized.
[d] Peptide products with these mass increases were formed after 48-h incubation with formaldehyde followed by incubation with $NaCNBH_3$.

using a 25-cm long × 50 μm inner diameter analytical column with Pepmap (5 μm; Dionex) at a flow rate between 100–125 nl/min. A linear gradient was started from 10% solvent B (0.1 M acetic acid in acetonitrile) to 60% solvent B in 25 min. Next, the columns were equilibrated in 100% solvent A for 10 min.

The analytes were measured in the MS^1 mode (*m/z* 400–2000) to determine the mass increase and conversion of peptides after incubation with formaldehyde. The heated capillary was set to 150 °C and electrospray voltage was set to 1.6–1.7 kV. A second LC/MS measurement was performed to obtain detailed sequence information. Therefore, the peptides were analyzed by data-dependent scanning comprising an MS^1 scan (*m/z* 400–2000) followed by collision-induced dissociation of the most abundant ion in the MS^1 spectra. The collision energy was set on 35%.

RESULTS AND DISCUSSION

Establishment of Reaction Conditions—A set of synthetic peptides (Table I) was used to investigate the reactivity of amino acid residues reacting with formaldehyde. The reaction was monitored over a 6 week period by LC/MS. In general, shorter exposure of the peptides sensitive to formaldehyde resulted in lower conversions. After a reaction time of 48 h, all modifications that were observed in this introductory study were detectable by using LC/MS analysis. Variations in the pH showed that reactions did not occur below pH 5 and that a maximal conversion rate was reached above pH 7. Furthermore, the conversion of the peptides was proportional with the reactant concentration. Based on these experiments, we used the following standard reaction conditions in the rest of this study (unless stated otherwise): 50 times excess of formaldehyde (and glycine) with regard to the peptide concentration, incubation at pH 7.2 and 35 °C for 48 h.

Formation of Methylol and Imine Adducts—Peptide 1 was designed with amino acids residues, which were expected not to react with formaldehyde (Table I). Indeed, LC/MS-analyses showed that peptide 1 was not modified after incubation with formaldehyde. On the other hand, peptides containing a cysteine (peptide 2), arginine (peptides 11 and 20), tryptophan (peptide 14), histidine (peptide 18), or lysine residue (peptide 19) gave products with a mass increase of 30 Da. A second modification in peptides containing a tryptophan or a lysine residue (peptides 14 and 19, respectively) was observed. This modification caused a mass increase of 12 Da. Unexpectedly, and in contrast with the results of histidine-containing peptide 18 and lysine-containing peptide 19 (showing mass increases of 12 and/or 30 Da), formaldehyde treatment of peptide 5 (containing histidine) and peptide 6 (containing lysine) did not yield detectable amounts of reaction products. Nonetheless, a formaldehyde-glycine adduct could be attached to the histidinyl in peptide 5, and the lysyl in peptide 6 could react with formaldehyde and $NaCNBH_3$. Thus, both residues were reactive with formaldehyde (see Table I). Therefore, we assume that the reaction equilibrium toward the methylol and imine adduct depends upon the amino acid sequence.

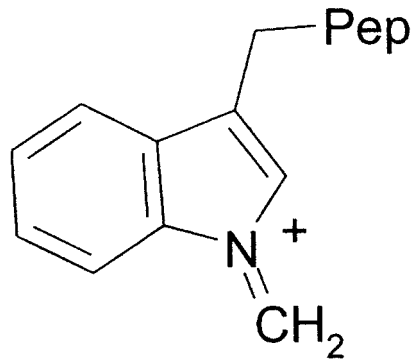


FIG. 1. **The imine adduct of a tryptophan residue formed after formaldehyde treatment.**

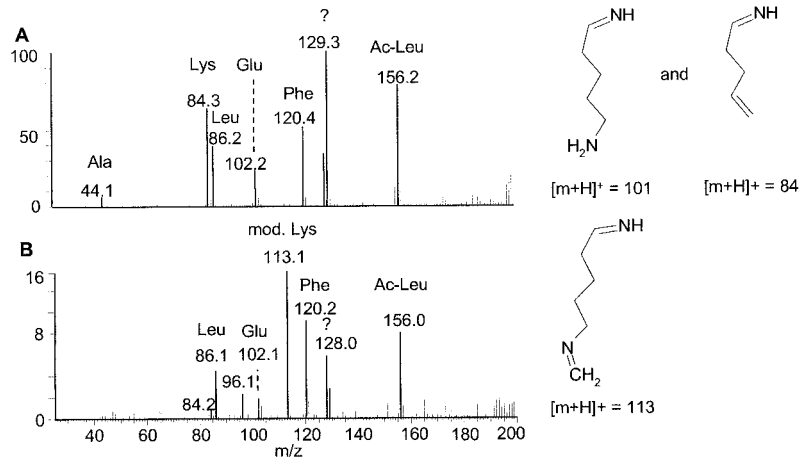
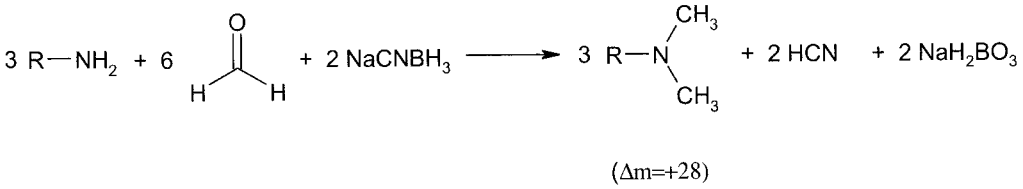
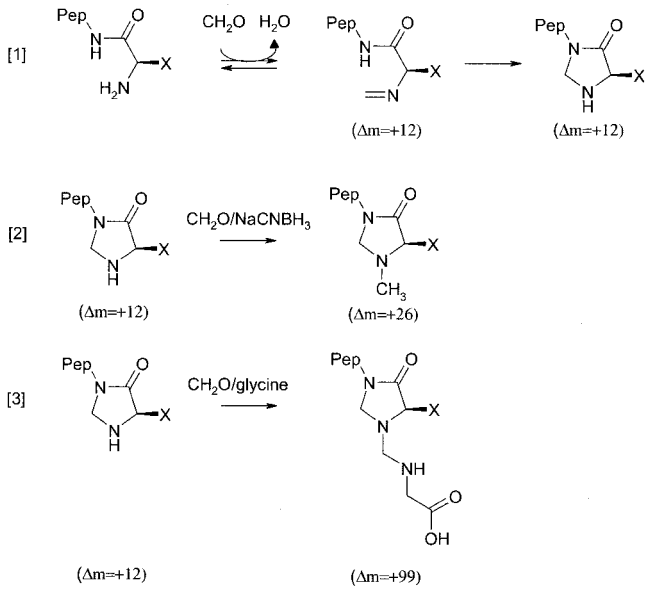


FIG. 2. **MS^2 spectra of peptide 19 containing a lysine residue.** Immonium ions of the lysine residue have typical masses of 101 and 84 Da. The mass of 101 Da is, in general, less frequently observed. Spectrum B shows the immonium ions of the modified peptide. An immonium ion with the particular mass of 113 corresponds to a modified lysine residue. Corresponding structures of the lysine immonium ions are shown on the *right*.



SCHEME 2. **Reduction of primary amino groups by adding formaldehyde and $NaCNBH_3$ (7, 8).**



SCHEME 3. **Modifications of peptide 16 containing a free N terminus.** A 4-imidazolidinone adduct was formed after adding formaldehyde to a peptide, probably by means of a Schiff-base intermediate [1]. The imidazolidinone could be reduced by adding $NaCNBH_3$ after 48-h incubation of peptide 16 with formaldehyde. An *N*-methyl-4-imidazolidinone was formed [2]. A glycine-formaldehyde adduct could be attached to the imidazolidinone [3].

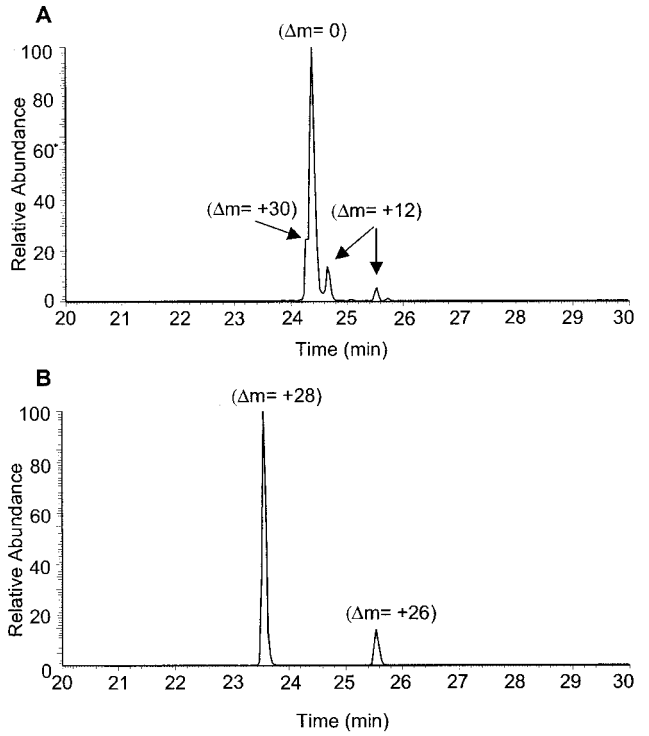


FIG. 3. **LC/MS chromatograms of formaldehyde-induced modifications in peptide 22 after 48-h incubation with formaldehyde (*A*) and after 48-h incubation with formaldehyde, followed by 48-h incubation with $NaCNBH_3$ (*B*).**

The increase of 30 Da is an indication of the formation of a methylol group (Scheme 1, reaction 1). Under standard reaction conditions, the conversion varied between 3–22%, depending upon the peptide. The formation of a methylol adduct to peptides is a reversible reaction, because the conversion of the arginyl peptide 20 was reduced form 17 to 4.5% after removal of free formaldehyde.

Structural Analysis of Reaction Products—MS^2 analyses were performed on formaldehyde-treated peptides to confirm that the methylol was located on cysteine, histidine, lysine, and arginine residues (peptides 2, 18, 19, and 20). The spectra revealed that peptide fragments were present with a mass increase of 30 Da, if they still contained a cysteine, arginine, or a histidine residue, whereas peptide fragments lacking these residues had the same mass as the corresponding MS^2 fragments of non-treated peptides. MS^2 measurements on the pep-

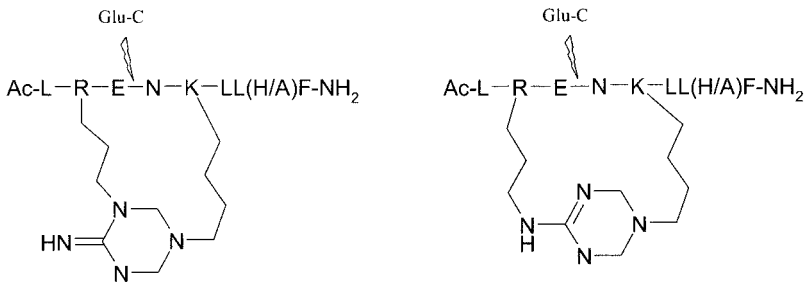


FIG. 4. **Modification of peptides 23 and 24.** Two different structures were presumably formed after incubation of these peptides with formaldehyde. Two methylene bridges were formed between the lysine residue and the arginine residue. The peptide bond at the C-terminal site of a glutamine residue can be cleaved with proteinase Glu-C.

tide containing a methylol on the lysine residue showed only peptide fragments with a mass increase of 12 Da instead of 30 Da, apparently because of dehydration of the methylol group (Scheme 1, reaction 2). The methylol located on a tryptophan residue could not be verified by MS^2 measurements, possibly because of the low conversion (3%). In conclusion, side chains of cysteine, histidine, lysine, arginine, and tryptophan residues can form methylol groups in the presence of formaldehyde.

Two possible reaction products could account for a mass increase of 12 Da found in formaldehyde-treated peptides 14 and 19: the formation of an imine or a methylene bridge (Scheme 1, reactions 2 and 3). These possibilities were studied by MS^2 measurements by generating immonium ions as a consequence of peptide fragmentation (24). MS^2 measurements performed on formaldehyde-treated peptide 14 showed that the mass increase of 12 Da was located on the tryptophan residue. The typical immonium ion of tryptophan (159 Da) was lost after the reaction with formaldehyde and a new fragment appeared (171 Da), indicating that the tryptophan residue was modified. No other new masses were detected, excluding the possibility that cross-links were formed between two residues. The proposed structure of the modified tryptophan residue is given in Fig. 1.

MS^2 measurements were also performed on formaldehyde-treated peptide 19 to determine the type of modification formed. The spectra revealed that, during formaldehyde incubation, a fragment with a mass of 84 Da disappeared and a fragment of 113 Da appeared. The fragment of 84 Da can be attributed to an immonium ion of an unmodified lysine residue. Another expected immonium ion of 101 Da was not found, but in general, this fragment is less frequently observed than the immonium ion of 84 Da, which lacks the ϵ-NH_2 group. The characteristic immonium ion of 113 Da, which was found after formaldehyde treatment, is indicative of the formation of a Schiff-base (Fig. 2). A second confirmation of the presence of a Schiff-base in peptide 19 was the reaction with $NaCNBH_3$, which was added 48 h after the incubation with formaldehyde. The ϵ-amino group of lysine was quantitatively converted to a dimethylated amine with a mass increase of 28 Da (Scheme 2). In conclusion, side chains of tryptophan and lysine residues can form imines during incubation with formaldehyde.

Intramolecular Cross-links—Peptide 16 containing a free N-terminal amino group was almost completely converted within 48 h into an adduct with a mass increment of 12 Da. According to the literature, formaldehyde (and acetaldehyde) can form a stable methylene bridge in such peptides, as determined by NMR and MS-measurements (25–27). The resulting ring structure is a 4-imidazolidinone (Scheme 3, reaction 1). To confirm the formation of a 4-imidazolidinone, we added $NaCNBH_3$ to the peptide after 48-h incubation with formaldehyde. This resulted in the formation of a peptide adduct with a mass increase of 26 Da (Scheme 3, reaction 2), indicating that an *N*-methyl-4-imidazolidinone had indeed formed. Normally, when adding formaldehyde and $NaCNBH_3$ simultaneously, N-terminal amino groups are reduced to a dimethylated amine. Indeed, a mass increase of 28 Da was then shown for peptide 16.

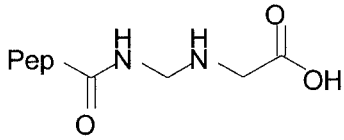
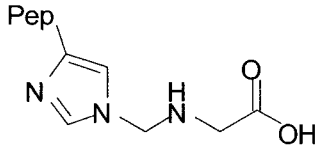
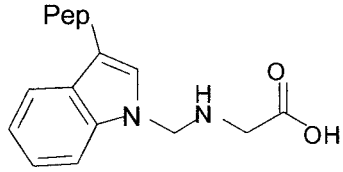
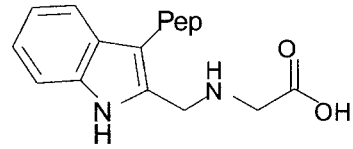
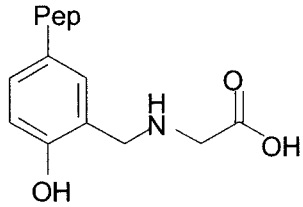
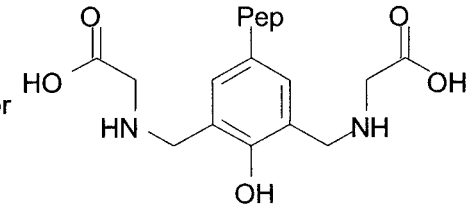
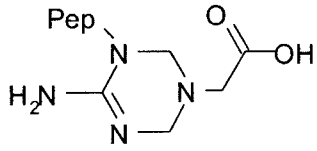
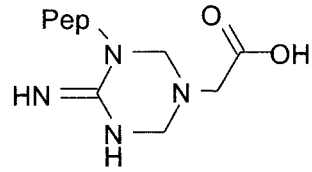
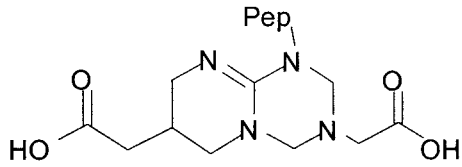
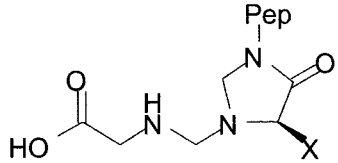
The formation of intramolecular cross-links was also expected for peptides 21–24, because they contain two (peptide 21–23) or three amino acid residues (peptide 24) that are reactive with formaldehyde; *i.e.* they contain lysine, arginine, and/or histidine residues. Under standard reaction conditions, peptide 21 showed one adduct with a mass increase of 30 Da after formaldehyde treatment. This suggests that one methylol adduct was formed, probably on the arginine or the histidine residue. A product with a mass increase of 60 Da was also expected, but could not be detected. When increasing the formaldehyde concentration to 500 mM, a reaction product was observed with a mass increase of 60 Da, indicating that two methylol groups were attached to the peptide. However, no intramolecular cross-link was formed in this peptide, because in that case, a mass increase of 12 Da was expected.

Besides the product with a mass increase of 30 Da, formaldehyde-treated peptide 22 showed two minor products each with a mass increase of 12 Da (Fig. 3). These minor products might be due to the formation of a Schiff-base located on the lysine residue or a methylene bridge between lysine and histidine residues. Addition of $NaCNBH_3$ to the formaldehyde-treated peptide yielded two products: small amounts of a peptide adduct with a mass increase of 26 Da and a larger amount with an increment of 28 Da. The mass increase of 28 Da can be explained by the formation of dimethylated lysine, whereas the increase of 26 Da presumably reflects a product with an intramolecular cross-link between the lysine and the histidine residue.

Both formaldehyde-treated peptides 23 and 24 showed two LC-peaks of adducts with a mass increase of 24 Da. These products could not be reduced by $NaCNBH_3$, which suggests that two methylene bridges had been formed between the side chains of lysine and arginine. The proposed structures are given in Fig. 4. The following experiment was performed to verify this hypothesis. Peptide 24 contains a glutamic acid residue between the arginine and the lysine residue (Table I), which allows cleavage of the peptides by endoproteinase Glu-C. So, the original and the formaldehyde-treated peptide 24 were both incubated with proteinase Glu-C. The original peptide was completely hydrolyzed by proteinase Glu-C, yielding two fragments with expected *m/z* of 770 and 459 Da. On the other hand, the formaldehyde-treated peptide 24 with a mass of 1234 Da was partially converted to a product with a mass of 1252 Da. The mass increase of 18 Da indicates that the peptide bond was hydrolyzed at the carboxylic site of the glutamic acid residue and that the two parts were still coupled to each other by means of cross-links between the lysine and arginine residues. This strongly supports the proposed structures given in Fig. 4.

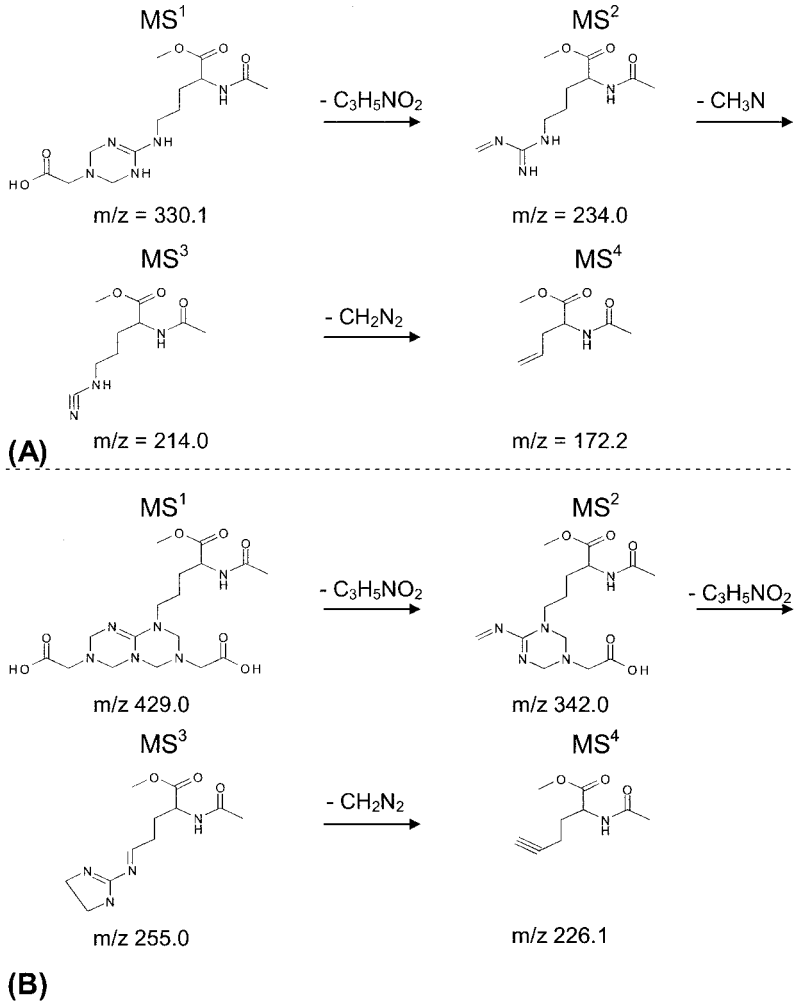
In contrast to previous studies with amino acids (18, 19), intermolecularly cross-linked peptides were not detected. This may be because of differences in reaction conditions and possible lower reactivity of peptides as compared with free amino acids. Unfortunately, the poor solubility of the peptides did not

TABLE II
Proposed structures of glycine-formaldehyde adducts attached to different amino acid residues present in peptides

Residue	Mass increase (Da)	Modification
Asparagine and glutamine	87	
Histidine	87	
Tryptophan	87	 and/or 
Tyrosine	87 and/or 174	 and/or 
Arginine	99	 and/or 
	198	
N-terminal amino group	99	

allow us to investigate whether cross-linking between two peptides occurs at higher concentrations. Increasing reaction time or mixing peptides with different reactive residues did not result in any intermolecular cross-linking.

Cross-links between Glycine and Peptides—From the previous experiments, it was shown that cross-links were formed between lysine and histidine residues or between lysine and arginine residues after formaldehyde treatment. To investigate whether formaldehyde can form methylene bridges between other amino acid residues, all peptides were individually incubated for 48 h with a 50-fold excess of formaldehyde and glycine. MS^1 analysis showed that glycine and formaldehyde reacted to form a reactive imine adduct with an ion mass of 88 Da. This glycine/formaldehyde adduct reacted with peptides



SCHEME 4. **MS analysis of two products formed during incubation of Ac-Arg-OME with formaldehyde and glycine.**[2]

TABLE III
Conversion of peptides to a glycine-formaldehyde adduct after 48-h incubation with formaldehyde[a]

Residue	Δm	Conversion
	Da	*%*
Asparagine	87	4.0 ± 1.8
Arginine	99	56 ± 13
	198	41 ± 14
Glutamine	87	3.6 ± 1.6
Histidine	87	6.6 ± 1.9
Tryptophan	87	4.7 ± 1.5
Tyrosine	87	62 ± 4
	174	5.2 ± 1.3
N-terminal amino group	12	76[b]
	99	13[b]

[a] Mean ± S.D.; $n = 3$.
[b] $n = 1$.

containing a histidine (peptide 5 and 18), asparagine (peptide 8), glutamine (peptide 10), or tryptophan residue (peptide 14), yielding products with a mass increase of 87 Da. This can be explained by the coupling of glycine to the peptides via a methylene bridge (Table II).

The peptide with a tyrosine residue (peptide 15) gave two products with an increase of 87 Da and 174 Da, which means that one or two glycine/formaldehyde adducts were coupled to the peptide, most probably to the ortho positions of the phenolic group (Table II). This type of reaction is known as the Mannich synthesis (21, 28).

Glycine was also coupled by formaldehyde to the N-terminal amino group of peptide 16, which gave the peptide a mass increase of 99 Da. The formation of this adduct presumably occurs in two steps: (*i*) a 4-imidazolidinone is formed and then (*ii*) the glycine is attached via a methylene bridge to this 4-imidazolidinone. The proposed structure is given in Scheme 3, reaction 3.

Arginine-containing peptides were also modified by formaldehyde and glycine. Peptide products were found with mass increases of 99 and 198 Da. The mass increase can be explained by the coupling of one or two glycine molecules to the peptide via two methylene bridges (see Table II). To verify that this modification occurs specifically on the arginine residue, an arginine derivative, Ac-Arg-OME, was treated with formaldehyde and glycine. The two main products from Ac-Arg-OME had the same mass increase of 99 and 198 Da. Their proposed structures are given in Scheme 4. MS^4 measurements were performed on both products with ion masses of 330 and 429 Da. The product with ion mass of 330 Da was degraded after four repeated fragmentations into fragment ions with *m/z* of 243, 214, and 172 Da. The product with ion mass of 429 Da was fragmented to ions with *m/z* of 342, 255, 226, and 184 Da. The possible structures of the fragment ions are given in Scheme 4.

Unexpectedly, formaldehyde did not cross-link detectable amounts of glycine to peptides 2, 6, and 19 containing a cysteine or a lysine residue, whereas these types of cross-links have been described in several articles (1, 8, 22, 29, 30). Pro-

[2] The products (MS^1 in *A* and *B*) were fragmented by the mass spectrometer in four successive steps. The *m/z* of the products and the fragments were measured.

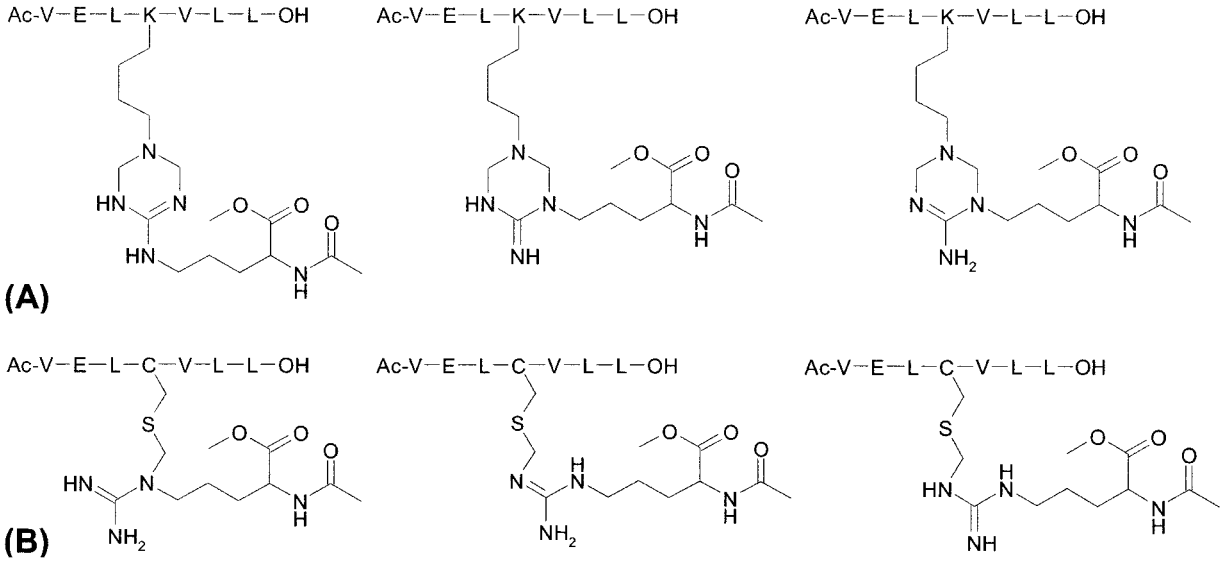


FIG. 5. **Modifications of peptide 6 (*A*) and peptide 2 (*B*) after incubation with formaldehyde and Ac-Arg-OME.**

longing the reaction time or increasing the formaldehyde and glycine concentration did not have any effect. The peptide samples were normally analyzed by LC/MS in an acidic environment. Because the methylene bridge between two amino groups might be unstable in an acidic environment (31), the samples were also measured at neutral pH by static nanoelectrospray ionization analysis. No modifications were found in these peptides. These outcomes are in contrast to the results of others who treated 1,3-diaminopropane and cysteine with formaldehyde. Methylene bridges were formed between the two amino groups in 1,3-diaminopropane and between the amino and thiol groups of cysteine (22, 30). However, both 1,3-diaminopropane and cysteine formed intramolecular cross-links, whereas in our experiments, intermolecular cross-links have to be formed between the peptide and glycine. Follow-up studies with a peptide containing two lysine residues may shed light upon these different observations.

MS analyses of peptides 21–24 treated with formaldehyde and glycine demonstrated that several products were formed with different masses (Table I). These peptides contain two or three residues that can react with formaldehyde. The determined masses can be explained as a combination of methylol adducts, imines, and intramolecular and intermolecular cross-links.

The conversion of peptides after glycine/formaldehyde treatment might predict the reaction rate of formaldehyde-induced intramolecular cross-links. Especially, intramolecular cross-links in proteins are initiated by the reaction of formaldehyde with lysine residues. Subsequently, the adducts probably form cross-links with reactive residues in their direct environment and especially with the residues that have the highest reactivity. Therefore, the conversion of peptides was determined after 48-h incubation with formaldehyde and glycine (Table III). The results show that the formaldehyde-glycine adduct was rapidly attached to free N-terminal amino groups, to arginine and tyrosine residues of peptides, and to a lesser extent to asparagine, glutamine, histidine, and tryptophan residues.

Cross-links between Ac-Arg-OME and Peptides—In theory, the cross-link reactions between glycine and peptides caused by formaldehyde can only occur via an imine. To verify this, peptides 5, 6, 11, and 14, containing a histidine, lysine, arginine, and tryptophan residue, respectively, were incubated with formaldehyde and Ac-Arg-OMe for 48 h. Ac-Arg-OME was only cross-linked to peptide 6 and gave the peptide a mass increase of 254 Da. This observation indicates that two methylene bridges were formed between the primary amine group of the lysine residue in peptide 6 and the arginine derivative (Fig. 5*A*). Peptides containing a histidine, arginine, or a tryptophan residue did not react with formaldehyde and Ac-Arg-OME, confirming that the reaction occurred by means of a Schiff-base.

Additionally, peptide 2 with a cysteine residue was treated with formaldehyde and Ac-Arg-OME to demonstrate that thiol groups can also form cross-links with arginine residues. A product with mass increase of 242 Da was found, indicating that one methylene bridge had been formed between the thiol group of peptide 2 and the arginine derivative (Fig. 5*B*). It has been reported in the literature that formaldehyde can react with the α-amino group and β-thiol group of N-terminal cysteines under formation of thiazolidine derivatives (32, 33). From these reports and our present data we conclude that cysteine residues can form cross-links at least with arginines and with N-terminal amino groups as a result of formaldehyde treatment.

CONCLUSIONS

This study has demonstrated that, depending on their sequence, peptides undergo a great diversity of chemical modifications after formaldehyde treatment. The modifications can be divided into three types: (*i*) methylol groups, (*ii*) Schiff-bases, and (*iii*) methylene bridges. The formation of methylol and Schiff-bases is reversible, and therefore these compounds are generally hard to detect. Still, methylol and Schiff-base derivatives could be demonstrated in several peptides by using LC/MS. They were located on residues with an amino or a thiol group.

The most important modification of peptides (and proteins) induced by formaldehyde is the formation of stable methylene bridges. In this study, we showed that only primary amino and thiol groups primarily react with formaldehyde and form cross-links in a second step with several other amino acid residues, *i.e.* with arginine, asparagine, glutamine, histidine, tryptophan, and tyrosine residues. In contrast to these cross-link reactions, no methylene bridges were formed between two primary amino groups. Moreover, Ac-Arg-OMe was not coupled to asparagine, glutamine, histidine, or tryptophan residues of a

peptide, indicating that only primary amino groups can form intermolecular cross-links with these reactive residues.

To discriminate between the formation of a Schiff-base or a methylene bridge, $NaCNBH_3$ was used. The reaction with formaldehyde and $NaCNBH_3$ is specific for primary amino groups; for peptides or proteins only, the N terminus and lysine residues are converted to dimethyl amino groups (7). In peptides, this conversion results in a mass increase of 28 Da. Cross-links between a lysine and a histidine residue were demonstrated by adding $NaCNBH_3$, resulting in a mass increase of 26 Da.

In conclusion, we have provided in the present study a detailed overview of possible chemical modifications of each individual amino acid residue caused by formaldehyde. Furthermore, the relative reactivity of the residues to form a particular cross-link was elucidated. Although identification of all intramolecular cross-links of formaldehyde-treated proteins still will be a tremendous job, the data from this study can be helpful to interpret peptide maps. In addition, if the local environment of the reactive residues is known (*e.g.* through x-ray crystallography or NMR studies), our data may be useful to predict the modifications in formaldehyde-treated proteins.

Acknowledgments—We thank Rob Liskamp for fruitful discussions, and Bert Zomer, Johan Kemmink, and Evert Evers for contributions to this work.

REFERENCES

1. Rappuoli, R. (1997) in *New Generation Vaccines* (Levine, M. M., Woodrow, G. C., Kaper, J. B., and Cobon, G. S., eds) 2nd Ed., pp. 417–435, Marcel Dekker, Inc., New York
2. Stapleton, J. T., and Lemon, S. M. (1997) in *New Generation Vaccines* (Levine, M. M., Woodrow, G. C., Kaper, J. B., and Cobon, G. S., eds) 2nd Ed., pp. 417–435, Marcel Dekker, Inc., New York
3. Leppla, S. H., Robbins, J. B., Schneerson, R., and Shiloach, J. (2002) *J. Clin. Invest.* **110,** 141–144
4. Murdin, A. D., Barreto, L., and Plotkin, S. (1996) *Vaccine* **14,** 735–746
5. Nencioni, L., Volpini, G., Peppoloni, S., Bugnoli, M., De Magistris, T., Marsili, I., and Rappuoli, R. (1991) *Infect. Immun.* **59,** 625–630
6. Metz, B., Jiskoot, W., Hennink, W. E., Crommelin, D. J. A., and Kersten, G. F. A. (2003) *Vaccine* **22,** 156–167
7. Jentoft, N., and Dearborn, D. G. (1983) *Methods Enzymol.* **91,** 570–579
8. Means, G. E., and Feeney, R. E. (1995) *Anal. Biochem.* **224,** 1–16
9. Gold, T. B., Smith, S. L., and Digenis, G. A. (1996) *Pharm. Dev. Technol.* **1,** 21–26
10. Kunkel, G. R., Mehrabian, M., and Martinson, H. G. (1981) *Mol. Cell. Biochem.* **34,** 3–13
11. Jackson, V. (1978) *Cell* **15,** 945–954
12. Feldman, M. Y. (1973) *Prog. Nucleic Acid Res. Mol. Biol.* **13,** 131–149
13. Fox, C. H., Johnson, F. B., Whiting, J., and Roller, P. P. (1985) *J. Histochem. Cytochem.* **33,** 845–853
14. Jayakrishnan, A., and Jameela, S. R. (1996) *Biomaterials* **17,** 471–484
15. Khor, E. (1997) *Biomaterials* **18,** 95–105
16. Mainil Varlet, P., Rieser, F., Grogan, S., Mueller, W., Saager, C., and Jakob, R. P. (2001) *Osteoarthritis Cartilage* **9,** Suppl. A, S6–S15
17. Cabral, J. M. S., and Kennedy, J. F. (1991) in *Protein Immobilization* (Taylor, R. F., ed) pp. 123–124, Marcel Dekker, Inc., New York
18. Fraenkel-Conrat, H., and Olcott, H. S. (1948) *J. Biol. Chem.* **174,** 827–843
19. Fraenkel-Conrat, H., and Olcott, H. S. (1948) *J. Amer. Chem. Soc.* **70,** 2673–2684
20. Blass, J., Bizzini, B., and Raynaud, M. (1965) *C. R. Acad. Sci. (Paris)* **261,** 1448–1449
21. Blass, J. (1966) *Bull. Soc. Chim. Fr.* **10,** 3120–3121
22. Kelly, D. P., Dewar, M. K., Johns, R. B., Wei-Let, S., and Yates, J. F. (1977) *Adv. Exp. Med. Biol.* **86A,** 641–647
23. Meiring, H. D., van der Heeft, E., ten Hove, G. J., and de Jong, A. (2002) *J. Sep. Sci.* **25,** 557–568
24. Biemann, K. (1990) in *Mass Spectrometry* (McCloskey, J. A., ed) Vol. 193, pp. 886–887, Academic Press, Inc., San Diego
25. Braun, K. P., Cody, R. B., Jones, D. R., and Peterson, C. M. (1995) *J. Biol. Chem.* **270,** 11263–11266
26. Fowles, L. F., Beck, E., Worrall, S., Shanley, B. C., and de Jersey, J. (1996) *Biochem. Pharmacol.* **51,** 1259–1267
27. Heck, A. J., Bonnici, P. J., Breukink, E., Morris, D., and Wills, M. (2001) *Chemistry* **7,** 910–916
28. Walker, J. F. (1964) *Formaldehyde*, 3rd Ed., Reinhold Publishing Corp., New York
29. Jiang, W., and Schwendeman, S. P. (2000) *Biotechnol. Bioeng.* **70,** 507–517
30. Kallen, R. G. (1971) *J. Am. Chem. Soc.* **93,** 6236–6248
31. Blass, J., Bizzini, B., and Raynaud, M. (1968) *Ann. Inst. Pasteur (Paris)* **115,** 881–898
32. Gesquière, J.-C., Diesis, E., and Tartar, A. (1990) *J. Chem. Soc. Chem. Commun.*, 1402–1403
33. Mitchell, M. A., Runge, T. A., Mathews, W. R., Ichhpurani, A. K., Harn, N. K., Dobrowolski, P. J., and Eckenrode, F. M. (1990) *Int. J. Pept. Protein Res.* **36,** 350–355