

Universiteit
Leiden
The Netherlands

Mass spectrometry-based degradomics analysis of toxoid vaccines

Michiels, T.J.M.

Citation
Michiels, T. J. M. (2021, September 9). *Mass spectrometry-based degradomics analysis of toxoid vaccines*. Retrieved from https://hdl.handle.net/1887/3209234

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3209234

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle https://hdl.handle.net/1887/3209234 holds various files of this Leiden University dissertation.

Author: Michiels, T.J.M.
Title: Mass spectrometry-based degradomics analysis of toxoid vaccines
Issue Date: 2021-09-09

MASS SPECTROMETRY-BASED DEGRADOMICS ANALYSIS OF TOXOID VACCINES

Thomas Michiels



MASS SPECTROMETRY-BASED DEGRADOMICS ANALYSIS OF TOXOID VACCINES

Thomas Michiels

Cover design: Thomas Michiels & Gildeprint
Cover photos: Wil Michiels & Lidy van Staalduinen
Thesis lay-out: Wendy Bour-van Telgen
Printing: Gildeprint

© Copyright, Thomas Michiels, 2021

ISBN: 978-94-6419-267-4

All rights reserved. No part of this book may be reproduced in any form or by any means without permission of the author.

MASS SPECTROMETRY-BASED DEGRADOMICS ANALYSIS OF TOXOID VACCINES

Proefschrift

ter verkrijging van

de graad van doctor aan de Universiteit Leiden,

op gezag van rector magnificus prof.dr.ir. H. Bijl,

volgens besluit van het college voor promoties

te verdedigen op donderdag 9 september 2021

klokke 13.45 uur

Door

Thomas Joost Martti Michiels

geboren te Roosendaal en Nispen

In 1991

Promotores:	Prof. Dr. G.F.A. Kersten
	Prof. Dr. W. Jiskoot
Co-promotor:	Dr. B. Metz

Promotiecommissie:	Prof. Dr. H. Irth, Universiteit Leiden (voorzitter)
	Prof. Dr. J.A. Bouwstra, Universiteit Leiden (secretaris)
	Prof. Dr. E.M. Topp, Purdue University
	Prof. Dr. C.A.C.M. van Els, Universiteit Utrecht
	Prof. Dr. C.F.M. Hendriksen, Universiteit Utrecht

The research described in this thesis was performed at the Institute for Translational Vaccinology, Intravacc (Bilthoven, the Netherlands) and the division BioTherapeutics of the Leiden Academic Centre for Drug Research (LACDR), Leiden University (Leiden, The Netherlands). The research was financially supported in part by the Vac2Vac project supported by the Innovative Medicines Initiative2 Joint Undertaking under grant agreement N-115924 and by the Ministry of Agriculture, Nature, and Food Quality, the Netherlands.

Table of contents

Chapter 1	General introduction and thesis outline	7
Chapter 2	Identification of formaldehyde-induced modifications in diphtheria toxin	21
Chapter 3	Formaldehyde treatment of proteins enhances proteolytic degradation by the endo-lysosomal protease cathepsin S	61
Chapter 4	Novel formaldehyde-induced modifications of lysine residue pairs in peptides and protein: identification and relevance to vaccine development	85
Chapter 5	Degradomics-based analysis of tetanus toxoids as a quality control assay	113
Chapter 6	Common reference-based tandem mass tag multiplexing for the relative quantification of peptides: design and application to degradome analysis of diphtheria toxoid	135
Chapter 7	Summary, general discussion and prospects	153
Appendix	Nederlandstalige samenvatting	166
	List of publications	168
	Curriculum vitae	170

1

GENERAL INTRODUCTION AND THESIS OUTLINE

Current events in vaccine development

The outbreak of COVID-19 has put vaccine development in the spotlight. The promise of protecting people without the economic downfall caused by lockdowns is appealing. Under enormous pressure with seemingly unlimited funding, the development of the first efficacious COVID-19 vaccines took less than a year [1]. Amongst others, suggestions and improvements are being made to the usually lengthy clinical trials [2]. The fastest vaccines to reach the market are based on existing vaccine concepts that were adapted to SARS-CoV-2. Hundreds of COVID-19 vaccines, spanning every potential vaccine concept, are currently in various stages of development. Every vaccine has its own characteristics and both comprehensive product characterization and the development of suitable quality control assays are fundamental to ensure the safety and efficacy of any vaccine.

Vaccine concepts

Over the past century several vaccine families have been developed. These families can be divided into four categories: live attenuated, subunit, vector and DNA/RNA vaccines, and inactivated vaccines.

Live attenuated viruses can infect the host and induce an immune response (and subsequent protection) but do not cause disease. The first form of vaccination employed this principle by inoculating (vaccinating) people with cowpox virus to protect against (human) smallpox. Although this vaccine was discontinued after the eradication of smallpox, several current vaccines still use live attenuated viruses (*e.g.*, measles, mumps, rubella and oral polio vaccines). The main advantage of live attenuated vaccines is their high and broad immunogenicity and potential for mucosal protection if administered mucosally [3]. There are, however, concerns regarding safety, in particular with respect to potential reversion to a virulent form. Moreover, the rational development of new live attenuated vaccines and evaluating the safety of the vaccine before the first clinical trials is more challenging than with other concepts [4].

For subunit vaccines, specific immunogenic antigens are either purified from a(n) (inactivated) pathogen or produced recombinantly in non-infectious eukaryotic production cells or bacteria, much like the production of recombinant therapeutic proteins, such as insulin and

monoclonal antibodies. The first recombinant vaccine based on this concept was hepatitis B vaccine, for which the surface antigen (HBsAg) was produced in yeast cells (replacing a vaccine based on HBsAg obtained from infected humans) [5]. Other examples of vaccines containing purified antigens include the conjugate vaccines, where an antigen is chemically bound to an immunogenic carrier protein (such as haemophilus B oligosaccharides conjugated to diphtheria toxoid [6]). Vaccines containing purified antigens have clear advantages because the risk of contamination with (other) pathogens or harmful substances such as bacterial endotoxins can be severely reduced. However, some subunit vaccines induce suboptimal immune responses [7,8]. The switch from killed whole cell pertussis vaccines to acellular pertussis vaccines (a subunit vaccine) likely is a contributing factor to the resurgence of pertussis in recent years [9].

In the past decades the concept of RNA- or DNA-based vaccines have gained attention. The main idea is that (messenger) RNA (or DNA) containing the sequence information of antigens can be translated in the cells of the vaccinee. This process mimics an infection and should ensure that the intracellularly produced antigens are displayed to the immune system, similar to the process in an infected cell. The main hurdle with RNA and DNA vaccines is getting the genetic information into the target cells. To facilitate this, viruses deprived of their ability to replicate, or harmless viruses can be used as vectors to deliver the genetic information of the intended antigen. This technology is very new and only recently (2019 and 2020) the first of such vaccines have been approved for use against Ebola [10,11]. Other strategies involve the use of mRNA, either naked or encapsulated in a delivery system such as lipid nanoparticles [12]. Adapting these vaccines to different antigens can be achieved relatively quickly, because different mRNA molecules have very similar physicochemical characteristics, making them suitable for rapid response against new diseases [1].

Inactivated vaccines contain a killed or detoxified pathogen or toxin. Several methods are being employed, but chemical inactivation, usually by treatment with aqueous formaldehyde or β-propiolactone, is the most common one. Licensed vaccines in this category include influenza, polio, diphtheria and tetanus vaccines.

Toxoid vaccines

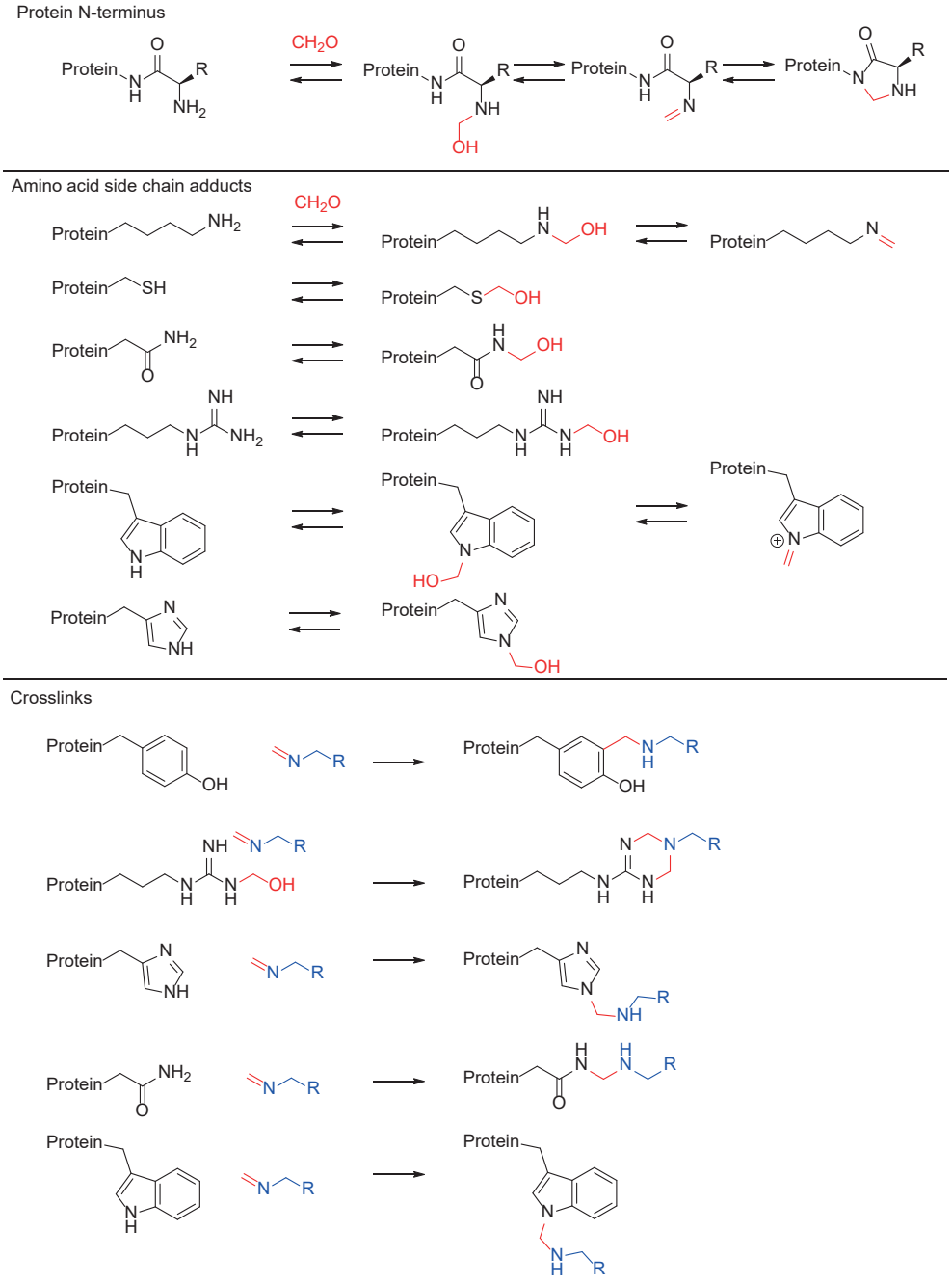
In this thesis, some of the oldest effective human vaccines –those prepared by formaldehyde-inactivation– are studied and characterized. Toxoid vaccines were discovered by serendipity. In 1891, antibodies obtained from the serum of animals that were injected with small amounts of diphtheria toxin were first used to successfully treat a patient suffering from diphtheria [13]. Following this success, more diphtheria toxin was needed for both research and for the production of more antibodies for the treatment of other patients. During such a diphtheria toxin production, a cultivation vessel was cleaned with formaldehyde but not all the residual formaldehyde was removed before adding new toxin to the vessel [14]. This toxin was then found to have a much lower toxicity in animals without detrimental effect on the antibody production. Further optimization based on formaldehyde-detoxification of the so-called *toxoid* (*i.e.,* toxin-like) by Gaston Ramon in 1923 eventually resulted in vaccines that were suitable for human use [15]. Nowadays, diphtheria and tetanus vaccines are still produced by formaldehyde treatment of the respective toxins.

Formaldehyde-inactivation has been very successful to obtain vaccines against bacterial toxins such as diphtheria, tetanus and pertussis toxins and viruses such as polio and influenza. Moreover, it has shown potential for new vaccines, such as those targeting enterovirus 71 [16], coxsackieviruses [17] and coronaviruses (*i.e.*, SARS-CoV-1[18]). Besides some unfounded criticism towards formaldehyde in vaccines [19] (the concentration limits in vaccines are lower than the endogenous formaldehyde concentrations found in the human body [20]), several failures have decreased its popularity for new vaccine concepts. The infamous Cutter incident with inactivated polio vaccine was related to the formaldehyde not reaching every polio virion, leading to improper inactivation, which resulted in 260 cases of paralytic polio in vaccinees [21]. These types of incidents can be avoided by proper quality control and designing a robust production process. Other serious incidents involve the clinical trials of formaldehyde-inactivated viruses which caused enhanced disease upon contracting the wildtype disease in recipients instead of offering protection. This was the case with formaldehyde-inactivated measles vaccine and a formaldehyde-inactivated RSV vaccine [22]. Understanding the biochemical and immunological fundamentals is essential for the development of new vaccines.

Because formaldehyde-inactivation of toxoid vaccines was invented by chance and not by design, the inactivation chemistry at the foundation of these vaccines was not immediately clear. Over the years, several groups have worked on understanding the interactions between formaldehyde and proteins [23-30]. An overview of common reactions and reaction products is depicted in **Scheme 1**. The reactions start with the formation of an hydroxymethyl on an amine, amide, guanidino group or thiol. For amines this product is in equilibrium with the corresponding imine. These imines can then act as electrophiles in subsequent (crosslinking) reactions. Crosslinks can be formed either between molecules in the matrix (such as amino acids) and the protein of interest, between various protein molecules, or within the same protein molecule. Even though amides (*e.g.*, the amide in the backbone of an N-terminus, glutamine residues or asparagine residues) are usually considered poor nucleophiles, reaction products with imines are commonly observed. The most commonly encountered crosslinks are between lysine residues and arginine or tyrosine residues [23]. Both K-R and K-Y crosslinks also occur as double crosslinks, in which two different lysine residues are linked to the same arginine or tyrosine residue. Overall, the conversion rates are low and the potential reaction products for a given amino acid are numerous. This results in very heterogeneous mixtures containing many different molecules when treating large proteins (such as diphtheria or tetanus toxin) with formaldehyde. This heterogeneity drastically complicates the analysis and characterization of toxoid-based vaccines.

Toxoid (and polio and hepatitis A and B) vaccines are often adsorbed to aluminum salts to enhance the immune response. The exact immunological mechanism remains poorly understood. The three most commonly suggested hypotheses are the so-called depot-effect (slow release of the antigen from the injection site over time), enabling phagocytosis of the adsorbed antigen by increasing the particle size and triggering local inflammation [31,32]. While after years of research no definitive answer has been given, the increase in efficacy of the vaccines that benefit from aluminum-based adjuvants is clear [33].

Scheme 1. Overview of formaldehyde-induced modifications described in the literature.



New assays for quality control of toxoid vaccines

The combination of product heterogeneity caused by the formaldehyde-inactivation process and the highly turbid nature of aluminum salt-containing vaccines makes characterization of the final product challenging. Historically, the two most important parameters of vaccines: safety and efficacy, have been studied by using animal tests, which are not affected by these challenges [34,35]. Efforts to reduce, refine or replace animal tests (3R concept) are being made [36]. To date, an important part of these efforts is to use a so-called consistency approach; if *in vitro* assays can show batch-to-batch consistency of vaccines or their intermediates, then comparison to a gold standard with confirmed potency and safety can circumvent additional animal tests [37]. *In vitro* assays have several advantages beside the obvious ethical considerations: the assay variability is generally smaller and they are considerably cheaper than *in vivo* assays. This results in the unique situation where every stakeholder (government, regulatory agencies and vaccine manufacturers) supports the development of alternatives to animal tests for the batch release of vaccines [38,39]. Several assays have been developed over the years, including antibody binding assays such as the flocculation test [40], ELISAs and biosensor assays [41,42] (non-adsorbed antigen fraction) or direct Alhydrogel formulation immunoassays (DAFIA, analysis of the adsorbed antigen fraction) [43]. These serological tests are aimed at the refinement of animal tests. Moreover, physicochemical techniques, such as SDS-PAGE, primary amino group determination, fluorescence spectroscopy and circular dichroism analysis, have been developed for the characterization of toxoids in a consistency approach [42]. The assays employed in a consistency approach are usually intended for the actual replacement of animal tests. Additionally, complementary assays, preferably those capable of analyzing the final product, are required to expand the panel of tests capable of testing various product characteristics to ensure that the product quality meets safety and efficacy requirements.

In search of a new animal-free assay to confirm the quality of toxoid vaccines, we looked into the adaptive immune system for inspiration. Globally the adaptive immune system can protect the body against pathogens with (a combination of) two types of responses: a humoral response with the production of antibodies, and a T-cell response. T-cells can be divided into cytotoxic T-cells expressing the CD8 receptor and T-helper cells expressing the CD4 receptor. To become effector T-cells, an antigen presenting cell has to present a

fragment of an antigen (a T-cell epitope) to the right T-cell. Almost all cells in the body present fragments of proteins (peptides, predominantly originating from within that particular cell) to naïve CD8 T-cells through Major Histocompatibility Complex I (MHC-I) molecules. Peptides displayed to the immune system in this way can show either endogenous self-proteins not resulting in a T-cell response, or unknown proteins which could be indicative of, for instance, a viral infection. The cytotoxic T-cell can then kill an infected or otherwise aberrant cell and, for instance, prevent the production of more virions. Contrarily, CD-4 T-cells are activated by peptides presented on MHC-II molecules, which are expressed by specific immune cells, with dendritic cells being the most important cell type. Dendritic cells take up potential antigens and process them in their endo-lysosomal vesicles to eventually present the peptides derived from these antigens on their MHC-II molecules. Activation of CD-4 T-cells is pivotal for the other arm of the adaptive immune response: humoral immunity [44]. Therefore, the antigen processing by dendritic cells is a key step in achieving a good immune response for formaldehyde-inactivated toxoid vaccines. Several groups have identified a correlation between the intracellular antigen degradation speed and the immunogenicity of the antigen. Antigens that were found to be more resistant to proteolytic degradation by dendritic cells were found to be more immunogenic [45-51]. We hypothesized that by mimicking this part of the immune system in a simplified way, we should be able to identify potential variations in the protein antigens used in vaccines. By studying the enzymatic proteolysis of antigens in such a *degradomics* approach, mimicking a small component of the immune system could be part of a panel consisting of several *in vitro* assays, replacing the need for *in vivo* studies.

Thesis outline

The chemical and structural heterogeneity of toxoid vaccines makes their analysis challenging. However, detailed insights on a molecular level can be obtained by mass spectrometry. Our initial focus was the identification of formaldehyde-induced modifications in diphtheria toxin, which is described in **Chapter 2**. Subsequently, the methods described in **Chapter 2** were applied to study what effects formaldehyde-induced modifications on model proteins have on their susceptibility to enzymatic proteolysis (**Chapter 3**). During the analysis of these model proteins, unknown formaldehyde-induced modifications were observed. The structural elucidation of these modifications, the discovery of a new type of crosslinks and various other subsequent reaction products are described in **Chapter 4**. The degradomics analysis

described in **Chapter 3** was applied to tetanus toxoids to distinguish heat-denaturated toxoids from their original state (**Chapter 5**). In order to reduce the analysis time and further improve the degradomics approach, an optimized strategy using Tandem Mass Tag multiplexing for the relative quantification of peptides was developed for the analysis of diphtheria toxoids (**Chapter 6**). Finally, **Chapter 7** provides a brief discussion on the results presented in this thesis and offers some perspectives on implementation of the findings for toxoid vaccine development, quality control and further research.

References

1. Krammer, F. SARS-CoV-2 vaccines in development. *Nature* **2020**, *586*, 516-527, doi:10.1038/s41586-020-2798-3.
2. van der Plas, J.L.; Roestenberg, M.; Cohen, A.F.; Kamerling, I.M.C. How to expedite early-phase SARS-CoV-2 vaccine trials in pandemic setting-A practical perspective. *Br J Clin Pharmacol* **2020**, 10.1111/bcp.14435, doi:10.1111/bcp.14435.
3. Parker, E.P.; Molodecky, N.A.; Pons-Salort, M.; O'Reilly, K.M.; Grassly, N.C. Impact of inactivated poliovirus vaccine on mucosal immunity: implications for the polio eradication endgame. *Expert Rev Vaccines* **2015**, *14*, 1113-1123, doi:10.1586/14760584.2015.1052800.
4. Minor, P.D. Live attenuated vaccines: Historical successes and current challenges. *Virology* **2015**, *479-480*, 379-392, doi:10.1016/j.virol.2015.03.032.
5. Emini, E.A.; Ellis, R.W.; Miller, W.J.; McAleer, W.J.; Scolnick, E.M.; Gerety, R.J. Production and immunological analysis of recombinant hepatitis B vaccine. *J Infect* **1986**, *13 Suppl A*, 3-9, doi:10.1016/s0163-4453(86)92563-6.
6. Schneerson, R.; Barrera, O.; Sutton, A.; Robbins, J.B. Preparation, characterization, and immunogenicity of Haemophilus influenzae type b polysaccharide-protein conjugates. *J Exp Med* **1980**, *152*, 361-376, doi:10.1084/jem.152.2.361.
7. Chasaide, C.N.; Mills, K.H.G. Next-Generation Pertussis Vaccines Based on the Induction of Protective T Cells in the Respiratory Tract. *Vaccines (Basel)* **2020**, *8*, doi:10.3390/vaccines8040621.
8. Brummelman, J.; Wilk, M.M.; Han, W.G.; van Els, C.A.; Mills, K.H. Roads to the development of improved pertussis vaccines paved by immunology. *Pathog Dis* **2015**, *73*, ftv067, doi:10.1093/femspd/ftv067.
9. Sheridan, S.L.; Frith, K.; Snelling, T.L.; Grimwood, K.; McIntyre, P.B.; Lambert, S.B. Waning vaccine immunity in teenagers primed with whole cell and acellular pertussis vaccine: recent epidemiology. *Expert Rev Vaccines* **2014**, *13*, 1081-1106, doi:10.1586/14760584.2014.944167.
10. FDA. First FDA-approved vaccine for the prevention of Ebola virus disease, marking a critical milestone in public health preparedness and response. Availabe online: https://www.fda.gov/news-events/press-announcements/first-fda-approved-vaccine-prevention-ebola-virus-disease-marking-critical-milestone-public-health (accessed on 11-Nov-2020).
11. EMA. Zabdeno Ebola vaccine (rDNA replication-incompetent). Availabe online: https://www.ema.europa.eu/en/medicines/human/EPAR/zabdeno (accessed on 11-Nov-2020).
12. Verbeke, R.; Lentacker, I.; De Smedt, S.C.; Dewitte, H. Three decades of messenger RNA vaccine development. *Nano Today* **2019**, *28*, 100766, doi:10.1016/j.nantod.2019.100766.
13. Plotkin, S.A.; Orenstein, W.A.; Offit, P.A.; Roper, M.H.; Wassilak, S.G.F.; Scobie, H.M.; Ridpath, A.D. A Short History of Vaccination. In *Vaccines*, 7th ed.; Elsevier: 2018; pp. 6-7.
14. Glenny, A.T.; Hopkins, B.E. Diphtheria Toxoid as an Immunising Agent. *Br J Exp Pathol* **1923**, *4*, 283-288.
15. Ramon, G. Sur le pouvoir floculant et sur les propriétés immunisantes d'une toxine diphtérique rendue anatoxique. *C R Acad Sci Paris* **1923**, 1338-1340.
16. Reed, Z.; Cardosa, M.J. Status of research and development of vaccines for enterovirus 71. *Vaccine* **2016**, *34*, 2967-2970, doi:10.1016/j.vaccine.2016.02.077.
17. 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., *et al.* A hexavalent Coxsackievirus B vaccine is highly immunogenic and has a strong protective capacity in mice and nonhuman primates. *Sci Adv* **2020**, *6*, eaaz2433, doi:10.1126/sciadv.aaz2433.

18. 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*, 1329-1338, doi:10.1086/522431.

19. Kata, A. A postmodern Pandora's box: Anti-vaccination misinformation on the Internet. *Vaccine* **2010**, *28*, 1709-1716, doi:https://doi.org/10.1016/j.vaccine.2009.12.022.

20. 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*, 2738-2743, doi:10.1016/j.vaccine.2013.03.071.

21. Plotkin, S.A.; Orenstein, W.A.; Offit, P.A.; Roper, M.H.; Wassilak, S.G.F.; Scobie, H.M.; Ridpath, A.D. General Aspects of Vaccination. In *Vaccines*, 7th ed.; Elsevier: 2018; p. 10.

22. 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*, 695-719, doi:10.1586/erv.12.38.

23. 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*, 815-822, doi:10.1021/bc050340f.

24. 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., *et al.* Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. *J Biol Chem* **2004**, *279*, 6235-6243, doi:10.1074/jbc.M310752200.

25. 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*, 126, doi:10.1038/s42004-019-0224-2.

26. 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*, 289-292.

27. 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, doi:10.1016/j.ijbiomac.2018.10.159.

28. 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*, 21-26, doi:10.3109/10837459609031414.

29. 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*, 60-67, doi:10.1016/j.aca.2010.07.040.

30. 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*, 168-183, doi:10.1016/j.aca.2008.04.049.

31. Ghimire, T.R. The mechanisms of action of vaccines containing aluminum adjuvants: an *in vitro* vs *in vivo* paradigm. *Springerplus* **2015**, *4*, 181, doi:10.1186/s40064-015-0972-0.

32. HogenEsch, H.; O'Hagan, D.T.; Fox, C.B. Optimizing the utilization of aluminum adjuvants in vaccines: you might just get what you want. *NPJ Vaccines* **2018**, *3*, 51, doi:10.1038/s41541-018-0089-x.

33. Lindblad, E.B. Aluminium adjuvants--in retrospect and prospect. *Vaccine* **2004**, *22*, 3658-3668, doi:10.1016/j.vaccine.2004.03.032.

34. Council of Europe. European Pharmacopoeia 10.0. In *Assay of tetanus vaccine (adsorbed)*, 2020; pp 275-278.

35. Council of Europe. European Pharmacopoeia 10.0. In *Tetanus vaccine (adsorbed)*, 2020; pp 1042-1043.

36. De Mattia, F.; Chapsal, J.M.; Descamps, J.; Halder, M.; Jarrett, N.; Kross, I.; Mortiaux, F.; Ponsar, C.; Redhead, K.; McKelvie, J., *et al.* The consistency approach for quality control of vaccines- a strategy to improve quality control and implement 3Rs. *Biologicals* **2011**, *39*, 59-65, doi:10.1016/j.biologicals.2010.12.001.

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

38. Halder, M.; Depraetere, H.; Delannois, F.; Akkermans, A.; Behr-Gross, M.E.; Bruysters, M.; Dierick, J.F.; Jungback, C.; Kross, I.; Metz, B., *et al.* Recommendations of the VAC2VAC workshop on the design of multi-centre validation studies. *Biologicals* **2018**, *52*, 78-82, doi:10.1016/j.biologicals.2018.01.003.

39. Bruysters, M.W.P.; Schiffelers, M.-J.; Hoonakker, M.; Jungbaeck, C.; Ragan, I.; Rommel, E.; van der Stappen, T.; Viviani, L.; Hessel, E.V.; Akkermans, A.M., *et al.* Drivers and barriers in the consistency approach for vaccine batch release testing: Report of an international workshop. *Biologicals* **2017**, *48*, 1-5, doi:https://doi.org/10.1016/j.biologicals.2017.06.006.

40. Lyng, J.; Bentzon, M.W. The quantitative estimation of diphtheria and tetanus toxoids. 1. The flocculation test and the Lf-unit. *J Biol Stand* **1987**, *15*, 27-37, doi:10.1016/0092-1157(87)90014-x.

41. Riches-Duit, R.; Hassall, L.; Rigsby, P.; Stickings, P. Evaluation of a capture antigen ELISA for the characterisation of tetanus vaccines for veterinary use. *Biologicals* **2019**, *61*, 8-14, doi:10.1016/j.biologicals.2019.08.003.

42. 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*, 156-167.

43. Westdijk, J.; Metz, B.; Spruit, N.; Tilstra, W.; van der Gun, J.; Hendriksen, C.; Kersten, G. Antigenic fingerprinting of diphtheria toxoid adsorbed to aluminium phosphate. *Biologicals* **2017**, *47*, 69-75, doi:10.1016/j.biologicals.2016.10.005.

44. Parham, P. Principles of Adaptive Immunity. In *The Immune System*, 3rd ed.; Taylor & Francis Group: 2009; pp. 71-84.

45. Carmicle, S.; Dai, G.; Steede, N.K.; Landry, S.J. Proteolytic sensitivity and helper T-cell epitope immunodominance associated with the mobile loop in Hsp10s. *J Biol Chem* **2002**, *277*, 155-160, doi:10.1074/jbc.M107624200.

46. Kim, A.; Hartman, I.Z.; Poore, B.; Boronina, T.; Cole, R.N.; Song, N.; Ciudad, M.T.; Caspi, R.R.; Jaraquemada, D.; Sadegh-Nasseri, S. Divergent paths for the selection of immunodominant epitopes from distinct antigenic sources. *Nat Commun* **2014**, *5*, 5369, doi:10.1038/ncomms6369.

47. Egger, M.; Jurets, A.; Wallner, M.; Briza, P.; Ruzek, S.; Hainzl, S.; Pichler, U.; Kitzmuller, C.; Bohle, B.; Huber, C.G., *et al.* Assessing protein immunogenicity with a dendritic cell line-derived endolysosomal degradome. *PLoS One* **2011**, *6*, e17278, doi:10.1371/journal.pone.0017278.

48. Delamarre, L.; Couture, R.; Mellman, I.; Trombetta, E.S. Enhancing immunogenicity by limiting susceptibility to lysosomal proteolysis. *J Exp Med* **2006**, *203*, 2049-2055, doi:10.1084/jem.20052442.

49. Machado, Y.; Freier, R.; Scheiblhofer, S.; Thalhamer, T.; Mayr, M.; Briza, P.; Grutsch, S.; Ahammer, L.; Fuchs, J.E.; Wallnoefer, H.G., *et al.* Fold stability during endolysosomal acidification is a key factor for allergenicity and immunogenicity of the major birch pollen allergen. *J Allergy Clin Immunol* **2016**, *137*, 1525-1534, doi:10.1016/j.jaci.2015.09.026.

50. Freier, R.; Dall, E.; Brandstetter, H. Protease recognition sites in Bet v 1a are cryptic, explaining its slow

processing relevant to its allergenicity. *Scientific Reports* **2015**, *5*, 12707, doi:10.1038/srep12707.

51. Kitzmuller, C.; Wallner, M.; Deifl, S.; Mutschlechner, S.; Walterskirchen, C.; Zlabinger, G.J.; Ferreira, F.; Bohle, B. A hypoallergenic variant of the major birch pollen allergen shows distinct characteristics in antigen processing and T-cell activation. *Allergy* **2012**, *67*, 1375-1382, doi:10.1111/all.12016.

2

IDENTIFICATION OF FORMALDEHYDE-INDUCED MODIFICATIONS IN DIPHTHERIA TOXIN

Bernard Metz [1,2], Thomas J.M. Michiels [1,3], Joost Uittenbogaard [1], Maarten Danial [1], Wichard Tilstra [1], Hugo D. Meiring [1], Wim E. Hennink [2], Daan J. A. Crommelin [2], Gideon F. A. Kersten [1,3], Wim Jiskoot [2, 3]

[1] Intravacc, Bilthoven, The Netherlands

[2] Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences (UIPS), Faculty of Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands

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

J Pharm Sci **2019**, *109* (1), 543-557.

Abstract

Diphtheria toxoid is produced by detoxification of diphtheria toxin with formaldehyde. This study was performed to elucidate the chemical nature and location of formaldehyde-induced modifications in diphtheria toxoid. Diphtheria toxin was chemically modified using four different reactions with the following reagents: (***1***) formaldehyde and $NaCNBH_3$, (***2***) formaldehyde, (***3***) formaldehyde and $NaCNBH_3$ followed by formaldehyde and glycine, and (***4***) formaldehyde and glycine. The modifications were studied by SDS-PAGE, primary amino group determination and LC-MS of chymotryptic digests. Reaction ***1*** resulted in quantitative dimethylation of all lysine residues. Reaction ***2*** caused intramolecular cross-links, including the NAD^+-binding cavity and the receptor-binding site. Moreover, A-fragments and B-fragments were cross-linked by formaldehyde on part of the diphtheria toxoid molecules. Reaction ***3*** resulted in formaldehyde-glycine attachments, including in shielded areas of the protein. The detoxification reaction typically used for vaccine preparation (reaction ***4***) resulted in a combination of intramolecular cross-links and formaldehyde-glycine attachments. Both the NAD^+-binding cavity and the receptor-binding site of diphtheria toxin were chemically modified. Although $CD4^+$ T-cell epitopes were affected to some extent, one universal $CD4^+$ T-cell epitope remained almost completely unaltered by the treatment with formaldehyde and glycine.

Introduction

Diphtheria and tetanus toxoids are very effective vaccine antigens, virtually eliminating associated diseases in vaccinated populations. Many countries have included these vaccines in their national immunization programs, which has drastically reduced the incidence and severity of diphtheria and tetanus [1]. Currently, a dozen companies around the world are producing diphtheria and tetanus vaccines. These toxoid vaccines were developed almost a century ago [2-4]. As a result, extensive data sets have been collected by companies and official medicines control laboratories to assure the quality of successive vaccine lots [5, 6]. Much experience with toxoid vaccines has been acquired by using the traditional potency and safety tests in animals. However, although the formaldehyde-induced chemical modifications of the antigen largely determine the quality of toxoid vaccines, little is known about the molecular structure of toxoids.

Interest has arisen in structural characterisation of diphtheria, tetanus and *Clostridium difficile* toxins [7-12] and toxoids [13-26]. Three reasons for this increased attention can be deduced from the literature, including (i) the development of analytical tests for quality control of toxoid-containing vaccines, (ii) the development of advanced vaccine formulations for toxoid vaccines and (iii) the use of toxoids as carrier proteins in conjugate vaccines:

(i) Several biophysical and immunochemical tests have been developed to assess the quality of toxoid-containing vaccines [16, 18, 19, 21, 23, 25, 27-37]. These tests are valuable for examination of the purity, safety, potency and stability of a product to reveal the impact of changes in a production process [38, 39].

(ii) Furthermore, the development of innovative vaccine formulations with diphtheria and tetanus toxoids has demanded detailed structural characterisation of the toxoids [24, 26, 40-43]. For example, controlled release or oral formulations with tetanus toxoid were designed by using polyester or chitosan microspheres [20, 42, 44-47]. Often, antigen instability in new vaccine delivery systems has been observed [15, 48-50].

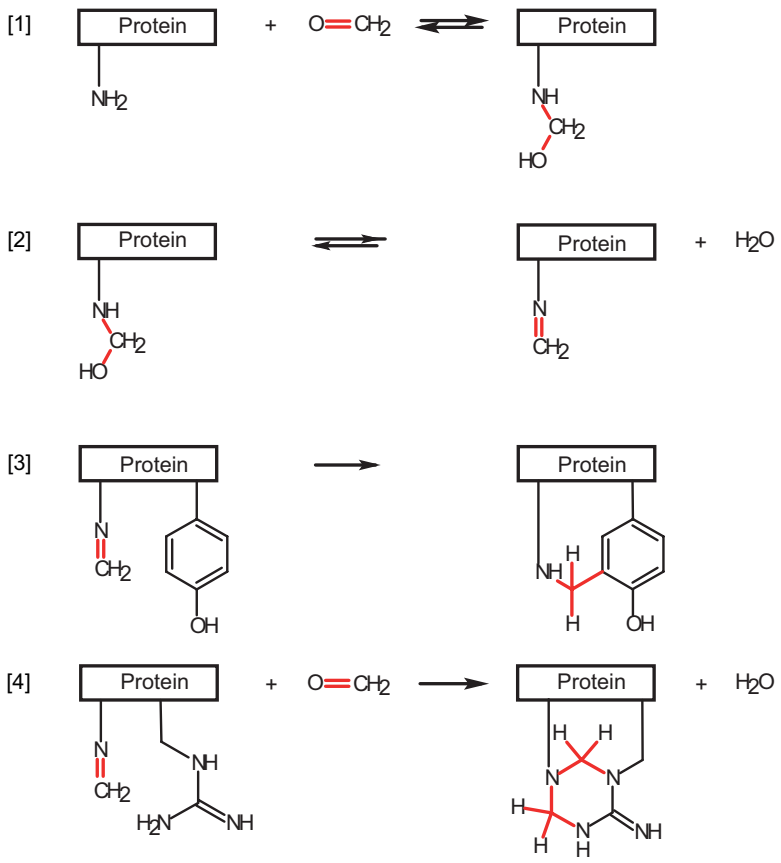
(iii) A third reason for renewed interest is the use of tetanus toxoid, diphtheria toxoid and a mutant of diphtheria toxoid (CRM_{197}) as carrier proteins in the production of polysaccharide conjugate vaccines. Multiple conjugate vaccines are available on the market [51]. Furthermore, several new conjugate vaccines are under development, *e.g.*, vaccines against *Shigella flexneri* and *Salmonella typhi* [52, 53].

Altogether, insight into the molecular structure of toxoids may help to improve the quality control of vaccines, support rational development of advanced vaccine formulations and facilitate the development of new conjugate vaccines.

Formaldehyde-mediated detoxification is an important step in the production of toxoid vaccines. To initiate the detoxification process, formaldehyde and extra glycine are usually added to diphtheria toxin-containing culture supernatant, which contains amino acids and metabolites. Formaldehyde treatment changes the toxicity, antigenicity and immunogenicity of diphtheria and tetanus toxoid [16,18,54]. Formaldehyde treatment causes chemical modifications in toxoid molecules. Remarkably, the chemical nature of each specific modification and the exact locations of the modified residues within diphtheria toxoid molecules have not yet been identified. For diphtheria toxin, formaldehyde converts the protein into a non-toxic toxoid, probably by permanently altering critical domains in the protein, *e.g.*, the catalytic site (NAD^+-binding cavity) and the receptor-binding site. Although detoxification causes the loss of some B-cell epitopes, the toxoid remains very immunogenic and induces a protective immune response by the generation of toxin-neutralising antibodies [16].

The reaction of formaldehyde with a protein starts with the formation of reversible methylol adducts on amino groups (Scheme 1). The methylol groups are partially dehydrated, yielding labile Schiff-bases. These Schiff-bases generate intramolecular cross-links with accessible, reactive amino acid residues, including arginine, asparagine, glutamine, histidine, tryptophan and tyrosine. Furthermore, formaldehyde can attach amino acids in solution to these reactive amino acid residues [55]. The conversion of reactive amino acid residues depends on their intrinsic reactivity and their accessibility for formaldehyde [55, 56].

The aim of the present study was to elucidate the chemical modifications in diphtheria toxoid as a result of the detoxification by formaldehyde and glycine. We investigated the reactivity of the individual lysine residues with formaldehyde, the formation of intramolecular cross-links and the attachment of formaldehyde-glycine moieties in diphtheria toxoid. Special emphasis was put on modifications in crucial parts of the molecule, *i.e.*, the catalytic site (NAD^+-binding



Scheme 1. The reaction of formaldehyde with proteins. The reaction 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 6 different amino acid residues, *e.g.*, with tyrosine [3] and arginine [4].

cavity), the receptor-binding site and $CD4^+$ T-cell epitope regions. The NAD^+-binding cavity is located in the catalytic domain of diphtheria toxin that transfers the ADP-ribose moiety of NAD^+ to elongation factor-2 (EF-2) [57]. The modification of EF-2 irreversibly inhibits the protein synthesis in the host cell leading to cell death. Three short peptide sequences in the diphtheria toxin molecule form the NAD^+-binding cavity, a loop from the residues 17 – 23, a β-strand followed by an α-helix from residues 50 – 67, and a β-strand from residues 147 – 150 (Figure 1A). Amino acid residues His^{21}, Tyr^{54}, Tyr^{65} and Glu^{148} participate in the binding of NAD^+ [58, 59]. Another important area in the toxin molecule is the receptor-binding site, which is formed by a loop of amino acid residues 511 – 530. This part of the receptor domain binds

to the heparin-binding epidermal growth factor-like precursor [60, 61]. The residues Tyr[514], Lys[516], Val[523], Asn[524], Lys[526] and Phe[530] participate in binding to the host cell receptor [60]. The crystal structure of the receptor-bound diphtheria toxin complex is elucidated (Figure 1B) [62]. Also, CD4$^+$ T-cell epitopes are identified by using blood from healthy subjects [63]. The CD4$^+$ T-cell epitopes are located in the B-fragment of diphtheria toxin: α-helices formed by residues 271 – 290, 321 – 340 and 331 – 350.

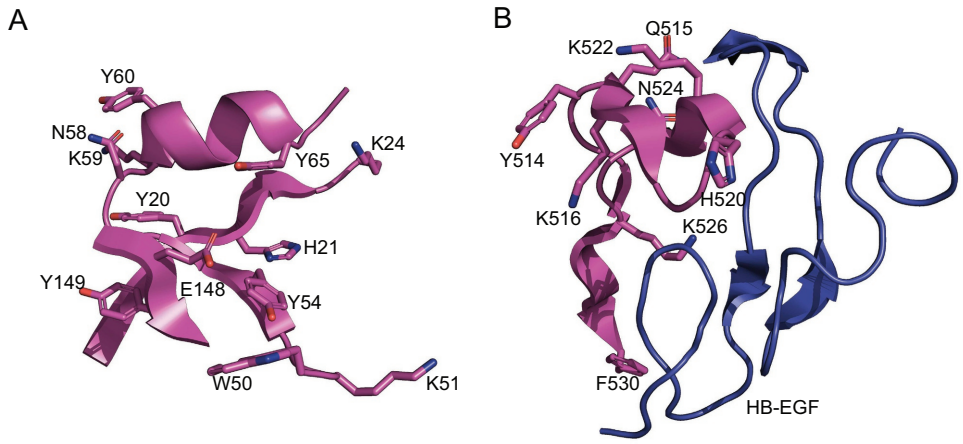


Figure 1. Catalytic and receptor binding sites in diphtheria toxin. The images represent two functional sites in diphtheria toxin: (A) the catalytic site (NAD+-binding cavity; PDB: 1TOX) and (B) the receptor-binding site (PDB: 1XDT) interacting with the binding site of the cell-surface receptor (HB-EGF). Searching for PDB codes was performed on https://www.rcsb.org/. The side chains are shown of those amino acid residues of diphtheria toxin that are potentially reactive with formaldehyde [55].

The NAD$^+$-binding cavity, the receptor-binding site and CD4+ T-cell epitopes contain formaldehyde reactive residues. The formaldehyde-induced modifications in NAD$^+$-binding cavity and the receptor-binding site of diphtheria toxin are probably responsible for complete detoxification. However, modifications in CD4$^+$ T-cell epitope regions of diphtheria toxin might reduce the immunogenicity of diphtheria vaccines. The chemical modifications in diphtheria toxin were analysed by SDS-PAGE, primary amino group determination (TNBS assay) and liquid chromatography – electrospray mass spectrometry (LC-MS) after digestion with chymotrypsin. In this study, we revealed the location and chemical nature of modifications that occur in diphtheria toxoid during a detoxification process by formaldehyde and glycine.

Materials and methods

Chemicals

Formaldehyde (37%) (CH_2O), Formaldehyde-D_2 (20%), (CD_2O), glycine, sodium cyanoborohydride ($NaCNBH_3$), sodium bisulphite, triethyl ammonium bicarbonate (TEAB) buffer pH 8.5 and dimethyl sulfoxide (DMSO) were obtained from Sigma Aldrich (Schnelldorf, Germany). Formic acid (99%), Sodium dihydrogen phosphate (NaH_2PO_4), disodium hydrogen phosphate (Na_2HPO_4) and sodium chloride were purchased from Merck (Darmstadt, Germany). Chymotrypsin was bought from Roche Diagnostics (Mannheim, Germany) and acetonitrile from Biosolve (Valkenswaard, The Netherlands).

Chemical treatment of diphtheria toxin

Prior to detoxification reactions, diphtheria toxin-containing culture supernatant (Intravacc, The Netherlands) was dialysed (MWCO 10 kDa; Slide-A-Lyzer Dialysis Cassette; Thermo Scientific; Rockford, IL, USA) extensively against PBS (0.15 M NaCl, 7.7 mM Na_2HPO_4 and 2.3 mMNaH_2PO_4, pH 7.2). Aqueous solutions of formaldehyde (CH_2O), deuterium-labelled formaldehyde (CD_2O), glycine and $NaCNBH_3$ were prepared at a concentration of 1.0 M in water. Four reactions (Figure 2) with diphtheria toxin were performed: (Reaction ***1***) diphtheria toxin with formaldehyde and $NaCNBH_3$, (Reaction ***2***) diphtheria toxin with formaldehyde (CH_2O or CD_2O), (Reaction ***3***) diphtheria toxin with formaldehyde (CH_2O) and $NaCNBH_3$ followed by the reaction with formaldehyde (CH_2O or CD_2O) and glycine, (Reaction ***4***) diphtheria toxin with formaldehyde (CH_2O or CD_2O) and glycine. The composition and conditions for reaction ***2*** are comparable to those used for vaccine production [16].

Reaction ***1*** – Formaldehyde (CH_2O) and $NaCNBH_3$ were added to diphtheria toxin (1.2 mg/ml) to final concentrations of 50 mM. After mixing, the solution was incubated for 2 h at 35 °C. Then, the sample was extensively dialysed against 10 mM PBS pH 7.2 (MWCO 10 kDa).

Reaction ***2*** – Formaldehyde (CH_2O or CD_2O) was added to diphtheria toxin (1.2 mg/ml) to final concentrations of 80 mM . After mixing, the solutions were incubated for one week at 37 °C. Then, the reaction was stopped by adding sodium bisulphite to a final concentration of 80 mM and subsequently extensively dialysed against PBS (MWCO 10 kDa). Sodium bisulphite reacted with free formaldehyde in solution ($Na_2HSO_3 + CH_2O \rightleftharpoons HOCH_2SO_3Na$). As a result,

the reversible methylol groups and Schiff-bases on diphtheria toxoid were largely removed, because the equilibrium shifted to the left (protein-NH_2 + CH_2O ⇌ protein-$NHCH_2OH$ ⇌ protein-NCH_2).

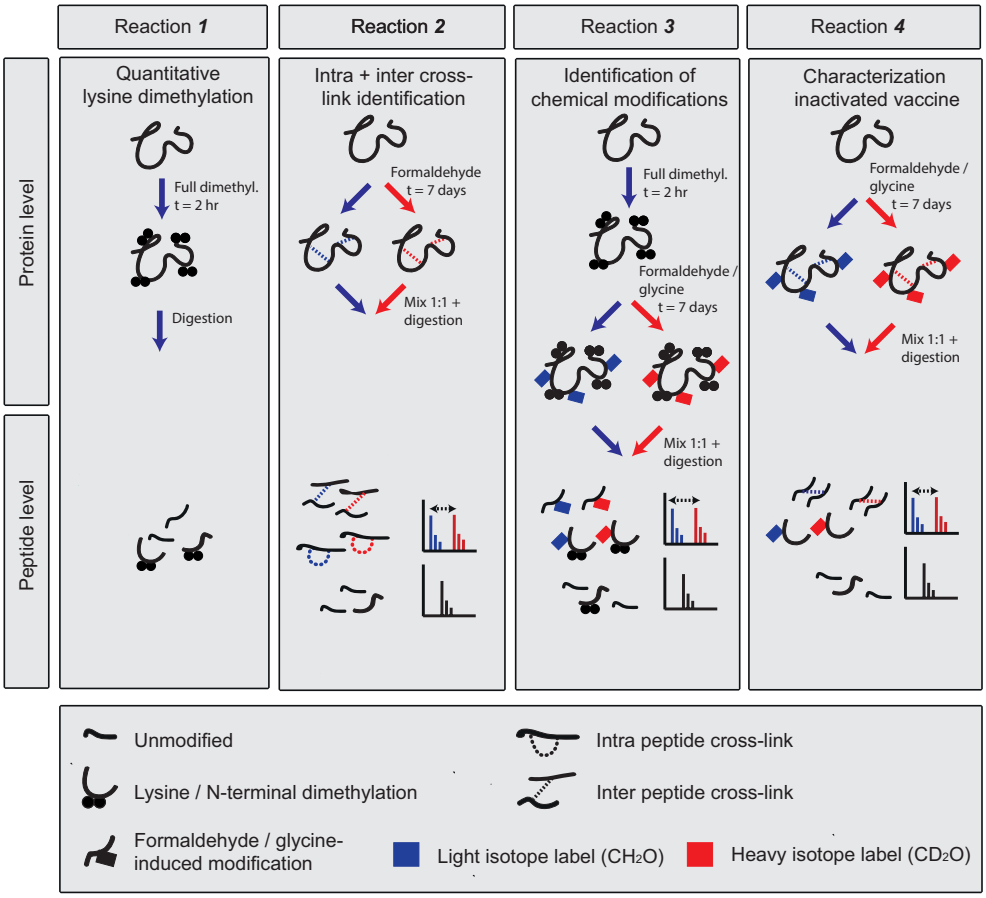


Figure 2. Characterization of formaldehyde-induced modifications in diphtheria toxoid by using stable isotope labelling strategies. Reaction ***1***: The accessibility of the lysine residues in diphtheria toxin after reductive dimethylation and digestion was determined by LC-MS/MS and database search analysis. Reaction ***2***: Target identification of cross-links in diphtheria toxoid formed upon formaldehyde treatment. Treatment with light (CH_2O) and heavy (CD_2O) formaldehyde resulted in intramolecular cross-links that appear as mass spectral doublets, while unmodified peptides appear as singlets. After digestion, the formaldehyde-modified peptides contain a cross-link within the peptide (intra-peptide cross-link) or between two peptides (inter-peptide cross-link). Reaction ***3***: Identification of modified residues by glycine attachments. First, lysine residues were blocked to prevent intra-peptide and inter-peptide cross-linking. As a result of treatment with light (CH_2O) or heavy (CD_2O) formaldehyde and glycine, chemically modified peptides appear as mass spectral doublets, while unmodified peptides appear as singlets. Reaction ***4***: The standard detoxification of diphtheria toxin. The protein is treated with light (CH_2O) or heavy (CD_2O) formaldehyde and glycine, resulting in the differential labelling of both formaldehyde-induced cross-links and formaldehyde-glycine attachments.

Reaction ***3*** – Diphtheria toxin (1.2 mg/ml) was incubated for 2 h at 37 °C with formaldehyde (CH_2O) and sodium cyanoborohydride. The final concentrations of formaldehyde and sodium cyanoborohydride in the samples were 50 mM. The sample was extensively dialysed against PBS. The obtained dimethylated diphtheria toxin was incubated for 1 week at 37 °C with formaldehyde (CH_2O or CD_2O) and glycine. The final concentrations of formaldehyde and glycine in the samples were 80 mM and the concentration of diphtheria toxin was 1.2 mg/ml. After the incubation, sodium bisulphite was added to a final concentration of 80 mM. Samples were extensively dialysed against PBS (MWCO 10 kDa).

Reaction ***4*** – Diphtheria toxin (1.2 mg/ml) was incubated for 1 week at 37 °C with formaldehyde (CH_2O or CD_2O) and glycine. The final concentrations of formaldehyde and glycine in the samples were 80 mM. The reaction was stopped by adding sodium bisulphite to a final concentration of 80 mM. Both samples were extensively dialysed against PBS (MWCO 10 kDa).

Finally, all samples were stored at 4 °C prior to analysis by SDS-PAGE, TNBS assay and LC-MS.

SDS-PAGE

SDS-PAGE was performed under reducing conditions, essentially as described by Sambrook *et al.* [64]. Protein samples were prepared by mixing 2 μg of the toxoid in the sample buffer (60 mM Tris, 70 mM SDS, 0.1 M dithiothreitol, 0.1 mM tetrabromophenol blue and 35% glycerol diluted in water) to a volume of 20 μl, and boiled for 10 min to denature the protein and to reduce disulphide bridges. The samples were loaded onto 12 % SDS-PAGE gels and electrophoretically separated. Molecular weight reference (broad range; Bio-Rad) was used for calibration. Protein bands were visualised by using Imperial Protein Stain (Pierce). The gels were scanned, and the intensity of protein bands was quantified using ImageJ 1.46r software (NIH, USA).

Protein assay

The protein concentration of dialysed diphtheria toxin or toxoid samples was determined by using the BCA protein assay according to the manufacturer's description (Thermo Fisher Scientific, The Netherlands).

TNBS assay

The concentration of primary amino groups present in diphtheria toxin or toxoid samples was determined by using a colorimetric assay with 2,4,6-trinitrobenzene sulphonic acid (TNBS) [65]. A reference (10-100 μM) was prepared from a stock solution of 1.0 mM glycine. The number of primary amino groups in a diphtheria toxin or toxoid molecule was calculated; primary amino groups (mM) / protein concentration (mM). The molecular mass of 58.3 kDa for the diphtheria toxin was used for the calculation.

Digestion by chymotrypsin

Diphtheria toxin and toxoids (from reaction ***1***, ***2, 3*** and ***4***) were individually digested by chymotrypsin (diphtheria toxin was used as a control.) To that end, 5 μl of a 1-M TEAB buffer, pH 8.5, and 2 μl of 1.0 mg/ml chymotrypsin were added to each sample containing 0.1 mg toxin or toxoid. Water was added to a final volume of 100 μl. Samples were incubated for 16 h at 37 °C. Subsequently, the samples were stored at-20 °C before LC-MS analysis.

LC-MS

Protein digests were analysed by nanoscale reversed phase liquid chromatography electrospray mass spectrometry, essentially as previously described by Meiring *et al.* [66], using an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific). The digests of reaction products ***1***, ***2***, ***3*** and ***4*** were mixed in equal amounts and diluted in water containing 5% (v/v) DMSO and 0.1% (v/v) formic acid to a concentration corresponding to 1.0 μM of the original protein concentration. An injection volume of 10 μl was used for analysis. Analytes were loaded on a trapping column (Reprosil-Pur C18-AQ 5 μm, Dr. Maisch GmbH, Germany); 20 mm long × 100 μm inner diameter) with solvent A (0.1% (v/v) formic acid in water) in 10 min at 5 μL/min. The analytes were separated by reversed phase chromatography on an analytical column (Reprosil-Pur C18-AQ 3 μm, Dr. Maisch GmbH, Germany); 27.5 cm long × 50 μm inner diameter) at a flow rate of 125 nL/min. A gradient was started with solvent B (0.1% (v/v) formic acid in acetonitrile): 8% to 34% in 65 min and 60% for 5 min. After the gradient, the columns were equilibrated in 100% solvent A for 10 min at 125 nL/min. The peptides were measured by data-dependent acquisition, comprising a MS scan (m/z 300–1500) in the orbitrap with a resolution of 120 000 (FWHM), followed by collision-induced dissociation (CID; ion trap) at top speed with a cycle time of 3 sec. The threshold value for these precursor ions was set at 25 000 counts. The normalised collision energy was set at

35%, the isolation width at 1.6 Da and the activation Q to 0.250. The maximum ion injection time for MS scans was set to 50 ms and for MS/MS scans to 150 ms. Precursor ions with +2 to +5 charge states were selected for MS/MS analysis. Dynamic exclusion was enabled (exclusion list with 500 entries) with repeat set to 1 and an exclusion duration of 45 s. The electron-transfer dissociation (ETD) reagent cation (202.0777) was used for internal mass calibration.

Peptides containing formaldehyde modifications typically appeared as mass spectral doublets as a result of the use of ”light” (CH_2O) and ”heavy” (CD_2O) formaldehyde. The doublets were retrieved from the mass spectra by using the software program MsXelerator (MsMetrix, Maarssen, The Netherlands). Samples were measured in triplicate. Doublets that were found with a relative intensity of 10^5 arbitrary units and present in at least two of the three replicates were selected for further evaluation. The obtained doublet lists were exported to text files and used as parent mass lists for targeted ETD and CID fragmentation. For ETD fragmentation, charge state dependent ETD parameters were used. For CID fragmentation the same settings were used as described above. Identification of the MS/MS spectra was performed with Peaks Studio 8.5 (Bioinformatics Solutions, Waterloo, ON, Canada) against the *Corynebacterium diphtheriae* proteome (taxonomic identifier 257309, 2267 entries) with the earlier-described formaldehyde modifications [55] as variable modifications. Mass spectra that were not automatically assigned to modified peptides were manually evaluated based upon the observed mass and the number of incorporated formaldehyde molecules.

***In silico* calculation of the solvent accessible surface area**

The *in silico* calculation of the solvent accessible surface area requires knowledge on the solvent radius and assumes that the solvent molecules are spheres. As such, the Connolly Molecular Area1 of the solvent molecule in Chem 3D Pro v11.0 was determined, followed by rearranging the equation for the area of a sphere:

$$r_{\text{solvent}} = \frac{1}{2}\sqrt{\frac{A}{\pi}} \quad (1)$$

where A is the Connolly Molecular Area and r_{solvent} is the molecular solvent radius. As presented in Supplementary Table 1, the value for r_{solvent} depended on the type of modification performed.

Table 1. Expected formaldehyde-induced cross-links in diphtheria toxin.

Position	Sequence	Distance reactive sites (Å)[1]	α-Carbon Distance (Å)[2]	Identified
	K – R cross-links			
K90 – R133	KVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIKRFGDGASR	13.2	6.1	no
K125 – R126	KR	10.7	3.8	yes
R170 – K172	RGK	8.9	5.6	yes
K172 – R173	KR	6.8	3.8	yes
K172 – R210	KRGQDAMYEYMAQACAGNRVRRSVGSSLSCINLDWDVIR	7.5	10.9	yes
R210 – K214	RDKTK	7.0	6.6	no
R407 – K534	RTGFQGESGHDIKITAENTPLPIAGVLLPTIPGKLDVNKSKTHISV NGRKIRMRCRAIDGDVTFCRPKSPVYVGNGVHANLHVAFHRS SSEKIHSNEISSDSIGVLGYQKTVDHTKVNSKLSLFFEIK	6.7	4.8	yes
K440 – R493	KLDVNKSKTHISVNGRKIRMRCRAIDGDVTFCRPKSPVYVGNGV HANLHVAFHR	10.7	5.5	no
K445 – R462	KSKTHISVNGRKIRMRCR	5.3	9.0	no
R493 – K498	RSSSEK	12.7	9.7	yes
	K – Y cross-links			
K24- Y27	KPGY	8.6, 10.4	5.8	yes
K24- Y65	KPGYVDSIQKGIQKPKSGTQGNYDDDWKGFYSTDNKYDAAGY	7.1, 8.4	7.5	no
K51- Y54	KGFY	10.0, 12.0	9.2	no
K59- Y60	KY	4.6, 5.3	3.8	yes
K59- Y181	KYDAAGYSVDNENPLSGKAGGVVKVTYPGLTKVLA LKVDNAETIKKELGLSLTEPLMEQVGTEEFIKRFGDGASR VVLSLPFAEGSSSVEYINNWE QAKALSVELEINFETRGKRGQDAMYEY	5.5, 6.6	9.4	no
K214- Y358	KTKIESLKEHGPIKNKMSESPNKTVSEEKAKQYLEEFHQTALEHPELS ELKTVTGTNPVFAGANYAAWAVNVAQVIDSETADNLEKTTAALSILP GIGSVMGIADGAVHHNTEEIVAQSIALSSLMVAQAIPLVGELV DIGFAAY	6.5, 8.5	10.0	no
Y380- K385	YSPGHK	5.7, 6.7	7.0	no
Y514- K516	YQK	4.6, 6.0	6.5	yes
	K – W cross-links			
W153- K157	WEQAK	10.6	6.8	yes

[1] Distance measured from N^{ε} atom on Lys to one or more *N* atoms of the guanidine group on Arg, *meta* C-H bond on Tyr, or to *N* atom on the indol group of Trp.

[2] Distance measured between the α-carbon atoms between two crosslinked residues.

The solvent accessible surface area (SASA) of monomeric diphtheria toxin (open form, PDB code: 1TOX [67]) was calculated by using the surface area per residue computation (relative solvent accessibility) in Pymol v. 2.1.1 (Schrodinger LLC), by setting the solvent radius to the values corresponding to formaldehyde or methylene-modified glycine (Supplementary Table 1). The generated residue – accessibility list was then exported to GraphPad Prism in which the plots were generated.

A solvent accessibility surface area calculation was also performed on the diphtheria toxin modified via *N,N*-dimethylation at the ε-amine group of all lysine residues (reaction ***1***). This was done by introducing the *N,N*-dimethylation of 1TOX PDB using the PyTM python script [68]. Following this modification, the surface accessibility surface area was calculated by using formaldehyde as solvent, as described above.

Secondary structure information

The secondary structure information added to the plots is based on the crystal structure of diphtheria toxin (PDB code 1TOX [67]).

Results and discussion

Production of diphtheria toxin and toxoid

Diphtheria toxoid is usually prepared by adding formaldehyde and particular amino acids to the culture supernatant after the cultivation of *Corynebacterium diphtheriae*. Besides high concentrations of diphtheria toxin, the culture supernatant contains other proteins produced or secreted by the bacteria. Diphtheria toxin is synthesised as a single protein but probably has to be nicked into an A and B fragment to exert full biological activity [69].

The dialysed toxin batch used in this study was examined by SDS-PAGE and mass spectrometry to determine its purity. SDS-PAGE demonstrated that diphtheria toxin used in this study was almost completely nicked (Figure 3, lane *0*). It showed a tiny band of the intact toxin at 58 Da and intense bands of the A and B fragment at 21 and 37 Da, respectively. The purity of the toxin was quantified by SDS-PAGE and appeared to be above 90%. To identify the proteins in the culture supernatant, chymotrypsin-digested material was analysed by LC-MS. Thirty-one different proteins were identified in the concentrated culture supernatant (Supplementary

Table 2). Based on the average response of the three most intensive peptides from each protein, the purity of the diphtheria toxin solution was estimated [70]. According to the LC-MS analysis, the culture supernatant contained 93 mole% diphtheria toxin. The contents of other proteins were between 0.05 and 1.3 mole%. The purity of diphtheria toxin was 96% based on protein weight (Supplementary Table 2). This purity was in line with the results from SDS-PAGE. The dialysed culture supernatant was used to study in detail the chemical modifications of diphtheria toxin after formaldehyde treatment.

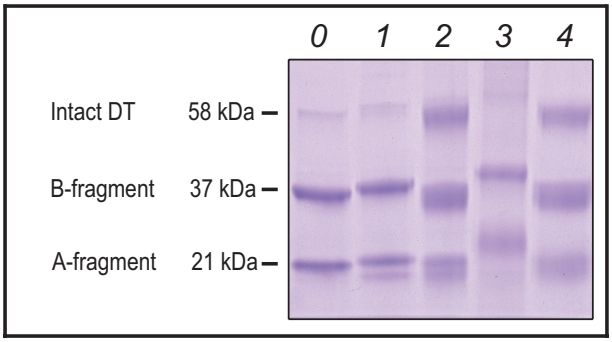


Figure 3. SDS-PAGE of diphtheria toxin (lane *0*) and four experimental toxoids. The diphtheria toxoids were prepared by reaction ***1*** (lane *1*), reaction ***2*** (lane *2*), reaction ***3*** (lane *3*), and reaction ***4*** (lane *4*) (see Materials & Methods for details).

Formaldehyde-reactive lysine residues

The modifications in diphtheria toxoid after formaldehyde and glycine treatment consist of intramolecular cross-links and formaldehyde-glycine attachments. The intramolecular cross-links occur between a lysine residue and a susceptible amino acid residue, *i.e.*, arginine, asparagine, glutamine, histidine, tryptophan or tyrosine [55]. To determine the accessibility of each individual lysine residue for formaldehyde, diphtheria toxin was treated with formaldehyde and $NaCNBH_3$ (reaction ***1***). In this reaction, the primary amino groups of lysine and N-terminal residues are converted to dimethylated structures with a mass increment of 28 Da [71]. Intact diphtheria toxin has 40 primary amino groups, whereas the toxin in the nicked form has one additional primary amino group.

SDS-PAGE showed three protein bands of the formaldehyde-treated diphtheria toxin (reaction ***2***) with slightly increased masses when compared to the untreated toxin as a result of the dimethylation (Figure 3, lane *1*). Unexpectedly, a fourth protein band was observed

at about 20 kDa. This 20-kDa band may represent some alternate fragmentation and not an impurity, as the diphtheria toxin preparation was highly pure (Supplementary Table 2). Another possibility might be that a formaldehyde-induced cross-link was introduced that caused a higher electrophoretic mobility due to incomplete unfolding by SDS-PAGE. However, no detailed information was collected in this study on the nature of this 20-kDa fragment.

The TNBS assay showed the presence of 29 primary amino groups in untreated diphtheria toxin (Figure 4), which is less than the expected value (39 lysine residues and 2 N-termini). Apparently, the other primary amine groups in diphtheria toxin were not reached by TNBS and/or the ε-amino groups of the lysine residues were slightly less reactive with TNBS than the α-amino groups of the glycine reference. Treatment of diphtheria toxin with formaldehyde and $NaCNBH_3$ (reaction ***1***) resulted in a drastic reduction (93%) of the number of primary amino groups (Figure 4). On average, 2 primary amino groups were present in diphtheria toxoid after reaction ***1***. This suggests that most lysine and N-terminal residues were accessible and modified by formaldehyde and $NaCNBH_3$ (reaction ***1***).

Furthermore, the modified lysine residues in diphtheria toxoid were identified by LC-MS analyses after digestion with chymotrypsin. Ninety-five percent (94.6%) of the total primary sequence of diphtheria toxin, including the complete NAD^+-binding and receptor-binding sites, was identified by LC-MS. Dimethylation (ΔM = +28 Da) was observed for all lysine residues present in diphtheria toxin. The conversion into dimethylated lysine residues, calculated based on data of the five most abundant lysine-containing peptides, was on average 99.6 ± 0.3%. In conclusion, all lysine residues were accessible for formaldehyde-induced modifications, which can lead to cross-links with other reactive amino acid residues in diphtheria toxin.

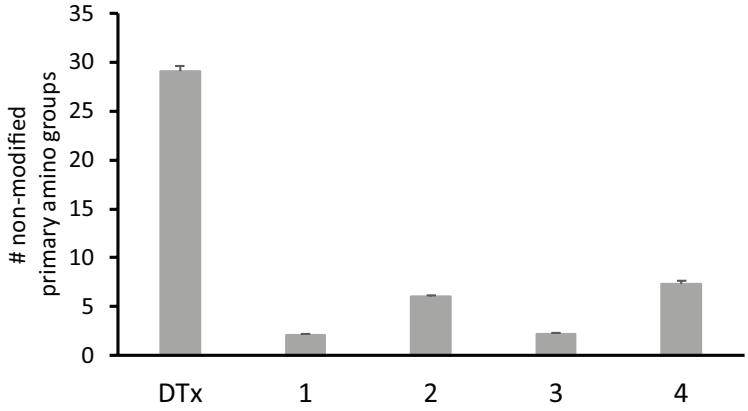


Figure 4. Average number of primary amino groups in diphtheria toxin (DTx) and several experimental toxoids (mean ± S.D.; n = 9). The modified toxoids were prepared by reactions ***1***, ***2***, ***3*** and ***4*** (see Materials & Methods for details).

Intramolecular cross-links in diphtheria toxin

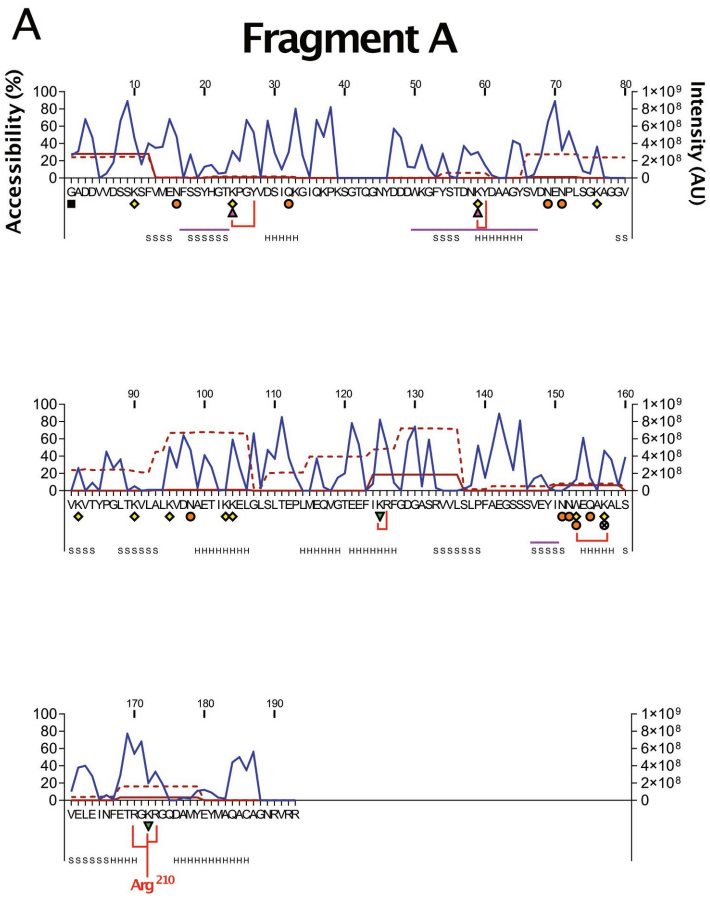
During vaccine production, diphtheria toxin is usually treated with formaldehyde in the presence of amino acids. In theory, 142 of the 535 residues in diphtheria toxin can react with formaldehyde (N-terminal, Arg, Asn, Gln, His, Lys, Tyr and Trp residues). Since many residues are partly converted, diphtheria toxoid may consist of a plethora of different reaction products. The stable formaldehyde-induced modifications can be divided into intramolecular cross-links and attachments of amino acids. In this study, two distinct reactions were performed to simplify the assignment of the formaldehyde-induced modifications in diphtheria toxoid, to induce either intramolecular cross-links (reaction ***2***) or formaldehyde-glycine attachments (reaction ***3***). In reaction ***2***, diphtheria toxin was treated with formaldehyde (CH_2O) or with deuterated formaldehyde (CD_2O) to introduce only intramolecular cross-links in the protein. The formaldehyde-glycine attachments (reaction ***3***) are described in the next section '*Formaldehyde-glycine attachments in diphtheria toxoid*'.

SDS-PAGE (Figure 3; lane *2*) performed with formaldehyde-treated diphtheria toxoid (reaction ***2***) showed four broadened protein bands of the apparently intact toxoid (59 kDa), the A-fragment (21 kDa) and the B-fragment (36 kDa). Although less clear than in lane *1* (Figure 2), a fourth protein band was observed in lane *2* at about 20 kDa. Furthermore, the protein bands were more spread out in lane *2* than in lane *0*. The broad bands indicate that the formaldehyde treatment resulted in a very heterogeneous product. The broadened

protein bands are probably caused by different intramolecular cross-links present in diphtheria toxoid molecules. Depending on the actually formed cross-links, diphtheria toxoid molecules were probably not completely unfolded by sodium dodecyl sulphate [16]. As a result, diphtheria toxoid and the fragments were visualised by SDS-PAGE as smeared protein bands. Furthermore, intermolecular cross-linking between toxoid molecules is unlikely, as indicated by the absence of any clear band larger than 59 kDa. However, the intensity of the apparently intact toxoid band (59 kDa) increased significantly compared to that of the toxin band (lane *0*). This observation can be explained by the formation of formaldehyde cross-links between the A-fragment and the B-fragment of diphtheria toxoid. The A-fragment and the B-fragment of diphtheria toxoid are in close proximity of each other, making the probability for formation of intramolecular cross-links by formaldehyde rather high. Intermolecular cross-links formed with small reactive molecules were not expected, because the toxin was extensively dialysed against PBS and only formaldehyde was added in reaction ***2***. In conclusion, the results deduced from SDS-PAGE indicate that cross-links mainly occur intramolecularly at these formaldehyde and protein concentrations (Figure 3; lane *2*).

In addition, the number of primary amino groups in diphtheria toxoid was drastically reduced after formaldehyde treatment (reaction ***2***). The TNBS assay revealed, on average, the presence of 6 primary amino groups in each diphtheria toxoid molecule.

Furthermore, LC-MS analysis was performed on chymotrypsin-digested mixtures of CH_2O-treated and CD_2O-treated toxoid. The use of CH_2O *vs.* CD_2O provided a method to discriminate formaldehyde-modified peptides (mass spectral doublets) from unmodified ones (mass spectral singlets). The analysis gave a protein coverage of about 95% for the whole sequence (96% of A-fragment and 95% of B-fragment). Sixty mass spectral doublets were observed with at least a relative intensity above 10^5 arbitrary units. Most peptide pairs had mass differences of 2 or 4 Da, indicating the presence of one or two formaldehyde-induced modifications, respectively (Supplementary Table 3). One spectral doublet had a mass difference of 6 Da. Fifty modified peptide sequences were assigned based on MS/MS data, their exact masses and the number of incorporated formaldehyde molecules. The formaldehyde-induced modifications in diphtheria toxoid (reaction ***2***) are described thereafter and depicted in Figure 5:



Modifications

- Cyclic N-terminus (+12.00)
- Crosslink K+Y (+12.00)
- Crosslink K+R (+24.00)
- Crosslink K+W (+12.00)
- Methylol (+30.01)
- Schiff base (+12.00)
- K+R, K+Y or K+W crosslink

Domains

- Catalytic domain
- Receptor domain
- T helper epitope

Chymotrypsin digestion

- Unmodified peptides
- Modified peptides

Secondary structure

- H α-helix
- S β-sheet

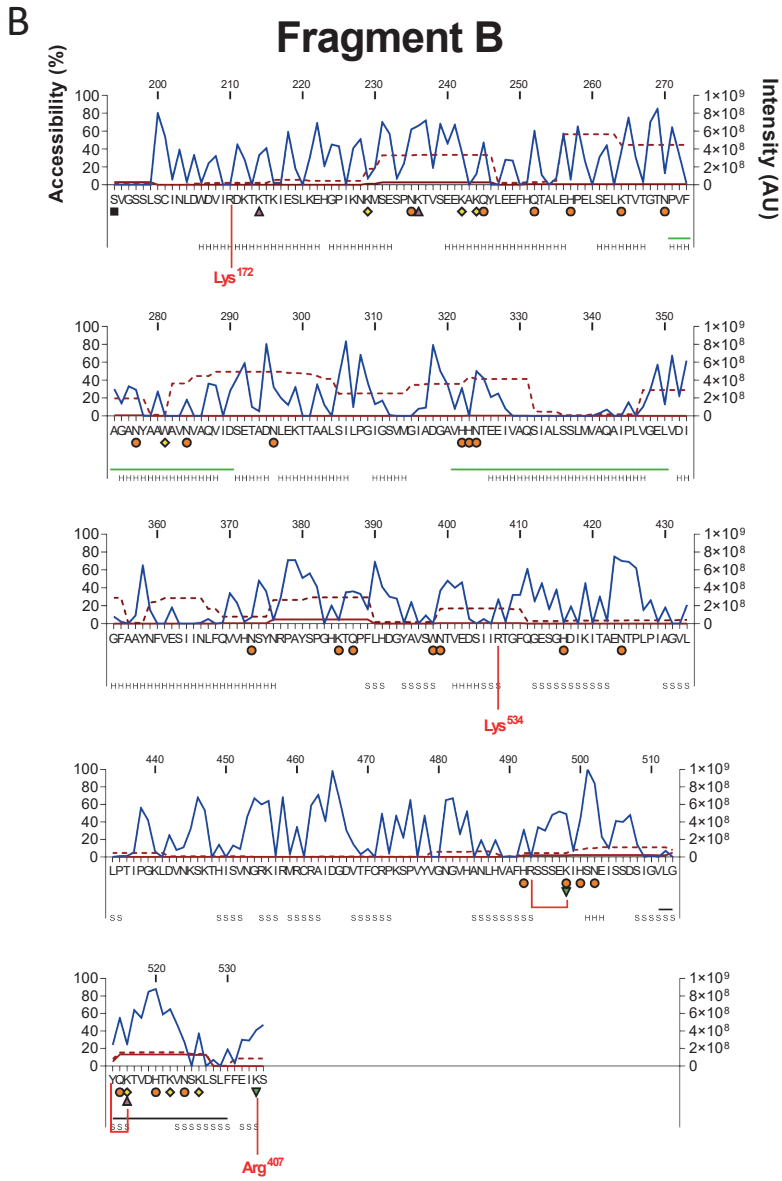


Figure 5. The modifications in diphtheria toxin caused by formaldehyde treatment (reaction 2). The blue line (y-axis left) indicates the theoretical solvent accessible surface area (accessibility) calculated using the molecular solvent radius of formaldehyde. The modified and odified peptides derived from chymotrypsin digestion are indicated with a solid and dotted dark red line, respectively of which the intensities are plotted (y-axis right). In the case of overlapping peptides, the intensities are summed for the overlapping amino acid positions. The primary structure is indicated on the x-axis along with the amino acid position number in increments of 10 residues at the top of each plot. The different amino acid modifications identified by MS analysis along with the secondary structure elements derived from the crystal structure PDB 1TOX. The catalytic site, receptor binding site and CD4+ T-cell epitope are indicated with purple, black and green underlines, respectively. The specific symbols are present in the figure legend. The red lines connect the residues between which a cross-link has been formed.

1) *4-imidazolididones* – In diphtheria toxoid, formaldehyde converted the N-termini of the A-fragment and B-fragment into cyclic products (Figure 5). These ring structures are also called 4-imidazolididones [55, 72, 73]. The adducts were revealed in the N-terminal peptides of the A-fragment (G^{1}ADDVVDSSKSF^{12}) and the B-fragment (S^{194}VGSSL^{199}), resulting in a mass increment of both peptides (ΔM = +12 Da).

2) *Methyl group* – One monomethylated lysine residue (K^{385}) was observed in diphtheria toxoid (reaction ***2***). Such modifications were observed previously in tetanus toxoid [17]. For this particular modification a mild reducing agent is needed to convert a Schiff-base into a monomethyl group. Probably, formic acid used prior to LC-MS analyses might act as a reducing agent (Eschweiler–Clarke reaction) [74]. Therefore, this modification was not depicted in Figure 5.

3) *Cross-links* – Seven unique peptides were identified that revealed formaldehyde-induced cross-links (ΔM = +24 Da) between lysine and arginine residues in diphtheria toxoid (Figure 5). Furthermore, eight MS/MS spectra indicated cross-links between lysine and a second amino acid residue, *i.e.*, a tyrosine and tryptophan residue. However, these cross-links could not be confirmed unambiguously, because Schiff-bases cause the same mass increment as an intramolecular cross-link (ΔM = +12 Da). Therefore, the digested peptides were treated with formaldehyde and $NaCNBH_3$ to confirm the presence of an intramolecular cross-link between lysine and tyrosine or between lysine and tryptophan residues. This reaction results in monomethylation of lysine residues in case of an intramolecular cross-link (ΔM = +12 Da (cross-link) +14 Da (methyl group) = +26 Da) and dimethylation in case of a Schiff-base (ΔM = +28 Da). The LC-MS analysis showed the existence of three intramolecular cross-links between a lysine and a tyrosine residue, and one between a lysine and a tryptophan residue (Supplementary Figure 1). In diphtheria toxoid, the formation of formaldehyde-induced cross-links between the A-fragment and the B-fragment was confirmed by SDS-PAGE. As such, we attempted to identify the amino acid residues that are likely to form these formaldehyde-induced cross-links. Based on the crystal structure of diphtheria toxin [7], two reactive residues of the A-fragment are in close proximity of reactive residues in the B-fragment of diphtheria toxin, *i.e.*, Lys^{172} – Arg^{210} and $Arg^{190,192,193}$ – Ser^{194}. The cross-link between residues Lys^{172} and Arg^{210} was confirmed by LC-MS (Figure 6). Contrarily, the presence of a cross-link

between one of three arginine residues (Arg[190,192,193]) and the N-terminal amino group of B-fragment (Ser[194]) was not found.

Cross-links were formed effectively when the distance between reactive residues was between 4.6 – 10.7 Å (Supplementary Figure 1). Within these distances, several other formaldehyde-induced cross-links could be expected, but were not identified. However, incomplete protein coverage, low concentrations or ionisation intensities of these modified peptides or a poor peptide fragmentation (CID or ETD), resulting in inconclusive MS/MS spectra which could lead to overseeing cross-link identifications. As such, the presence of expected cross-links cannot be ruled out.

4) *Schiff-bases* – Nineteen peptides from diphtheria toxoid containing a Schiff-base were identified. The Schiff-bases were located on 15 lysine residues, 2 tryptophan residues and 1 histidine residue. Schiff-bases were found despite bisulphite treatment and extensive dialysis, performed to revert the reversible Schiff-bases and methylol moieties in diphtheria toxoid. The Schiff bases and methylol groups were more stable than expected.

5) *Methylol groups* – Data analysis revealed the presence of 26 reversible methylol groups present in diphtheria toxoid (ΔM = +30 Da). The methylol groups were located on 17 distinct asparagine, 6 glutamine, 4 histidine, 2 lysine residues and 1 tryptophan residue.

Next to formaldehyde-induced modifications, oxidation of methionine and deamidation of asparagine residues were observed by LC-MS. However, the extent of oxidation and deamidation was not increased upon exposure to 37 °C for 7 days in reaction ***2*** (results not shown).

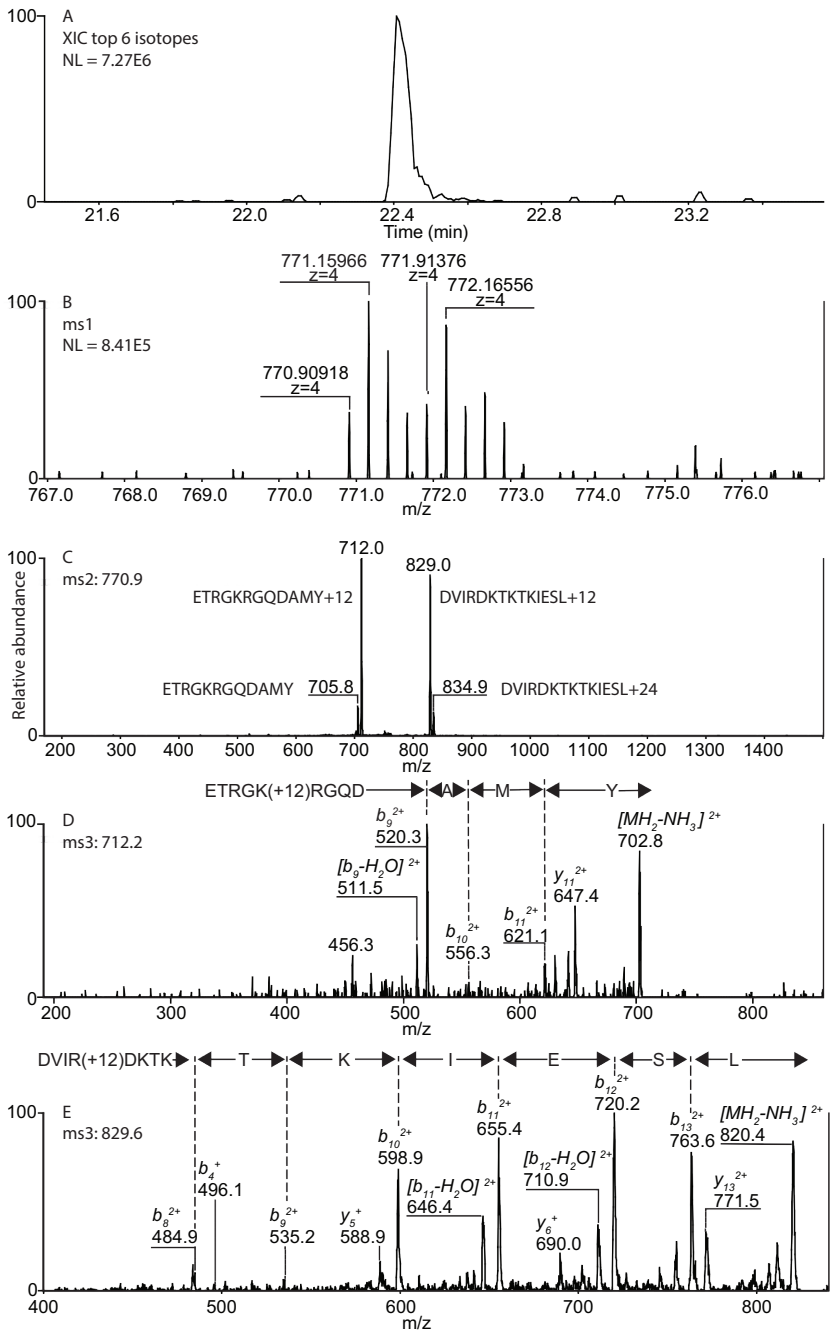


Figure 6. The intramolecular cross-link identified between A-fragment and B-fragment of diphtheria toxoid. LC/MS analyses performed on a chymotrypsin digest of diphtheria toxoid prepared by reaction ***2***: (A) chromatographic separation of the cross-linked dipeptide, and (B) the observed mass spectral doublet. (C) MS2 analyses revealed the presence of a cross-link between two peptides: (D) ETRGKRGQDAMY (712.0 Da) and (E) DVIRDKTKTKIESL (829.0 Da). The peptide sequences of D and E were identified by MS3 analysis.

After formaldehyde treatment, the reactive residues in diphtheria toxoid are only partially converted. To get an impression of the degree of modifications, the intensities of the mass spectral singlets and doublets detected by LC-MS were plotted independently in descending order (Figure 7A). Seventeen percent of all hits were mass spectral doublets, *i.e.*, 17% of the peptide hits contained a formaldehyde-induced modification. Furthermore, the intensities of all singlets and doublets were added up, respectively. Based on the total sum of intensities of the singlets and the doublets, an average conversion by formaldehyde was calculated (Σ intensities of doublets / Σ intensities of singlets + doublets). The average conversion of the amino acid residues in diphtheria toxoid by reaction ***2*** was 11.6%.

In conclusion, the formaldehyde-treatment (reaction ***2***) resulted in 5 different types of modifications, including intramolecular cross-links in diphtheria toxoid.

Formaldehyde-glycine attachments in diphtheria toxoid

Diphtheria toxin was treated with formaldehyde and $NaCNBH_3$ (to prevent intramolecular cross-linking) and subsequently with formaldehyde (CH_2O) and glycine or with deuterium-labelled formaldehyde (CD_2O) and glycine to introduce formaldehyde-glycine attachments to receptive amino acid residues, *i.e.*, arginine, asparagine, glutamine, histidine, tryptophan or tyrosine (reaction ***3***). Reaction ***3*** was performed to simplify the assignment of formaldehyde-glycine attachments.

Diphtheria toxoid with only formaldehyde-glycine attachments (reaction ***3***) revealed that the A-fragment and B-fragment of diphtheria toxoid had an apparently increased mass when compared to the untreated diphtheria toxin (Figure 3; lane *3 vs.* lane *0*, respectively). Moreover, SDS-PAGE revealed that the protein bands were somewhat broadened compared to those of diphtheria toxin (lane *0*), especially the bands of the A-fragment and apparently intact diphtheria toxoid. This is probably due to the incorporation of different numbers of formaldehyde-glycine moieties in diphtheria toxoid molecules.

In addition, the average number of primary amino groups in diphtheria toxoid was considerably reduced (down to 8% relative to the number in diphtheria toxin) after chemical treatment (reaction ***3***). The TNBS assay revealed the presence of 2 remaining primary amino groups on average in a diphtheria toxoid molecule. The number of amino groups detected in diphtheria toxoid (reaction ***3***) was comparable to the number of primary amino groups after dimethylation (reaction ***1***).

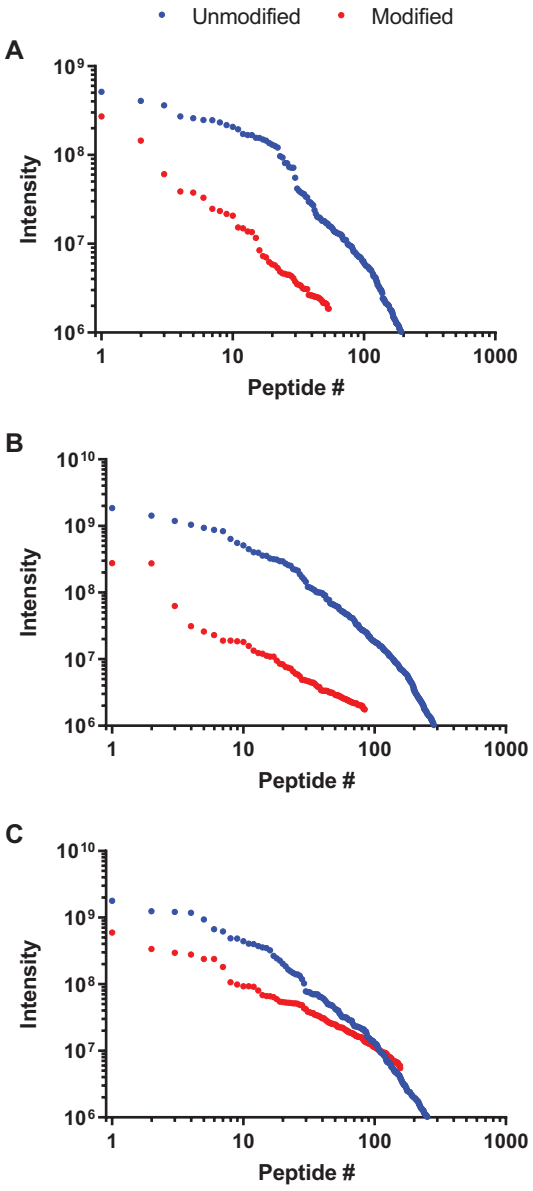


Figure 7. Relative peak intensities of mass spectral singlets (blue) and doublets (red) of peptides are presented and ranked in descending order. The singlets originate from unmodified peptides, whereas doublets are formed by the treatment with light (CH_2O) and heavy (CD_2O) formaldehyde. The relative peak intensities of peptides were determined with LC-MS (between 10^5 and 10^{10} arbitrary units) after treatment of diphtheria toxin with: (A) formaldehyde (reaction ***2***), (B) formaldehyde and glycine after dimethylation (reaction ***3***) or (C) formaldehyde and glycine (reaction ***4***) and after digestion with chymotrypsin.

Furthermore, LC-MS analysis of the chymotrypsin-digested product of reaction ***3*** gave a protein coverage of 88% (72% of A-fragment and 96% of B-fragment). The analysis revealed 156 peptides with formaldehyde-glycine attachments (Supplementary Table 4). These formaldehyde-modified peptides were recognised by their mass spectral doublets by the use of CH_2O and CD_2O. Sequences of these modified peptides were assigned and several peptides were detected with multiple modifications. The modifications were identified by MS/MS analysis or allocated based on the observed masses and the number of incorporated formaldehyde molecules. Based on MS/MS analysis, 52 residues in diphtheria toxoid were identified containing a formaldehyde-glycine attachment (Figure 8). The majority of the moieties was attached to arginine and tyrosine residues, but modifications of asparagine, glutamine, histidine and tryptophan residues were observed as well. In theory, 103 out of 535 amino acid residues in dimethylated diphtheria toxoid (reaction ***3***) can react with formaldehyde and glycine. However, arginine and tyrosine residues can react with two formaldehyde-glycine moieties [55]. Therefore, 137 formaldehyde-glycine moieties can be attached to one dimethylated diphtheria toxoid molecule. In addition to formaldehyde-glycine attachments, methylol groups and Schiff-bases were observed on particular amino acid residues, *i.e.*, on asparagine, glutamine, histidine and tryptophan residues.

To get an impression of the conversion of reactive residues by formaldehyde and glycine (reaction ***3***), the intensities of the modified (mass spectral doublets) and non-modified peptides (mass spectral singlets) were plotted separately in descending order (Figure 7B). The intensities of doublets relative to singlets gives an impression on the conversion of the reactive amino acid residues in diphtheria toxoid molecules by formaldehyde and glycine. It revealed that 30% of all hits were mass spectral doublets, demonstrating the presence of a formaldehyde-induced modification in the peptide. The observations indicate that the average conversion of amino acid residues by formaldehyde and glycine was 21.6% (Σ intensities of doublets / Σ intensities of singlets + doublets).

The conversion of individual reactive residues in diphtheria toxoid by formaldehyde and glycine (reaction ***3***) depends on multiple factors, *e.g.*, concentrations of reagents (diphtheria toxin, formaldehyde, glycine), reaction time and pH. Moreover, the intrinsic reactivity and accessibility of residues in diphtheria toxin probably determine the conversion degree. The reactivity depends on the type of amino acid residue, as demonstrated before [55]. The

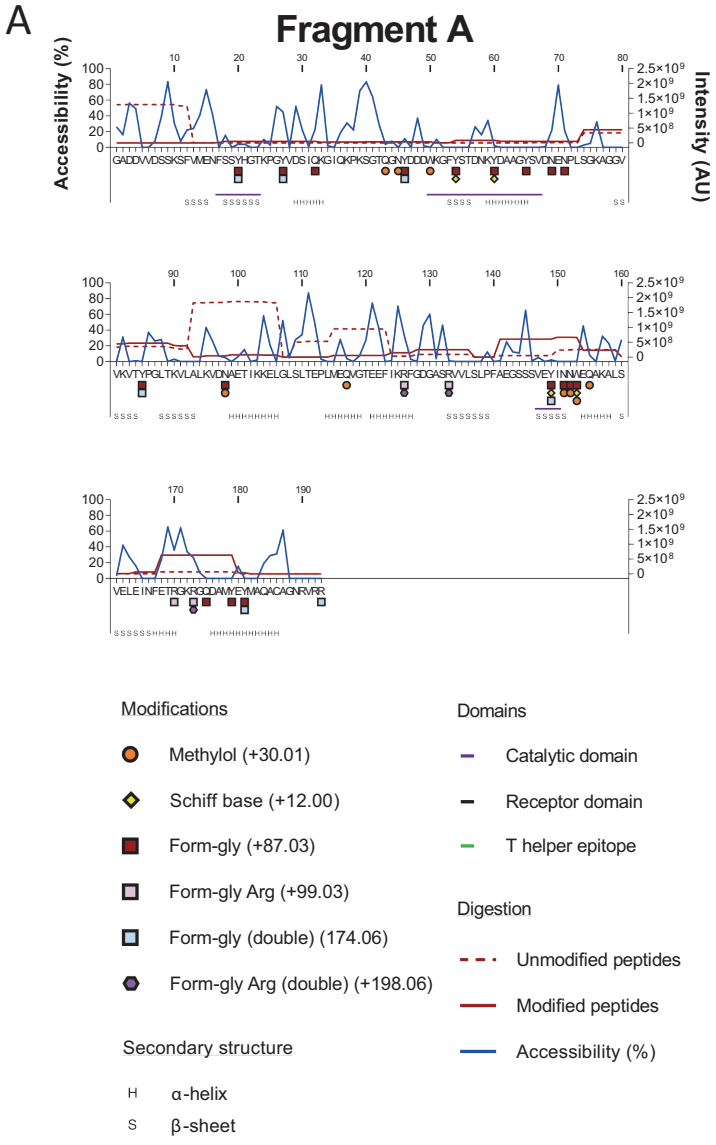
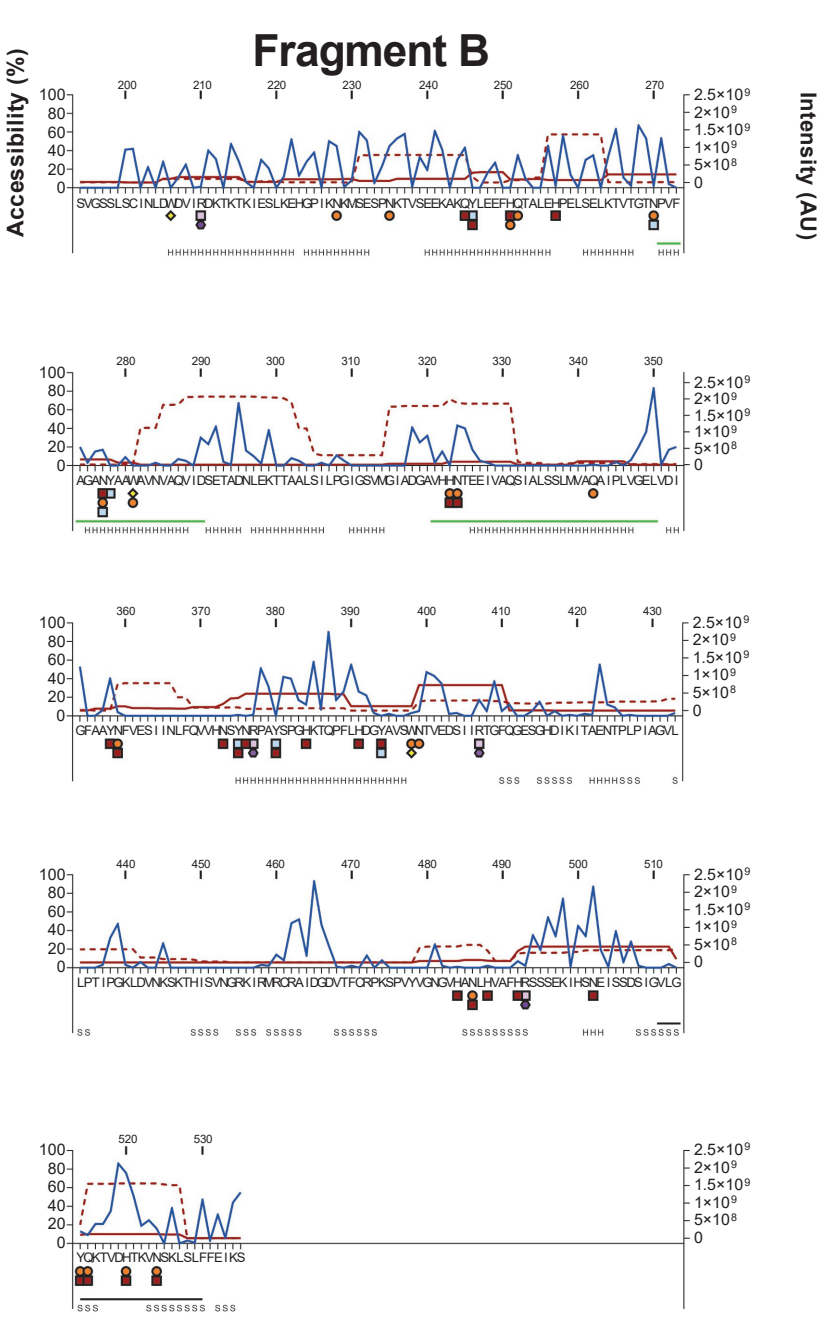


Figure 8. The modifications in diphtheria toxin caused by formaldehyde-glycine treatment after dimethylation of lysine residues (reaction ***3***). The blue line (y-axis left) indicates the theoretical solvent accessible surface area (accessibility) calculated using the molecular solvent radius of the formaldehyde-glycine Schiff base following dimethylation of the 1TOX crystal structure. The modified and unmodified peptides derived from chymotrypsin digestion are indicated with a solid and dotted dark red line, respectively of which the intensities are plotted (y-axis right). In the case of overlapping peptides the intensities are summed for the overlapping amino-acid positions. The primary structure is indicated on the x-axis along with the amino acid position number in increments of 10 residues at the top of each plot. The different amino acid modifications identified by MS analysis along with the secondary structure elements derived from the crystal structure PDB 1TOX. The catalytic site, receptor binding site and $CD4^+$ T-cell epitope are indicated with purple, black and green underlines, respectively. The specific symbols are present in the figure legend.

B



present study revealed that several rather inaccessible but reactive residues (<25% accessible) contained formaldehyde-glycine-induced modifications (Figure 8). The accessibility of the modified residues was calculated based on the crystal structure of diphtheria toxin [67]. Unexpectedly, no relationship (R^2=0.11) was observed between the conversion of a particular type of residue (*i.e.*, arginine and tyrosine) and its accessibility (Supplementary Figure 2). Contrarily, the results of the current study do not match our previous study with insulin in which the conversion by formaldehyde and glycine was dependant on the accessibility [56]. However, the correlation 'conversion *vs.* accessibility' was based on only four tyrosine residues present in insulin.

In conclusion, glycine molecules were attached to diphtheria toxin by formaldehyde. The majority of formaldehyde-glycine attachments were located on the arginine and tyrosine residues in diphtheria toxoid, irrespective of their position in the molecule.

Detoxification of diphtheria toxin by formaldehyde and glycine

The detoxification of diphtheria toxin with a mixture of formaldehyde and glycine resulted in a very heterogeneous toxoid (reaction ***4***). The heterogeneity of diphtheria toxoid could be visualised by SDS-PAGE (Figure 3; lane *4*), showing three broadened protein bands as compared to those of diphtheria toxin (lane *0*), corresponding to the entire diphtheria toxoid (approximate molecular weight of 58 kDa), A-fragment (22 kDa) and B-fragment (37 kDa). After chemical treatment (reaction ***4***), the intensity of the apparent intact diphtheria toxoid band was increased when compared to diphtheria toxin (lane *0*). The broadened protein bands were probably caused by intramolecular cross-links, as was shown for formaldehyde-treated diphtheria toxoid (reaction ***2***) by SDS-PAGE (Figure 3; lane *2*). Moreover, different numbers of formaldehyde-glycine moieties (*i.e.*, intermolecular cross-links) may have been attached to diphtheria toxin molecules as well, which might have contributed to the broadening of the protein bands. In addition, the TNBS assay showed 75% reduction in the number of primary amino groups, with on average only 7 primary amino groups per toxoid molecule.

Multiple formaldehyde-induced modifications (reaction ***4***) in diphtheria toxoid were determined by using LC-MS analysis (Supplementary Table 5). The chemical nature and the location of formaldehyde-glycine-induced modifications were identified in diphtheria toxoid (Figure 9).

The stable formaldehyde-induced modifications could be divided into intramolecular cross-links (reaction ***2***) and glycine attachments (reaction ***3***). Moreover, reversible methylol groups and Schiff bases were identified. LC-MS analysis provided a protein coverage of 92% (86% of A-fragment and 95% of B-fragment).

We focused predominantly on the formaldehyde-induced modifications in three areas of diphtheria toxin which are important for its toxicity and immunogenicity: (i) the NAD^+-binding cavity in the catalytic domain, (ii) the loop in the receptor domain involved in receptor binding and (iii) T-cell epitopes. Chemical modifications in the NAD^+-binding cavity and receptor binding site contribute to complete detoxification of the diphtheria toxin. However, other modifications might contribute to the reduction of toxicity as well, *e.g.*, due to conformational changes and cross-linking of A-fragment and B-fragment.

Catalytic site – The NAD^+-binding cavity of the diphtheria toxin molecule consists of three short peptide sequences which are folded together (residues: 17 – 23, 50 – 67, and 147 – 150). In theory, 10 amino acid residues of the NAD^+-binding cavity are reactive with formaldehyde alone or with formaldehyde and glycine (Tyr^{20}, His^{21}, Trp^{50}, Lys^{51}, Tyr^{54}, Asn^{58}, Lys^{59}, Tyr^{60}, Tyr^{65}, Tyr^{149}). Amino acid residues His^{21}, Tyr^{54} and Tyr^{65} are involved in the binding of NAD^+ (Figure 1). Three formaldehyde-glycine attachments were found on tyrosine residues (Tyr^{54}, Tyr^{60} and Tyr^{149}). The expected cross-link observed previously between residues Lys^{59} and Tyr^{60} (in reaction ***3***) was not identified after treatment with formaldehyde and glycine (reaction ***4***). Based on this data, we can conclude that formaldehyde-glycine attachments are the most common modifications present in the NAD^+-binding cavity, which will contribute to the inhibition of NAD^+-binding by diphtheria toxin.

Receptor-binding site – A second location studied in more detail for formaldehyde-induced modifications was the receptor-binding site of diphtheria toxin. With LC-MS analyses, modifications were identified at the receptor-binding site of diphtheria toxin (Figure 9). The peptides contain an intramolecular cross-link between Lys^{516} and Tyr^{514}. Furthermore, formaldehyde-glycine attachments were formed at amino acid residues Tyr^{514}, His^{520} and Asn^{524}. For residues Tyr^{514} and Asn^{524}, participation in receptor binding has been demonstrated [60]. Moreover, residue Lys^{522} was cross-linked by formaldehyde to another residue (*e.g.*, Tyr^{514} and His^{520}) in the peptide (Gly^{513} – Leu^{527}) based on the mass increment of 12 Da found on this lysine residue.

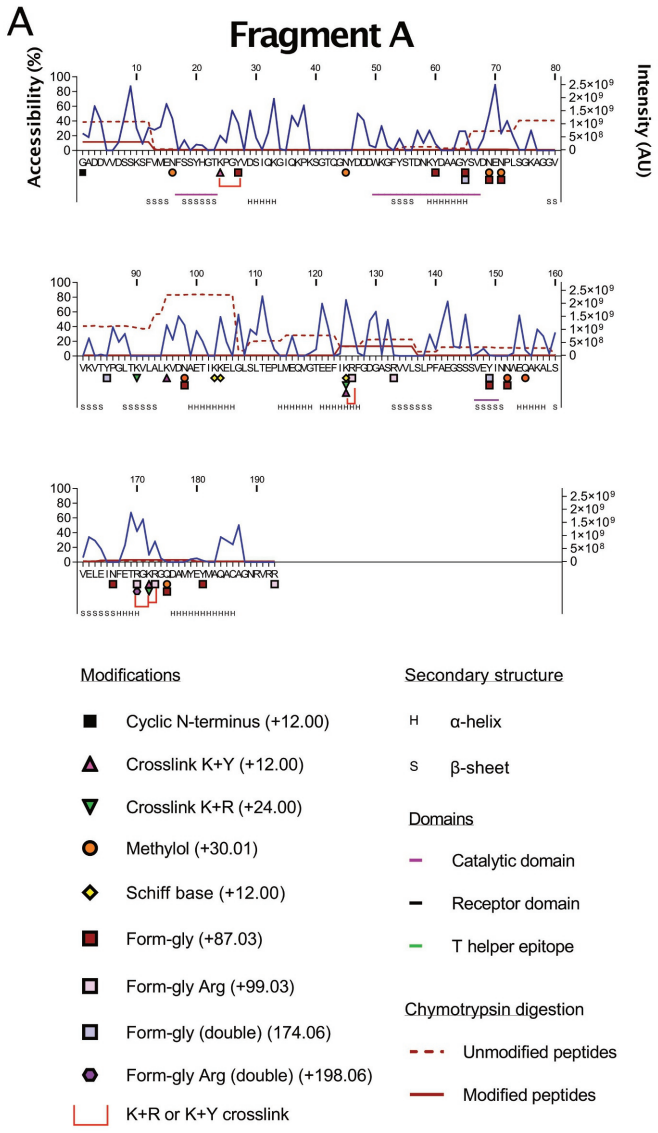
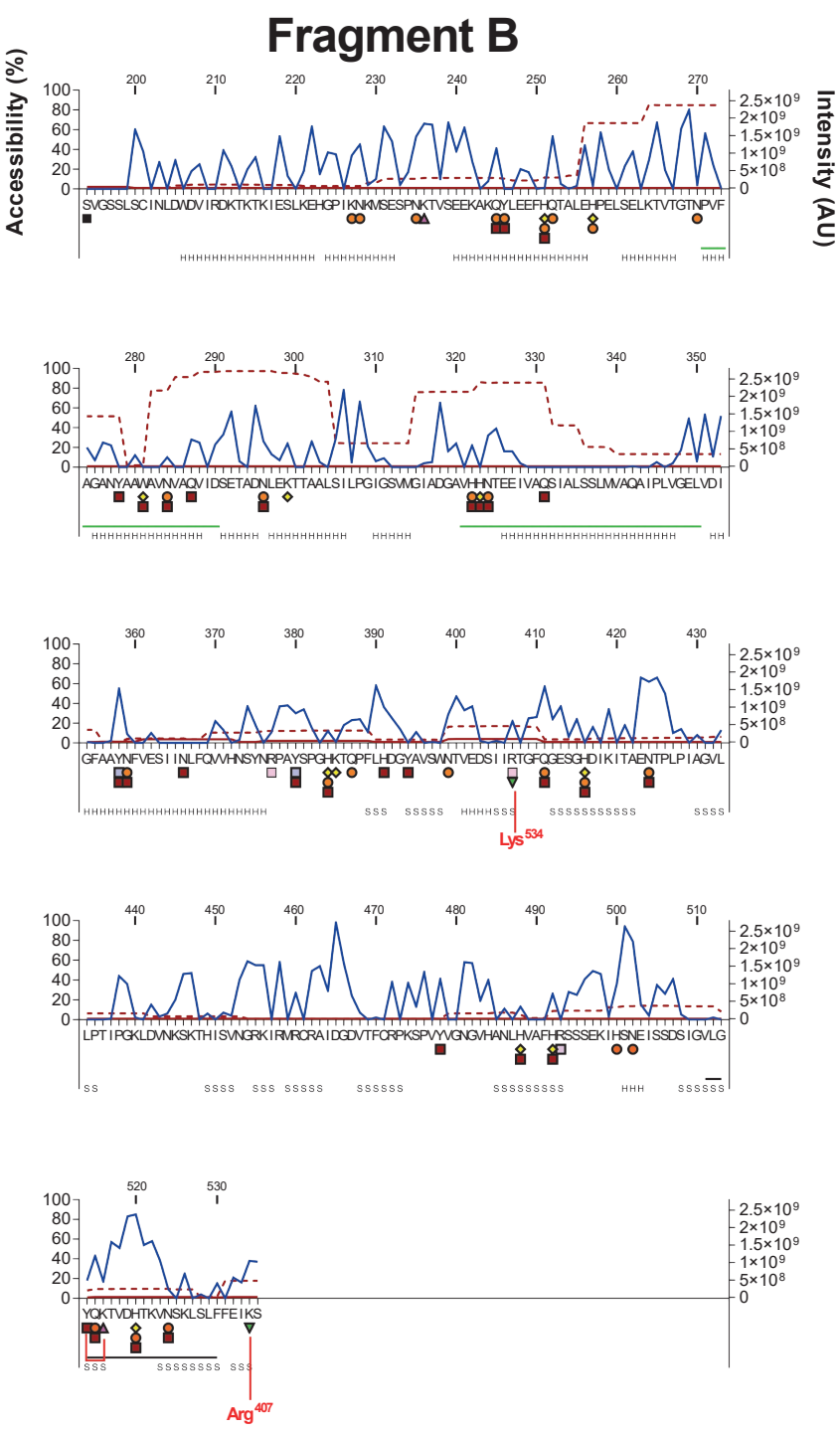


Figure 9. The modifications in diphtheria toxin caused by formaldehyde-glycine treatment (reaction ***4***). The blue line (y-axis left) indicates the theoretical solvent accessible surface area (accessibility), calculated using the molecular solvent radius of the Schiff base of formaldehyde and glycine. The modified and unmodified peptides derived from chymotrypsin digestion are indicated with a solid and dotted dark red line, respectively of which the intensities are plotted (y-axis right). In the case of overlapping peptides the intensities are summed for the overlapping amino-acid positions. The primary structure is indicated on the x-axis along with the amino acid position number in increments of 10 residues at the top of each plot. The different amino acid modifications identified by MS analysis along with the secondary structure elements derived from the crystal structure PDB 1TOX. The catalytic site, receptor binding site and $CD4^+$ T-cell epitope are indicated with purple, black and green underlines, respectively. The specific symbols are present in the figure legend. The red lines are connecting the residues between which a cross-link has been formed.

B



T-cell epitopes – Diphtheria toxoid contains three $CD4^+$ T-cell epitopes which are recognised by 70-82% of the diphtheria-vaccinated humans [63]. All $CD4^+$ T-cell epitopes are located in the B-fragment of diphtheria toxin residues 271 – 290, 321 – 340 and 331 – 350, with two overlapping epitopes. Several formaldehyde-glycine induced modifications were found in epitope 271 – 290 and epitope 321 – 340. However, the epitope 331 – 350 contains only one reactive residue (Gln^{331}). A formaldehyde-glycine attachment to this glutamine residue occurred, but in the vast majority of the diphtheria toxoid molecules this epitope remained unchanged (Figure 9). The latter is probably important for vaccine potency, as this particular epitope is recognised by 82% of the human population (n=100) [63].

The conversion of each reactive amino acid residue in a regular diphtheria toxoid, induced by formaldehyde and glycine (reaction ***4***), is difficult to determine. An impression is given by comparing the intensities of the mass spectral doublets with the mass spectral singlets (Figure 7C). The study revealed that 17% of the peptides contained formaldehyde-induced modifications. An average conversion of amino acid residues in diphtheria toxoid by formaldehyde and glycine (reaction ***4***) was calculated based on the total sum of intensities of the singlets and the doublets (Σ intensities of doublets / Σ intensities of singlets + doublets). The average conversion of the amino acid residues in diphtheria toxoid was 6.2%. The average conversion by a standard detoxification reaction (reaction ***4***) was less than the other two reaction conditions (reactions ***2*** and ***3***).

Based on these data, we conclude that both the NAD^+-binding cavity and the receptor-binding site are affected by intramolecular cross-links or by formaldehyde-glycine attachments during the standard detoxification reaction (reaction ***4***). A combination of formaldehyde-induced modifications in diphtheria toxoid will ensure complete detoxification of each toxoid molecule. Despite the many modifications in several $CD4^+$ T-cell epitopes, one major epitope (at residue numbers 321-340) remained largely unmodified (Figure 10).

Concluding remarks

The formaldehyde-induced modifications in diphtheria toxoid were investigated with a detailed analysis of the NAD^+-binding cavity and receptor-binding site. In both areas of the toxoid molecule, intramolecular cross-links and formaldehyde-glycine attachments were found. Moreover, cross-links occurred between the A-fragment and B-fragment of diphtheria toxoid. The conversion of these sites contributes to the inactivation of diphtheria toxin (Figure 10).

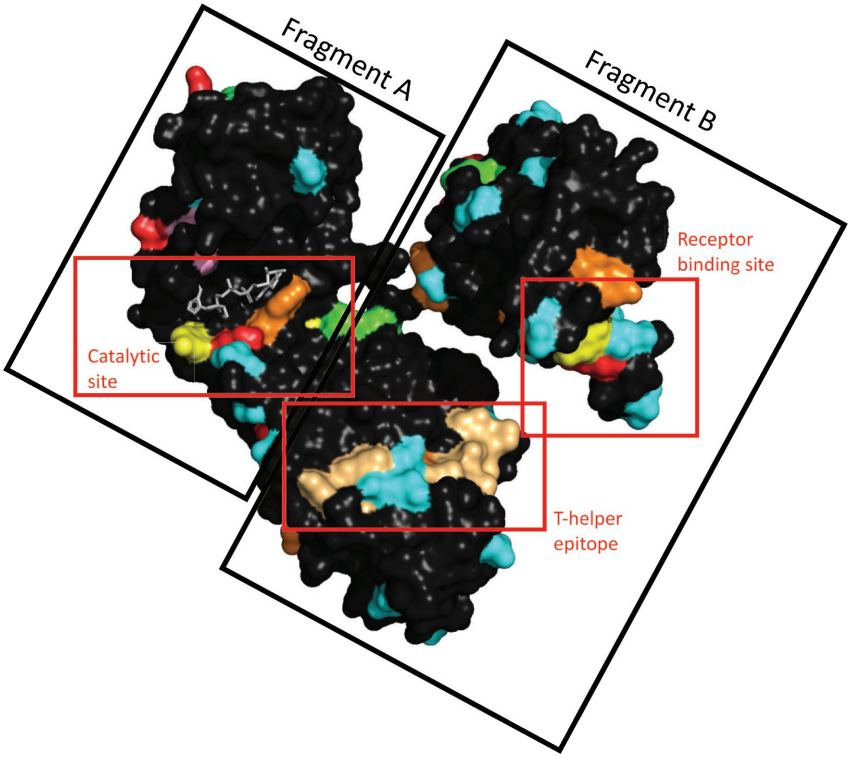


Figure 10. Illustration of formaldehyde-glycine-induced modifications on diphtheria toxin (PDB 1TOX). Amino acid residues that have cross-links being confirmed by mass spectrometric analysis are coloured red (Lys), green (Arg), yellow (Tyr) and pink (Trp). Formylation-glycine modifications on tryptophan and tyrosine are coloured orange. Methylol, Schiff-base or formaldehyde-glycine modifications on His, Gln and Asn are coloured cyan. Formylation-glycine modifications on Arg are coloured green. Non-modified amino acids are coloured black. Pymol v2.1.1 (Schrodinger LLC) was used to create this illustration.

The study revealed that all primary amino groups in diphtheria toxin are accessible to formaldehyde (demonstrated by reaction ***1***). Only 11 out of 39 lysine residues formed formaldehyde-induced intramolecular cross-links with neighbouring amino acid residues (reaction ***2***). The formaldehyde-induced intramolecular cross-links (Table 1) were formed between residues in close proximity of each other (distances of α-carbon atoms 3.8-10.9 Å). However, several other intramolecular cross-links that were expected, according to the crystal structure of diphtheria toxin, could not be confirmed. Possibly, these cross-links were not constructed because these residues are very mobile in solution or because their orientation is incompatible with cross-linking. In addition to intramolecular cross-links, formaldehyde and glycine-induced modifications were introduced during the detoxification reaction by formaldehyde and glycine. The modifications are most frequently found on arginine and tyrosine residues. Remarkably, relatively inaccessible but reactive residues were also modified by formaldehyde and glycine. In a few particular cases the conversion of inaccessible residues was rather high (above 50%; Figure 10). Unlike our expectation, the accessibility did not evidently steer the conversion of reactive residues. Probably, the formaldehyde-glycine moieties could reach rather inaccessible residues due to conformational dynamics within the diphtheria toxin molecule.

The detoxification process changes the antigenicity and immunogenicity of diphtheria toxoid [16]. The formaldehyde-induced modifications located on or near the immunodominant B-cell and T-cell epitopes of diphtheria toxin are of special interest. These modifications might affect the antigenicity and immunogenicity significantly. The immunodominant $CD4^+$ T-cell epitopes have been identified and examined [63]. Our study revealed that one dominant $CD4^+$ T-cell epitope remained almost completely unaltered upon formaldehyde-glycine treatment. This particular epitope is recognised by 82% of the subjects (n=100). Unfortunately, the immunodominant B-cell epitopes are unknown. Our previous study revealed that formaldehyde treatment results in the impairment of particular B-cell epitopes in diphtheria toxin. The impairment was measured by biosensor analysis using a set of monoclonal antibodies [16]. Therefore, formaldehyde-induced modifications in these immunodominant B-cell epitopes cannot be excluded. A different detoxification matrix might result in a distinct antigenicity and immunogenicity. Therefore, the detoxification process could determine the efficacy of the diphtheria vaccine in humans considerably.

The strategy followed in this study was suitable to identify the different types of formaldehyde-induced modifications in diphtheria toxoid. However, in many vaccine production processes of diphtheria toxoid, the detoxification of diphtheria toxin takes place in a matrix of culture supernatant with various amino acid compositions. The attachments of different amino acids to the reactive residues of diphtheria toxin will add to the heterogeneity of the resulting toxoid. The identification of modified residues in such a diphtheria toxoid can be very difficult and laborious, even with the sophisticated mass spectrometers and software tools available today. Nevertheless, the methods can be used to analyse such products in comparability studies, *e.g.*, to support registration of these products after process or formulation improvements. The work demonstrates that with current powerful analytical techniques, it is possible to approach classical vaccines as if they were well-defined biologicals.

Abbreviations

DTT: DL-dithiotreitol
LC-MS: Tandem liquid chromatography – electrospray ionisation mass spectrometry
MS: Mass spectrometry
m/z: Mass-over-charge ratio
NAD^+: Nicotinamide adenine dinucleotide
TNBS: 2,4,6-trinitrobenzene sulfonic acid
CID: Collision-induced dissociation
ETD: Electron-transfer dissociation

Acknowledgements

We thank Geert Mommen for his fruitful discussions and support to this work. This project was funded by the Ministry of Agriculture, Nature and Food Quality.

References

1. Liang, J.L., *et al.*, *Prevention of Pertussis, Tetanus, and Diphtheria with Vaccines in the United States: Recommendations of the Advisory Committee on Immunization Practices (ACIP).* MMWR Recomm Rep, 2018. **67**(2): p. 1-44.
2. Glenny, A.T. and B.E. Hopkins, *Diphtheria toxoid as an immunising agent.* Br J Exp Pathol, 1923. **4**(283): p. 283-287.
3. Ramon, G., *Sur la toxine et sur l' anatoxine diphtériques.* Ann Inst Pasteur., 1924. **38**: p. 1-10.
4. Ramon, G. and P. Descombey, *Sur l'Immunisation antitétanique et sur la production de l'antitoxine tétanique.* Compt Rend Soc Biol, 1925. **93**: p. 508-9.
5. Corbel, M.J., *et al.*, *WHO Working Group on revision of the Manual of Laboratory Methods for Testing DTP Vaccines-Report of two meetings held on 20-21 July 2006 and 28-30 March 2007, Geneva, Switzerland.* Vaccine, 2008. **26**(16): p. 1913-21.
6. Organization, W.H., *Manual for Quality Control of Diphtheria, Tetanus and Pertussis Vaccines*. 2013: Expanded Programme on Immunization (EPI) of the Department of Immunization, Vaccines and Biologicals
7. Choe, S., *et al.*, *The crystal structure of diphtheria toxin.* Nature, 1992. **357**(6375): p. 216-22.
8. Fotinou, C., *et al.*, *The crystal structure of tetanus toxin Hc fragment complexed with a synthetic GT1b analogue suggests cross-linking between ganglioside receptors and the toxin.* J Biol Chem, 2001. **276**(34): p. 32274-81.
9. Breidenbach, M.A. and A.T. Brunger, *2.3 A crystal structure of tetanus neurotoxin light chain.* Biochemistry, 2005. **44**(20): p. 7450-7.
10. Masuyer, G., J. Conrad, and P. Stenmark, *The structure of the tetanus toxin reveals pH-mediated domain dynamics.* EMBO Rep, 2017. **18**(8): p. 1306-1317.
11. Sauve, S., G. Gingras, and Y. Aubin, *NMR study of mutations of glycine-52 of the catalytic domain of diphtheria toxin.* J Pharm Biomed Anal, 2018. **150**: p. 72-79.
12. Mishra, R.P.N., *et al.*, *Structural and immunological characterization of E. coli derived recombinant CRM197 protein used as carrier in conjugate vaccines.* Biosci Rep, 2018. **38**(5).
13. Paliwal, R. and E. London, *Comparison of the conformation, hydrophobicity, and model membrane interactions of diphtheria toxin to those of formaldehyde-treated toxin (diphtheria toxoid): formaldehyde stabilization of the native conformation inhibits changes that allow membrane insertion.* Biochemistry, 1996. **35**(7): p. 2374-9.
14. Schwendeman, S.P., *et al.*, *Stabilization of tetanus and diphtheria toxoids against moisture-induced aggregation.* Proc Natl Acad Sci U S A, 1995. **92**(24): p. 11234-8.
15. Johansen, P., H.P. Merkle, and B. Gander, *Physico-chemical and antigenic properties of tetanus and diphtheria toxoids and steps towards improved stability.* Biochim Biophys Acta, 1998. **23**(2): p. 425-36.
16. Metz, B., *et al.*, *Physicochemical and immunochemical techniques predict the quality of diphtheria toxoid vaccines.* Vaccine, 2003. **22**(2): p. 156-167.
17. Thaysen-Andersen, M., *et al.*, *Investigation of the detoxification mechanism of formaldehyde-treated tetanus toxin.* Vaccine, 2007. **25**(12): p. 2213-27.
18. Metz, B., *et al.*, *Physicochemical and immunochemical assays for monitoring consistent production of tetanus toxoid.* Biologicals, 2013. **41**(4): p. 231-7.
19. Regnier, M., *et al.*, *Structural perturbation of diphtheria toxoid upon adsorption to aluminium hydroxide adjuvant.* Vaccine, 2012. **30**(48): p. 6783-8.

20. Shukla, A., *et al.*, *Alginate coated chitosan microparticles mediated oral delivery of diphtheria toxoid. Part A. Systematic optimization, development and characterization.* Int J Pharm, 2015. **495**(1): p. 220-233.

21. Alsarraf, H., *et al.*, *Biophysical comparison of diphtheria and tetanus toxins with the formaldehyde-detoxified toxoids, the main components of diphtheria and tetanus vaccines.* Virulence, 2017. **8**(8): p. 1880-1889.

22. Bayart, C., *et al.*, *The combined use of analytical tools for exploring tetanus toxin and tetanus toxoid structures.* J Chromatogr B Analyt Technol Biomed Life Sci, 2017. **1054**: p. 80-92.

23. Kalbfleisch, K., *et al.*, *Identity, Structure and Compositional Analysis of Aluminum Phosphate Adsorbed Pediatric Quadrivalent and Pentavalent Vaccines.* Comput Struct Biotechnol J, 2019. **17**: p. 14-20.

24. Samra, H.S., *et al.*, *The effects of substituted cyclodextrins on the colloidal and conformational stability of selected proteins.* J Pharm Sci, 2010. **99**(6): p. 2800-18.

25. Salnikova, M.S., *et al.*, *Physical characterization of clostridium difficile toxins and toxoids: effect of the formaldehyde crosslinking on thermal stability.* J Pharm Sci, 2008. **97**(9): p. 3735-52.

26. Salnikova, M.S., *et al.*, *Preformulation studies of Clostridium difficile toxoids A and B.* J Pharm Sci, 2008. **97**(10): p. 4194-207.

27. Iwaki, M., *et al.*, *Toxoid flocculation assay by laser light-scattering.* J Immunol Methods, 2007. **318**(1-2): p. 138-46.

28. Prieur, S., *et al.*, *Development of an in vitro potency test for tetanus vaccines: an immunoassay based on Hc fragment determination.* Dev Biol (Basel), 2002. **111**: p. 37-46.

29. Bolgiano, B., *et al.*, *Monitoring of diphtheria, pertussis and tetanus toxoids by circular dichroism, fluorescence spectroscopy and size-exclusion chromatography.* Dev Biol, 2000. **103**: p. 51-9.

30. Winsnes, R., *et al.*, *Collaborative study for the validation of serological methods for potency testing of diphtheria toxoid vaccines-part 1.* Pharmeuropa Bio, 2004. **2003**(2): p. 35-68.

31. Sesardic, D., *et al.*, *Collaborative study for the validation of serological methods for potency testing of diphtheria toxoid vaccines - extended study: correlation of serology with in vivo toxin neutralisation.* Pharmeuropa Bio, 2004. **2003**(2): p. 69-76.

32. Metz, B., *et al.*, *Quality control of routine, experimental and real-time aged diphtheria toxoids by in vitro analytical techniques.* Vaccine, 2007. **25**(39-40): p. 6863-6871.

33. Coombes, L., *et al.*, *Development and use of a novel in vitro assay for testing of diphtheria toxoid in combination vaccines.* J Immunol Methods, 2009. **350**(1-2): p. 142-9.

34. Coombes, L., *et al.*, *In vitro antigen ELISA for quality control of tetanus vaccines.* Biologicals, 2012.

35. Westdijk, J., *et al.*, *Antigenic fingerprinting of diphtheria toxoid adsorbed to aluminium phosphate.* Biologicals, 2017. **47**: p. 69-75.

36. Hickey, J.M., *et al.*, *Analytical Comparability Assessments of 5 Recombinant CRM197 Proteins From Different Manufacturers and Expression Systems.* J Pharm Sci, 2018. **107**(7): p. 1806-1819.

37. Moller, J., *et al.*, *Proteomics of diphtheria toxoid vaccines reveals multiple proteins that are immunogenic and may contribute to protection of humans against Corynebacterium diphtheriae.* Vaccine, 2019. **37**(23): p. 3061-3070.

38. Chirino, A.J. and A. Mire-Sluis, *Characterizing biological products and assessing comparability following manufacturing changes.* Nat Biotechnol, 2004. **22**(11): p. 1383-91.

39. Federici, M., *et al.*, *Analytical lessons learned from selected therapeutic protein drug comparability studies.* Biologicals, 2013. **41**(3): p. 131-47.

40. Sasiak, A.B., *et al.*, *Comparison of in vitro and in vivo methods to study stability of PLGA microencapsulated*

tetanus toxoid vaccines. Vaccine, 2000. **19**(7-8): p. 694-705.

41. Jaganathan, K.S., *et al.*, *Development of a single dose tetanus toxoid formulation based on polymeric microspheres: a comparative study of poly(D,L-lactic-co-glycolic acid) versus chitosan microspheres.* Int J Pharm, 2005. **294**(1-2): p. 23-32.

42. Alonso, M.J., *et al.*, *Biodegradable microspheres as controlled-release tetanus toxoid delivery systems.* Vaccine, 1994. **12**(4): p. 299-306.

43. Jung, T., *et al.*, *Loading of tetanus toxoid to biodegradable nanoparticles from branched poly(sulfobutyl-polyvinyl alcohol)-g-(lactide-co-glycolide) nanoparticles by protein adsorption: a mechanistic study.* Pharm Res, 2002. **19**(8): p. 1105-13.

44. Kersten, G.F., *et al.*, *Single shot with tetanus toxoid in biodegradable microspheres protects mice despite acid-induced denaturation of the antigen.* Vaccine, 1996. **14**(17-18): p. 1627-32.

45. van der Lubben, I.M., *et al.*, *Chitosan microparticles for mucosal vaccination against diphtheria: oral and nasal efficacy studies in mice.* Vaccine, 2003. **21**(13-14): p. 1400-8.

46. Ahire, V.J., *et al.*, *Chitosan Microparticles as Oral Delivery System for Tetanus Toxoid.* Drug Dev Ind Pharm, 2007: p. 1-13.

47. Arthanari, S., *et al.*, *Chitosan-HPMC-blended microspheres as a vaccine carrier for the delivery of tetanus toxoid.* Artif Cells Nanomed Biotechnol, 2016. **44**(2): p. 517-23.

48. Xing, D.K., *et al.*, *Physicochemical and immunological studies on the stability of free and microsphere-encapsulated tetanus toxoid in vitro.* Vaccine, 1996. **14**(13): p. 1205-13.

49. Determan, A.S., *et al.*, *Protein stability in the presence of polymer degradation products: consequences for controlled release formulations.* Biomaterials, 2006. **27**(17): p. 3312-20.

50. Gordon, S., *et al.*, *Chitosan hydrogels containing liposomes and cubosomes as particulate sustained release vaccine delivery systems.* J Liposome Res, 2012. **22**(3): p. 193-204.

51. Broker, M., *et al.*, *Polysaccharide conjugate vaccine protein carriers as a "neglected valency" - Potential and limitations.* Vaccine, 2017. **35**(25): p. 3286-3294.

52. Martin, L.B., *et al.*, *Status of paratyphoid fever vaccine research and development.* Vaccine, 2016. **34**(26): p. 2900-2902.

53. Mani, S., T. Wierzba, and R.I. Walker, *Status of vaccine research and development for Shigella.* Vaccine, 2016. **34**(26): p. 2887-2894.

54. Rappuoli, R., *New and improved vaccines against diphtheria and tetanus*, in *New generation vaccines*, M.M. Levine, *et al.*, Editors. 1997, Marcel Dekker, inc.: New York. p. 417-435.

55. Metz, B., *et al.*, *Identification of formaldehyde-induced modifications in proteins: reactions with model peptides.* J Biol Chem, 2004. **279**(8): p. 6235-43.

56. Metz, B., *et al.*, *Identification of formaldehyde-induced modifications in proteins: reactions with insulin.* Bioconjug Chem, 2006. **17**(3): p. 815-22.

57. Van Ness, B.G., J.B. Howard, and J.W. Bodley, *ADP-ribosylation of elongation factor 2 by diphtheria toxin. Isolation and properties of the novel ribosyl-amino acid and its hydrolysis products.* J Biol Chem, 1980. **255**(22): p. 10717-20.

58. Papini, E., *et al.*, *Histidine 21 is at the NAD+ binding site of diphtheria toxin.* J Biol Chem, 1989. **264**(21): p. 12385-8.

59. Blanke, S.R., K. Huang, and R.J. Collier, *Active-site mutations of diphtheria toxin: role of tyrosine-65 in NAD binding and ADP-ribosylation.* Biochemistry, 1994. **33**(51): p. 15494-500.

60. Shen, W.H., *et al.*, *Participation of lysine 516 and phenylalanine 530 of diphtheria toxin in receptor recognition.* J Biol Chem, 1994. **269**(46): p. 29077-84.

61. Naglich, J.G., *et al.*, *Expression cloning of a diphtheria toxin receptor: identity with a heparin-binding EGF-like growth factor precursor.* Cell, 1992. **69**(6): p. 1051-61.

62. Louie, G.V., *et al.*, *Crystal structure of the complex of diphtheria toxin with an extracellular fragment of its receptor.* Mol Cell, 1997. **1**(1): p. 67-78.

63. Diethelm-Okita, B.M., *et al.*, *Universal epitopes for human CD4+ cells on tetanus and diphtheria toxins.* J Infect Dis, 2000. **181**(3): p. 1001-9.

64. Sambrook, J., E.F. Fritsch, and T. Maniatis, *Molecular cloning*. second ed. 1989, New York: Cold Spring Harbor Laboratory Press.

65. Habeeb, A.F.S.A., *Determination of free amino groups in proteins by trinitrobenzenesulfonic acid.* Anal. Biochem., 1966. **14**: p. 328-336.

66. Meiring, H.D., *et al.*, *Nanoscale LC-MS$^{(n)}$: technical design and applications to peptide and protein analysis.* J Sep Sci, 2002. **25**: p. 557-568.

67. Bell, C.E. and D. Eisenberg, *Crystal structure of nucleotide-free diphtheria toxin.* Biochemistry, 1997. **36**(3): p. 481-8.

68. *Pytms*. 2015; Available from: https://pymolwiki.org/index.php/Pytms.

69. Carroll, S.F., J.T. Barbieri, and R.J. Collier, *Diphtheria toxin: purification and properties.* Methods Enzymol, 1988. **165**: p. 68-76.

70. Silva, J.C., *et al.*, *Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition.* Mol Cell Proteomics, 2006. **5**(1): p. 144-56.

71. Jentoft, N. and D.G. Dearborn, *Protein labeling by reductive alkylation.* Methods Enzymol, 1983. **91**: p. 570-579.

72. Braun, K.P., *et al.*, *A structural assignment for a stable acetaldehyde-lysine adduct.* J Biol Chem, 1995. **270**(19): p. 11263-6.

73. Fowles, L.F., *et al.*, *The formation and stability of imidazolidinone adducts from acetaldehyde and model peptides. A kinetic study with implications for protein modification in alcohol abuse.* Biochem Pharmacol, 1996. **51**(10): p. 1259-67.

74. Eschweiler, *Ersatz von an Stickstoff gebundenen Wasserstoffatomen durch die Methylgruppe mit Hülfe von Formaldehyd.* . Chemische Berichte, 1905. **38**: p. 880-882.

3

FORMALDEHYDE TREATMENT OF PROTEINS ENHANCES PROTEOLYTIC DEGRADATION BY THE ENDO-LYSOSOMAL PROTEASE CATHEPSIN S

Thomas J. M. Michiels [1,2], Hugo D. Meiring [2], Wim Jiskoot [1], Gideon F. A. Kersten [1,2], Bernard Metz [2]

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

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

Scientific Reports **2020**, *10* (1), 11535.

Abstract

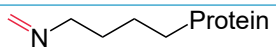
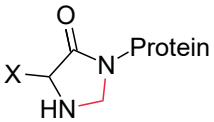
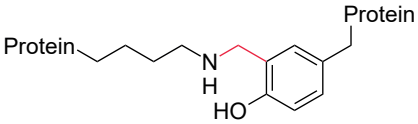
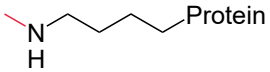
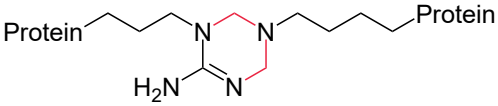
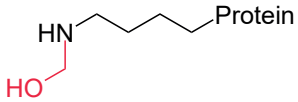
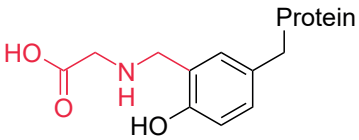
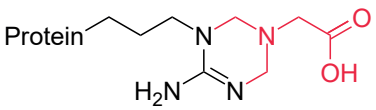
Enzymatic degradation of protein antigens by endo-lysosomal proteases in antigen-presenting cells is crucial for achieving cellular immunity. Structural changes caused by vaccine production process steps, such as formaldehyde inactivation, could affect the sensitivity of the antigen to lysosomal proteases. The aim of this study was to assess the effect of the formaldehyde detoxification process on the enzymatic proteolysis of antigens by studying model proteins. Bovine serum albumin, β-lactoglobulin A and cytochrome c were treated with various concentrations of isotopically labelled formaldehyde and glycine, and subjected to proteolytic digestion by cathepsin S, an important endo-lysosomal endoprotease. Degradation products were analysed by mass spectrometry and size exclusion chromatography. The most abundant modification sites were identified by their characteristic MS doublets. Unexpectedly, all studied proteins showed faster proteolytic degradation upon treatment with higher formaldehyde concentrations. This effect was observed both in the absence and presence of glycine, an often-used excipient during inactivation to prevent intermolecular crosslinking. Overall, subjecting proteins to formaldehyde or formaldehyde/glycine treatment results in changes in proteolysis rates, leading to an enhanced degradation speed. This accelerated degradation could have consequences for the immunogenicity and the efficacy of vaccine products containing formaldehyde-inactivated antigens.

Introduction

Enzymatic degradation of antigens is a crucial step in the process of acquiring cellular immunity, *e.g.,* through the induction of antigen specific T-helper cells or cytotoxic T-cells. The endo-lysosomal protease activity is lower for immune cells with high antigen presentation capacity, such as dendritic cells, than for cells with lower antigen presentation capacity, such as neutrophils [1]. Several groups have found a correlation between slow proteolytic antigen degradation and increased immunogenicity of the studied antigen [2-9], although it does not hold up for all antigens [10]. This correlation suggests that protease resistance could be an important factor for vaccine efficacy. Currently, numerous efforts are being made to replace, reduce and refine (3Rs) the use of animal tests [11,12]. One approach to achieve this is the so-called consistency approach [13]. This approach is based on the principle that if a panel of *in-vitro* tests can prove that a vaccine product is produced in a consistent manner, reduction or replacement of quality control animal tests is possible. An *in-vitro* test that could follow changes in enzymatic degradation kinetics of protein antigens may thus be used to monitor vaccine batch quality in a consistency approach.

Many antigens in inactivated vaccine products, such as bacterial toxins and poliovirus, are inactivated by using a mixture of formaldehyde and amino acids. This treatment results in modifications to the protein, as reported previously by Metz *et al.* (Table 1) [14,15]. These modifications may alter the immunogenicity of the antigen. For instance, for diphtheria toxoid the immunogenicity increases upon increased exposure to formaldehyde [16]. Previous research on pertussis antigens indicates a slower *in-vitro* proteolysis of formaldehyde-treated proteins than untreated counterparts [17]. However, that study used trypsin as an enzyme, which cleaves after lysine and arginine residues. Both of these amino acids are susceptible to formaldehyde modification, especially the latter being converted in substantial amounts [15]. Formaldehyde modification of trypsin's cleavage sites is very likely to affect the substrate's degradation kinetics, as lysine modifications inhibit digestion by trypsin, but this might be less relevant for lysosomal degradation in antigen-presenting cells (APCs) [18]. Thus, these results might not be indicative of the effect formaldehyde treatment has on the endo-lysosomal degradation kinetics of the antigen.

Table 1 Overview of the most common formaldehyde-induced modifications of proteins

Example structure	Reactive residues	ΔMass (Da)	Conversion rate (%)[a]
	K, Q, N, W, H	+12.000	–[b]
	Free *N* termini	+12.000	76
	K, Y	+12.000	–[b]
	K	+14.016	–[b c]
	K, R	+24.000	–[b]
	K, Q, N, W, H	+30.011	–[b]
	H, W, Q, N Y	+87.032 +87.032	3.6 – 6.6 62
	R	+99.032 +198.064[d]	56 41

[a] Conversions as reported by Metz *et al.* [15]. [b] Conversion rates not reported. [c] Product reported by Trezl *et al.* [23]. [d] Double +99 modification.

To assess the effect of formaldehyde treatment on enzymatic protein processing, three formaldehyde-treated model proteins were prepared: bovine serum albumin (BSA), β-lactoglobulin A and cytochrome c. These proteins were chosen for their structural diversity, commercial availability and their low toxicity (as compared to toxins). These proteins were treated with increasing concentrations of formaldehyde and glycine. They were then subjected to proteolytic digestion by cathepsin S, an important endo-lysosomal enzyme [19]. The

degradation of the intact proteins was monitored by size exclusion chromatography (SEC) and the resulting peptides were identified and quantified using nanoscale liquid chromatography-mass spectrometry (LC-MS).

Results and Discussion

Enzymatic degradation of formaldehyde-treated model proteins by cathepsin S

Formaldehyde and glycine-treated cytochrome c, BSA and β-lactoglobulin A were subjected to enzymatic degradation by cathepsin S. The digestion was followed by using SEC and LC-MS. The formaldehyde and glycine-treated proteins eluted slightly earlier with SEC than their untreated counterparts. The treated proteins showed broader chromatographic peaks, which were not fully resolved, and a higher dimer content than the untreated proteins (Fig. S1). To allow for more robust quantification, peak height rather than peak area was used, as it was difficult to reliably determine the peak edge of formaldehyde-treated samples at the end of the digestion. Regarding the degradation kinetics, SEC analysis (Fig. 1) showed that the higher the formaldehyde/glycine concentrations the protein was exposed to, the faster the decrease of intact protein content was upon subsequent exposure to cathepsin S. This effect is most clear for cytochrome c and BSA, but is also observed for β-lactoglobulin.

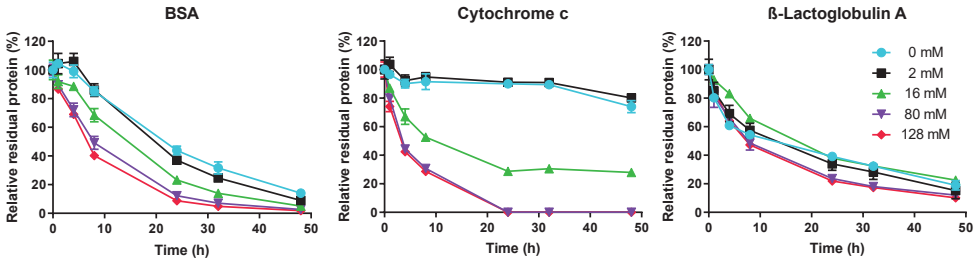


Figure 1. SEC analysis of intact formaldehyde- and glycine-treated proteins as function of cathepsin S digestion time. Error bars represent SD of three enzyme reactions.

Peptides that are sensitive to formaldehyde-induced modifications would not allow for direct comparison between samples, because the ratio modified/unmodified peptide depends on the formaldehyde/glycine concentrations. Furthermore, variety in MS response factors between the modified and unmodified peptides would have to be taken into account, as the proton affinity of an unmodified side chain would likely be different to that of a modified side

chain [20]. Other difficulties with direct comparison arise from cross-linked dipeptides, which cannot be identified automatically by the PEAKS software (or equivalent software packages), and from the fact that cross-linked dipeptides and adducts give challenging Collision Induced Dissociation (CID) fragmentation spectra with neutral losses. The latter problem results in exclusion based on Peptide to Spectrum Match (PSM) score. The challenging neutral losses from CID fragmentation can often be avoided by using Electron Transfer Dissociation (ETD) fragmentation, but this requires multiply charged (z>2) species at relatively high abundance [21]. To overcome these limitations and avoid bias, only the formation kinetics of some of the most abundant peptides that are considered inert, *i.e.*, not modifiable by formaldehyde, are plotted in Figure 2 [15]. These representative peptides were selected to include peptides with various kinetic profiles. Length variants with the same core sequence were excluded.

LC-MS analysis showed that inert peptides were formed faster after exposure to formaldehyde and glycine. For instance, the cytochrome c peptide MIFA reached a maximum degradation speed upon treatment with 16 mM formaldehyde/glycine (Fig. 2A); higher formaldehyde/glycine concentrations did not alter the kinetics further. Another cytochrome c peptide, EETLM, showed a more gradual increase in formation rate with increasing formaldehyde/glycine concentrations (Fig. 2B). Cytochrome c peptide GLFG showed an optimum degradation speed at 16 mM formaldehyde/glycine (Fig. 2C), although even exposure to the highest concentrations still enhanced the formation of this peptide compared to the control (*i.e.*, cathepsin S-treated cytochrome c that was not previously exposed to formaldehyde/glycine). The β-lactoglobulin peptide MAASDISLL is an example of an intermediate product that likely was degraded to smaller peptides by the enzyme. This peptide was formed faster with higher formaldehyde/glycine concentrations, with the peptide's maximum intensity shifting to earlier time points (Fig. 2D). BSA showed relatively few unreactive peptides compared to the smaller cytochrome c and ß-lactoglobulin. This is mainly due to the high number of cysteines (35 per BSA molecule) on top of the other reactive amino acids. The usual reduction/alkylation was not performed to avoid unwanted side reactions (with the formaldehyde modifications) and protein unfolding. While cystines are not reactive towards formaldehyde [15], the cysteines formed under the reaction conditions impede quantification of the peptides that contain them. Peptides TPTLVE and APELL (Fig. 2E and 2F) were characteristic for the BSA peptides.

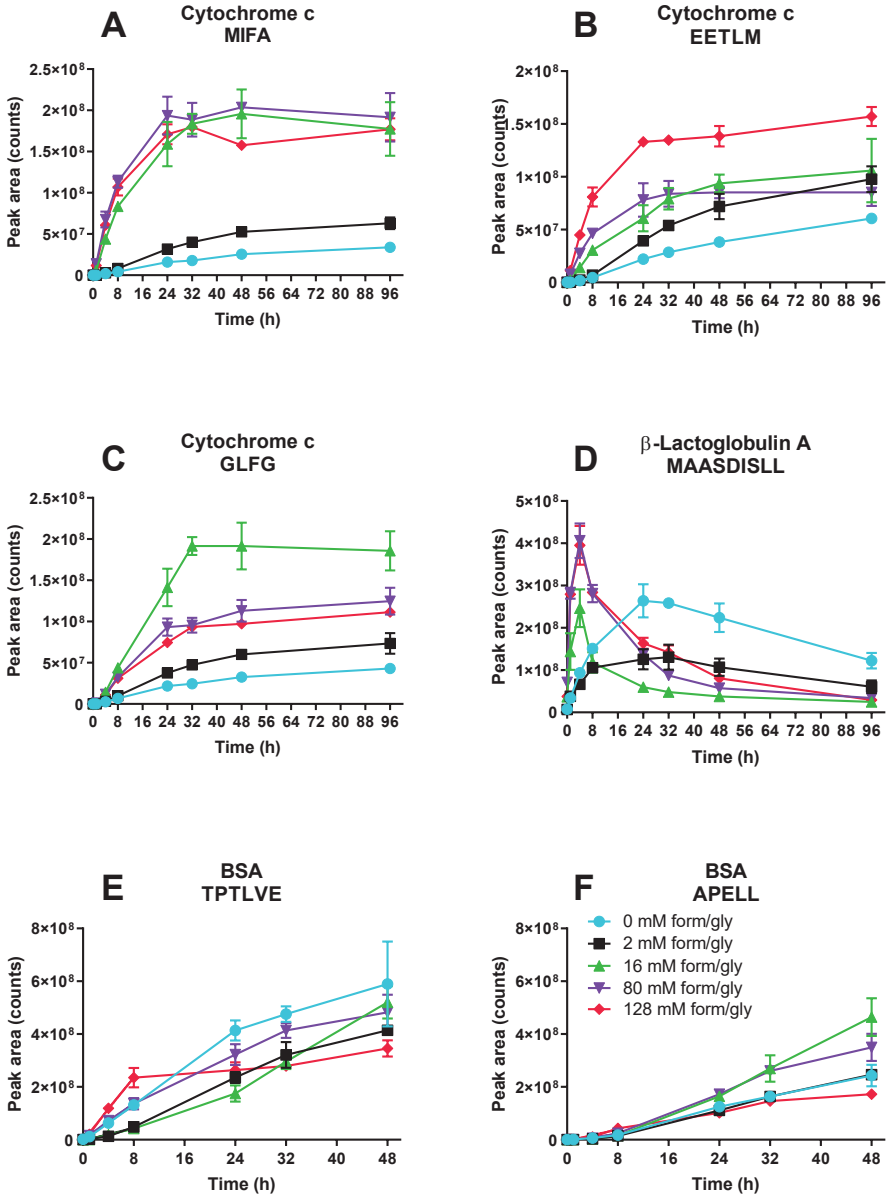


Figure 2. Illustrative examples of LC-MS analysis of degradation products that do not contain formaldehyde reactive amino acids and are derived from formaldehyde- and glycine-treated proteins, as function of cathepsin S digestion time. Error bars represent SD of three enzyme reactions. (A) Kinetics of the cytochrome c peptide MIFA. (B) Kinetics of the cytochrome c peptide EETLM. (C) Kinetics of the cytochrome c peptide GLFG. (D) Kinetics of the β-lactoglobulin A peptide MAASDISLL. (E) Kinetics of the BSA peptide TPTLVE. (F) Kinetics of the BSA peptide APELL.

TPTLVE showed a decrease in formation speed compared to the control when the formaldehyde and glycine concentration was low (2 and 16 mM), but higher concentrations (80 mM and 128 mM) showed a recovery of the original degradation speed. The APELL degradation profile is somewhat similar to cytochrome c's GLFG kinetics, with an optimum at intermediate concentrations (16 and 80 mM), although the differences in degradation rate between the treatments are small.

Role of glycine in the increased enzymatic degradation upon formaldehyde/glycine treatment

From the above, it cannot be concluded to what extent the presence of glycine contributes to the observed increase in enzymatic degradation. Therefore, we performed a follow-up experiment in which the most susceptible protein, cytochrome c, was treated with only formaldehyde (128 mM, **CYT-F128-1**) or with formaldehyde (128 mM) and glycine (128 mM, **CYT-FG128-1**) for one week, to evaluate whether formaldehyde-glycine adducts affect the increased proteolytic degradation rate.

Reaction **CYT-FG128-1** resulted in an increase in the formation of the GLFG peptide compared to both control reaction **CYT-FG0-1** and reaction **CYT-F128-1** (Fig. 3A). This indicates involvement of one or more formaldehyde-glycine adducts in the proteolytic process leading to the formation of GLFG. For the peptides EETLM and QAPGFT, however, **CYT-F128-1** and **CYT-FG128-1** resulted in a similar kinetic profile. So, modifications caused by formaldehyde itself were responsible for the increase in proteolytic degradation rate (Fig. 3C and D). For the MIFA peptide, both treatments showed a large increase in peptide formation rate compared to the control sample, although the formaldehyde-/glycine-treated sample showed a larger effect than the sample treated with only formaldehyde (Fig. 3B). Altogether, these results indicate that, depending on the proteolytic cleavage site, the formation of formaldehyde adducts and/or crosslinks and formaldehyde-glycine adducts contribute to the altered proteolytic peptide formation kinetics.

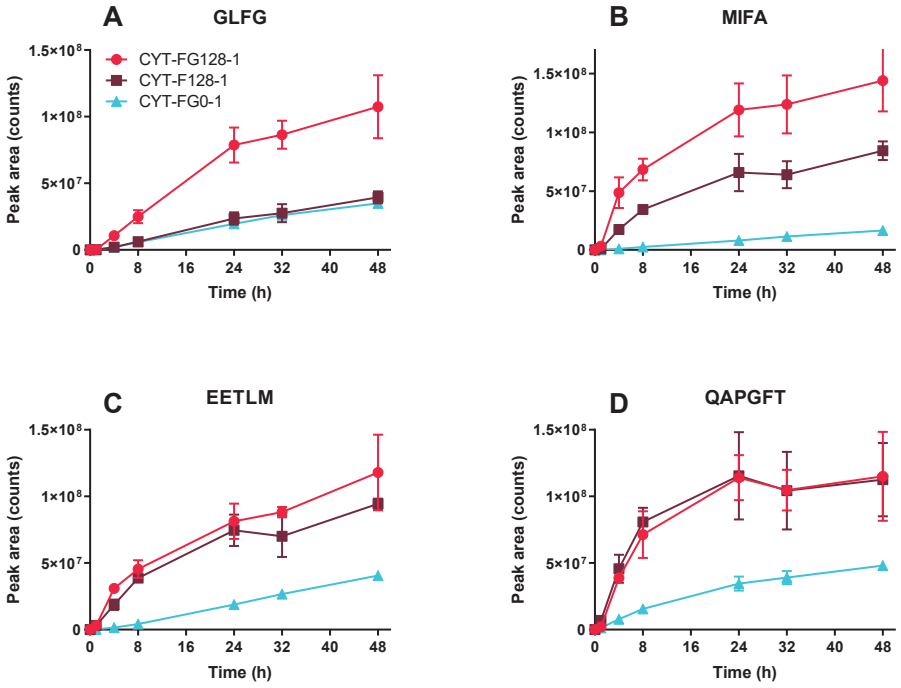


Figure 3. Effect of the presence of glycine during formaldehyde treatment on the subsequent enzymatic degradation rate of cytochrome c upon digestion by cathepsin S. Peptide formation was quantified by using LC-MS for samples/conditions **CYT-FG0-1**, **CYT-F128-1** and **CYT-FG128-1**. Peptide peak areas are corrected for the internal standard peak area. Error bars represent the SD of three LC-MS runs. (A) Kinetics of the cytochrome c peptide GLFG. (B) Kinetics of the cytochrome c peptide MIFA. (C) Kinetics of the cytochrome c peptide EETLM. (D) Kinetics of the cytochrome c peptide QAPGFT.

Location of the most abundant modifications and their solvent accessibility

By using stable isotope labelling for both formaldehyde (CD_2O) and glycine (^{13}C), doublet peaks in the MS1 spectra of modified peptides formed by cathepsin S digestion were obtained from samples **CYT-F128-1** and **CYT-FG128-1** (Fig. S2). Identification of these doublet peaks allowed us to elucidate the majority of the most abundant amino acid residue modifications and compare these to the residues' solvent accessibility. In combination with this data, the summed area responses of unmodified peptides originating from the protein can be plotted against the protein's amino acid sequence to distinguish areas of the protein where the enzymatic degradation is affected by the formaldehyde/glycine treatment (Fig. 4). This plot can then be used to reveal specific modifications responsible for altered enzymatic degradation kinetics. If more peptide is modified, the potential maximum amount of unmodified peptide decreases.

Several regions show clear differences in peptide abundance between the different treatments, which corresponds with differences in the modifications near these regions. For instance, the unmodified *N*-terminal peptides remaining in the formaldehyde-/glycine-treated sample showed a strong decrease in abundance compared to the control sample. This site contains three lysine residues that are susceptible to modification, especially imine formation (Δmass = +12 Da (Δm+12)) and methylation (Δm+14). Remarkably, R38 was modified both into a formaldehyde-glycine Δm+99 adduct and into a Δm+24 lysine crosslink and combinations thereof, despite having a relatively low (ca. 25%) accessibility. The region between Y48 and K60 also showed a low abundance of unmodified peptides for formaldehyde-/glycine-treated protein. For this region, crosslinks between the lysine residues and the tyrosine residue, imine formation on the lysine residues, and methylol formation on the asparagine residue were identified. Another important difference between **CYT-FG0-1** and **CYT-F128-1** and between **CYT-FG0-1** and **CYT-FG128-1** was observed in the region between N70 to K79. Here, the main identified modification was the crosslink between K79 and Y74. The preceding peptide EYL was more abundant in the samples exposed to formaldehyde, therefore this region showed a sharp increase in the overall peptide abundance compared to the control sample, despite Y67 being a potentially reactive residue. No modifications of this residue were identified, implying that the changes in its abundance must originate from modifications elsewhere, possibly the adjacent K79-Y74 crosslink. The unmodified residues near the C-terminus, such as Y97, K99 and K100, had similar profiles for **CYT-FG128-1** compared to **CYT-FG0-1**, with the former being slightly more intense. However, **CYT-F128-1** showed a dramatic decrease in unmodified peptide abundance in this region, indicating major changes compared to the other samples. Modifications on the more buried residues of a protein are more likely to affect protein folding than residues that are more solvent-exposed. Interestingly, the two arginine residues and all tyrosine residues involved in modifications have low solvent accessibility, implicating potential structural changes after modification. Overall, the most striking difference between **CYT-FG0-1** and both **CYT-F128-1** and **CYT-FG128-1** is that exactly the regions with unreactive or unreacted amino acids were far more abundant for **CYT-F128-1** and **CYT-FG128-1**. Metz *et al.* identified the most common modifications caused by formaldehyde/glycine treatment [14,15]. The modifications found in cytochrome c correspond to these modifications, with the addition of methylation of lysine residues in an Eschweiler-Clarke reaction [22]. This reaction has also previously been described for aqueous N^{α}-Acetyl-L-lysine upon exposure to formaldehyde [23].

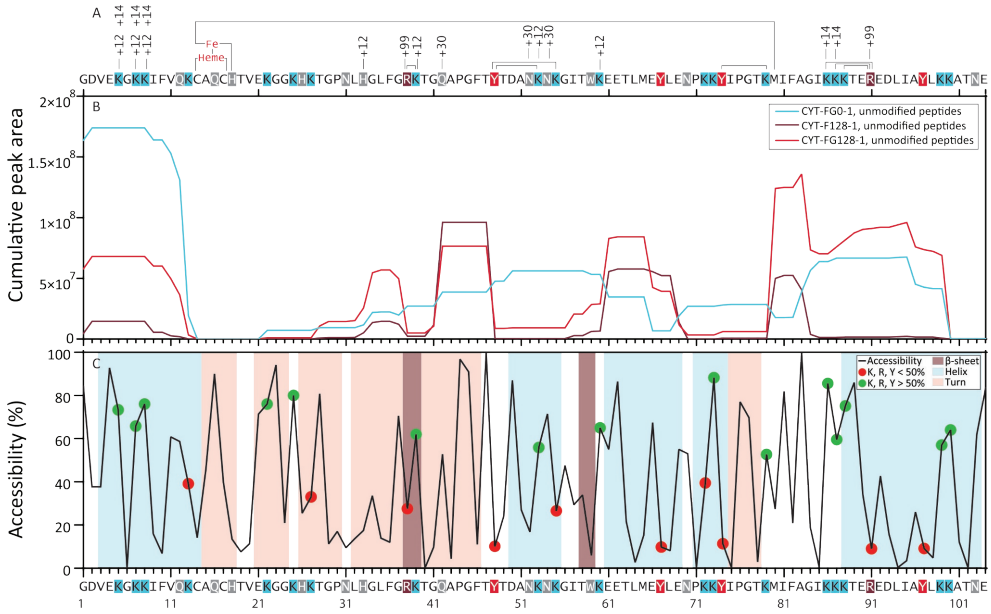


Figure 4. Graphical representation of formaldehyde-induced modifications and enzymatic degradation within the cytochrome c sequence. (A) All identified formaldehyde-induced modifications are annotated in this graph, crosslinks are represented by lines linking the crosslinked residues, imine formation is annotated with +12, methylation is annotated with +14, methylol adducts are annotated with +30, formaldehyde-glycine adducts on arginine residues are annotated with +99. See Table 1 for chemical structures of the modifications. Histidine 18 and methionine 81 are axial ligands to the haem iron, also indicated by a solid line. Lysine residues are marked blue, tyrosine residues are marked red, arginine residues are marked brown, and other potentially reactive amino acid residues are marked grey. (B) Peptide intensity relative to the amino acid sequence. The peptide areas that contained a certain amino acid were summed. Only peptides without formaldehyde modifications were used. The data used was obtained from the 8 hour time point of the cathepsin S digestion. Data is plotted for CYT-FG0-1, CYT-F128-1 and CYT-FG128-1. (C) Accessibility calculated using GETAREA on 1HRC [29]. Secondary structures were calculated using STRIDE on 1HRC [30]. Green dots represent accessible tyrosine, lysine, or arginine residues (>50%). Red dots represent less accessible tyrosine lysine or arginine residues (<50%).

Modification of R38 results in faster degradation of surrounding sequence

Some modifications that could explain the accelerated formation of the peptide GLFG upon exposure to formaldehyde and glycine were identified. In the cytochrome c sequence, GLFG is directly followed by R38, allowing for formaldehyde-glycine adducts. The peptide RKTGQAPGFT, immediately following GLFG in the cytochrome c sequence, was identified along with various related, modified peptides. The Δm+99 arginine modification (*i.e.*, 2 formaldehyde groups and 1 glycine group) and the Δm+111 (+99 on R and +12 on K) modifications were identified. The spectra contained the expected doublet peaks associated with addition of two formaldehyde molecules and one glycine molecule, and three formaldehyde molecules plus one glycine

molecule, respectively. Both products showed the characteristic neutral loss of 87.03 Da upon CID fragmentation. The MS1 peak intensities of these peaks were similar after 8 hours of digestion for all four peptides (Fig. S3).

Faster cleavage at GLFG↓RKTGQAPGFT is necessary to get faster formation of GLFG. Faster formation of the R(+99)KTGQAPGFT peptide compared to the unmodified peptide would indicate that formation of the Δm+99 modification contributed to the faster formation of GLFG in formaldehyde/glycine-treated samples. The intensities of the modified RKTGQAPGFT peptides during the various timepoints were normalised to compare the kinetic profile of these peptides (Fig. 5). Peptides that contained a Δm+99 adduct (*i.e.*, 2 formaldehyde groups and 1 glycine group) or Δm+111 adduct (99 + 12) were formed much faster than the native peptide or the peptide with the +12 adduct, the last two showing nearly identical formation kinetics. This further supports the finding that glycine adducts on R38 enhance proteolytic cleavage by cathepsin S, whereas modifications that do not involve glycine near this cleavage site (*e.g*, at K39) do not.

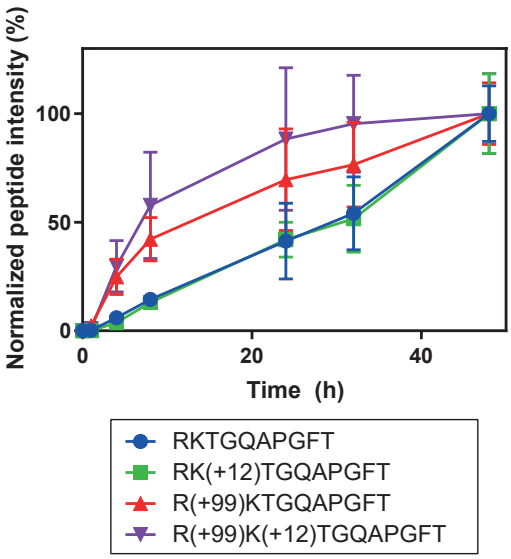


Figure 5. Normalized data of RKTGQAPGFT related peptides in **CYT-FG128-1** subjected to cathepsin S degradation. Peptide areas were first corrected for the internal standard peptide angiotensin-III. Peptides containing Δm+99 (formaldehyde-glycine) adducts show a faster formation than peptides without such adducts. Error bars represent SD of three LC-MS runs.

Enzymatic degradation of a modified synthetic peptide

The increased formation of the GLFG peptide upon digestion by cathepsin S of cytochrome c could have two reasons: local unfolding of the protein could make the cleavage site more accessible to the enzyme, or the enzyme could have a preference for the new, modified, arginine side chain. To rule out the latter, a synthetic peptide, Ac-GLFGRKTG, was subjected to formaldehyde/glycine treatment. The mixture of resulting peptides was further purified by SPE to remove free formaldehyde and glycine, and subsequently subjected to enzymatic degradation. Peptides containing a crosslink between the lysine and arginine (*i.e.*, a Δm+24 adduct) were much more resistant to proteolytic degradation, despite the cleavage site being at the N terminal side of the arginine. Peptides containing the Δm+99 adduct were not degraded substantially faster than the native peptide. These results strongly suggest that the observed increase in GLFG formation in the formaldehyde/glycine-treated cytochrome c is due to changes in protein folding (discussed below) induced by the chemical modifications rather than by the adduct formation *per se* (Fig. 6).

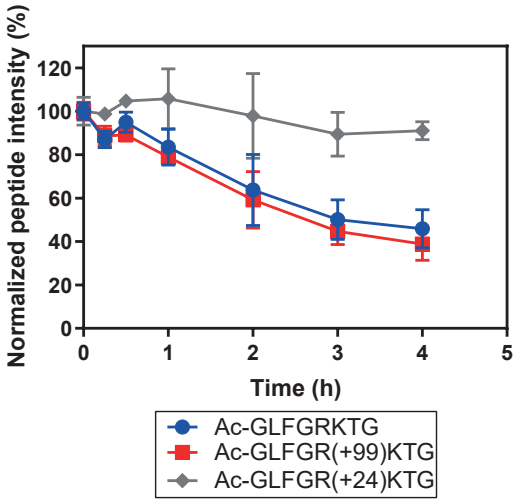


Figure 6. LC-MS peak area of the synthetic peptide Ac-GLFGRKTG and its modified analogues as function of cathepsin S digestion time. Peak areas are corrected by using an internal standard. The native peptide and its Δm+99 analogue show similar kinetics, whereas the crosslinked Δm+24 peptide shows a much slower degradation speed. Error bars represent SD of a triplicate of three peptide solutions.

Changes in secondary structure and thermostability of formaldehyde-treated cytochrome c. To be able to compare conformational differences between untreated and formaldehyde-/glycine-treated cytochrome c, CD spectra of **CYT-FG0-1**, **CYT-F128-1**, and **CYT-FG128-1** were measured and the ratio between the 208 and 222 nm signals was determined. **CYT-FG128-1** and **CYT-F128-1** resulted in different CD spectra compared both to **CYT-FG0-1** and to each other, indicating treatment-induced changes in secondary structure (Fig. 7). To compare the various samples, the ratio between the CD signals at 208 and 222 nm was determined. For **CYT-FG0-1** this ratio was 0.85, **CYT-FG128-1** and **CYT-F128-1** had ratios of 1.16 and 1.41, respectively. The unfolding temperature was determined by measuring CD at 220 nm. **CYT-FG0-1** unfolded near 78 °C while both **CYT-F128-1**, **CYT-FG128-1** did not unfold below 94 °C, indicating that these samples were more thermostable than native cytochrome c. In general, studies that correlate protein thermostability to stability towards proteolytic enzymes have shown that the more thermostable a protein is, the slower it is degraded by proteases [24,25]. However, these studies mostly use native proteins or proteins with natural amino acid mutations, whereas the structural changes caused by formaldehyde treatment could alter the conformation to make it more accessible to proteases; this treatment can also rigidify this altered conformation, which could lead to enhanced thermostability.

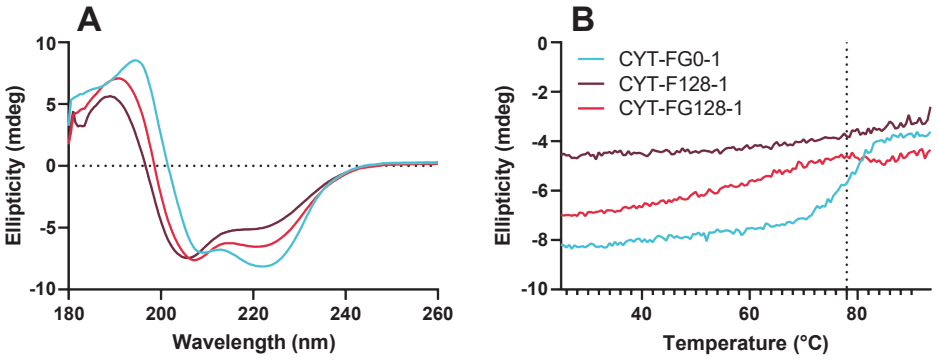


Figure 7. Circular dichroism spectroscopy analysis of **CYT-FG0-1**, **CYT-F128-1** and **CYT-FG128-1**. (A) Averaged (n=3) and smoothed (5 points) CD spectra of **CYT-FG0-1**, **CYT-F128-1** and **CYT-FG128-1** in 20 mM phosphate buffer, pH 7.4. (B) CD analysis of protein unfolding temperature (unsmoothed) measured at 220 nm.

Contrary to previous publications [17] and our expectations, the incubation of model proteins with glycine and formaldehyde enhances the protein's degradation speed upon exposure to cathepsin S. Cathepsin S has two important advantages over the previously used trypsin. Firstly, the lysine and arginine residues are among the most affected amino acids upon formaldehyde

treatment, which are the exact amino acid residues after which trypsin cleaves. It is very likely that modified lysine and arginine residues are not readily recognised by trypsin anymore. Secondly, the cathepsin S used in our study is a more biologically relevant protease in the context of antigen processing. Moreover, because of its preference for hydrophobic-branched amino acids in the P2, P1' and P3' position of the substrate, it does not have the same limitation regarding altered cleavage sites as trypsin does [26]. These hydrophobic-branched amino acids are not affected by formaldehyde, so only the effect of other neighbouring amino acids (primary structure) and overall protein folding (secondary and tertiary structure) contribute to the degradation kinetics.

Overall, subjecting proteins to formaldehyde or formaldehyde and glycine results in changes in proteolysis rates, leading to an enhanced degradation speed. For specific formaldehyde-glycine adducts we have shown that it is not due to changes in primary structure, but most likely due to changes in protein conformation. This effect is not limited to formaldehyde-glycine adducts, as it was also observed when the proteins were treated with only formaldehyde. Furthermore, using a model peptide, we have shown that formaldehyde-induced crosslinks increase protease resistance. This shows that while some modifications increase proteolytic resistance, the majority of the modifications actually enhance the susceptibility of the protein towards enzymatic proteolysis. Admittedly, the effects seen with the model proteins chosen in this study may not hold true for every protein. However, given the structural diversity of the model proteins used (*e.g.*, size, function, lysine content, disulphide bonds, proteolytic resistance), it is expected that many other proteins will behave similarly. This could have consequences for the immunogenicity and thus the efficacy of vaccine products, since endo-lysosomal degradation results in *in situ* formation of T-cell epitopes that provide induction of T-helper cells and cytotoxic T-cells [27]. With regard to vaccine production the chemical modifications caused by the detoxification processes, used for, *e.g.*, diphtheria and tetanus toxins or poliovirus vaccines, could alter antigen processing. This emphasises the importance of ensuring good batch to batch consistency, as variation in the detoxification process might result in variation of the efficacy of the final vaccine product. As variations in the detoxification process can alter the protease resistance of antigens, assays that measure the stability of a vaccine to enzymatic proteolysis have the potential to confirm a consistent detoxification process and in turn contribute to the reduction or replacement of animal tests when combined with other *in vitro* assays.

Experimental procedures

Preparation of formaldehyde and glycine treated proteins

Three model proteins were treated with formaldehyde and glycine, essentially as described by Metz *et al.* [14]. Bovine serum albumin (BSA, Serva), Bovine β-lactoglobulin A (Sigma-Aldrich) and equine cytochrome c (Sigma-Aldrich) were dissolved in phosphate buffered saline (PBS, pH 7.2, Gibco) to obtain 1.5 mg/mL stock solutions. To 700 µL of these stock solutions per reaction condition, glycine (1 M solution in PBS) and formaldehyde (1 M solution in PBS) were added to achieve 0, 2, 16, 80 or 128 mM formaldehyde and glycine in the final solution. Additional PBS was then added to obtain a total volume of 1400 µL. Samples were placed in an oven at 37 °C for 6 weeks. A buffer exchange to sodium citrate (100 mM pH 6.0) was then carried out by using Amicon 3-kDa spinfilters (Merck). Protein concentrations were determined by A280 measurements on a SynergyMX using a Take3 microplate (BioTek). Samples were stored at -20 °C until further treatment.

Preparation of cytochrome c treated with isotopically labelled formaldehyde and glycine

Cytochrome c was dissolved in PBS to obtain a 1.25-mg/mL stock solution. Formaldehyde and deuterated formaldehyde solutions were diluted in PBS to obtain 1 M solutions. Glycine and glycine-1-^{13}C were dissolved in water to obtain 1 M solutions. To aliquots of the cytochrome c solution, either unlabelled formaldehyde and glycine, D_2-formaldehyde and glycine-1-^{13}C, unlabelled formaldehyde, or D_2-formaldehyde solutions were added to achieve 1 mg/mL cytochrome c with 128 mM of the additives. A control sample was prepared with addition of PBS. All samples were then placed in an oven at 37 °C for 7 days. A buffer exchange to sodium citrate (100 mM pH 6.0) was then carried out by using Amicon 3-kDa spinfilters (Merck). Protein concentrations were determined by measuring A280 on a SynergyMX using a Take3 microplate (Biotek).

Enzymatic degradation of proteins

Protein samples (table 2) were subjected to proteolytic degradation by *E. coli*-derived recombinant human cathepsin S (Merck). Reactions were carried out in triplicate, in 200 µL total volume containing 100 mM sodium citrate buffer (pH 6.0), 2 mM ethylenediaminetetraacetic acid (EDTA) and 2 mM dithiothreitol (DTT), with 40 µg protein and 20 ng cathepsins S. Samples were placed in an oven at 37 °C. For the 6-week incubation samples, an aliquot of 10 µL was

transferred to 90 µL PBS containing 1 µM E-64 (Sigma-Aldrich) as cysteine protease inhibitor for subsequent SEC analysis upon reaching the desired time point. For all samples, an aliquot of 5 µL was transferred to 95 µL water containing 5 vol% DMSO, 0.1 vol% formic acid and 1 fmol/µL angiotensin-I (Sigma-Aldrich), angiotensin-III (Sigma-Aldrich) and oxytocin (Sigma-Aldrich) for LC-MS analysis. Samples for SEC analysis were stored at 4 °C; LC-MS samples were stored at-20 °C.

Table 2 Reaction conditions of protein formaldehyde treatment

Reaction condition code	Protein	Formaldehyde concentration (mM)	Glycine concentration (mM)	Incubation time (weeks)
BSA-FG0-6	Bovine serum albumin	0	0	6
BSA-FG2-6	Bovine serum albumin	2	2	6
BSA-FG16-6	Bovine serum albumin	16	16	6
BSA-FG80-6	Bovine serum albumin	80	80	6
BSA-FG128-6	Bovine serum albumin	128	128	6
BLG-FG0-6	ß-lactoglobulin A	0	0	6
BLG-FG2-6	ß-lactoglobulin A	2	2	6
BLG-FG16-6	ß-lactoglobulin A	16	16	6
BLG-FG80-6	ß-lactoglobulin A	80	80	6
BLG-FG128-6	ß-lactoglobulin A	128	128	6
CYT-FG0-6	Cytochrome c	0	0	6
CYT-FG2-6	Cytochrome c	2	2	6
CYT-FG16-6	Cytochrome c	16	16	6
CYT-FG80-6	Cytochrome c	80	80	6
CYT-FG128-6	Cytochrome c	128	128	6
CYT-FG0-1	Cytochrome c	0	0	1
CYT-F128-1	Cytochrome c	128[a]	0	1
CYT-FG128-1	Cytochrome c	128[a]	128[b]	1

[a)] The reaction was performed with 128 mM CH_2O and 128 mM CD_2O separately, and the reaction products were mixed 1:1 based on protein concentration afterwards. [b)] The reaction was performed with 128 mM glycine and 128 mM glycine-1-^{13}C separately, and the reaction products were mixed 1:1 based on protein concentration afterwards. CD_2O and glycine-1-^{13}C were used in the same reaction.

Preparation and enzymatic degradation of formaldehyde-modified synthetic peptide Ac-GLFGRKTG

Ten mg of the synthetic peptide Ac-GLFGRKTG (Pepscan) was dissolved in water (1 mL). Of this solution, 100 µL was added to 125 µL glycine solution (1 M), followed by 665 µL water, 100 µL PBS (10X Cal Biochem) and 10 µL formaldehyde solution (37 wt% 12.3 M). This mixture was placed in an oven at 37 °C for 24 hours. Excess formaldehyde was removed by using a C18 solid phase extraction (SPE) cartridge (Waters) with an ASPEC GX-271 automated SPE system (Gilson). The sample was loaded slowly (50 µL/min), washed with water containing 0.1 vol% formic acid (200 µL/min, 600 µL) and eluted with 60 vol% acetonitrile in water and 0.1 vol% formic acid (100 µL/min, 600 µL). This fraction was then collected and dried in a vacuum centrifuge (Eppendorf Concentrator plus) and reconstituted in 1 mL water to obtain an approximate 1-mg/mL mixture of Ac-GLFGRKTG and its modified analogues. The products were checked by LC-MS and were in agreement with previously published modifications [15]. Subsequently, 4 µL (4 µg peptide equivalent) of the peptide mixture was diluted in sodium citrate buffer (pH 6.0 100 mM) containing DTT (2 mM) and EDTA (2 mM) to a total volume of 200 µL. Then, cathepsin S (0.69 µL, 1 µg) was added and the sample was placed in an oven at 37 °C for 4 hours. This was performed in triplicate. At various time points (0 h, 15 min, 30 min, 1 h, 2 h, 3 h, 4 h), 1-µL aliquots were taken and diluted in 999 µL water containing 5 vol% DMSO, 0.1 vol% formic acid and 1 fmol/µL angiotensin-I, angiotensin-III and oxytocin and stored at-80 °C until LC-MS analysis.

LC-MS analysis

The partially digested proteins were analysed by reversed phase nanoscale LC-MS on an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher), essentially as described by Meiring et al [28]. For each analysis, 10 µL sample was injected and loaded onto the trapping column (Reprosil-Pur C18-AQ 5 µm, Dr. Maish, Germany; 20 mm long × 100 µm inner diameter). The trapping column was subsequently washed with 100% solvent A (0.1 vol% formic acid in water) for 10 minutes at a flow rate of 5 µL/min. After trapping, the flow was switched to the analytical column (Reprosil-Pur C18-AQ 3 µm, Dr. Maish, Germany; 25.6 cm long × 50 µm inner diameter) with a flow rate of 100 nL/min. A gradient with increasing solvent B (0.1 vol% formic acid in acetonitrile) was started from 8 vol% solvent B to 34 vol% B in 30 minutes, followed by a steeper gradient to 58 vol% in 5 minutes. The column was then washed with 85% B for 5 minutes and equilibrated for the succeeding analysis with 100 vol% A for 10

minutes. Using nanoelectrospray ionisation (ESI) the column was coupled to the Orbitrap Fusion Lumos Tribrid. Data Dependent Acquisition (DDA) was used, with MS1 scans measured in the 300-1500 m/z mass range, in the orbitrap at 120,000 FWHM resolution with a cycle time of 2 seconds, a maximum injection time of 100 ms and automatic gain control (AGC) target of $2 \cdot 10^5$. Fragmentation was achieved by using both collision-induced dissociation (CID) and electron transfer dissociation (ETD). The isolation window was set to 1.6 Da with an offset of 0.25 Da. The CID normalised collision energy was 35% with the activation Q set to 0.250. CID was used for abundant ($> 1 \cdot 10^6$) precursor ions with a charge state of 1, and precursor ions with charge states >1 and an intensity of at least $5 \cdot 10^5$. ETD reaction times were based on the calibration settings. ETD was used for abundant peptides ($>1 \cdot 10^6$) with charge states ≥ 2. MS2 scans were measured in the orbitrap at 15,000 FWHM resolution with an AGC target of $1 \cdot 10^5$, employing CID and ETD (100 ms and 200 ms injection times, respectively).

MS data analysis

The raw data was analysed by using PEAKS 8.5 (Bioinformatics Solutions Inc.) using the DENOVO, PEAKSDB, PTM and SPIDER and Label-Free Quantification (LFQ) modules. Known formaldehyde-glycine adducts were included as variable modifications in the DENOVO and PEAKSDB search [14,15]. The error tolerances were set to 3 ppm parent mass error tolerance and 0.01 Da fragment mass error tolerance The resulting peptide areas were corrected for the angiotensin-III internal standard. After 48 h of digestion, the isotopically labelled digestion samples **CYT-F128-1** and **CYT-FG128-1** were analysed with MsXelerator (MsMetrix) to obtain a list of isotopic doublets of peptides that should contain formaldehyde or formaldehyde-glycine adducts [21]. The 48 h samples were then measured again with a targeted analysis. Subsequently, another PEAKS search was performed to identify these modified products. MS spectra that PEAKS could not identify were manually interpreted. Accessibility was calculated by using GETAREA on 1HRC [29]. The location of α-helices, β-sheets and turns within cytochrome c were calculated by using STRIDE on 1HRC [30].

SEC

Degradation kinetics were followed by determining the maximum peak height of the intact protein with SEC using an ACQUITY UPLC 125 Å 1.7 µm 4.6 mm x 150 mm, 1K-80K Protein BEH SEC Column (Waters) and a Waters H-Class ACQUITY UPLC, with UV detection at 205 nm and column oven at 40 °C. The mobile phase was 100 mM sodium phosphate buffer at pH

6.8 for BSA and β-lactoglobulin, and 50 mM phosphate buffer with 450 mM NaCl at pH 8 for cytochrome c. The injection volume for all samples was 5 µL. The flow rate was 0.3 mL/min.

Circular dichroism (CD)

Prior to CD analysis, a buffer exchange through dialysis (3-kDa, Thermo Fisher) on **CYT-FG0-1**, **CYT-F128-1** and **CYT-FG128-1** was performed to achieve 20 mM phosphate buffer, pH 7.4. CD spectra were acquired on a Chirascan Circular Dichroism Spectrometer (Applied Photophysics). CD spectra were measured from 260 to 180 nm with steps of 0.5 nm. Spectra were measured in triplicate, averaged and smoothed over 5 points. For temperature ramping experiments performed in 100 mM sodium citrate buffer, pH 6.0, CD was measured at 220 nm and the temperature was increased 1 °C/min from 25 °C up to 94.5 °C.

Acknowledgements

We thank Joost Uittenbogaard for his advice on the identification of the formaldehyde-induced modifications of the studied proteins, and Marjolein Zohlandt for reviewing the manuscript. This project was funded by the Ministry of Agriculture, Nature and Food Quality, the Netherlands.

Author contributions statement

T.M. conceived and conducted experiments. H.M. provided technical LC-MS support and advice. T.M. wrote the original manuscript. All authors were involved in data interpretation and reviewing the manuscript. B.M., G.K. and W.J. were involved in conceptualisation, funding acquisition and supervision.

Additional information

Competing interest

The authors declare no competing interests.

References

1. Delamarre, L., Pack, M., Chang, H., Mellman, I. & Trombetta, E. S. Differential lysosomal proteolysis in antigen-presenting cells determines antigen fate. *Science* **307**, 1630-1634, doi:10.1126/science.1108003 (2005).
2. Carmicle, S., Dai, G., Steede, N. K. & Landry, S. J. Proteolytic sensitivity and helper T-cell epitope immunodominance associated with the mobile loop in Hsp10s. *J Biol Chem* **277**, 155-160, doi:10.1074/jbc.M107624200 (2002).
3. Kim, A. *et al.* Divergent paths for the selection of immunodominant epitopes from distinct antigenic sources. *Nat Commun* **5**, 5369, doi:10.1038/ncomms6369 (2014).
4. Egger, M. *et al.* Assessing protein immunogenicity with a dendritic cell line-derived endolysosomal degradome. *PLoS One* **6**, e17278, doi:10.1371/journal.pone.0017278 (2011).
5. Delamarre, L., Couture, R., Mellman, I. & Trombetta, E. S. Enhancing immunogenicity by limiting susceptibility to lysosomal proteolysis. *J Exp Med* **203**, 2049-2055, doi:10.1084/jem.20052442 (2006).
6. Ackaert, C. *et al.* The impact of nitration on the structure and immunogenicity of the major birch pollen allergen Bet v 1.0101. *PLoS One* **9**, e104520, doi:10.1371/journal.pone.0104520 (2014).
7. Freier, R., Dall, E. & Brandstetter, H. Protease recognition sites in Bet v 1a are cryptic, explaining its slow processing relevant to its allergenicity. *Sci Rep* **5**, 12707, doi:10.1038/srep12707 (2015).
8. Schulten, V. *et al.* Pru p 3, the nonspecific lipid transfer protein from peach, dominates the immune response to its homolog in hazelnut. *Allergy* **66**, 1005-1013, doi:10.1111/j.1398-9995.2011.02567.x (2011).
9. Kitzmuller, C. *et al.* A hypoallergenic variant of the major birch pollen allergen shows distinct characteristics in antigen processing and T-cell activation. *Allergy* **67**, 1375-1382, doi:10.1111/all.12016 (2012).
10. Kitzmuller, C. *et al.* Correlation of sensitizing capacity and T-cell recognition within the Bet v 1 family. *J Allergy Clin Immunol* **136**, 151-158, doi:10.1016/j.jaci.2014.12.1928 (2015).
11. De Mattia, F. *et al.* The vaccines consistency approach project: an EPAA initiative. *Pharmeur Bio Sci Notes* **2015**, 30-56 (2015).
12. Halder, M. *et al.* Recommendations of the VAC2VAC workshop on the design of multi-centre validation studies. *Biologicals* **52**, 78-82, doi:10.1016/j.biologicals.2018.01.003 (2018).
13. De Mattia, F. *et al.* The consistency approach for quality control of vaccines- a strategy to improve quality control and implement 3Rs. *Biologicals* **39**, 59-65, doi:10.1016/j.biologicals.2010.12.001 (2011).
14. Metz, B. *et al.* Identification of formaldehyde-induced modifications in proteins: reactions with insulin. *Bioconjug Chem* **17**, 815-822, doi:10.1021/bc050340f (2006).
15. Metz, B. *et al.* Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. *J Biol Chem* **279**, 6235-6243, doi:10.1074/jbc.M310752200 (2004).
16. 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* **22**, 156-167 (2003).
17. di Tommaso, A. *et al.* Formaldehyde treatment of proteins can constrain presentation to T cells by limiting antigen processing. *Infect Immun* **62**, 1830-1834 (1994).
18. Golghalyani, V., Neupartl, M., Wittig, I., Bahr, U. & Karas, M. ArgC-Like Digestion: Complementary or Alternative to Tryptic Digestion? *J Proteome Res* **16**, 978-987, doi:10.1021/acs.jproteome.6b00921 (2017).
19. Colbert, J. D., Matthews, S. P., Miller, G. & Watts, C. Diverse regulatory roles for lysosomal proteases in the immune response. *Eur J Immunol* **39**, 2955-2965, doi:10.1002/eji.200939650 (2009).

20. Cramer, C. J., Johnson, J. L. & Kamel, A. M. Prediction of Mass Spectral Response Factors from Predicted Chemometric Data for Druglike Molecules. *J Am Soc Mass Spectrom* **28**, 278-285, doi:10.1007/s13361-016-1536-4 (2017).

21. Metz, B. *et al.* Identification of formaldehyde-induced modifications in diphtheria toxin. *J Pharm Sci*, doi:10.1016/j.xphs.2019.10.047 (2019).

22. Eschweiler, W. Ersatz von an Stickstoff gebundenen Wasserstoffatomen durch die Methylgruppe mit Hülfe von Formaldehyd. *Berichte der deutschen chemischen Gesellschaft* **38**, 880-882, doi:10.1002/cber.190503801154 (1905).

23. 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* **214**, 289-292 (1983).

24. Daniel, R. M., Cowan, D. A., Morgan, H. W. & Curran, M. P. A correlation between protein thermostability and resistance to proteolysis. *Biochem J* **207**, 641-644 (1982).

25. Ahmad, S., Kumar, V., Ramanand, K. B. & Rao, N. M. Probing protein stability and proteolytic resistance by loop scanning: a comprehensive mutational analysis. *Protein Sci* **21**, 433-446, doi:10.1002/pro.2029 (2012).

26. Neil D. Rawlings, G. S. S. *Handbook of Proteolytic Enzymes*. Vol. 2 1824 (Academic Press, 2013).

27. Rock, K. L., Reits, E. & Neefjes, J. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends Immunol* **37**, 724-737, doi:10.1016/j.it.2016.08.010 (2016).

28. 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* **25**, 557-568, doi:doi:10.1002/1615-9314(20020601)25:9<557::AID-JSSC557>3.0.CO;2-F (2002).

29. Fraczkiewicz, R. & Braun, W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *Journal of Computational Chemistry* **19**, 319-333, doi:10.1002/(sici)1096-987x(199802)19:3<319::aid-jcc6>3.0.co;2-w (1998).

30. Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins: Structure, Function, and Bioinformatics* **23**, 566-579, doi:10.1002/prot.340230412 (1995).

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.

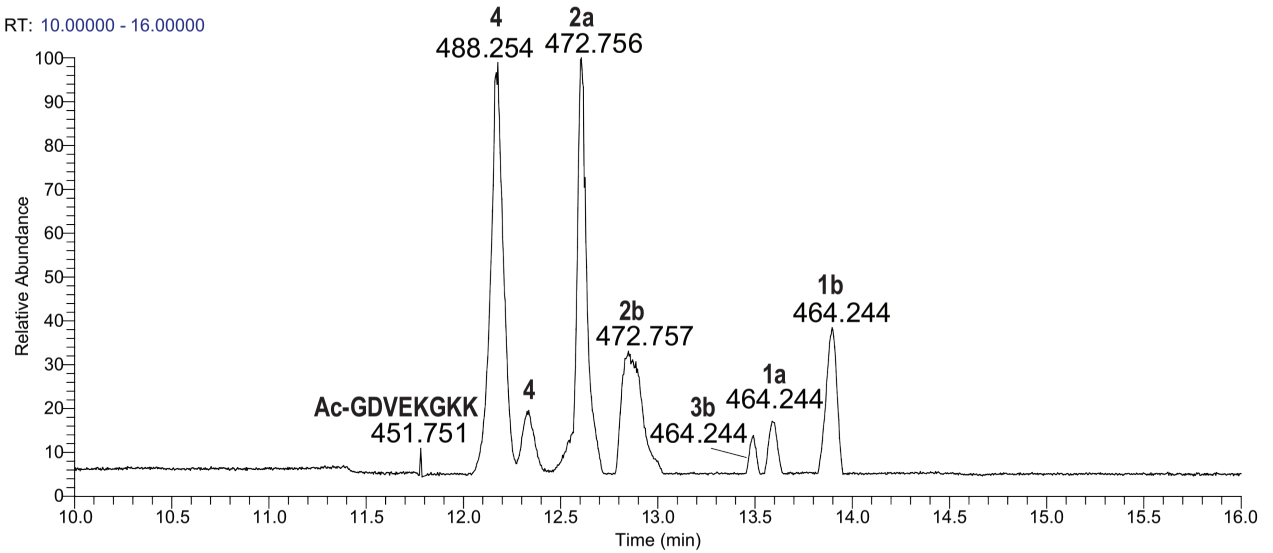
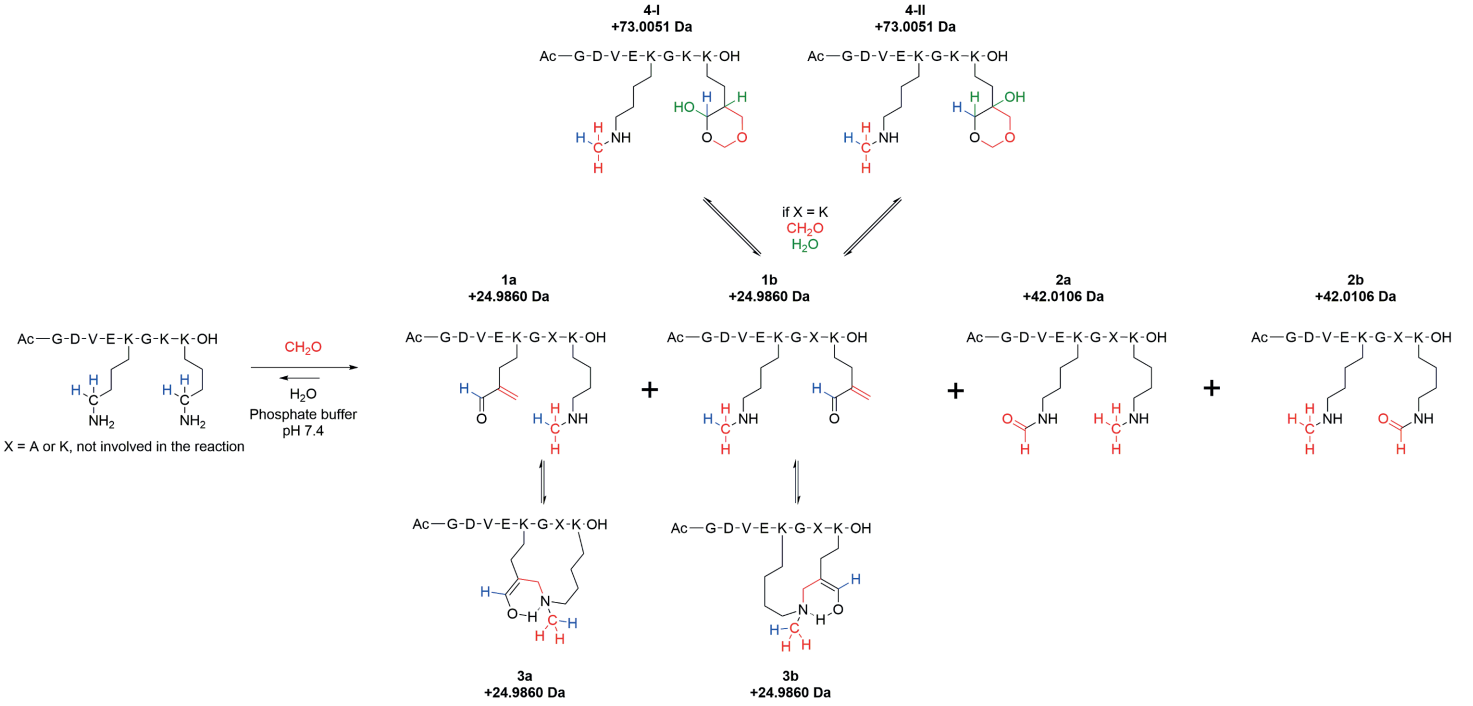


Figure 1. Representative nanoscale LC-MS base peak (m/z 400-500) chromatogram of synthetic Ac-GDVEKGKK treated with aqueous CH_2O for 46 hours (z=2 for all peaks). The exact regioisomers of the two peaks corresponding to product **4** could not be assigned.



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).

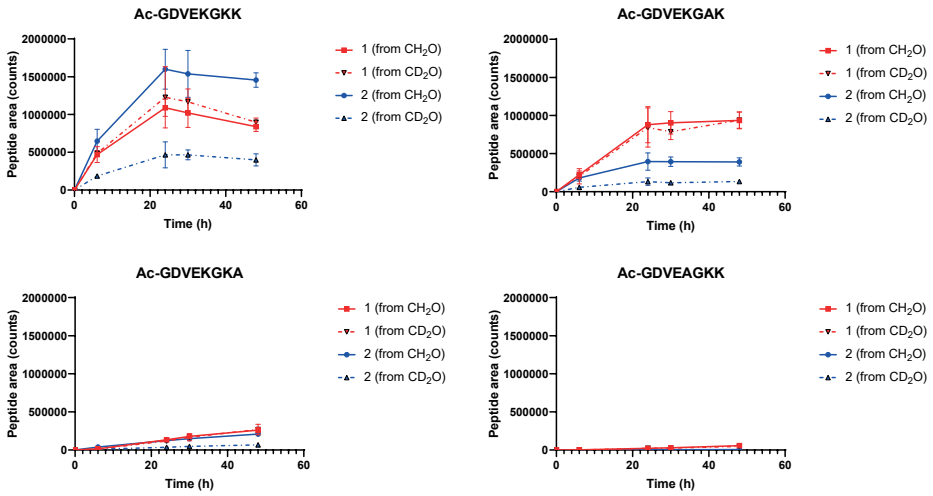


Figure 2. Comparison of the reaction between synthetic analogues of peptide Ac-GDVEKGKK (where one of the various lysine residues is replaced with an alanine residue) and the use of CH_2O or CD_2O on the formation of products **1** and **2**. The area of the regioisomers has been summed.

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.

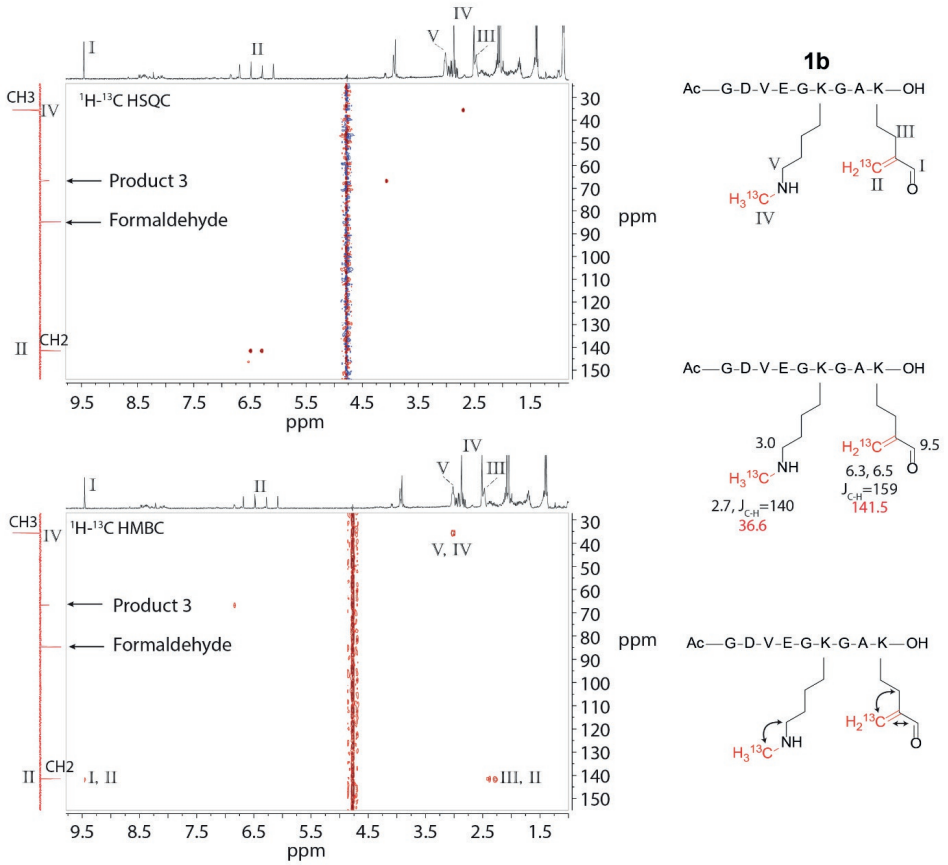


Figure 3. ^{1}H-^{13}C HSQC and ^{1}H-^{13}C HMBC of purified **1b** obtained from a reaction with Ac-GDVEKGAK with ^{13}C labeled formaldehyde (adducts containing ^{13}C are marked in red, ^{13}C NMR shifts are also marked in red). F1: DEPT-135 and F2: Robust5-^{1}H (for water suppression) spectra.

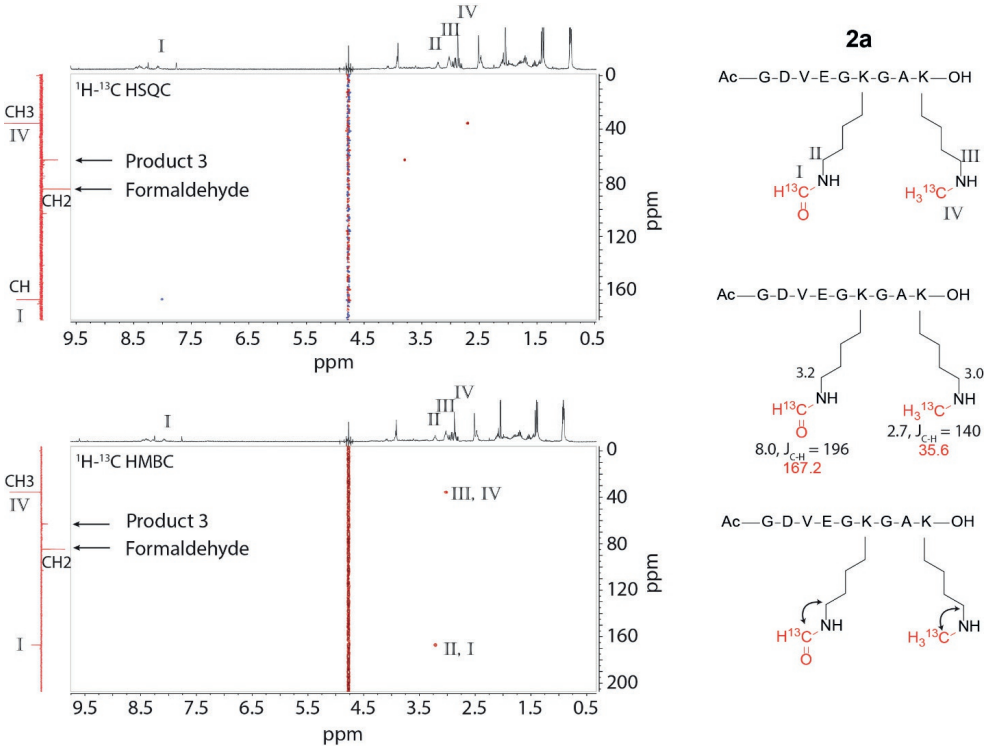


Figure 4. ^{1}H-^{13}C HSQC and ^{1}H-^{13}C HMBC of purified **2a** obtained from a reaction with Ac-GDVEKGAK with ^{13}C labeled formaldehyde (adducts containing ^{13}C are marked in red, ^{13}C NMR shifts are also marked in red). F1: DEPT-135 and F2: Robust5-^{1}H (for water suppression) spectra.

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.

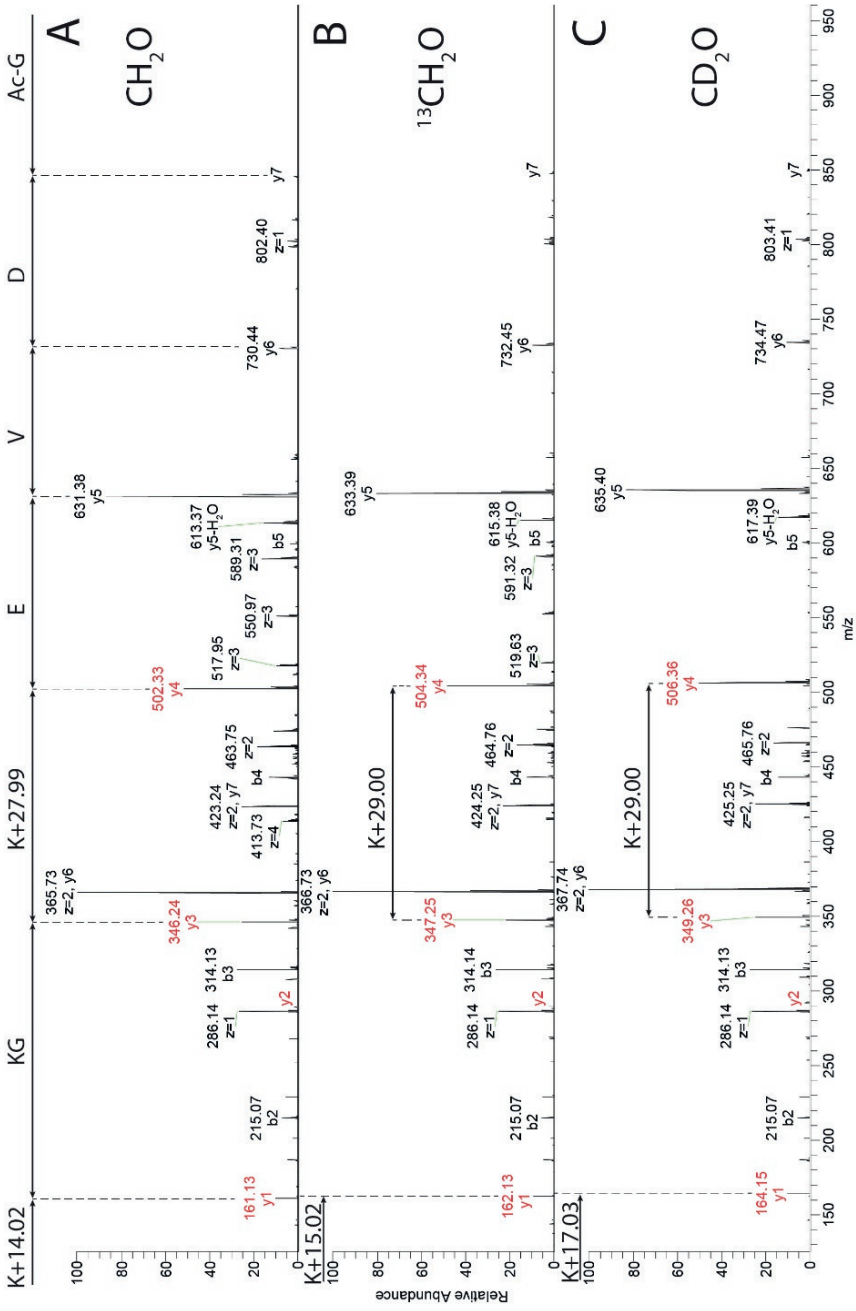


Figure 5. High resolution MS2 of **2a**. MS2 spectrum of Ac-GDVEKGKK treated with A) CH_2O, B) $^{13}CH_2O$, C) CD_2O.

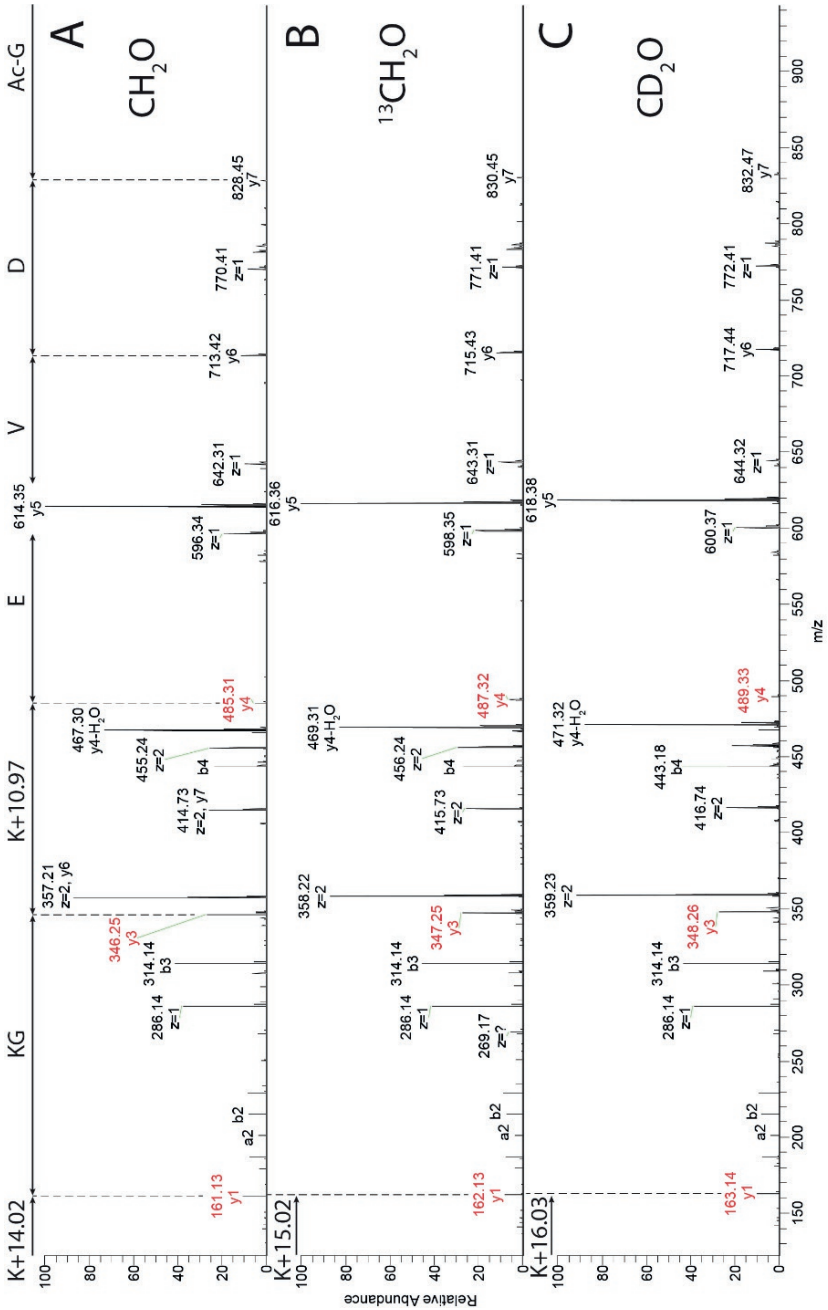


Figure 6. High resolution MS2 of **1a**. MS2 spectrum of Ac-GDVEKGKK treated with A) CH_2O, B) $^{13}CH_2O$, C) CD_2O.

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.

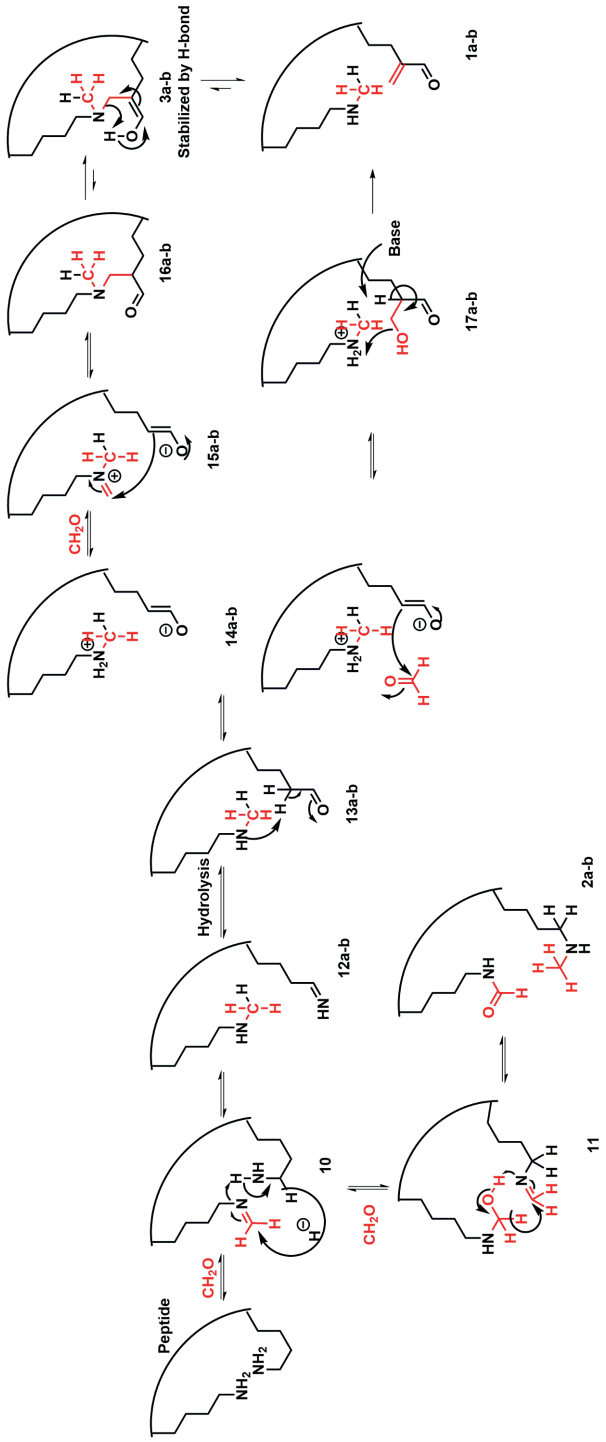
Discussion

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.

5

DEGRADOMICS-BASED ANALYSIS OF TETANUS TOXOIDS AS A QUALITY CONTROL ASSAY

Thomas J. M. Michiels [1,2], Wichard Tilstra [2], Martin R. J. Hamzink [2], Justin W. de Ridder [2], Maarten Danial [2], Hugo D. Meiring [2], Gideon F. A. Kersten [1,2], Wim Jiskoot [1] and 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

Vaccines **2020**, *8* (4), 712.

Abstract

Currently, batch release of toxoid vaccines, such as diphtheria and tetanus toxoid, requires animal tests to confirm safety and immunogenicity. Efforts are being made to replace these tests with *in vitro* assays in a consistency approach. Limitations of current *in vitro* assays include the need for reference antigens and most only being applicable to drug substance, not to the aluminum adjuvant-containing and often multivalent drug product. To overcome these issues, a new assay was developed based on mimicking the proteolytic degradation processes in antigen-presenting cells with recombinant cathepsin S, followed by absolute quantification of the formed peptides by liquid chromatography-mass spectrometry. Temperature-exposed tetanus toxoids from several manufacturers were used as aberrant samples and could easily be distinguished from the untreated controls by using the newly developed degradomics assay. Consistency of various batches of a single manufacturer could also be determined. Moreover, the assay was shown to be applicable to $Al(OH)_3$ and $AlPO_4$-adsorbed tetanus toxoids. Overall, the assay shows potential for use in both stability studies and as an alternative for *in vivo* potency studies by showing batch-to-batch consistency of bulk toxoids as well as for aluminum-containing vaccines.

Introduction

Tetanus is a severe disease caused by *Clostridium tetani*. Since their development in 1924, tetanus toxoid (TTd) vaccines have dramatically reduced the incidence and the case fatality rate of tetanus [1]. Because tetanus is transmitted by *C. tetani* spores from dirt or soil infecting wounds and not through infected patients, vaccination will always remain an important tool to avoid resurgence of the disease. Similar to the first TTd vaccines [2], the vaccines used today are still based on the principle of inactivating tetanus toxin with formaldehyde. Although these vaccine products are not as pure or well defined as other biologicals (*e.g.*, recombinant insulin or therapeutic antibodies) [3–5], they have been used for decades and are safe and efficacious. Currently, *in vivo* tests are required for batch release to confirm safety (*i.e.*, effective detoxification) and immunogenicity [3,6]. Efforts are being made to implement a consistency approach, where manufacturers that can consistently produce the drug product would only have to prove the potency of a limited number of batches using an animal model and verify consistency of every subsequent batch with *in vitro* assays [7–9]. This could significantly reduce the number of animals needed to verify the potency of toxoid vaccine batches.

In search of a new animal-free assay to confirm the quality of a TTd batch, we looked into the adaptive immune system for inspiration. An important aspect of the adaptive immune system is cellular immunity, in which antigen processing by antigen presenting cells (APCs) plays an important role. After the antigen has been taken up by APCs, proteases in the lysosomes cleave the antigenic proteins and form peptides which are eventually presented by MHC molecules to T-cells. A correlation between slower antigen processing and a stronger immune response suggests that the antigen processing kinetics could be used as a metric for immunogenicity [10–16]. We hypothesized that by mimicking proteolytic antigen degradation *in vitro* with the lysosomal endoprotease cathepsin S, the obtained degradation kinetics have the potential to verify the quality of the TTd. A previous study on model proteins showed that differences in formaldehyde-inactivation could be picked up in such an assay [17].

Our aim was to develop a new and reliable mass spectrometry based *in vitro* assay to show TTd production consistency and (bulk) product stability by measuring cathepsin S degradation. The main advantage of such an assay over other *in vitro* assays is that it does not need a biological reference but can use well defined synthetic peptides as a standard. To verify that our test could identify aberrant products, heat exposed plain tetanus toxoids (drug

substance) as well as control batches of drug substance from various manufacturers were analyzed and the results of the degradomics assay were compared to more conventional techniques (*e.g.*, circular dichroism, fluorescence, and antigenicity). Finally, the bulk TTd was adsorbed to $Al(OH)_3$ and $AlPO_4$ to explore the potential of the assay on analyzing aluminum-adsorbed TTd. TTd vaccines are stable for years, but temperature exposure can accelerate loss of potency [18,19]. Exposure to 55 °C for three days has been reported to decrease TTd potency by 48%, whereas exposure to 53 °C for four days only decreased potency by 17%, indicating a turning point in TTd stability near these temperatures [20]. For our aberrant samples, we chose temperatures that would cover this range.

Materials and Methods

Preparation of toxoids

Tetanus toxoids were obtained from three sources: In-house produced research grade TTd and GMP grade TTd from two manufacturers from the VAC2VAC consortium (Table 1). The toxoids were split in triplicates and thoroughly dialyzed (10-kDa MWCO, Thermo Scientific) against 10 mM phosphate, pH 7.2 (obtained as 1 M solution from Sigma Aldrich). Subsequently, protein concentrations were determined by BCA and 5-mL samples at 100 µg/mL were then stored at various temperatures (Table 1). TTd-B1.3–TTd-B1.5 were not dialyzed prior to adsorption. The toxoids were adsorbed to $Al(OH)_3$ (Brenntag, final concentration 1 mg/mL Al) and $AlPO_4$ (Brenntag, final concentration 0.66 mg/mL Al) in presence of 50 mM phosphate, pH 7.2, to obtain a final TTd concentration of 20 Lf/mL. After two days of mixing at 4 °C, the adsorption efficiency was determined relative to the control samples. Subsequent temperature exposure was performed in accordance to Table 1.

Table 1. Overview of the tetanus toxoid batches used in this study.

Name	Source	Experiment	Temperature Exposure (Duration)
TTd-A1.1	In-house	Peptide selection and evaluation of kinetics	4, 50, 55, 60, 65 °C, (2 days) 37 °C, (30 days)
TTd-A1.2	In-house	Quantification with isotopically labeled standard	4, 37, 45, 50, 55, 60, 65 °C, (2 days)
TTd-B1.1	Manufacturer B	Quantification with isotopically labeled standard	4, 37, 45, 50, 55, 60, 65 °C, (2 days)
TTd-C1	Manufacturer C	Quantification with isotopically labeled standard	4, 37, 45, 50, 55, 60, 65 °C, (2 days)
TTd-B1.2	Manufacturer B	Evaluation of batch consistency	4 °C
TTd-B1PC	Manufacturer B	Evaluation of batch consistency positive control	55 °C, (2 days)
TTd-B2	Manufacturer B	Evaluation of batch consistency	4 °C
TTd-B3	Manufacturer B	Evaluation of batch consistency	4 °C
TTd-B4	Manufacturer B	Evaluation of batch consistency	4 °C
TTd-B5	Manufacturer B	Evaluation of batch consistency	4 °C
DTd	Manufacturer B Diphtheria toxoid	Evaluation of batch consistency negative control	4 °C
TTd-B1.3	Manufacturer B	Adsorbed toxoids control (no adjuvant added)	4, 37, 45, 50, 55, 60, 65 °C, (2 days)
TTd-B1.4	Manufacturer B	Adsorbed toxoids $Al(OH)_3$ adsorbed	4, 37, 45, 50, 55, 60, 65 °C, (2 days)
TTd-B1.5	Manufacturer B	Adsorbed toxoids $AlPO_4$ adsorbed	4, 37, 45, 50, 55, 60, 65 °C, (2 days)

Enzymatic degradation and LC-MS analysis

Two types of enzymatic degradation experiments were performed. In the first experiment, both the formation kinetics and identity of the TTd derived peptides were determined, followed by selecting three temperature sensitive peptides. The selection criteria for these peptides were: high overall peak area, low degrees of asparagine deamidation and methionine oxidation, no peptide length variants of the same part of the sequence, and the peptides having increasing intensities during the course of the digestion. In the second type of experiments the previously selected peptides were quantified after 20 h of digestion. Partial enzymatic digestions were carried out with 10 µg or 2 Lf TTd with 0.2 µg recombinant human cathepsin S (*E. coli*, activity: 80 U/µg, Merck) in the presence of 2 mM dithiothreitol, 2 mM ethylenediaminetetraacetic acid, and 100 mM sodium citrate, pH 5.0, at 37 °C. In the peptide selection and kinetics experiment, aliquots were diluted 20x in water containing 0.1 vol% formic acid with 1 fmol/µL angiotensin-III (Sigma Aldrich) after 0, 1, 4, 9, 26, and 32 h of incubation. When using stable isotope labeled (SIL) internal standards (Pepscan), at 20 h, 100 µL 0.1 vol% formic acid containing the following peptides was added: E-64 (Sigma Aldrich, protease inhibitor, 0.01 mM); EDNNITLK*, NLDRILR* and ASNWYFNHLK* at 25 fmol/µL; ASDWYFNHLK*, ASNWYFDHLK*, ASDWYFDHLK*, EDNDITLK*, EDDNITLK*, EDDDITLK*

and DLDRILR* at 2.5 fmol/µL (K* contains 6 ^{13}C and 2 ^{15}N, R* contains 6 ^{13}C and 4 ^{15}N). Subsequently, the samples were purified by solid phase extraction (Sep-Pak C18, Waters) by washing with water containing 0.1 vol% formic acid (1 mL) followed by elution of the peptides with 600 µL 60 vol% acetonitrile containing 0.1% vol% formic acid. The eluent was dried and reconstituted in 15 µL DMSO followed by 285 µL 0.1 vol% formic acid. Samples were analyzed on a nanoLC-MS system on a Fusion Lumos Tribrid (ThermoFisher) as described elsewhere [17,21]. A 100-µm (I.D.) × 20-mm (L) trapping column packed with 5-µm Reprosil-Pur C18-AQ particles followed by a 50-µm (I.D.) × 31.5-cm (L) analytical column packed with 3-µm Reprosil-Pur C18-AQ particles were used for the separation of the peptides. The peptide mixtures were first trapped onto the trap column for 10 min at a flow rate of 5 µL/min (100% eluent A (0.1 vol% formic acid in water). The subsequent analytical separation was performed at a column flow rate of approximately 125 nL/min, achieved with a precolumn flow restrictor, all in conjunction with an Agilent 1290 Infinity HPLC system. A gradient of 10 to 40% eluent B (acetonitrile containing 0.1 vol% formic acid) in 10 min, followed by a 5-min gradient to 85% B, was used for the analysis of peptides in conjunction with a SIL standard. For label-free quantification, a gradient of 10 to 35% B in 25 min followed by a 35 to 45% B in 10 min gradient was used. In-house-prepared gold/carbon-coated spray tips were used for electro-spray ionization. For the detection, an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher, Waltham (Ma), United States of America) was used with the following global settings for the label-free quantification: spray voltage 2100 V, Data Dependent Acquisition (DDA) with a maximum cycle time of 1.5 s between subsequent MS1 scans; MS1 scans were performed using the Orbitrap detector at a resolution of 30,000, maximum injection time of 50 ms and automatic gain control (AGC) target of 200,000. MS2 scans were performed in the Orbitrap detector at 15,000 resolution. For MS2 fragmentation, the included charge states were 1 to 8. Singly charged ions were only fragmented (by Collision Induced Dissociation (CID)) if their intensity was higher than 500,000 counts. Precursor ions with charge states of 2 to 4 were fragmented (CID) when the intensity was higher than 50,000 counts. Precursor ions with charge states 3 to 8 were also subjected to Electron Transfer Dissociation (ETD) if their intensities were higher than 200,000 counts. For experiments where SIL internal standard peptides were added, the following settings were adjusted to maximize MS1 performance: MS1 resolution 50,000, cycle time 1.0 s, MS2 detection in the ion trap, included charge states were 2 to 6 and only CID fragmentation was performed. Label free quantification and identification of peptides for the analysis of

TTd-A1.1 was done with the QUAN module in PEAKS Studio X (Bioinformatics Solutions). Precursor mass tolerance was set to 5 ppm, fragment mass tolerance was set to 0.02 Da, enzyme was set to unspecific, variable modifications considered were: deamidation (NQ), oxidation (M) and potential formaldehyde adducts[22] (cyclic N terminus (+12), imine (+12) and methylol formation (+30.01), crosslinks between R and K (+24) and crosslinks between K and Y (+12)). A maximum of three variable PTMs were considered per peptide. No charge state filtering was applied to include the z = 1 peptides formed by cathepsin S digestion. The Quan module includes peptides with a false discovery rate of <1%. Peptide lists were exported and corrected for the relative abundance of the added angiotensin-III (Table S1). Quantification of the three selected peptides was achieved by comparing peak height in Xcalibur Quan (Thermo Fisher, Waltham (MA), United States of America), the deamidated forms were quantified by their total peak area to include iso-aspartate (Table S2 and Table S3).

Protein folding: tryptophan fluorescence was measured on a fluorescence spectrometer (LS55, PerkinElmer). One mL sample in a 1-cm quartz cuvette was measured with a high gain from 300 to 450 nm with a scan speed of 50 nm/minute and an excitation wavelength of 295 nm. Subsequently, Bis-ANS fluorescence was measured by addition of a 4,4′-Dianilino-1,1′-Binaphthyl-5,5′-Disulfonic Acid (Bis-ANS, Invitrogen, Carlsbad (CA), United States of America) solution to a final concentration of 10 µM to the non-adsorbed TTd. Emission fluorescence spectra were recorded after 20 min incubation at room temperature on a fluorescence spectrometer (LS55, PerkinElmer, Waltham (Ma), United States of America) from 430 to 580 nm with a scan speed of 50 nm/minute and an excitation wavelength of 395 nm. Far-UV circular dichroism spectra (185 to 260 nm in steps of 0.5 nm) were recorded on a Chirascan circular dichroism spectrometer (Applied Photophysics, Leatherhead, United Kingdom). Measurements were performed on 400 µL TTd sample in a 1-mm quartz cuvette using a band width of 1 nm and a response time of 0.5 s at each wavelength point. Data of far-UV spectra were averaged over 10 subsequent scans per replicate. Higher order structure analysis was performed with qBiC (v1.1.2, Applied Photophysics, Leatherhead, United Kingdom) in accordance with the manufacturer's recommendations. Note: samples were always compared to the 4 °C controls from the same source; Z-scores cannot be compared between different products.

Antibody binding

TTd binding to monoclonal antibody 10/134 (NIBSC) was measured on a Biacore T100 (GE Healthcare) similar to methods described elsewhere for diphtheria toxoids [23]. 0.1 mg/mL goat-anti-rat IgG UNLB (Southern Biotech) was immobilized on a CM5 chip (GE Healthcare) and used to subsequently capture mAb 10/134 at a flow of 10 µL/min for 5 min. Maximum binding of the TTd was measured after a 5 min flow of 5 µL/min. Responses were normalized relative to the 4 °C controls.

Monomer quantification

The TTd monomer content was determined by size exclusion chromatography. Samples were centrifuged for 60 min at 17,000 g to remove large aggregates prior to analysis. Four µL of the supernatant was injected onto a ACQUITY UPLC Protein BEH SEC Column (200A, 1.7 µm, 4.6 mm × 300 mm, 10k–500k, Waters) with phosphate buffered saline (Gibco) at a 0.3 mL/min flowrate for 18 min. Detection was performed by using UV absorbance at 215 nm.

Results and Discussion

Preparation of Aberrant Toxoids

Development of stability indicating tests and batch-to-batch consistency assays for biologicals requires known aberrant products, which the test should be able to distinguish from high quality products. Acquiring real toxoids intended for human (or veterinary) use that did not meet current quality control (QC) criteria was not feasible because manufacturers with a robust production process generate no failed batches and any potential manufacturers that struggle with consistency would be unlikely to cooperate in studies such as these. As such, the best option is to alter TTd drug products in such a way that they should fail the current QC criteria. Previous studies have shown that tetanus toxoid rapidly loses potency upon exposure to 55 °C for three days [20]. Therefore, an accelerated degradation protocol was implemented whereby the TTd samples were exposed to temperatures covering this range as a means to obtain aberrant samples. In-house produced non-adsorbed tetanus toxoid (TTd-A1.1) was used to conduct an initial evaluation of our degradomics assay concept. The toxoid was subjected to elevated temperatures: 37 °C for 30 days and 50, 55, 60, 65 °C for two days. Afterwards, the toxoids (TTd-A1.1) were stored at 4 °C for several months prior to analysis. During this time period, visible aggregates formed in the toxoids that were

exposed to the three highest temperatures, where the severity of the aggregation increased with increasing temperature. For the follow-up experiments (comparison of batches from different manufacturers and comparison of various batches of one manufacturer) samples were exposed to all temperatures for two days, with subsequent storage at 4 °C and follow-up analysis within a month. During this time, no visible aggregates were formed. It is likely that the damage to the protein folding triggered further accelerated aggregation. Although this results in differences between the samples used in the initial screening and those used in the final assay, the alterations could be picked up for both TTd-A1.1 and the follow-up experiments.

Degradomics Analysis

Temperature exposed TTd-A1.1 was subjected to cathepsin S digestion and the formation of peptides was monitored by nanoscale LC-MS. These peptides (not modified by formaldehyde) were identified and quantified by peak area with PEAKS Studio X. To reduce inter-measurement variation, the peak areas were corrected by addition of a fixed amount of angiotensin-III as an internal standard. In total, 396 tetanus toxin-derived peptides were identified and used for the initial comparison. In this comparison, the total peptide intensity was summed per time point (Figure 1). At exposure to 55 °C and above, the overall area of the peptides increased dramatically. While this initial approach gave insight into the overall reaction kinetics, degradation reactions in various parts of the protein are not detected and potential opposite effects of exposure to elevated temperature could cancel out.

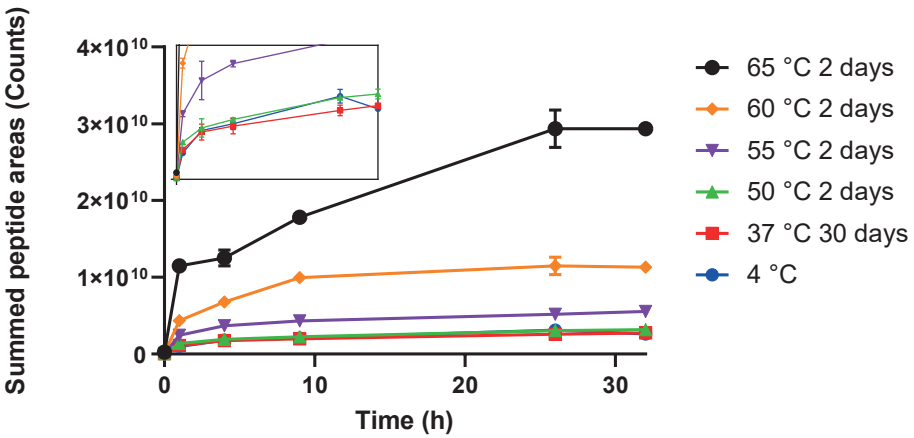


Figure 1. Summed areas of peptides formed during partial digestion of temperature exposed TTd-A1.1 by cathepsin S. Inset: zoomed graph to allow comparison between TTd exposed to the lower temperature ranges. Peptides were quantified by label-free quantification. Error bars represent SD of three measurements.

To further improve the sensitivity of the assay and to find specific areas of the protein that underwent thermal degradation, peptides were sorted according to their 50 °C to 4 °C ratio measured at the 4 h time point. Further selection of the peptides was based on the overall peak area to include the most abundant peptides. Peptides with high degrees of asparagine deamidation and methionine oxidation were excluded, as well as peptide length variants of the same part of the sequence. The final criterion was that the peptides should have increasing intensities during the course of the digestion. The reason is that some peptides are cleaved for a second time during digestion, reducing the amount of the precursor peptide and complicating the quantification. The three peptides that complied with our criteria were E^{1066}DNNITLK^{1073}, N^{1220}LDRILR^{1226}, and A^{1286}SNWYFNHLK^{1295} (Figure 2). These peptides all reside in the tetanus toxoid heavy chain (Figure S1). NLDRILR and ASNWYFNHLK are in very close proximity and part of the same β-sheet, which explains their similar sensitivity to temperature exposure; apparently this part of the molecule is easily affected by temperature exposure. Peptide EDNNITLK is part of an unrelated large β-sheet in another part of the molecule. To further optimize analysis time and assay performance, synthetic stable isotope labeled (SIL) peptides were used as internal standard for more robust, absolute, quantification and instead of plotting the kinetics, the amount of peptide at a set time point (20 h) was analyzed and compared (a representative chromatogram is depicted in Figure S2). Addition of SIL

internal standards for these peptides eliminates most of the inter-assay variation, and most likely avoids the need for expensive high-resolution mass spectrometers and allows further transfer of the method to more conventional liquid chromatography systems. Because of the relatively high abundance of asparagine residues in tetanus toxin (9.21%), it was not possible to select peptides without asparagine residues and completely avoid deamidation. To overcome this drawback, SIL internal standard peptides with aspartate residues instead of asparagine residues were used. Because deamidation results in a mixture of aspartate and *iso*-aspartate [24], the peak area was used to quantify all deamidated forms (which are not fully resolved) instead of quantifying by peak height (useful to compare the same MS1 scan of coeluting peptides). Determining the exact degree of deamidation by degradomics analysis is difficult, because this would require complete digestion of the protein by a selective protease. However, our current analysis is sufficient to demonstrate that the selected peptides were not deamidated below the concentrations obtained with the 4 °C reference sample.

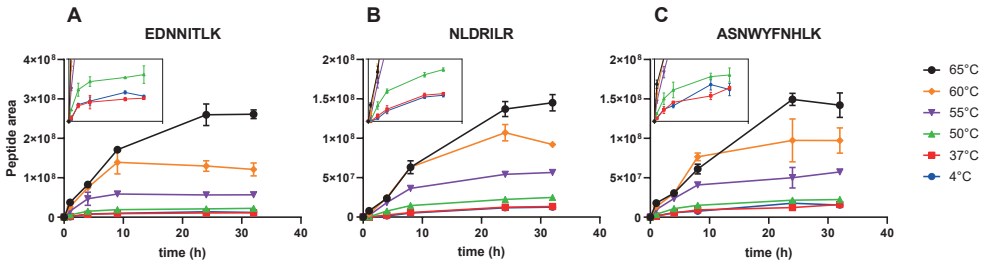


Figure 2. Formation of peptides (**A**) EDNNITLK, (**B**) NLDRILR, and (**C**) ASNWYFNHLK over time obtained from cathepsin S-mediated digestion of TTd-A1.1 after exposure to various temperatures (2 (50, 55, 60, 65 °C) or 30 days (37 °C)). Peptides were quantified by label-free quantification. Insets: zoomed graphs show the differences between exposure to 4 and 37 °C, and 50 °C. Error bars represent SD of three measurements.

Absolute quantification of the peptides obtained after 20 h of enzymatic degradation was then applied to heat-treated bulk tetanus toxoids from two manufacturers to confirm that our new degradomics-based assay can be utilized to detect aberrant toxoids regardless of their source. The toxoids from both manufacturers and the reference toxoid showed an increased enzyme-mediated formation of the selected peptides after exposure to 45 °C for two days compared to their 4 °C counterparts (Figure 3). As expected, small differences in the absolute concentrations of formed peptide were observed between the toxoids from different sources, but the overall trends were remarkably similar. Identical protein concentrations were used, so the differences most likely originate from dissimilarities in

protein composition and purity. Unexpectedly, exposure to 60 °C and 65 °C decreased the amount of peptide present compared to 55 °C, although the concentrations were still well above those of the 4 °C controls. This decrease was most prominent for EDNNITLK and NLDRILR. Our initial screening experiments (which used TTd-A1.1 with increasing amounts of visible aggregates at the highest temperatures), did not show this trend. A probable explanation is that certain intermediate unfolded states of the protein are slightly more resistant to enzymatic proteolysis than those predominantly formed upon exposure up to 55 °C, and further degradation and the formation of visible aggregates could result in faster enzymatic proteolysis. The absolute amount of deamidated peptides observed was proportional to the increase in temperature exposure, but does not explain the decrease in intensity at the highest temperatures. Deamidation of EDNNITLK and NLDRILR was limited (maxima at 65 °C between 13–24 and 6–16%, respectively), however, exposure to 65 °C resulted in 33 to 34% deamidation of ASNWYFNHLK, although care should be taken with interpretation of this data; the enzymatic degradation is not complete and not selective, which are requirements for accurate determination of the extent of asparagine deamidation. Overall quantification of E^{1066}DNNITLK^{1073}, N^{1220}LDRILR^{1226}, and A^{1286}SNWYFNHLK^{1295} released from partial cathepsin S digestion could easily distinguish non-adsorbed tetanus toxoids stored at 4 °C from those exposed to 45 °C and higher, for both our research toxoid and toxoids provided by two manufacturers of vaccines.

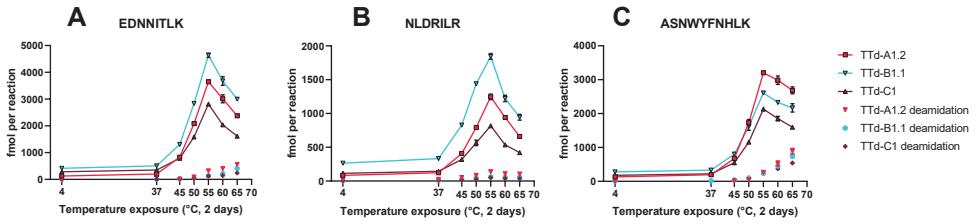


Figure 3. Absolute quantification of (**A**) EDNNITLK, (**B**) NLDRILR, and (**C**) ASNWYFNHLK formed by partial digestion (at t = 20 h) with cathepsin S of tetanus toxoids (TTd-A1.2, TTd-B1.1 and TTd-C1) that were pre-exposed to elevated temperatures. Connected points indicate the sum of the native peptides and their deamidated forms, single points represent the deamidated forms only. Error bars represent the SD of three temperature exposed aliquots of the same TTd batch.

To evaluate the potential of this new assay for batch-to-batch consistency testing, four additional batches from manufacturer B were analyzed. As such, a reference sample (TTd-B1.2 untreated, stored at 4 °C) and an aberrant sample (TTd-B1PC stored at 55 °C for

two days) were prepared from TTd-B1. The concentrations of the enzymatically released peptides were very similar among the different batches (Figure 4) and the aberrant sample can easily be distinguished from the toxoids stored at 4 °C. Acceptable limits should be set by manufacturers for their specific products, but overall these results show that the assay is suitable for batch-to-batch consistency testing.

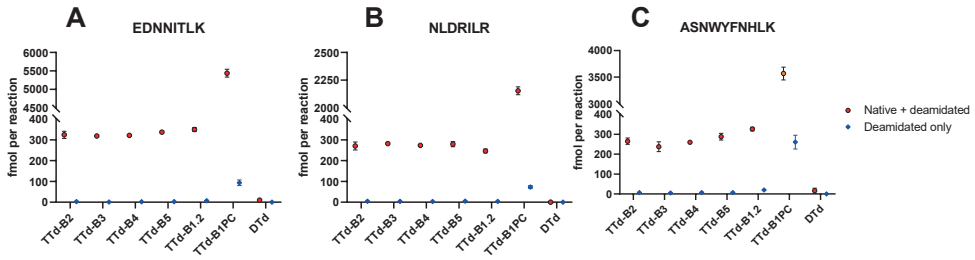


Figure 4. Temperature sensitive formation of peptides (**A**) EDNNITLK, (**B**) NLDRILR, and (**C**) ASNWYFNHLK by partial digestion with cathepsin S of tetanus toxoid batches of the same manufacturer. TTd-B1.2 was used as a reference and is two years older than TTd-B2–TTd-B5, an aliquot of TTd-B1.2 exposed to 55 °C for two days is used as a positive control (TTd-B1PC). Diphtheria toxoid of the same manufacturer was used as a negative control. Error bars represent the SD of three enzymatic digestions of aliquots of the same batch.

Evaluation of Heat-Treated Toxoids by Traditional Techniques

In order to put the temperature-induced alterations to the tetanus toxoids observed with the degradomics analysis into perspective, the toxoids were also analyzed by more traditional techniques. The parameters studied included: antigen binding (biosensor analysis), protein folding (circular dichroism, tryptophan fluorescence and Bis-ANS fluorescence), and aggregation (size-exclusion chromatography).

The decrease in antibody binding of TTd showed similar trends for the batches of all three manufacturers: a moderate decrease in antibody binding compared to the 4 °C control was observed after exposure to 50 °C (10–18% reduction), the epitope was severely affected upon exposure to 55 °C for two days (68–79% reduction), and exposure to higher temperatures completely destroyed the epitope (Figure 5A).

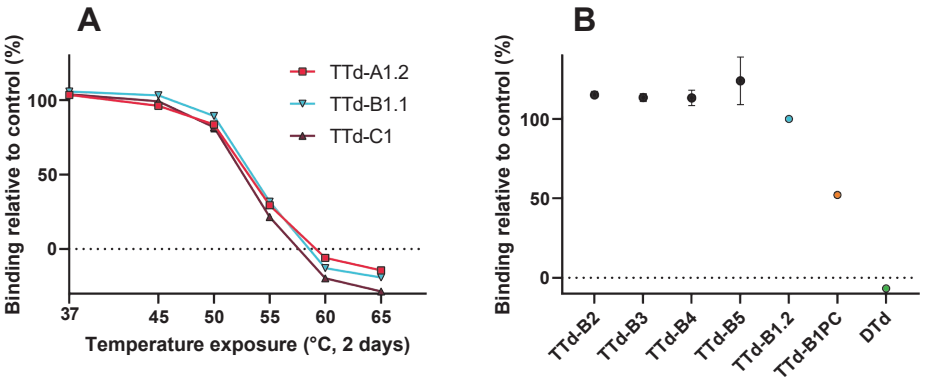


Figure 5. Biosensor analysis of tetanus toxoids. (**A**) Decrease in antibody binding relative to the 4 °C control. (**B**) Comparison of antibody binding between the various batches relative to TTd-B1.2. Error bars represent the SD of three aliquots of the same batch.

Protein folding analysis by circular dichroism (CD), tryptophan fluorescence and Bis-ANS fluorescence all showed temperature dependent alterations of the protein structure (Figure 6). CD spectra were compared using qBiC, a software package intended for protein higher order structure (HOS) analysis. This analysis was very sensitive and could already pick up differences between the 4 °C reference and the toxoids that were exposed to 37 °C for two days, subjecting the toxoids to higher temperatures resulted in larger spectral differences. Moreover, differences between the four new batches and the older reference batch could also be detected. Analysis of the toxoids by tryptophan and Bis-ANS fluorescence also showed differences after exposure to 45 °C.

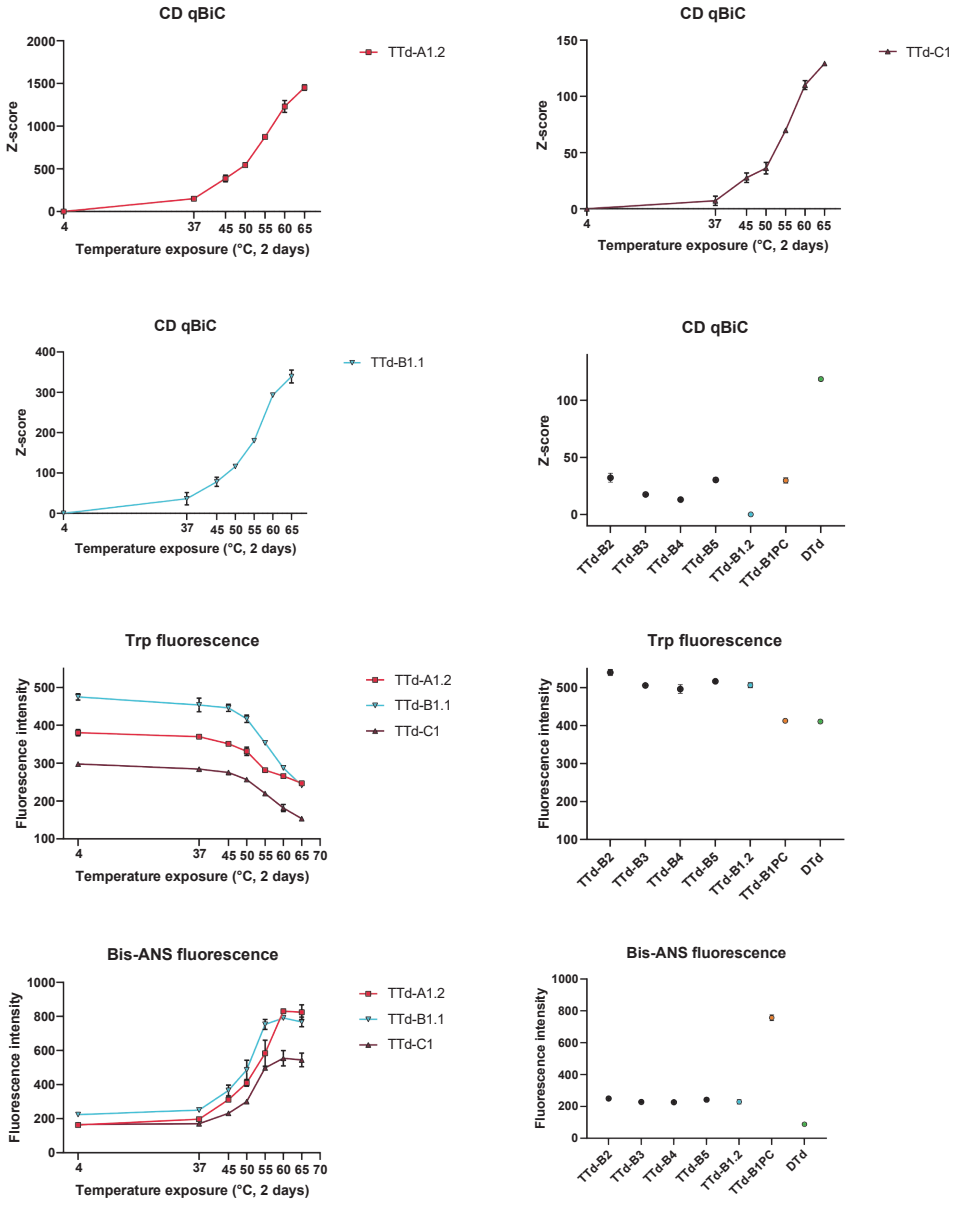


Figure 6. Evaluation of protein folding by circular dichroism (CD), tryptophan fluorescence and Bis-ANS fluorescence. CD spectra were analyzed using qBiC to compare differences in higher order structure. Error bars represent the SD of three aliquots of the same batch.

Finally, protein aggregation was inferred from size exclusion chromatography. Exposure to temperatures above 50 °C rapidly decreased the monomer content (with negligible effects on total protein concentration as determined by BCA prior to pre-SEC centrifugation) (Figures 7 and S3). This is consistent with the trend observed with all other techniques.

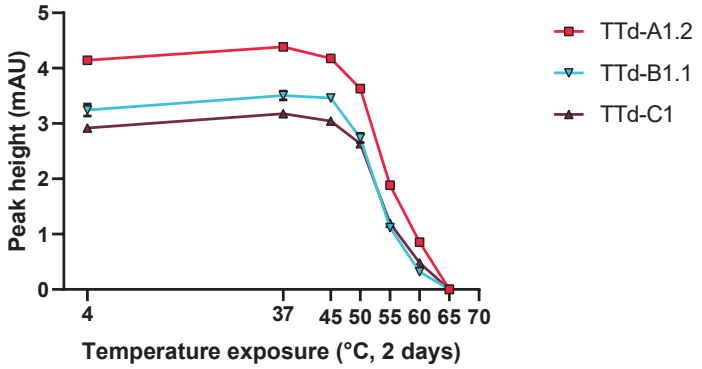


Figure 7. Monomer quantification of temperature exposed tetanus toxoids by size exclusion chromatography. Error bars represent the SD of three aliquots of the same batch.

Although our degradomics analysis and most other techniques picked up differences to the toxoids after exposure to 45 °C for two days, the epitope where mAb 10/134 binds was still mostly intact after exposure to 50 °C, where loss of the epitope started at exposure to 55 °C. While this is only one epitope, it is consistent with known decreases in *in vivo* potency found by others [20]. Despite our degradomics assay being very sensitive and picking up changes before potency of the toxoid is completely lost, the superior performance of the degradomics assay compared to the animal tests should avoid rejection of good batches.

$Al(OH)_3$ and $AlPO_4$ Adsorbed Tetanus Toxoid

Tetanus toxoids are usually adsorbed to $Al(OH)_3$, $AlPO_4$ or a mixture of the two. Analysis of antigens adsorbed to aluminum-based adjuvants is challenging with many traditional techniques, because the samples are highly turbid, the proteins are not completely in solution and several vaccines contain multiple antigens. Although showing batch consistency through testing of the drug substance (before adsorption) could contribute to the replacement of *in vivo* tests of the drug product, an assay that could be directly applied to the drug product would be highly favorable. To evaluate the feasibility of utilizing our assay under these

conditions, TTd of manufacturer B was adsorbed to $Al(OH)_3$ and $AlPO_4$ separately, to achieve adsorbed tetanus toxoids with 10 Lf/mL TTd and 1 mg/mL or 0.66 mg/mL Al^{3+}, respectively. Along with a control (10 Lf/mL TTd without aluminum-based adjuvant), these samples were mixed overnight and adsorption was determined by relative decrease in protein concentration of the supernatant compared to the control. Adsorption of $Al(OH)_3$ was more effective (48%, SD 2%) than $AlPO_4$ (8%, SD 1%). Subsequently, the samples were split and the aliquots were subjected to elevated temperatures for two days and then analyzed with our degradomics assay. Exposure of the $AlPO_4$-adsorbed TTd resulted in similar kinetic profiles for peptides ASNWYFNHLK and NLDRILR, whereas EDNNITLK was slightly more sensitive to temperature exposure when the TTd was combined with $AlPO_4$ (Figure 8). $Al(OH)_3$ adsorbed TTd was more sensitive to temperature exposure for all three peptides, even though the 4 °C starting points were very similar, indicating that adsorption itself did not alter the degradation kinetics of these peptides. The increased temperature sensitivity observed with TTd adsorbed to $Al(OH)_3$ compared to the control and the $AlPO_4$ adsorbed TTd is likely due to the higher percentage of adsorption, where the adsorbed product's structure is altered in such a way that it is less stable. Overall, the potential for drug product testing with the enzymatic degradation assay is a clear advantage over many other *in vitro* assays, which are usually limited to analysis of drug substance.

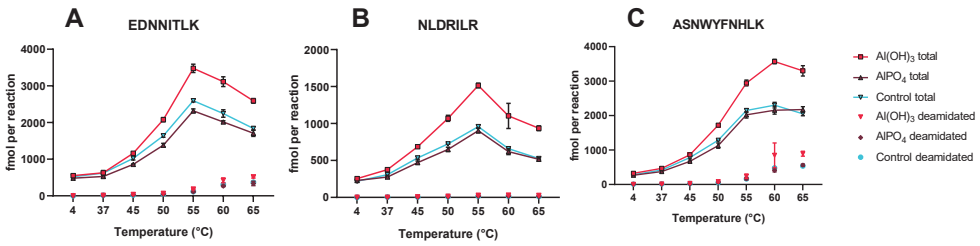


Figure 8. Temperature sensitive formation of peptides (**A**) EDNNITLK, (**B**) NLDRILR, and (**C**) ASNWYFNHLK by cathepsin S from $Al(OH)_3$ and $AlPO_4$ adsorbed tetanus toxoids. Connected lines show the sum of the native and the deamidated peptides, single points indicate the sum of all variants of the deamidated peptides. Error bars represent SD of three individually adsorbed samples.

In general, the degradomics assay does not require a biological reference for every experiment, but a number of synthetic peptides, which can be well defined, quantified, and are readily available. This is a major advantage compared to other biochemical techniques which would require antibodies or toxoids. Although the activity of the recombinant cathepsin S is critical, activity tests with fluorescent probes offer an easy method for quality control of the enzyme [25].

Conclusions

A new type of degradomics-based assay was developed for the analysis of aluminum-adsorbed and non-adsorbed tetanus toxoids, which could be applicable to other protein-antigens as well. By analyzing the formation of peptides from treatment of tetanus toxoid by recombinant human cathepsin S, an important enzyme involved in antigen processing *in vivo*, aberrant tetanus toxoids could be distinguished from control toxoids. Overall, with some further optimization, implementation and validation in a QC environment, the assay may have potential for use in both stability and batch-to-batch consistency studies as an alternative for *in vivo* potency studies.

Supplementary Materials

The following are available online at https://www.mdpi.com/2076-393X/8/4/712/s1, Figure S1: Crystal structure of tetanus toxin where the selected peptides that were formed faster from cathepsin S digestion after temperature exposure are highlighted, Figure S2: Representative chromatogram and spectra of the tetanus toxoid degradomics analysis., Figure S3: Representative size exclusion chromatography chromatogram of TTd-B1.1 after exposure to 4, 37, 45, 50, 55, 60, or 65 °C for two days. Table S1 contains the exported PEAKS data used for Figures 1 and 2. Table S2 contains the quantification data used for Figures 3 and 4. Table S3 contains the quantification data used for Figure 8. The data presented in this study are openly available in the MassIVE mass spectrometry repository at https://doi.org/doi:10.25345/C58F6R.

Author Contributions

Conceptualization, T.J.M.M., M.D., H.D.M., G.F.A.K., W.J. and B.M.; methodology, T.J.M.M.; validation, T.J.M.M. and H.D.M.; formal analysis, T.J.M.M., W.T. and J.W.d.R.; investigation, T.J.M.M., W.T. and M.R.J.H.; data curation, T.J.M.M. and H.D.M.; writing—original draft preparation, T.J.M.M.; writing—review and editing, T.J.M.M. W.T., M.R.J.H., J.W.d.R., M.D., H.D.M., G.F.A.K., W.J. and B.M.; visualization, T.J.M.M. and M.D.; supervision, G.F.A.K., W.J. and B.M.; project administration, M.D.; funding acquisition, B.M.

Funding

This work was performed within the Vac2Vac project supported by the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement N-115924.

Conflicts of Interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Plotkin, S.A.; Orenstein, W.A.; Offit, P.A.; Roper, M.H.; Wassilak, S.G.F.; Scobie, H.M.; Ridpath, A.D. Tetanus Toxoid. In *Vaccines*, 7th ed.; Elsevier: Amsterdam, The Netherlands, 2018; p. 1061.
2. WHO. Production and Control of Tetanus Vaccine: Module III: Principles of Tetanus Vaccine Production. WHO/VSQ/94.4. Availabe online: http://whqlibdoc.who.int/hq/1994/WHO_VSQ_GEN_94.4.pdf (accessed on 26 March 2020).
3. Council of Europe. European Pharmacopoeia 10.0. In *Tetanus Vaccine (Adsorbed)*; Council of Europe: Strasbourg, France, 2020; pp. 1042–1043.
4. Council of Europe. European Pharmacopoeia 8.0. In *Insulin Aspart*; Council of Europe: Strasbourg, France, 2014; pp. 2485–2486.
5. Council of Europe. European Pharmacopoeia 8.0. In *Monoclonal Antibodies for Human Use*; Council of Europe: Strasbourg, France, 2014; pp. 753–755.
6. Council of Europe. European Pharmacopoeia 10.0. In *Assay of Tetanus Vaccine (Adsorbed)*; Council of Europe: Strasbourg, France, 2020; pp. 275–278.
7. Bruysters, M.W.; Schiffelers, M.-J.; Hoonakker, M.; Jungbaeck, C.; Ragan, I.; Rommel, E.; Van Der Stappen, T.; Viviani, L.; Hessel, E.V.; Akkermans, A.M.; *et al.* Drivers and barriers in the consistency approach for vaccine batch release testing: Report of an international workshop. *Biologicals* **2017**, *48*, 1–5, doi:10.1016/j.biologicals.2017.06.006.
8. De Mattia, F.; Hendriksen, C.; Buchheit, K.H.; Chapsal, J.M.; Halder, M.; Lambrigts, D.; Redhead, K.; Rommel, E.; Scharton-Kersten, T.; Sesardic, T.; *et al.* The vaccines consistency approach project: An EPAA initiative. *Pharmeuropa Bio Sci. Notes* **2015**, *2015*, 30–56.
9. De Mattia, F.; Chapsal, J.-M.; Descamps, J.; Halder, M.; Jarrett, N.; Kross, I.; Mortiaux, F.; Ponsar, C.; Redhead, K.; McKelvie, J.; *et al.* The consistency approach for quality control of vaccines—A strategy to improve quality control and implement 3Rs. *Biologicals* **2011**, *39*, 59–65, doi:10.1016/j.biologicals.2010.12.001.
10. Carmicle, S.; Dai, G.; Steede, N.K.; Landry, S.J. Proteolytic Sensitivity and Helper T-cell Epitope Immunodominance Associated with the Mobile Loop in Hsp10s. *J. Biol. Chem.* **2002**, *277*, 155–160, doi:10.1074/jbc.m107624200.
11. Kim, A.; Hartman, I.Z.; Poore, B.; Boronina, T.; Cole, R.N.; Song, N.; Ciudad, M.T.; Caspi, R.R.; Jaraquemada, D.; Sadegh-Nasseri, S. Divergent paths for the selection of immunodominant epitopes from distinct antigenic sources. *Nat. Commun.* **2014**, *5*, 5369, doi:10.1038/ncomms6369.
12. Egger, M.; Jürets, A.; Wallner, M.; Briza, P.; Ruzek, S.; Hainzl, S.; Pichler, U.; Kitzmüller, C.; Bohle, B.; Huber, C.G.; *et al.* Assessing Protein Immunogenicity with a Dendritic Cell Line-Derived Endolysosomal Degradome. *PLoS ONE* **2011**, *6*, e17278, doi:10.1371/journal.pone.0017278.
13. Delamarre, L.; Couture, R.; Mellman, I.; Trombetta, E.S. Enhancing immunogenicity by limiting susceptibility to lysosomal proteolysis. *J. Exp. Med.* **2006**, *203*, 2049–2055, doi:10.1084/jem.20052442.
14. Machado, Y.; Freier, R.; Scheiblhofer, S.; Thalhamer, T.; Mayr, M.; Briza, P.; Grutsch, S.; Ahammer, L.; Fuchs, J.E.; Wallnoefer, H.G.; *et al.* Fold stability during endolysosomal acidification is a key factor for allergenicity and immunogenicity of the major birch pollen allergen. *J. Allergy Clin. Immunol.* **2016**, *137*, 1525–1534, doi:10.1016/j.jaci.2015.09.026.

15. Freier, R.; Dall, E.; Brandstetter, H. Protease recognition sites in Bet v 1a are cryptic, explaining its slow processing relevant to its allergenicity. *Sci. Rep.* **2015**, *5*, 12707, doi:10.1038/srep12707.

16. Kitzmüller, C.; Wallner, M.; Deifl, S.; Mutschlechner, S.; Walterskirchen, C.; Zlabinger, G.J.; Ferreira, F.; Bohle, B. A hypoallergenic variant of the major birch pollen allergen shows distinct characteristics in antigen processing and T-cell activation. *Allergy* **2012**, *67*, 1375–1382, doi:10.1111/all.12016.

17. 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. *Sci. Rep.* **2020**, *10*, 11535,, doi:10.1038/s41598-020-68248-z.

18. Doekhie, A.; Dattani, R.; Chen, Y.-C.; Yang, Y.; Smith, A.; Silve, A.P.; Koumanov, F.; Wells, S.A.; Edler, K.J.; Marchbank, K.J.; *et al.* Ensilicated tetanus antigen retains immunogenicity: *In vivo* study and time-resolved SAXS characterization. *Sci. Rep.* **2020**, *10*, 1–9, doi:10.1038/s41598-020-65876-3.

19. Milstein, J.B. *Temperature Sensitivity of Vaccines*; WHO/IVB/06.10; World Health Organization: Geneva, Switzerland, 2006.

20. Cohen, H.; Van Ramshorst, J.D.; Tasman, A. Consistency in potency assay of tetanus toxoid in mice. *Bull. World Heal. Organ.* **1959**, *20*, 1133–1150.

21. Meiring, H.D.; Van Der Heeft, E.; Hove, G.J.T.; De Jong, A.P.J.M. Nanoscale LC-MS(n): Technical design and applications to peptide and protein analysis. *J. Sep. Sci.* **2002**, *25*, 557–568, doi:10.1002/1615-9314(20020601)25:93.0.co;2-f.

22. Metz, B.; Kersten, G.F.A.; Hoogerhout, P.; Brugghe, H.F.; Timmermans, H.A.M.; De Jong, A.; Meiring, H.; Hove, J.T.; Hennink, W.E.; Crommelin, D.J.A.; *et al.* Identification of Formaldehyde-induced Modifications in Proteins. *J. Biol. Chem.* **2004**, *279*, 6235–6243, doi:10.1074/jbc.m310752200.

23. 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*, 156–167, doi:10.1016/j.vaccine.2003.08.003.

24. Catak, S.; Monard, G.; Aviyente, V.; Ruiz-LópezM.F. Deamidation of Asparagine Residues: Direct Hydrolysis versus Succinimide-Mediated Deamidation Mechanisms. *J. Phys. Chem. A* **2009**, *113*, 1111–1120, doi:10.1021/jp808597v.

25. Kirschke, H.; Wiederanders, B. [34] Cathepsin S and related lysosomal endopeptidases. In *Methods in Enzymology*; Academic Press, Cambridge (MA), United States of America: 1994; Volume 244, pp. 500–511.

6

COMMON REFERENCE-BASED TANDEM MASS TAG MULTIPLEXING FOR THE RELATIVE QUANTIFICATION OF PEPTIDES: DESIGN AND APPLICATION TO DEGRADOME ANALYSIS OF DIPHTHERIA TOXOID

Thomas J.M. Michiels [1,2], Madelief A. van Veen [1], Hugo D. Meiring [2], Wim Jiskoot [1], Gideon F.A. Kersten [1,2], 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

J Am Soc Mass Spectrom **2021**, *32* (6), 1490-1497.

Abstract

Currently, animal tests are being used to confirm the potency and lack of toxicity of toxoid vaccines. In a consistency approach, animal tests could be replaced if production consistency (compared to known good products) can be proven in a panel of *in vitro* assays. By mimicking the *in vivo* antigen processing in a simplified *in vitro* approach, it may be possible to distinguish aberrant products from good products. To demonstrate this, heat-exposed diphtheria toxoid was subjected to partial digestion by cathepsin S (an endoprotease involved in antigen processing) and the peptide formation/degradation kinetics were mapped for various heated toxoids. To overcome the limitations associated with the very large number of samples, we used common reference-based tandem mass tag (TMT) labeling. Instead of using one label per condition with direct comparison between the set of labels, we compared multiple labeled samples to a common reference (a pooled sample containing an aliquot of each condition). In this method, the number of samples is not limited by the number of unique TMT labels. This TMT multiplexing strategy allows for a 15-fold reduction of analysis time, while retaining the reliability advantage of TMT labeling over label-free quantification. The formation of the most important peptides could be followed over time and compared among several conditions. The changes in enzymatic degradation kinetics of diphtheria toxoid revealed several suitable candidate peptides for use in a quality control assay that can distinguish structurally aberrant diphtheria toxoid from compliant toxoids.

Introduction

Degradomics analysis of antigens and allergens has been successfully used for predicting T-cell epitopes[1] and for studying allergens and antigens[2-9]. Usually, an antigen of interest is subjected to limited proteolysis in an *in vitro* setting which mimics antigen processing of the immune system. *In vitro* proteolysis can be done in isolated antigen-presenting cells (APCs) [10], or with endolysosomal extracts of APCs [11-12], a mixture of APC-derived enzymes and proteins [1], or recombinant enzymes [9]. In most applications the formed peptides are identified after a set time point by using liquid chromatography-tandem mass spectrometry (LC-MS/MS) for protein identification.

We have previously developed a degradomics-based analysis that can be used as a quality control (QC) assay for tetanus toxoids as an alternative for the animal tests that are currently being used to confirm potency and lack of toxicity of these vaccines [13]. The rationale for this type of assay is that if the assay (or a panel of assays) can confirm batch-to-batch consistency, the potency and safety profile of the batches are also consistent. Toxoids are prepared by formaldehyde-inactivation of toxins, which results in very heterogeneous mixtures due to a myriad of chemical modifications of the antigen [14]. These modifications are most common in arginine, tyrosine and lysine residues, but are certainly not limited to these amino acid residues. This heterogeneity is further exacerbated by adsorption to colloidal aluminum salts. These salts enhance immunogenicity, but lead to challenging characterization of the final vaccine. In the previously described assay, tetanus toxoids were exposed to elevated temperatures to simulate aberrant batches. Subsequently, the samples were subjected to digestion with recombinant cathepsin S. By using label-free quantification of the peptides formed at various time points, the kinetics of the individual peptides were mapped. In subsequent analyses, selected peptides that showed temperature-induced differences in their formation/degradation kinetics were quantified by addition of synthetic isotopically labeled standards, which could reliably distinguish denatured products from unaltered products. We intend to apply similar degradomics analysis to other antigens where animal tests are still being used for QC, starting with diphtheria toxoid (DTd). The diphtheria toxin is a 58 kDa protein consisting of an A and a B fragment, which are connected by a disulfide bridge. In this study, the purified bulk was analyzed. Digestion of DTd with chymotrypsin (comparable to cathepsin S: a clear preference for cleavage sites but not completely specific)

results in the identification of approximately 150 to 300 peptides with modern equipment [14], but the number of very abundant peptides is more limited [15], which is comparable to digestions of other purified proteins of a similar size. The first step in the development of a degradomics assay involves mapping the kinetics of as many of the peptides that are formed by cathepsin S digestion as possible. Once the degradation process has been mapped, the assay can be simplified for use in QC by quantification of specific stability-indicating peptides. Despite its simple sample preparation, the previously used unbiased label-free quantification for mapping the enzymatic degradation of tetanus toxoid has several disadvantages. In particular, the analysis time of approximately a week to analyze all samples is long, with the available instrument time usually being a limiting factor. Further disadvantages inherent to label-free quantification are inter-run variability (worsened by the long overall analysis time) and susceptibility to ion suppression or enhancement [16].

To overcome these issues, we looked into isobaric labeling-based relative quantification (reviewed by Rauniyar *et al.* and Arul *et al.* [17-18]), and specifically the use of tandem mass tag (TMT) labeling [19]. The main drawback of standard isobaric labeling is the limited number of available labels/channels. Efforts are being made to increase the number of channels by designing new labels, such as TMTpro 16-plex [20], but this approach will eventually be limited by the size of the molecules, the limits of isotope incorporation and the mass spectrometric specifications (in particular its resolving power). Instead of using more channels to cover both of our variables (temperature exposure and enzymatic digestion time of the protein), we used a strategy in which different channels are used for one variable (temperature exposure) and in which one or more dedicated channels are used for a pooled sample (referred to as the *common reference* (CR), containing an aliquot of each of the samples). This results in a control that can be used for relative quantification between all samples, because the common reference content is identical in each sample (Figure 1). Similar pooling strategies are common in quantitative DNA or RNA studies [21], and are used in the field of proteomics in conjunction with dimethyl labeling [22] and TMT labeling [23-25]. In this study we report the use of the TMT-CR multiplexing strategy for mapping the enzymatic degradation kinetics of DTd solutions that were exposed to elevated temperatures to simulate faulty batches.

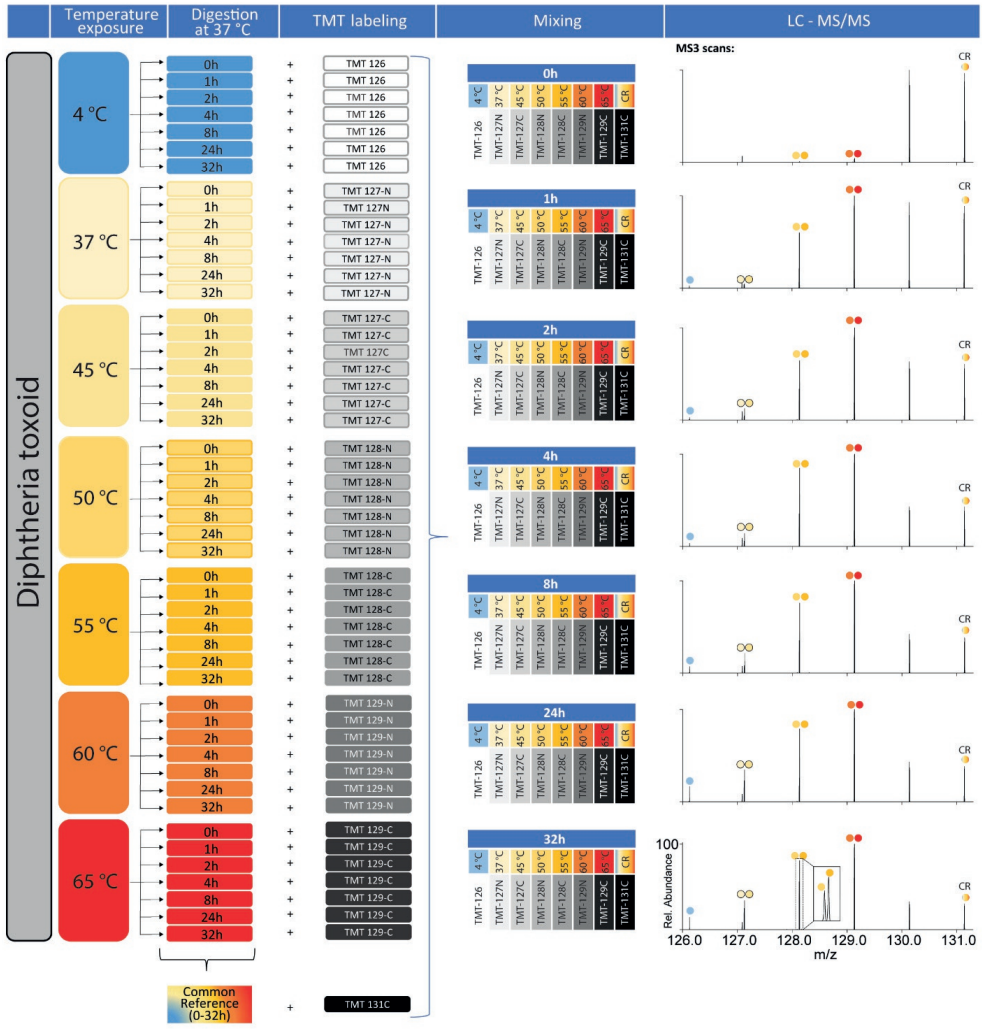


Figure 1. Schematic overview of the common reference-based tandem mass tag labeling strategy. Diphtheria toxoid was exposed to various temperatures and subsequently digested by cathepsin S at 37 °C. The enzymatic reaction was stopped at various digestion time points. Aliquots of each sample were pooled to form a common reference sample. Subsequently, each sample was labeled with a unique TMT label for every temperature (the same TMT-channels were used for the same exposure temperature), and a unique label for the pooled common reference sample (TMT^{11}-131C). Since each digestion sample (time point and temperature) was prepared in triplicate, there was an opportunity for extra pooled control samples, made by labeling a pooled sample containing just one triplicate. These additional controls were labeled with TMT^{11}-130N, TMT^{11}-130C and TMT^{11}-131N, one for each triplicate. For simplicity these are not depicted in the schematic overview (except for the MS3 spectrum). After labeling, the various heat-exposed samples were mixed with the other heat-exposed samples of the same time point and with the common reference. The mixed samples were then measured by nanoscale LC-MS, identified by MS1 and MS2, and quantified by the reporter ions generated in MS3. The reporter ions can be compared to the common reference reporter ion (TMT^{11}-131C) for every separate analysis, allowing for comparison between different runs (*i.e.*, different time points and replicates). The peptide used for the MS3 spectra in this example is YPGLT.

Materials & Methods

Preparation of aberrant diphtheria toxoids

Diphtheria toxoid was obtained from a manufacturer within the IMI-funded VAC2VAC consortium [26]. The protein concentration (determined by BCA (Thermo Fisher)) was adjusted to 100 µg/mL. The toxoid was then thoroughly dialyzed (Slide-A-Lyzer Dialysis cassettes 10000 MWCO, Thermo Scientific) against a phosphate buffer (10 mM, pH 7.2, prepared from a 1-M solution from Sigma Aldrich). The protein concentration was confirmed to still be approximately 100 µg/mL by BCA. Subsequently, aliquots of the toxoid were incubated at different temperatures. Three samples were prepared per incubation temperature. The samples were incubated at 4 °C, 37 °C, 45 °C, 50 °C, 55 °C, 60 °C and 65 °C for 2 days.

Digestion conditions

Five µg heated DTd was digested with cathepsin S (0.1 µg) in 100 µL sodium citrate buffer (100 mM, pH 5.0) containing 2 mM dithiothreitol and 2 mM ethylenediaminetetraacetic acid. For each temperature (in triplicate) and time point, a separate reaction was carried out. The digestion took place at 37 °C and was stopped by addition of 50 µL 0.1 mM E-64 solution (a cysteine protease inhibitor, Sigma Aldrich).

Labeling

Prior to labeling, Solid Phase Extraction (SPE) was performed by using 50-mg Sep-Pak C18 cartridges (Waters) in conjunction with a Gilson GX-271 ASPEC robot. The digestion solution was loaded and washed with ammonium carbonate solution (10 mM, pH 10). The high pH is critical to remove DTT, which interferes with TMT labeling. Subsequently, the peptides were eluted with 60 vol% acetonitrile, collected and dried in a vacuum centrifuge. After drying, the samples were re-dissolved in 100 µL phosphate buffer (100 mM, pH 7.4) and a 10-µL aliquot was taken of each sample. These samples were used to prepare the pooled common reference samples: the main common reference consisted of an aliquot from every sample, three additional common references consisted of the pooled aliquots of the individual triplicates (CR1, CR2 and CR3). To 45 µL of the remaining digests, 5 µL of a solution of the synthetic peptide Ac-GDVEAGKK (20 fmol/µL, purchased from Pepscan, the Netherlands) was added as an internal standard to correct for any labeling or measurement bias. Every sample was labeled by addition of 4.5 µL of a TMT label dissolved in acetonitrile (7.3 µg/µL)

as depicted in Figure 1, and then incubated for 1 hour at room temperature. Then, 8 µL 5 vol% hydroxylamine was added and the resulting mixture was incubated for 1 hour at room temperature. Subsequently, the samples were mixed as depicted in Figure 1. Samples of the various temperature treatments with the same digestion time were pooled along with the common references. CR1, CR2 and CR3 were labeled with labels TMT^{11}-130N, TMT^{11}-130C and TMT^{11}-131N. These three channels could be used for potential troubleshooting. The common reference of every sample (so, a mixture of every sample of all replicates) was labeled with label TMT^{11}-131C. This common reference was used for relative quantification of every sample. After mixing, the samples were again subjected to SPE, but at low pH with 0.1 vol% formic acid for the initial washing and 60% acetonitrile with 0.1 vol% formic acid for the peptide collection. After drying, the peptides were dissolved in 550 µL 0.1 vol% formic acid containing 1 fmol/µL angiotensin-I, angiotensin-III and oxytocin as system suitability controls.

LC-MS

The TMT^{11}-labeled peptides were analyzed by reversed phase nanoscale LC-MS using a vented column system as described by Meiring et al [27]. A 100-µm I.D. x 20-mm L trapping column packed with 5 µm Reprosil-Pur C18-AQ particles, followed by a 50-µm I.D. x 32.4-cm L analytical column packed with 3 µm Reprosil-Pur C18-AQ particles were connected to an Agilent 1290 Infinity HPLC system. The samples were injected onto the trapping column and washed for 10 minutes with 0.1 vol% formic acid in water at a column flow rate of 5 µL/min. Subsequently, the peptides were separated on the analytical column by a 30-minute gradient from 6 vol% to 56 vol% acetonitrile containing 0.1 vol% formic acid at a column flow rate of 125 nL/min. The analytical column was coupled to a Fusion Lumos Tribrid mass spectrometer (Thermo Fisher) by electrospray ionization (spray tip prepared in-house). A targeted inclusion list was used to ensure quantification of the internal standard peptide, as well as some of the most abundant digestion products (DSIIR, GTNPVF, DGASRVV, DGASRVVL, HPELS, YPGLT, ESPNKTVS, VDENPLS). Detailed instrumental settings can be found in the Supplemental Information.

Data processing

The data was analyzed by Proteome Discoverer 2.4 SP1 (Thermo Fisher) to obtain the relative intensities of the internal standard peptide. Proteome Discoverer allows the user to set the

TMT label modification to dynamic, which was required as the peptide Ac-GDVEAGKK was predominantly labeled on one lysine residue. The quan spectra were exported and filtered for the correct retention time and correction factors for every condition were determined. Subsequently, the data was processed by PEAKS X (Bioinformatics Inc.) to quantify the DTd-derived peptides. The parent mass error tolerance was set to 5.0 ppm, the fragment mass error to 0.6 Da, and methionine oxidation and deamidation of asparagine and glutamine residues were considered as dynamic modifications. The quantitation module used 20 ppm mass tolerance for the reporter ions and a peptide confidence cut-off of -10logP > 15.0. The peptide list was exported to Excel (Table S1) and after correction for the intensity of the Ac-GDVEAGKK reporter ions, the relative intensity of each reporter compared to the common reference was determined. To estimate the quantification quality (Figure 2), points that were either 0 or missing were counted as missing. For the plots in Figure 3, any point where no CR channel was available (most likely no MS3 spectrum acquired), no point is plotted and the line representing the average was automatically calculated with the remainder of the points. Final graphs were prepared by using GraphPad Prism 8.1.2.

Results & Discussion

Diphtheria toxoid solutions were heated (37, 45, 50, 55, 60, 65 °C) for two days to obtain aberrant samples. The kinetics of the subsequent partial digestion of heated and 4 °C-stored DTd solutions by cathepsin S was evaluated at seven time points. A total of 232 peptides were identified (confidence cut-off -10logP > 15.0) and quantified, resulting in 70% coverage of the DTd sequence. However, most of these peptides were not measured at all of the 462 data points. To produce a meaningful kinetic plot for a specific peptide, its presence should be detectable at a sufficient number of data points. To visualize for which peptides a meaningful kinetic plot may be obtainable, the percentage of quantifiable TMT signals, *i.e.*, the reporter ion coverage, was calculated for each of the 232 peptides (Figure 2A). For example, if at t = 0h the intensity threshold for a single MS3 scan is not met because not enough peptide is present and no MS3 scan is obtained, this will decrease the reporter ion coverage by 2.3 percentage point (11/462 data points missing).

In total, 118 peptides had a reporter ion coverage of more than 50% of the data points. For these peptides, good kinetic profiles could be obtained, as shown in Figure 2B for peptide

ILPG. Even peptides with a reporter ion coverage as low as 19% yielded useful kinetic information, as illustrated by peptide VAQVIDSETADNLE (Figure 2C). So, most of the peptides that were identified and quantified yielded useful information, emphasizing the robustness of the method. It may be beneficial to make an inclusion list to ensure that all peptides of interest are quantified in each sample. For instance, a screening run of the common reference sample could be analyzed first. However, this does remove some of the unbiased nature of the assay. To avoid missing unexpected peptides this screening would have to be done for every batch that is being analyzed. Furthermore, a low average intensity may result in excluding a peptide because peptides with a low MS1 intensity will require very long injection times to reach a sufficient ion count for MS3. However, low average intensity is commonly seen for intermediate cleavage products (*e.g.*, LTEPLME in Figure 3), which are abundant only in a limited time frame but still provide useful insights in the enzymatic degradation process. Care should be taken to ensure that the use of an inclusion list fits the purpose of the experiment.

Prior to analyzing diphtheria toxoid samples, optimization experiments were performed to ensure maximum reproducibility, in particular with respect to differences in relative reporter ion intensity between MS3 scans. By direct infusion of TMT^6 labeled synthetic peptide Ac-GDVEAGKK various settings were screened and low amounts of ions in the Orbitrap were found to be detrimental to the inter-scan reproducibility. Relatively high automatic gain control (AGC) settings combined with both sufficiently long injection times and MS1 intensity thresholds were required to reach these targets (Figure S1). To correct for any labeling efficiency, mixing variation or channel response differences, a fixed amount of internal standard peptide was added to each sample. This resulted in good reproducibility between injections of the same sample (Figure S2 and Figure S3).

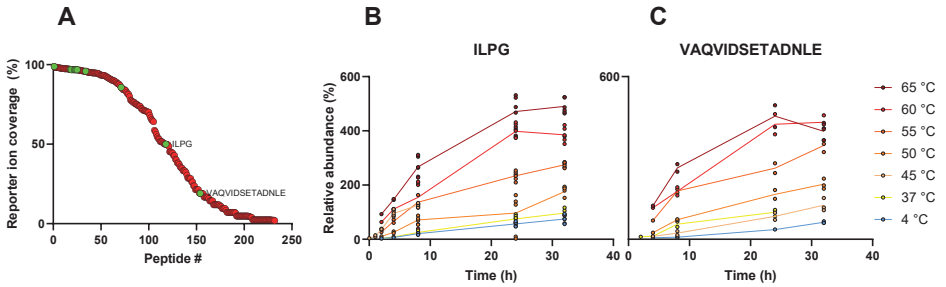


Figure 2. Peptide quantification quality. A) The identified and quantified peptides sorted by TMT reporter ion coverage, expressed as the percentage of the total number of data points (462) where the reporter was detected. The selected peptides shown in Figure 3 are marked green. B) Example of a kinetic plot of a peptide with 50% reporter ion coverage (ILPG). C) Example of a kinetic plot of a peptide with only 19% reporter ion coverage (VAQVIDSETADNLE). For panels B and C, up to 6 data points per timepoint at a given temperature are expected (triplicates measured twice). The relative abundance compared to the average intensity of a particular peptide over all points (the common reference) is plotted over time for panels B and C.

A representative selection of peptides indicative of the formation of aberrant DTd upon temperature exposure is depicted in Figure 3. Peptides originating from different parts of the toxoid were all formed faster when exposed to higher temperatures. During the studied degradation timespan, this resulted in higher areas under the abundance-time curve. Three types of kinetics could be observed. In the first type, peptides such as VTYPGLT and LTEPLME are formed rapidly, but are also degraded further into smaller peptides at the later time points. The peptide LTEPLM is of the same type, but the second phase, where degradation is faster than formation, is slower than for LTEPLME. Although we cannot determine with absolute certainty if a peptide is formed from the intact DTd or from an intermediate peptide, it is likely that the rapid formation and decrease of VTYPGLT is at least in part responsible for the formation of YPGLT. YPGLT, DSIIR and GTNPVF are part of the second type of kinetic profiles, where rapid formation in the first hours of digestion is followed by a slower second formation phase or steady state. These peptides are usually short and lack the hydrophobic branched amino acids in the middle or N terminal side of the peptide. Although cathepsin S should not be considered a completely selective endoprotease like trypsin, it has a strong preference for valine and leucine residues in the P2 position (and P1' and P3' or aromatic amino acids at P3') of the substrate [28]. Peptides lacking this cleavage preference, or those that are too short, are often not cleaved further. This results in most peptides having a valine or leucine as the penultimate amino acid. Peptides with valine and leucine are rarer in the middle and further towards the N terminal side of the peptide, because such peptides are

prone to cathepsin S digestion at this location. The last type of kinetic profile involves a steadier increase of the peptide over time, such as observed with peptides DGASRVV, HPELS or VDNENPLS. Because TMT quantification only gives relative intensities of the same peptide, a representative chromatogram of the TMT labeled and mixed sample after 32 hours of digestion is depicted in Figure 4, to put the abundance of the peptides into perspective. A selection of the peptides shown in Figure 3 is highlighted in Figure 4, and these are among the most abundant peptides in the chromatogram.

Although it is not possible to directly correlate the formation kinetics of a single peptide or a group of peptides to *in vivo* efficacy or toxicity, these kinetics could be used for the development of an assay that measures consistency between batches without the use of animal studies. The current animal tests have notoriously high variability and insensitivity which makes direct comparison between new assays and animal tests difficult [29]. Historic data has shown that toxoid vaccines are very stable when stored correctly, but exposure to elevated temperatures – similar to those used in the current study – can decrease the potency [30]. Depending on how sensitive the chosen peptides are, the differences in kinetics can distinguish even changes in the toxoid induced by exposure to 37 °C compared to the controls. The formaldehyde-inactivation process also causes changes in structural stability [31-32] and changes the susceptibility to enzymatic degradation by cathepsin S [9]. It is therefore likely that a variety of batch-to-batch differences can also be detected in this type of assay.

From our current data, it is clear that peptides can be selected that could be used in a potential degradomics-based QC assay, similar to the assay described for tetanus toxoids [13]. The selected peptides can be quantified, and deviation from the control value is indicative of denaturation of the toxoid. Important peptide selection criteria would be: (*i*) based on quantification of the peptide it should be possible to distinguish heated samples from native samples, (*ii*) the concentration of the peptide candidate should increase over time, (*iii*) the peptide should be sufficiently abundant in the sample, (*iv*) peptides containing unstable amino acid residues (*e.g.*, asparagine) should be avoided, if possible. Based on these criteria, peptides YPGLT, DSIIR and DGASRVV would be our suggested candidates for a degradomics-based QC assay for DTd. We recommend the use of our TMT-CR multiplexing strategy to identify peptides that fulfill the previously mentioned criteria. To subsequently develop a routine assay for batch-to-batch comparison and/or product stability testing, stable

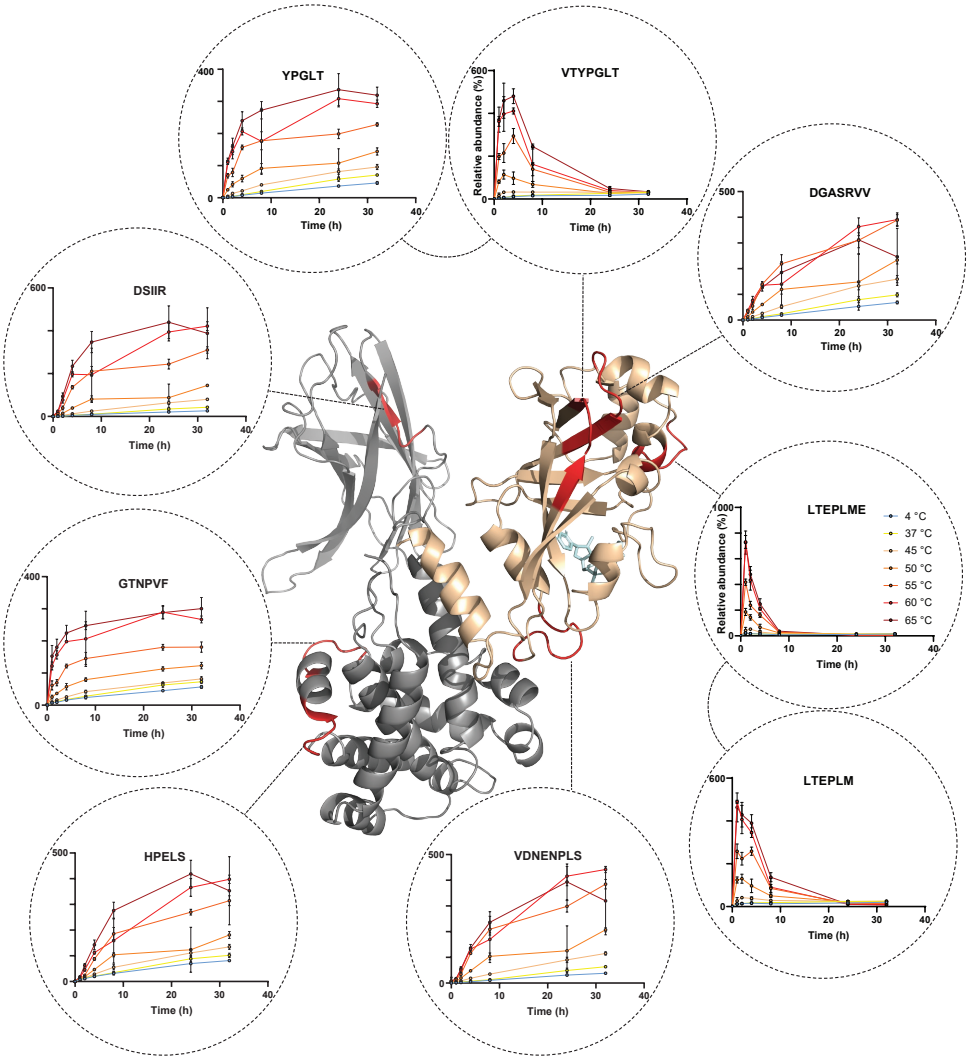


Figure 3. Kinetic profiles of a selection of representative peptides annotated to the diphtheria toxin crystal structure (PDB: 1DDT). The following color codes have been used: tan regions are part of the A fragment, grey regions are part of the B fragment and red is used to highlight the peptides. The relative abundance compared to the average intensity of a particular peptide over all points (the common reference) is plotted over time. Error bars represent the SD of the digestions of a diphtheria toxoid sample incubated and digested in triplicate that was measured in duplicate (*i.e.*, 6 data points).

isotopically labeled internal (SIL) standard peptides can be used. Quantifying these three peptides after 20 hours of exposure to cathepsin S instead of measuring several time points is sufficient if the kinetic profiles have been mapped before. Acceptable quantification criteria should be set by the manufacturer for their specific products in subsequent studies and should, for instance, take into account their batch-to-batch variation. After confirming that peptides YPGLT, DSIIR and DGASRVV can be used to pick up aberrant batches of a particular manufacturer's product, a full validation in accordance with the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) guidelines should be carried out [33]. The use of SIL standards is more common in QC than using TMT labels, because it is easy to monitor and it is transferable to more accessible Triple Stage Quadrupole mass spectrometers coupled to conventional liquid chromatography [16], and will make the assay easier to validate.

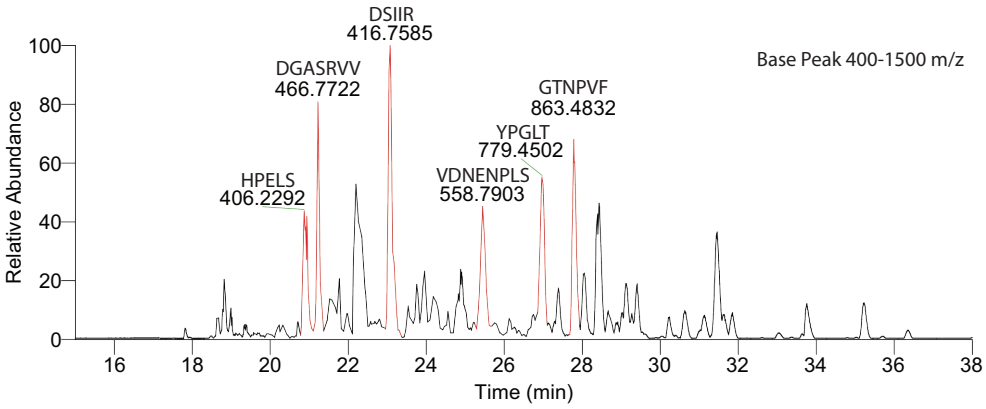


Figure 4. Base peak chromatogram of a pooled and TMT-labeled diphtheria toxoid after 32 h of cathepsin S digestion. The N-termini of the annotated peptides are TMT-labeled. The annotated peptides are those depicted in Figure 3 with increasing intensities over time.

Overall, the TMT-CR labeling strategy presented in this study allowed us to map the degradation kinetics of DTd when exposed to the cysteine protease cathepsin S. In order to distinguish mass spectrometer-errors from labeling or sample pretreatment errors, our 147 conditions were measured twice in a total of 42 runs for the sake of this study. However, measuring the triplicates only once would have been sufficient to map the degradation kinetics. In our case the common reference-based labeling strategy reduced the theoretical analysis time from almost a week to an overnight analysis. When employing all labels in the TMT 11-plex kit, a theoretical time reduction of 10x over label-free quantification is possible, and a 15x reduction with TMTpro 16-plex, without being limited to one condition per channel. This dramatically reduces the analysis time and allows for a direct, relative comparison of different samples. The use of isobaric mass tagging in conjunction with the use of a common reference, as shown in this study, has a lot of potential and should be considered when reliable relative quantification of many parameters is required.

Acknowledgements

This work was performed within the Vac2Vac project supported by the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement N-115924. We thank Wichard Tilstra for his assistance in the sample preparation, and Marjolein Zohlandt for reviewing the manuscript.

Supporting Information (available online)

Table S1 contains a complete list of peptides with their relative intensities compared to the common reference. A complete list of the MS settings is available. Supplementary Figure 1 and 2 illustrate the reproducibility between injections of the same samples. Supplementary Figure 3 shows the MS3 scan's precision with various scan parameters.

References

1. Hartman, I. Z.; Kim, A.; Cotter, R. J.; Walter, K.; Dalai, S. K.; Boronina, T.; Griffith, W.; Lanar, D. E.; Schwenk, R.; Krzych, U.; Cole, R. N.; Sadegh-Nasseri, S., A reductionist cell-free major histocompatibility complex class II antigen processing system identifies immunodominant epitopes. *Nat Med* **2010,** *16* (11), 1333-40.
2. Egger, M.; Jurets, A.; Wallner, M.; Briza, P.; Ruzek, S.; Hainzl, S.; Pichler, U.; Kitzmuller, C.; Bohle, B.; Huber, C. G.; Ferreira, F., Assessing protein immunogenicity with a dendritic cell line-derived endolysosomal degradome. *PLoS One* **2011,** *6* (2), e17278.
3. Schulten, V.; Nagl, B.; Scala, E.; Bernardi, M. L.; Mari, A.; Ciardiello, M. A.; Lauer, I.; Scheurer, S.; Briza, P.; Jurets, A.; Ferreira, F.; Jahn-Schmid, B.; Fischer, G. F.; Bohle, B., Pru p 3, the nonspecific lipid transfer protein from peach, dominates the immune response to its homolog in hazelnut. *Allergy* **2011,** *66* (8), 1005-13.
4. Ackaert, C.; Kofler, S.; Horejs-Hoeck, J.; Zulehner, N.; Asam, C.; von Grafenstein, S.; Fuchs, J. E.; Briza, P.; Liedl, K. R.; Bohle, B.; Ferreira, F.; Brandstetter, H.; Oostingh, G. J.; Duschl, A., The impact of nitration on the structure and immunogenicity of the major birch pollen allergen Bet v 1.0101. *PLoS One* **2014,** *9* (8), e104520.
5. Machado, Y.; Freier, R.; Scheiblhofer, S.; Thalhamer, T.; Mayr, M.; Briza, P.; Grutsch, S.; Ahammer, L.; Fuchs, J. E.; Wallnoefer, H. G.; Isakovic, A.; Kohlbauer, V.; Hinterholzer, A.; Steiner, M.; Danzer, M.; Horejs-Hoeck, J.; Ferreira, F.; Liedl, K. R.; Tollinger, M.; Lackner, P.; Johnson, C. M.; Brandstetter, H.; Thalhamer, J.; Weiss, R., Fold stability during endolysosomal acidification is a key factor for allergenicity and immunogenicity of the major birch pollen allergen. *J Allergy Clin Immunol* **2016,** *137* (5), 1525-34.
6. Freier, R.; Dall, E.; Brandstetter, H., Protease recognition sites in Bet v 1a are cryptic, explaining its slow processing relevant to its allergenicity. *Scientific Reports* **2015,** *5*, 12707.
7. Kitzmuller, C.; Zulehner, N.; Roulias, A.; Briza, P.; Ferreira, F.; Fae, I.; Fischer, G. F.; Bohle, B., Correlation of sensitizing capacity and T-cell recognition within the Bet v 1 family. *J Allergy Clin Immunol* **2015,** *136* (1), 151-8.
8. Kitzmuller, C.; Wallner, M.; Deifl, S.; Mutschlechner, S.; Walterskirchen, C.; Zlabinger, G. J.; Ferreira, F.; Bohle, B., A hypoallergenic variant of the major birch pollen allergen shows distinct characteristics in antigen processing and T-cell activation. *Allergy* **2012,** *67* (11), 1375-82.
9. 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.
10. Ghimire, T. R.; Benson, R. A.; Garside, P.; Brewer, J. M., Alum increases antigen uptake, reduces antigen degradation and sustains antigen presentation by DCs *in vitro*. *Immunol Lett* **2012,** *147* (1-2), 55-62.
11. Delamarre, L.; Couture, R.; Mellman, I.; Trombetta, E. S., Enhancing immunogenicity by limiting susceptibility to lysosomal proteolysis. *J Exp Med* **2006,** *203* (9), 2049-55.
12. Delamarre, L.; Pack, M.; Chang, H.; Mellman, I.; Trombetta, E. S., Differential lysosomal proteolysis in antigen-presenting cells determines antigen fate. *Science* **2005,** *307* (5715), 1630-4.
13. Michiels, T. J. M.; Tilstra, W.; Hamzink, M. R. J.; de Ridder, J. W.; Danial, M.; Meiring, H. D.; Kersten, G. F. A.; Jiskoot, W.; Metz, B., Degradomics-Based Analysis of Tetanus Toxoids as a Quality Control Assay. *Vaccines* **2020,** *8* (4), 712.
14. 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**.

15. DeLange, R. J.; Williams, L. C.; Drazin, R. E.; Collier, R. J., The amino acid sequence of fragment A, an enzymically active fragment of diphtheria toxin. III. The chymotryptic peptides, the peptides derived by cleavage at tryptophan residues, and the complete sequence of the protein. *J Biol Chem* **1979,** *254* (13), 5838-42.

16. Bozovic, A.; Kulasingam, V., Quantitative mass spectrometry-based assay development and validation: from small molecules to proteins. *Clin Biochem* **2013,** *46* (6), 444-55.

17. Rauniyar, N.; Yates, J. R., 3rd, Isobaric labeling-based relative quantification in shotgun proteomics. *J Proteome Res* **2014,** *13* (12), 5293-309.

18. Arul, A. B.; Robinson, R. A. S., Sample Multiplexing Strategies in Quantitative Proteomics. *Anal Chem* **2019,** *91* (1), 178-189.

19. Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K.; Hamon, C., Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **2003,** *75* (8), 1895-904.

20. Thompson, A.; Wolmer, N.; Koncarevic, S.; Selzer, S.; Bohm, G.; Legner, H.; Schmid, P.; Kienle, S.; Penning, P.; Hohle, C.; Berfelde, A.; Martinez-Pinna, R.; Farztdinov, V.; Jung, S.; Kuhn, K.; Pike, I., TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-Based Isobaric 16-Plex Tandem Mass Tag Reagent Set. *Anal Chem* **2019,** *91* (24), 15941-15950.

21. Kim, H.; Zhao, B.; Snesrud, E. C.; Haas, B. J.; Town, C. D.; Quackenbush, J., Use of RNA and genomic DNA references for inferred comparisons in DNA microarray analyses. *Biotechniques* **2002,** *33* (4), 924-30.

22. an de Waterbeemd, B.; Mommen, G. P.; Pennings, J. L.; Eppink, M. H.; Wijffels, R. H.; van der Pol, L. A.; de Jong, A. P., Quantitative proteomics reveals distinct differences in the protein content of outer membrane vesicle vaccines. *J Proteome Res* **2013,** *12* (4), 1898-908.

23. Plubell, D. L.; Wilmarth, P. A.; Zhao, Y.; Fenton, A. M.; Minnier, J.; Reddy, A. P.; Klimek, J.; Yang, X.; David, L. L.; Pamir, N., Extended Multiplexing of Tandem Mass Tags (TMT) Labeling Reveals Age and High Fat Diet Specific Proteome Changes in Mouse Epididymal Adipose Tissue. *Mol Cell Proteomics* **2017,** *16* (5), 873-890.

24. Dou, Y.; Kawaler, E. A.; Cui Zhou, D.; Gritsenko, M. A.; Huang, C.; Blumenberg, L.; Karpova, A.; Petyuk, V. A.; Savage, S. R.; Satpathy, S.; Liu, W.; Wu, Y.; Tsai, C. F.; Wen, B.; Li, Z.; Cao, S.; Moon, J.; Shi, Z.; Cornwell, M.; Wyczalkowski, M. A.; Chu, R. K.; Vasaikar, S.; Zhou, H.; Gao, Q.; Moore, R. J.; Li, K.; Sethuraman, S.; Monroe, M. E.; Zhao, R.; Heiman, D.; Krug, K.; Clauser, K.; Kothadia, R.; Maruvka, Y.; Pico, A. R.; Oliphant, A. E.; Hoskins, E. L.; Pugh, S. L.; Beecroft, S. J. I.; Adams, D. W.; Jarman, J. C.; Kong, A.; Chang, H. Y.; Reva, B.; Liao, Y.; Rykunov, D.; Colaprico, A.; Chen, X. S.; Czekanski, A.; Jedryka, M.; Matkowski, R.; Wiznerowicz, M.; Hiltke, T.; Boja, E.; Kinsinger, C. R.; Mesri, M.; Robles, A. I.; Rodriguez, H.; Mutch, D.; Fuh, K.; Ellis, M. J.; DeLair, D.; Thiagarajan, M.; Mani, D. R.; Getz, G.; Noble, M.; Nesvizhskii, A. I.; Wang, P.; Anderson, M. L.; Levine, D. A.; Smith, R. D.; Payne, S. H.; Ruggles, K. V.; Rodland, K. D.; Ding, L.; Zhang, B.; Liu, T.; Fenyo, D.; Clinical Proteomic Tumor Analysis, C., Proteogenomic Characterization of Endometrial Carcinoma. *Cell* **2020,** *180* (4), 729-748 e26.

25. Satpathy, S.; Jaehnig, E. J.; Krug, K.; Kim, B. J.; Saltzman, A. B.; Chan, D. W.; Holloway, K. R.; Anurag, M.; Huang, C.; Singh, P.; Gao, A.; Namai, N.; Dou, Y.; Wen, B.; Vasaikar, S. V.; Mutch, D.; Watson, M. A.; Ma, C.; Ademuyiwa, F. O.; Rimawi, M. F.; Schiff, R.; Hoog, J.; Jacobs, S.; Malovannaya, A.; Hyslop, T.; Clauser, K. R.; Mani, D. R.; Perou, C. M.; Miles, G.; Zhang, B.; Gillette, M. A.; Carr, S. A.; Ellis, M. J., Microscaled proteogenomic methods for precision oncology. *Nat Commun* **2020,** *11* (1), 532.

26. VAC2VAC Consortium website. http://www.vac2vac.eu/ (accessed October 16th 2020).

27. 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.

28. Neil D. Rawlings, G. S. S., *Handbook of Proteolytic Enzymes*. Academic Press: 2013; Vol. 2, p 4094.

29. Stalpers, C. A. L.; Retmana, I. A.; Pennings, J. L. A.; Vandebriel, R. J.; Hendriksen, C. F. M.; Akkermans, A. M.; Hoefnagel, M. H. N., Variability of *in vivo* potency tests of Diphtheria, Tetanus and acellular Pertussis (DTaP) vaccines. *Vaccine* **2021**.

30. Milstein, J. B. *Temperature Sensitivity of Vaccines*; WHO/IVB/06.10; World Health Organization: 2006.

31. 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.

32. Paliwal, R.; London, E., Comparison of the conformation, hydrophobicity, and model membrane interactions of diphtheria toxin to those of formaldehyde-treated toxin (diphtheria toxoid): formaldehyde stabilization of the native conformation inhibits changes that allow membrane insertion. *Biochemistry* **1996,** *35* (7), 2374-9.

33. ICH Q2(R1) Validation of Analytical Procedures: Text and Methodology. https://ich.org/page/quality-guidelines (accessed 19th of April 2021).

7

SUMMARY, GENERAL DISCUSSION AND PROSPECTS

Summary

Since their discovery, toxoid vaccines have proven to be reliable, safe, and effective. Inactivation of toxins, such as diphtheria and tetanus toxin, with formaldehyde results in very heterogeneous modified toxin molecules, called toxoids. The complexity of toxoid vaccines is exacerbated by adsorption of the toxoid to colloidal aluminum salts, leading to a colloidal system with a heterogeneous size distribution. Most conventional techniques to analyze biologicals, such as ELISA, UV absorption spectroscopy and size-exclusion chromatography, are incompatible with turbid analytes. Moreover, the heterogeneity makes it difficult to analytically characterize every antigen species in the vaccine; the mean result of an assay may hide the presence of certain antigen species that could significantly contribute to the safety and efficacy of the vaccine. Historically, functional tests for measuring immunity and safety in animals have been used to mitigate these issues. Up until today, such animal tests are being employed as batch release assays to test both safety and potency [1,2]. Efforts to replace or reduce these *in vivo* assays have been made based on a so-called consistency approach [3,4]: if the consistency between batches can be proven by using *in vitro* assays, no –or fewer– additional *in vivo* assays would be required. In line with this approach, we aimed to develop a new *in vitro* assay mimicking *in vivo* antigen processing and degradation. The field studying proteases, their substrates and their products has been termed *degradomics* [5]. Because differences in antigen processing *in vivo* are associated with differences in immunogenicity [6-12], mapping the enzymatic proteolysis of antigen and the peptides that are formed in a degradomics assay could be a viable alternative to animal tests.

Although toxoids are heterogeneous mixtures of formaldehyde-modified protein molecules, a well-designed and consistent production process should result in products with consistent quality. *i.e.*, potency, safety and stability. Characterization of the detoxification process and the resulting changes in the protein can support the development of new assays and can be used to compare different products. In **chapter 2,** formaldehyde-induced chemical modifications of diphtheria toxoid were mapped by mass spectrometry. To overcome the issues associated with the toxoid's heterogeneity, a bottom-up proteomics approach was applied, *i.e.*, analyzing the enzymatically digested toxoid. The diphtheria toxin was prepared with regular or deuterated formaldehyde, which then allowed software tools to compile a list of peptides with formaldehyde adducts. After identification of the modified peptides,

the location of the formaldehyde-induced modifications in the protein was obtained. The NAD^+-binding cavity and the receptor-binding site, important for the toxin's toxicity, were found to be affected by formaldehyde. Additionally, an important CD4 T-cell epitope was unaffected. Many adducts were crosslinks between glycine (present in the inactivation matrix) and the toxoid. If the glycine used in our studies would be replaced by other amino acids (or other amine-containing chemicals in general), as is sometimes the case, this will result in a toxoid that is different on a molecular level. Therefore, care should be taken when choosing (or changing) the matrix components. The heterogeneity of the toxoids makes it impossible to definitively identify a single formaldehyde modification responsible for loss of toxicity or potential loss of immunogenicity. Nevertheless, a thorough characterization of these products with an approach as presented in **chapter 2** could help to compare different products and assess their comparability.

The methods used in **chapter 2** can be applied for the identification of formaldehyde-induced modifications in other proteins as well. In **chapter 3** the effect of formaldehyde-induced modifications on the enzymatic proteolysis of bovine serum albumin, β-lactoglobulin and cytochrome c by cathepsin S is described. Contrary to our expectations, exposure to higher concentrations of formaldehyde made these model proteins more susceptible to proteolytic degradation. The most abundant chemical modifications could be mapped. Some specific modifications in cytochrome c –in a particular formaldehyde-glycine adduct– were found to be responsible for faster degradation of certain parts of the protein. However, even proteins subjected to formaldehyde in the absence of additional amino acids showed increased degradation overall. The differences in enzymatic degradation kinetics obtained after incubation with various concentrations of formaldehyde highlight the influence of the various aspects of the inactivation process, such as the importance of using a standardized amount of formaldehyde for antigen detoxification. Our newly developed degradomics assay could pick up these formaldehyde-induced changes in proteins of interest.

Unexpected mass increases of peptides derived from formaldehyde-treated cytochrome c were observed during the studies of **chapter 3**. **Chapter 4** focuses on the structural elucidation of these formaldehyde-induced modifications. Two proximate lysine residues were able to undergo (*i*) formaldehyde-induced deamination and formation of α,β-unsaturated aldehydes and methylation on the two lysine residues, and (*ii*) formaldehyde-induced methylation and

formylation of two adjacent lysine residues. These modifications result in intramolecular cross-links between two lysine residues in a single peptide, but also form cross-linked dimers and trimers. Knowledge of these potential formaldehyde-modifications in vaccine products could aid in the development and registration of new vaccines and contribute to the characterization of current vaccines.

Chapter 5 describes the application of the degradomics approach for the analysis of tetanus toxoids. Several tetanus toxoids were obtained from different manufacturers and subjected to elevated temperatures to simulate faulty batches. First, the stressed samples were subjected to cathepsin S digestion, and the kinetics of peptide formation was evaluated. Subsequently, three suitable stability-indicating peptides were selected and quantified after a fixed digestion time. Our improved degradomics assay could pick up temperature-induced changes in these products. Importantly, the method worked for both non-adsorbed and adsorbed tetanus toxoids. Therefore, analysis with our method is not restricted to intermediate (non-adsorbed) bulk toxoids but can also be used to assess the quality of the final (adsorbed) drug product. Our degradomics assay is both sensitive and precise, making it a good potential alternative for the current animal tests.

To be able to easily apply the degradomics assay to other antigens, the analysis of the kinetic profiles of all peptides originating from enzymatic degradation was further optimized in **chapter 6**. By using tandem mass tags (TMT), multiple samples could be compared to a pooled control sample: the *common reference*. This greatly reduced the analysis time and decreased assay variation, enabling easier implementation of our degradomics approach to other antigens and vaccines.

General discussion and prospects

The main objective of this thesis was to develop a new *in vitro* assay that can confirm product consistency and potentially can replace the current *in vivo* assays used for the batch release of toxoid vaccines. Our envisioned assay was inspired by the immune system, where changes in antigen degradation in antigen-presenting cells have been correlated with altered immune responses; slower degradation has been found to correlate to higher immunogenicity [6-12]. From this starting point several approaches were possible: antigen degradation by dendritic cells (DCs) has been studied (*i*) *in vivo* and (*ii*) *ex vivo* [9], (*iii*) in cell lines [8], in (*iv*) lysosomal extracts derived from DCs [8], and (*v*) in assays where only key enzymes and proteins are added [6,13]. For future routine use, assay robustness is key. With decreasing biological and immunological fidelity, the complexity of the assay also decreases. For instance, it is much easier to produce a recombinant enzyme in a consistent manner than it is to obtain lysosomal extracts from DCs in a consistent manner. Therefore, we opted to develop a degradomics assay based on recombinant human cathepsin S. Arguably, showing consistency with a precise method is more important for a vaccine release assay than showing a (slightly) more direct correlation with immunogenicity with an imprecise method.

An ideal *in vitro* assay would be a direct measure for a vaccine's immunogenicity. However, in reality this is not feasible for any single *in vitro* assay. Because of the complexity of the immune system, the currently implemented alternatives for certain immunological animal tests always focus on a specific aspect of the immune system. For example, the polio D-antigen ELISA release assay [14] is used to test polio vaccines based on antigenicity, whereas the intended effect of the vaccine is to induce a protective immune response (immunogenicity) and not to bind antibodies (antigenicity). Similarly, alternatives to rabbit pyrogen tests involve cell-based assays based on monocyte activation [15]: again the readout is related but not identical. The degradomics assay attempts to mimic antigen processing, which is an important step in inducing a CD4 T-cell response, but it is also limited in that it only covers a part of the T-cell response and not the –at least– equally important B-cell response needed for the production of protective antibodies. Moreover, the assay as presented in **chapter 5** does not include other important proteins involved in MHC-II epitope selection and processing, such as other proteases, HLA-DM (involved in peptide selection) or HLA cell surface receptors. These additional proteases can be added as demonstrated by Isamu Hartman *et al.*, but

this would severely increase the assay's complexity [13]. From a scientific point of view, it may be interesting to evaluate the correlation between the immunogenicity of a vaccine and the readout of our degradomics assay. Tetanus toxoids exposed to elevated temperatures have shown decreased potency in previous studies [16]. Indeed, the results described in **chapter 5** show enhanced enzymatic proteolysis upon heat treatment of tetanus toxoids, which supports the hypothesis that antigens that are more prone to proteolysis are less immunogenic. In general, tetanus and diphtheria toxoids were found to be very protease resistant, especially when compared to the rapid enzymatic degradation after the toxoids were exposed to elevated temperatures. Formaldehyde-treatment of model proteins resulted in faster proteolytic degradation, so it would be interesting to both determine the immunogenicity of model proteins and their formaldehyde-treated counterparts. If proteolysis correlates with immunogenicity, the formaldehyde-treated proteins will not be more immunogenic compared to their native model proteins. Furthermore, it would be interesting to determine the effect formaldehyde-treatment of diphtheria and tetanus toxin has on their enzymatic degradation. Previous research has shown that formaldehyde-treatment enhances the immunogenicity of diphtheria toxoid compared to diphtheria toxin [17], so slower degradation would be expected despite our findings with model proteins. To find a stronger correlation between protease resistance and immunogenicity, antigens should be altered in such a way that the protein's B-cell epitopes are minimally affected but antigen proteolysis is either enhanced or decreased, similar to the minor changes of the antigens compared in a study by Egger *et al.* [8]. Analysis of a similar set of vaccine antigens in an *in vivo* immunogenicity study could then be used to confirm the correlation. However, from an ethical point of view it seems counterintuitive to perform an animal experiment with the goal to reduce the number of animal tests. Moreover, from a quality control point-of-view, showing product consistency –rather than demonstrating a correlation with *in vivo* immunogenicity– is key. There are always more insights to be gained from additional animal tests, but their use should be limited if they are not strictly necessary to reach the predefined study objectives, as is the case in this thesis.

Compared to other *in vitro* alternatives that are being developed or have been developed, the degradomics approach has several advantages. Most other assays, such as physicochemical techniques [17,18], ELISAs [19] and cell-based assays [20-23], require biological reference samples for comparison. Limited shelf life and batch-to-batch consistency of such standards are

important considerations, although, for instance, various toxoid standards are available from the National Institute for Biological Standards and Control (NIBSC). If the formation of specific peptides is analyzed with a degradomics analysis, absolute quantification (*i.e.*, moles per antigen quantity at a given reaction time) can be achieved with the help of synthetic internal standards, and the assay does not rely on biological reference samples. Production, characterization and validation of these standards is much easier than the use of biological reference standards, and should facilitate technology transfer to labs world-wide. Another advantage of the degradomics assay is that it evaluates specific parts of the protein and can identify changes in a small subset of proteins. In other assays, such as physicochemical methods, the average response of the antigen is measured. Small changes in large proteins could remain unnoticed because they might have little influence on the average response. Analysis of parts of the protein that are particularly sensitive to inconsistency (including production- and storage-induced changes) could result in a more sensitive assay. Further advantages include the orthogonal nature of the assay compared to the other *in vitro* alternatives and the very limited amount of sample required for mass-spectrometric analysis. The use of mass spectrometry, however, does come with drawbacks; it requires relatively expensive specialized equipment and specialized personnel. Moreover, the wealth of data generated makes it challenging to evaluate everything, and the much-needed focus on a small subset of the data could introduce bias, particularly when setting up the assay for a new antigen. Finally, the degradomics assay can be used to test antigens in the presence of aluminum salts, so the final drug product can be assayed, something that has been challenging for other assays. Overall, the complementarity with other assays makes the degradomics approach a very suitable candidate for a panel consisting of multiple tests, which could be used for both stability and release assays, because its pros and cons are complementary to other *in vitro* tests and different parameters of the product are analyzed. For instance, when the results of a degradomics assay are combined with those of an antigenicity assay such as an ELISA, both correlations to the important CD4 response (antigen processing) and binding of the B-cell to the antigen (and antigen processing in the B-cell itself) are covered.

In this thesis, a proof of principle for an *in vitro* degradomics approach for quality control of toxoid vaccines has been established. However, several other research questions, which could be answered in subsequent studies, remain. These topics can be divided into research questions related to the vaccine(s) and those related to the assay itself. For instance, the

studies in **chapter 3**, which addressed the effects of formaldehyde-inactivation on model proteins, could be followed up by studies on toxin inactivation and its effect on enzymatic proteolysis. Formaldehyde inactivation is a key step in toxoid vaccine production processes, both for safety and efficacy, so it would be ideal if our assay could pick up unwanted changes in the inactivation process. The working degradomics assay is described for tetanus toxoid in **chapter 5**, and general proof of principle is described for diphtheria toxoid in **chapter 6**. A subsequent step would be to show that the assay works for an actual (combination) vaccine containing all excipients, adjuvant and other antigens. Moreover, the assay could be further optimized to include the commonly included polio and pertussis antigens, so that all antigens in the combination vaccine can be assayed at once. This raises the question whether the degradomics assay is capable of evaluating more complicated antigens, such as tetanus toxoid-haemophilus influenzae type b polysaccharide conjugates. It seems likely that the polysaccharide-conjugated tetanus toxoid will react similarly to the unconjugated toxoid, *i.e.*, exposure to higher temperatures leads to faster enzymatic degradation. However, it remains to be seen if the degradomics assay will provide useful insights, since the polysaccharide contains the relevant B-cell epitopes, although an adequate CD4 response to the carrier protein is important for long term efficacy [24]. Further research into the degradomics assay itself may shed more light on the usefulness of this assay for conjugate vaccines. For more complex antigens such as live attenuated vaccines and vector-based vaccines, mimicking the MHC-II processing pathway seems illogical. Other assays (such as virus titration) are more suitable for the analysis of those products. For further studies it would be interesting to evaluate the degradation kinetics of the antigens when treated with other proteases, such as trypsin. It is likely that for antigens that are not treated with formaldehyde, which affects the lysine and arginine residues needed for trypsin digestion, similar kinetic trends will be observed when using trypsin (or other proteases) instead of cathepsin S. Finally, the discovery of novel formaldehyde modifications described in **chapter 4** suggests that several related undiscovered modifications could be present in formaldehyde-treated antigens. Identification and characterization of these products in real vaccines would be another important research topic and could aid in the development and registration of new formaldehyde-inactivated vaccines, as well as support the replacement of animal tests in current vaccines.

Altogether, besides the additional knowledge about formaldehyde-adducts (**chapters 2, 3** and **4**), the degradomics approach has resulted in a sensitive assay that may be able to pick

up subtle changes in the protein folding of antigens used in vaccines. Actual implementation of the degradomics assay described in this thesis and those proposed by others for the batch-release of vaccines requires further validation by vaccine manufacturers and cooperation by regulatory agencies. However, all parties involved should have sufficient incentives (*e.g.*, reduction in assay variation, reduction in costs) to ensure that further reduction and replacement of animal tests is achieved in the coming years. Moreover, the data presented in this thesis could provide useful insights into the characteristics of new vaccines and provide alternatives to animal tests before *in vivo* release tests are implemented.

References

1. Council of Europe. European Pharmacopoeia 10.0. In *Assay of tetanus vaccine (adsorbed)*, 2020; pp 275-278.
2. Council of Europe. European Pharmacopoeia 10.0. In *Tetanus vaccine (adsorbed)*, 2020; pp 1042-1043.
3. De Mattia, F.; Hendriksen, C.; Buchheit, K.H.; Chapsal, J.M.; Halder, M.; Lambrigts, D.; Redhead, K.; Rommel, E.; Scharton-Kersten, T.; Sesardic, T., *et al.* The vaccines consistency approach project: an EPAA initiative. *Pharmeur Bio Sci Notes* **2015**, *2015*, 30-56.
4. De Mattia, F.; Chapsal, J.M.; Descamps, J.; Halder, M.; Jarrett, N.; Kross, I.; Mortiaux, F.; Ponsar, C.; Redhead, K.; McKelvie, J., *et al.* The consistency approach for quality control of vaccines- a strategy to improve quality control and implement 3Rs. *Biologicals* **2011**, *39*, 59-65, doi:10.1016/j.biologicals.2010.12.001.
5. Lopez-Otin, C.; Overall, C.M. Protease degradomics: a new challenge for proteomics. *Nat Rev Mol Cell Biol* **2002**, *3*, 509-519, doi:10.1038/nrm858.
6. Carmicle, S.; Dai, G.; Steede, N.K.; Landry, S.J. Proteolytic sensitivity and helper T-cell epitope immunodominance associated with the mobile loop in Hsp10s. *J Biol Chem* **2002**, *277*, 155-160, doi:10.1074/jbc.M107624200.
7. Kim, A.; Hartman, I.Z.; Poore, B.; Boronina, T.; Cole, R.N.; Song, N.; Ciudad, M.T.; Caspi, R.R.; Jaraquemada, D.; Sadegh-Nasseri, S. Divergent paths for the selection of immunodominant epitopes from distinct antigenic sources. *Nat Commun* **2014**, *5*, 5369, doi:10.1038/ncomms6369.
8. Egger, M.; Jurets, A.; Wallner, M.; Briza, P.; Ruzek, S.; Hainzl, S.; Pichler, U.; Kitzmuller, C.; Bohle, B.; Huber, C.G., *et al.* Assessing protein immunogenicity with a dendritic cell line-derived endolysosomal degradome. *PLoS One* **2011**, *6*, e17278, doi:10.1371/journal.pone.0017278.
9. Delamarre, L.; Couture, R.; Mellman, I.; Trombetta, E.S. Enhancing immunogenicity by limiting susceptibility to lysosomal proteolysis. *J Exp Med* **2006**, *203*, 2049-2055, doi:10.1084/jem.20052442.
10. Machado, Y.; Freier, R.; Scheiblhofer, S.; Thalhamer, T.; Mayr, M.; Briza, P.; Grutsch, S.; Ahammer, L.; Fuchs, J.E.; Wallnoefer, H.G., *et al.* Fold stability during endolysosomal acidification is a key factor for allergenicity and immunogenicity of the major birch pollen allergen. *J Allergy Clin Immunol* **2016**, *137*, 1525-1534, doi:10.1016/j.jaci.2015.09.026.
11. Freier, R.; Dall, E.; Brandstetter, H. Protease recognition sites in Bet v 1a are cryptic, explaining its slow processing relevant to its allergenicity. *Scientific Reports* **2015**, *5*, 12707, doi:10.1038/srep12707.
12. Kitzmuller, C.; Wallner, M.; Deifl, S.; Mutschlechner, S.; Walterskirchen, C.; Zlabinger, G.J.; Ferreira, F.; Bohle, B. A hypoallergenic variant of the major birch pollen allergen shows distinct characteristics in antigen processing and T-cell activation. *Allergy* **2012**, *67*, 1375-1382, doi:10.1111/all.12016.
13. Hartman, I.Z.; Kim, A.; Cotter, R.J.; Walter, K.; Dalai, S.K.; Boronina, T.; Griffith, W.; Lanar, D.E.; Schwenk, R.; Krzych, U., *et al.* A reductionist cell-free major histocompatibility complex class II antigen processing system identifies immunodominant epitopes. *Nat Med* **2010**, *16*, 1333-1340, doi:10.1038/nm.2248.
14. Kouiavskaia, D.; Puligedda, R.D.; Dessain, S.K.; Chumakov, K. Universal ELISA for quantification of D-antigen in inactivated poliovirus vaccines. *J Virol Methods* **2020**, *276*, 113785, doi:10.1016/j.jviromet.2019.113785.
15. da Silva, C.C.; Presgrave, O.A.F.; Hartung, T.; de Moraes, A.M.L.; Delgado, I.F. Applicability of the Monocyte Activation Test (MAT) for hyperimmune sera in the routine of the quality control laboratory: Comparison with the Rabbit Pyrogen Test (RPT). *Toxicology in Vitro* **2016**, *32*, 70-75, doi:10.1016/j.tiv.2015.12.004.
16. Cohen, H.; Van Ramshorst, J.; Tasman, A. Consistency in potency assay of tetanus toxoid in mice. *Bull World*

Health Organ **1959**, *20*, 1133-1150.

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*, 156-167.

18. 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*, 6863-6871, doi:10.1016/j.vaccine.2007.07.009.

19. Riches-Duit, R.; Hassall, L.; Rigsby, P.; Stickings, P. Evaluation of a capture antigen ELISA for the characterisation of tetanus vaccines for veterinary use. *Biologicals* **2019**, *61*, 8-14, doi:10.1016/j.biologicals.2019.08.003.

20. van den Biggelaar, R.; van Eden, W.; Rutten, V.; Jansen, C.A. Macrophage Activation Assays to Evaluate the Immunostimulatory Capacity of Avibacterium paragallinarum in A Multivalent Poultry Vaccine. *Vaccines (Basel)* **2020**, *8*, doi:10.3390/vaccines8040671.

21. Hoonakker, M.E.; Verhagen, L.M.; van der Maas, L.; Sloots, A.; Hendriksen, C.F.M. Reporter cell lines for detection of pertussis toxin in acellular pertussis vaccines as a functional animal-free alternative to the *in vivo* histamine sensitization test. *Vaccine* **2017**, *35*, 1152-1160, doi:10.1016/j.vaccine.2017.01.011.

22. Vandebriel, R.; Hoefnagel, M.M. Dendritic cell-based *in vitro* assays for vaccine immunogenicity. *Hum Vaccin Immunother* **2012**, *8*, 1323-1325, doi:10.4161/hv.21350.

23. Hoefnagel, M.H.; Vermeulen, J.P.; Scheper, R.J.; Vandebriel, R.J. Response of MUTZ-3 dendritic cells to the different components of the Haemophilus influenzae type B conjugate vaccine: towards an *in vitro* assay for vaccine immunogenicity. *Vaccine* **2011**, *29*, 5114-5121, doi:10.1016/j.vaccine.2011.05.050.

24. Kelly, D.F.; Moxon, E.R.; Pollard, A.J. Haemophilus influenzae type b conjugate vaccines. *Immunology* **2004**, *113*, 163-174, doi:10.1111/j.1365-2567.2004.01971.x.

Appendix

NEDERLANDSTALIGE SAMENVATTING
LIST OF PUBLICATIONS
CURRICULUM VITAE

Nederlandstalige samenvatting

De COVID-19-pandemie heeft het belang van vaccins extra benadrukt. Voor het ontwikkelen van nieuwe vaccins en het optimaliseren van bestaande vaccins is het niet alleen belangrijk om nieuwe vaccinconcepten en analysetechnieken te ontwikkelen, maar evengoed nuttig om te leren van al bestaande, succesvolle vaccins, zoals die tegen difterie en tetanus.

Een belangrijk aspect van alle geneesmiddelen is de kwaliteitscontrole. Waar bij de meeste therapeutische geneesmiddelen *in vitro* analyses volstaan, worden bij vaccins van oudsher veel dierproeven gebruikt. Voor difterie- en tetanusvaccins vinden er, bijna een eeuw na de introductie van deze vaccins, nog dierproeven plaats om de veiligheid en werkzaamheid van elke vaccinpartij aan te tonen. Het in deze dissertatie beschreven onderzoek was gericht op het ontwikkelen van nieuwe *in vitro* testen ter vervanging en vermindering van deze dierproeven.

Difterie- en tetanusvaccins zijn zogeheten toxoïdvaccins. Bijna een eeuw na hun ontwikkeling worden deze vaccins nog op vergelijkbare wijze gemaakt. Difterie en tetanus worden veroorzaakt door toxines die door bacteriën uitgescheiden worden. Door deze toxines te behandelen met formaldehyde kunnen deze onschadelijk worden gemaakt. Deze onschadelijk gemaakte toxines (toxoïden) kunnen worden gebruikt om in mensen (maar ook in dieren) immuniteit op te wekken tegen de zeer giftige toxines. Toxoïdvaccins behoren tot de meest succesvolle vaccins, met zeer goede effectiviteit en veiligheid.

De behandeling met formaldehyde zorgt voor een grote verscheidenheid aan chemische modificaties van verschillende delen van de toxinemoleculen. In hoofdstuk 2 van dit proefschrift worden de belangrijkste chemische modificaties van met formaldehyde behandeld difterietoxine in kaart gebracht. In hoofdstuk 4 beschrijven we een niet eerder beschreven reactie van formaldehyde met twee lysineresiduen. Deze fundamentele onderzoeken naar de werkingsmechanismen van het inactivatieproces kunnen als basis dienen voor het analyseren van met formaldehyde behandelde antigenen.

Voor het vervangen van dierproeven door een *in vitro* test hebben we geprobeerd om een belangrijke stap van het immuunsysteem na te bootsen. Een vroege stap in het opwekken van immuniteit is het activeren van CD4+ T-helper cellen. Dat gebeurt door activatie via een

antigeen-presenterende cel (APC). De APC neemt het antigeen (bijvoorbeeld een toxoïd) op en breekt dit met behulp van enzymen af tot kleinere peptiden. Deze peptiden worden vervolgens via de major histocompatibility complex class II moleculen (MHC-II) aan T-helper cellen gepresenteerd en zorgen met name voor CD4 T-celactivatie. Een kritische stap in dit proces is de enzymatische afbraak van het antigeen door de APC. Als het antigeen wordt afgebroken tot kleinere peptiden kan het gepresenteerd worden door APC's, maar als het te snel wordt afgebroken tot losse aminozuren dan blijft er niks over dat gepresenteerd kan worden. De hypothese aan de basis van dit onderzoek was dat voor een optimale immuunrespons een optimale afbraaksnelheid van het antigeen nodig is. Afwijkingen van het antigen, ontstaan door bijvoorbeeld veranderingen in het productieproces (met name bij de belangrijke behandeling met formaldehyde) of tijdens de opslag, zouden kunnen leiden tot een verandering in afbraaksnelheid. De enzymen betrokken bij de afbraak zouden bijvoorbeeld een bepaalde knipplaats niet meer kunnen herkennen, of door een ontvouwing van een deel van het antigeen zou er een nieuwe knipplaats beschikbaar kunnen komen. In hoofdstuk 3 is gekeken naar de invloed van formaldehydebehandeling van modeleiwitten op de afbraak door één van de belangrijkste enzymen in een APC: cathepsin S. Tegen de verwachting in werden de eiwitten sneller door cathepsin S afgebroken nadat ze door formaldehyde behandeld waren, maar de effecten waren wel specifiek voor de locatie binnen het eiwit. Sommige formaldehydemodificaties zorgden voor langzamere enzymatische afbraak, maar de meeste modificaties zorgden juist voor snellere afbraak van de modeleiwitten. In hoofdstuk 5 en hoofdstuk 6 is deze strategie verder toegepast voor het analyseren van tetanus- en difterietoxoïden. De in deze hoofstukken ontwikkelde test, gebaseerd op enzymatische antigeenafbraak met cathepsin S, bleek zeer gevoelig voor het detecteren van structurele veranderingen in de antigenen. Het was zelfs mogelijk de test uit te voeren in de aanwezigheid van colloïdale aluminiumzouten. Deze aluminiumzouten worden als adjuvans gebruikt voor het verbeteren van de immuunrespons, maar bemoeilijken de meeste conventionele analysemethoden.

In combinatie met andere *in vitro* testen geeft deze enzymatische degradatietest nieuwe inzichten in de kwaliteit van vaccins. Het vervangen van dierproeven is met deze *in vitro* testen binnen handbereik, en difterie- en tetanusvaccins zijn zeer geschikte kandidaten om dit als eerste voor elkaar te krijgen. De nieuwe test zou in de toekomst ook vanaf de ontwikkeling van nieuwe op eiwitantigeen gebaseerde vaccins toegepast kunnen worden; het voorkómen van een *in vivo* test is immers makkelijker dan hem achteraf vervangen.

List of Publications

Yang, X.; Dong, G.; **Michiels, T. J. M.**; Lenselink, E. B.; Heitman, L.; Louvel, J.; IJzerman, A.P., A Covalent Antagonist for the Human Adenosine A2A Receptor. *Purinergic Signal* **2017,** *13 (2), 191-201.*

Yang, X.; **Michiels, T. J. M.**; de Jong, C.; Soethoudt, M.; Dekker, N.; Gordon, E.; van der Stelt, M.; Heitman, L. H.; van der Es, D.; IJzerman, A.P., An Affinity-Based Probe for the Human Adenosine A2A Receptor. *J Med Chem* **2018,** *61 (17), 7892-7901.*

Ben Haïm, C., Depré, D.P.M., Medina, F.D., **Michiels, T.J.M.**, Wagschal, S.A., Processes and Intermediates for Preparing a BTK Inhibitor, Applicant: Janssen Pharmaceutica, Patent: WO2018065504

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* **2020,** *109 (1), 543-557.*

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. *Sci Rep* **2020,** *10 (1), 11535.*

Michiels, T. J. M.; Schoneich, C.; Hamzink, M. R. J.; Meiring, H. D.; Kersten, G. F. A.; Jiskoot, W.; Metz, B., Novel Formaldehyde-Induced Modifications of Lysine Residue Pairs in Peptides and Proteins: Identification and Relevance to Vaccine Development. *Mol Pharm* **2020,** *17 (11), 4375-4385.*

Michiels, T. J. M.; Tilstra, W.; Hamzink, M. R. J.; de Ridder, J. W.; Danial, M.; Meiring, H. D.; Kersten, G. F. A.; Jiskoot, W.; Metz, B., Degradomics-Based Analysis of Tetanus Toxoids as a Quality Control Assay. *Vaccines (Basel)* **2020,** *8 (4), 712.*

Michiels, T. J. M.; van Veen, M. A.; Meiring, H. D.; Jiskoot, W.; Kersten, G. F. A.; Metz, B., Common Reference-Based Tandem Mass Tag Multiplexing for the Relative Quantification of Peptides: Design and Application to Degradome Analysis of Diphtheria Toxoid. *J Am Soc Mass Spectrom* **2021,** *32 (6), 1490-1497.*

Curriculum Vitae

Thomas Michiels was born on October 25th, 1991, in Roosendaal, the Netherlands. After graduating high school in 2011, he studied Bio-Pharmaceutical Sciences at Leiden University. After a chemistry-oriented bachelor, he continued with a chemistry-oriented Bio-Pharmaceutical Sciences master. His master's nine-month research project focused on the design and synthesis of covalent adenosine A_{2A} receptor antagonists at the Medicinal Chemistry department of the LACDR under the supervision of Dr. Julien Louvel. Subsequently, he conducted his six-month internship at Janssen Pharmaceutica in Beerse, Belgium, under the supervision of Dr. Simon Wagschal. This work focused on optimizing the synthesis of enantiopure precursors for a BTK inhibitor by using ruthenium-catalyzed chemo-enzymatic dynamic kinetic resolution. He obtained his master's degree in 2016 and started his PhD research in November that year, under the supervision of Prof. Dr. Gideon Kersten, Prof. Dr. Wim Jiskoot and Dr. Bernard Metz. The PhD project was a collaboration between the Leiden Academic Centre for Drug Research (LACDR) and the Institute for Translational Vaccinology (Intravacc). He currently works as a scientist at Intravacc in the Product Characterization and Formulation department.

