Journal of Chromatography B 1054 (2017) 80–92

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/jchromb

The combined use of analytical tools for exploring tetanus toxin and tetanus toxoid structures

Caroline Bayart[a,b], Sébastien Peronin[c], Elisa Jean[b], Joseph Paladino[b], Philippe Talaga[d], Marc Le Borgne[a,*]

[a] Université de Lyon, Université Claude Bernard Lyon 1, Faculté de Pharmacie − ISPB, EA 4446 Bioactive Molecules and Medicinal Chemistry, SFR Santé Lyon-Est CNRS UMS3453 −INSERM US7, 8 avenue Rockefeller, F-69373, Lyon Cedex 8, France
[b] Reaction and Coupling Chemistry Laboratory, MTech, Sanofi Pasteur, 31/33 quai Armand Barbès, 69250 Neuville-sur-Saône, France
[c] Structural Analysis Laboratory, MTech, Sanofi Pasteur, 31/33 quai Armand Barbès, 69250 Neuville-sur-Saône, France
[d] Analytical Research and Development Department, Sanofi Pasteur, Campus Mérieux,1541 Avenue Marcel Mérieux, 69280 Marcy L'Etoile, France

ARTICLE INFO

Keywords:
Tetanus toxin
Tetanus toxoid
Formaldehyde
Analytical tools
Amino acid
Vaccine

ABSTRACT

Aldehyde detoxification is a process used to convert toxin into toxoid for vaccine applications. In the case of tetanus toxin (TT), formaldehyde is used to obtain the tetanus toxoid (TTd), which is used either for the tetanus vaccine or as carrier protein in conjugate vaccines. Several studies have already been conducted to better understand the exact mechanism of this detoxification. Those studies led to the identification of a number of formaldehyde-induced modifications on lab scale TTd samples. To obtain greater insights of the changes induced by formaldehyde, we used three industrial TTd batches to identify repeatable modifications in the detoxification process. Our strategy was to combine seven analytical tools to map these changes. Mass spectrometry (MS), colorimetric test and amino acid analysis (AAA) were used to study modifications on amino acids. SDS-PAGE, asymmetric flow field flow fractionation (AF4), fluorescence spectroscopy and circular dichroism (CD) were used to study formaldehyde modifications on the whole protein structure. We identified 41 formaldehyde-induced modifications across the 1315 amino acid primary sequence of TT. Of these, five modifications on lysine residues were repeatable across TTd batches. Changes in protein conformation were also observed using SDS-PAGE, AF4 and CD techniques. Each analytical tool brought a piece of information regarding formaldehyde induced-modifications, and all together, these methods provided a comprehensive overview of the structural changes that occurred with detoxification. These results could be the first step leading to site-directed TT mutagenesis studies that may enable the production of a non-toxic equivalent protein without using formaldehyde.

1. Introduction

Tetanus toxin (TT[1]) is a potent neurotoxin produced by *Clostridium tetani* bacteria, and in 2013 was the cause of over 58,000 deaths worldwide [1]. TT binds to motor neurons using specific receptors [2], and is then internalized and transported into the cell body using axonal retrograde transport. In the spinal cord, TT blocks the release of inhibitory neurotransmitters by cleaving synaptobrevin-2, leading to hyperactivity of the motor neurons and consequently spastic paralysis [3].

TT is a 150.7 kDa protein composed of a 52.4 kDa light chain, responsible for synaptobrevin cleavage, and a 98.3 kDa heavy chain [4,5]. The heavy chain can be further subdivided into two domains, the N-terminal domain (46.7 kDa), responsible for cell penetration [6], and the C-terminal domain (51.6 kDa), also called the tetanus toxin fragment C (TTFC), which governs TT neuronal specific binding [6,7]. TT has yet to be fully crystallized, and as such its entire structure remains unknown. Currently, only the light chain and the TTFC 3D structures have been resolved [2,8].

Vaccination against tetanus disease caused by *Clostridium tetani* has been used since the 1930s. The vaccine is composed of the detoxified TT protein, tetanus toxoid (TTd); which is obtained by formaldehyde

* Corresponding author.
E-mail address: marc.le-borgne@univ-lyon1.fr (M.L. Borgne).

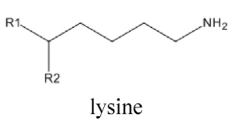
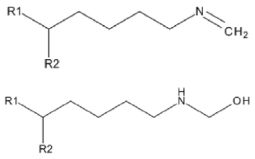
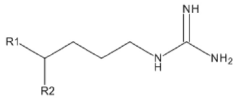
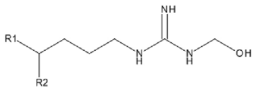
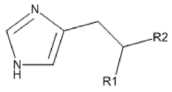
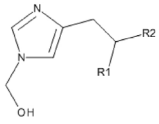
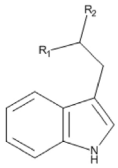
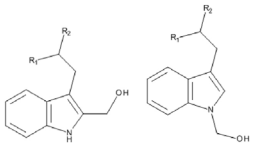
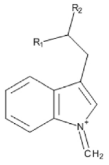
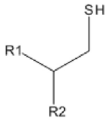
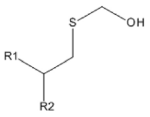
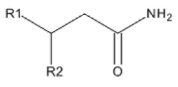
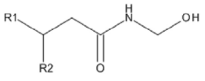
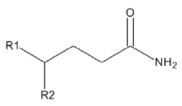
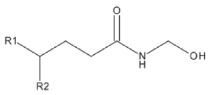
[1] The abbreviations used are: AAA, amino acid analysis; AF4, asymmetrical flow field-flow fractionation; AGC, automatic gain control; CID, collision induced dissociation; dRI, differential refractive index; DLS, dynamic light scattering; ESI, electrospray ionization; FTICR, Fourier transformed ion cyclotron resonance; FWHM, full width at half maximal resolution; HC, heavy chain; LC, light chain; MALS, multi-angle light scattering; MS/MS, tandem mass spectrometry; Mw, molecular weight; NCE, normalized collision energy; QELS, quasi elastic light scattering; Rg, gyration radii; Rh, hydrodynamic radius; SEC, size exclusion chromatography; TIC, total ions currents; TNBS, trinitrobenzene sulfonate; TT, tetanus toxin; TTd, tetanus toxoid; TTFC, tetanus toxin fragment C; UPLC, ultra-performance liquid chromatography.

http://dx.doi.org/10.1016/j.jchromb.2017.04.009
Received 27 January 2017; Received in revised form 27 March 2017; Accepted 2 April 2017
Available online 04 April 2017

treatment of TT. TTd protein is also widely used as carrier protein in conjugate vaccines. Due to TTd immunological properties, the T-cell independent response is converted to a T-cell dependent one, which boosts the immunity response directed against the conjugated bacterial polysaccharide [9,10].

The formaldehyde detoxification process induces several TT amino

Table 1
Formaldehyde-induced modifications on amino acids and their corresponding mass changes. Groups R1 and R2 represent N and C-terminal functions.

Amino acid	Modified amino acid	Mass modification	Ref.
 lysine		Schiff base +12.000 Da Methylol adduct: +30.011 Da	[12,13]
 arginine		Methylol adduct: +30.011 Da	[14]
 histidine		Methylol adduct: +30.011 Da	[15]
 tryptophan		Methylol adduct: +30.011 Da	[16]
		Imine adduct: +13.008 Da	[12]
 cysteine		Methylol adduct: +30.011 Da	[17]
 asparagine		Methylol adduct: +30.011 Da	[12]
 glutamine		Methylol adduct: +30.011 Da	[12]

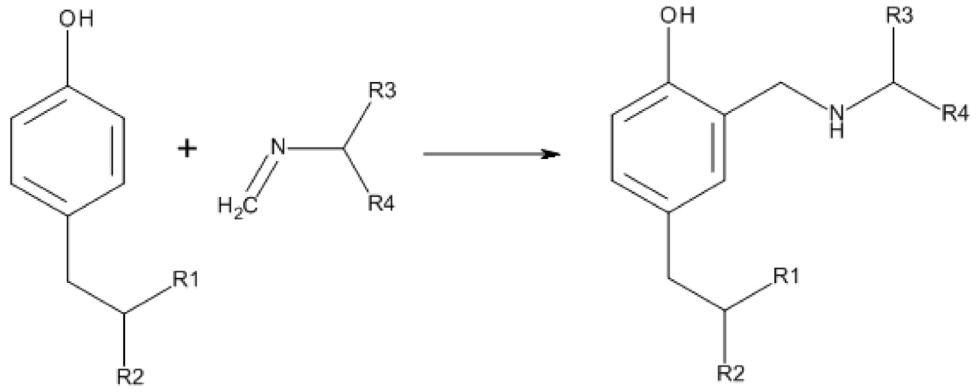


Fig. 1. Formaldehyde-induced modification on tyrosine residues: cross-link bridge formation. Groups R1, R2, R3 and R4 represent N and C-terminal functions.

acids and structure modifications, removing its toxicity while keeping its immunogenicity [11]. The formaldehyde-induced modifications of TT amino acids have previously been studied [12–17], and seven are well described in the literature (Table 1). The first step of the formaldehyde-induced modifications on amino acids is the addition of a formaldehyde molecule forming a methylol adduct. This methylol adduct can be further dehydrated leading to the creation of other species such as Schiff base or imine adducts on lysine and tryptophan residues, respectively [12,13]. Another modification, the creation of cross-links, can be mainly found on tyrosine residues [12]. This cross-link formation is believed to occur between a tyrosine residue and a lysine containing a Schiff base modification (Fig. 1). These cross-links can appear inside the protein (intra-protein cross-links) or between several different proteins (inter-protein cross-links). As the number of cross-linked molecules formed is likely to be high, this modification has not yet been fully characterized in TTd.

Thaysen-Andersen et al. [11] identified 26 partial Schiff base modifications on TTd using mass spectrometry (MS) (involving ten lysine residues, seven arginine residues, three tyrosine residues, two isoleucine residues, and one residue each of proline, alanine, tryptophan and aspartic acid), and other partial Schiff base modifications on nineteen lysine residues after the addition of a reductive agent ($NaCNBH_3$). The most modified amino acid appeared to be lysine with 29 modifications. Surprisingly, at least one Schiff base modification was found on arginine, tyrosine, isoleucine, proline, alanine, tryptophan and aspartic acid residues, as this modification requires a primary amino group. Additionally, amino acid analysis (AAA) showed similar composition of all amino acids except tyrosine and lysine residues [11]. The amount of tyrosine was reduced to 4% (instead of 6% in TT), depending on the amount of added formaldehyde. Lysine quantification remained complicated. Cross-links formation was observed on some TTd batches using SDS-PAGE [11]. Other studies reported the comparison of TT and TTd using circular dichroism (CD) and fluorescence analysis experiments [18,19]. No significant difference was observed in CD whereas some were seen in fluorescence analysis, which may indicate a change in protein folding between TT and TTd. All these previous studies have served to underscore the main formaldehyde-induced modifications, but they were not sufficient to determine which modification was necessary to induce the lack of toxicity.

In this study, we characterize three industrial batches of TTd used as protein carrier in conjugate vaccines. As these are used in commercial vaccines, the amount of formaldehyde needed for detoxification was fixed to ensure the non-toxicity of the protein. We used a set of seven different analytical tools to scrutinize formaldehyde-induced modifications in TTd, and to identify repeatable modifications. MS was used to identify methylol adducts and Schiff base modifications (Table 1). Colorimetric tests [20–22] and AAA [23] were used to determine modifications on other amino acids (e.g. tyrosine residues). SDS-PAGE analyses were conducted to study intra- and inter-protein cross-links (Fig. 1). Fluorescence spectroscopy and CD [18,19] were used to provide insight into changes in protein folding. Asymmetric flow field flow fractionation (AF4) was used to investigate the quaternary structure of the proteins.

2. Experiment

2.1. Chemicals and reagents

Tetanus toxin (FA593524) and toxoid (FA501115, batch 1; FA533368, batch 2; FA488119, batch 3) were produced at Sanofi Pasteur (Marcy l'Etoile, France). TTFC (TTC-041709) was obtained from PX'Therapeutics (Grenoble, France). AccQ•Tag™ reagent, borate buffer, eluent A concentrate AccQ•Tag™ Ultra, eluent B AccQ•Tag™ Ultra and Rapigest SF were purchased from Waters (Milford, MA, USA). Trinitrobenzene sulfonate 5% (w/v) in H_2O (TNBS), sodium borate decahydrate, L-lysine, L-tyrosine, Folin Denis' reagent, NaOH, amino acids used for blank preparation in tyrosine quantification, DL-2-aminobutyric acid, dithiothreitol, iodoacetamide, sodium nitrate, sodium azide, phosphate buffered saline (PBS) 10X and trifluoroacetic acid were purchased from Sigma-Aldrich (St. Louis, MO, USA). HCl, methanol, acetic acid and acetonitrile were purchased from Carlo ERBA (Val-de-Reuil, France). Sodium tartrate and Coomassie blue solution were purchased from Fisher scientific (Illkirch, France). Formic acid, copper sulfate, sodium carbonate and Trypsin/LysC were purchased from VWR (Radnor, PA, USA), Acros organics (Geel, Belgium), Sanofi Aventis (Paris, France) and Promega (Madison, WI, USA), respectively. Sample buffer 4X, reducing buffer 20X and on Criterion XT Precast Gel 4-12% Bis-Tris gel were purchased from Biorad (Hercules, CA, USA). Water was purified with a Millipore (Bedford, MA, USA) Milli-Q system. All solvents were HPLC grade.

Before each analysis, protein concentrations were evaluated using Lowry test as previously described [24].

2.2. Colorimetric tests

Accessible lysine residues of were quantified using TNBS; sodium tetraborate and TNBS at a final concentration of 26 mM and 0.85 mM, respectively, were added to 1.42 nmol of protein. Accessible tyrosine residues were quantified using Folin Denis' reagent; 100 µl of Folin Denis' reagent (2 times diluted in water before adding) and 1 ml of alkaline copper solution (sodium tartrate 0.43 mM, copper sulfate 1 ml, sodium carbonate 189 mM and NaOH 0.16 M) were added to 1.04 µmol of protein. Accessible cysteine residues were quantified using Ellman's reagent, 1.2 nmol of protein was mixed with 0.42 mmol/l of Ellman's reagent (5,5'-dithio-bis (2-nitrobenzoic acid)) previously dissolved in 0.1 M phosphate buffer pH 8. For each of these assessments the reaction was monitored using V-630 UV–vis spectrophotometer (Jasco); lysine monitored at 420 nm in kinetic mode at 20 °C, tyrosine monitored at

665 nm in a fixed wavelength mode after exactly 15 min of reaction and cysteine monitored at 412 nm in a fixed wavelength mode after 90 min of reaction. Quantification was calculated using L-lysine, L-tyrosine or acetyl-cysteine as standard. Each protein was analyzed in triplicate. Results were accepted if the calibration curve correlation coefficient was superior to 0.98 (0.985 for tyrosine) and if the difference between each triplicate was less than 30%. TT and the three batches of TTd were analyzed in the same set of analyses for comparison.

2.3. Amino acid analysis (AAA)

5 µmol of protein was hydrolyzed with HCl 6 M at 110 °C over 16 h. Samples were then dried under nitrogen gas and redissolved in 3 ml HCl 20 mM. Amino acids from hydrolyzed proteins were then derivatized using AccQ•Tag™ reagent: 10 µl of internal standard (DL-2-aminobutyric acid, 100 pmol/µl), 60 µl of borate buffer and 20 µl of AccQ•Tag™ reagent were added to 10 µl of hydrolyzed sample. The solution was immediately mixed and heated at 55 °C for 10 min. After cooling down to ambient temperature, samples were injected in ultra performance liquid chromatography (UPLC) system.

UPLC was performed on Acquity UPLC H-Class Bio system equipped with a fluorescence detector Acquity FLR (Waters, Milford, MA, USA). Amino acid separation was performed using AccQ•Tag™ Ultra Amino Acid Analysis Column, BEH C18 (100 mm × 2.1 mm, 1.7 µm, 130 Å) (Waters, Milford, MA, USA) in 10.2 min. The flow rate was set to 0.7 ml/min, the injection volume to 1 µl and the column temperature to 43 °C. Fluorescence detection parameters were fixed as followed: λ_{ex} 266 nm (extinction wavelength), λ_{em} 473 nm (emission wavelength), sampling rate 10 points/s, filtered time constant 0.2 sec, gain 1. Amino acid elution was achieved using a gradient of solvent A (Eluent A concentrate AccQ•Tag™ Ultra from Waters), solvent B (Eluent B AccQ•Tag™ Ultra from Waters diluted 10 times), solvent C (ultrapure water) and solvent D (Eluent B AccQ•Tag™ Ultra from Waters): A 10-9.9%, B 0%, C 90-90.1%, D 0% in 0.29 min; A 9.9-9.0%, B 0-80%, C 90.1-11%, D 0% in 5.2 min; A 8%, B 80-15.6%, C 11-57.9%, D 0-18.5% in 1.61 min; A 8-7.8%, B 15.6-0%, C 57.9-70.9%, D 18.5-21.3% in 0.39 min; A 7.8-4%, B 0%, C 70.9-36.3%, D 21.3-59.7% in 0.30 min and A 4-10%, B 0%, C 36.3-90%, D 59.7-0% in 0.09 min.

Data acquisition and system control were performed on Empower 2 software (Waters).

2.4. Mass spectrometry

2.4.1. Sample Processing

TT and TTd protein (100 µg) were first denaturated and reduced using 20 µl of Rapigest 1% (m/v) and dithiothreitol 10 mM, incubated at 56 °C for 20 min. Reduced proteins were then alkylated using 17 mM of iodoacetamide at room temperature for 1 h. Digestion was performed using 4 µg of Trypsin/LysC at 37 °C overnight. Reaction was stopped by adding trifluoroacetic acid at a final concentration of 1% and incubating at 37 °C for 45 min. The supernatant was diluted four times in 20% methanol in water and used for LC-MS/MS analysis.

2.4.2. Liquid chromatography tandem mass spectrometry

All mass spectrometry analyses were performed under Xcalibur data system using a LTQ-FT Ultra mass spectrometer (Thermo Fisher Scientific) coupled to an Agilent series 1200 high-performance liquid chromatography (HPLC) equipped with high pressure binary pump G1312B. Digests (10 µl) were loaded on reverse phase UPLC Acquity CSH130 column 150 mm × 2.1 mm, 1.7 µm, 130 Å (Waters, Milford, MA, USA) using a G1367C autosampler (refrigerated at 4 °C). Peptides were separated in 100 min using a gradient of solvent A (5% acetonitrile, 0.1% formic acid, 95% water) and solvent B (80% acetonitrile, 0.1% formic acid, 20% water). The gradient conditions were: 100-90% of solvent A in 17 min, 90-70% of solvent A in 48 min, 70-50% of solvent A in 7 min, 50-25% of solvent A in 4 min and 25-100% of solvent A in 2 min at a constant flow rate of 250 µl/min. Separated peptides were then ionized using an Ion Max Electrospray (ESI) source equipped with a H-ESI II probe, heated at 60 °C and with a source voltage of 3 kV. The mass spectrometer was operated in data-dependent mode top-3 with MS performed in the Fourier transform ion cyclotron resonance (FTICR) analyzer at a resolution of 50,000 full width at half maximal resolution (FWHM) @ m/z 400; mass range *m/z* 230-2000; automatic gain control (AGC) target 5×10^5, maximum injection time 100 ms. MS/MS was performed in linear ion trap; MS/MS analysis consisted of collision induced dissociation (CID); AGC 1×10^4; normalized collision energy (NCE) 30; maximum injection time 200 ms. The three most abundant ions were selected for MS/MS analysis (top-3). Dynamic exclusion was also applied with only one repeat count and exclusion duration of 15 s.

2.4.3. Data analysis

Proteome discoverer™ 1.4.1.14 with SEQUEST (Thermo Fisher Scientific) was used for protein identification against protein database containing tetanus toxin sequence (uniprot accession number P04958). Precursor mass tolerance was set to 6 ppm and the fragment mass tolerance to 0.8 Da. Modifications were all set as dynamic modifications: carbamidomethylation (+57.021 Da) on cysteine; oxidation (+15.995 Da) on methionine; deamidation (+0.984 Da) on asparagine and glutamine; Schiff base (+12.000 Da) on lysine; methylol adducts (+30.011 Da) on cysteine, histidine, lysine, arginine, tryptophan, asparagine and glutamine. Data were filtered using a target decoy of 0.01 false discovery rate (FDR) for high confidence and a FDR of 0.05 for moderate confidence. For data treatment, only high confidence peptides were kept and modifications were validated by manual inspection using Qual Browser.

2.5. SDS-PAGE analysis

TT, TTd and TTFC (around 8 µg each) were mixed with 12.5 µl of sample buffer 4X in non-reductive conditions and with 12.5 µl of sample buffer 4X and 2.5 µl of reductive buffer 20X. Solutions were heated at 100 °C for 10 min and then loaded on Criterion XT Precast Gel 4-12% Bis-Tris gel. 60 V, 130 V and 195 V were applied to the gel for 30 min, 1 h and 30 min, respectively. The gel was then stained using fixation solution (50% methanol, 7% acetic acid) followed by Coomassie blue solution and destained using water.

2.6. AF4-UV-MALS-RI analysis

AF4 analyses were performed using Agilent series 1200 HPLC combined with Wyatt eclipse system. G1367C auto sampler (refrigerated at 4 °C) and binary pump G1312 B were driven by Agilent system and flow distribution by the eclipse system. Three detectors were used: Agilent UV detector (wavelength 280 nm), a Wyatt HELEOS-II online multi-angle light scattering (MALS, 665 nm) detector having the online DLS module (QELS, quasi elastic light scattering), and a Wyatt T-rEX differential refractive index (RI, 658 nm, 25 °C) detector. For the separation, a channel of 201 mm equipped with 350 µm spacer and a regenerated cellulose membrane with a 8 kDa cut-off was used. The eluent used was sodium nitrate ($NaNO_3$) 50 mM, sodium azide (NaN_3) 300 ppm. 100 µl of proteins (around 3–4 g/l) were loaded. An initial focus-flow of 1.5 ml/min was applied for 3 min followed by an exponential decreasing cross-flow gradient from 3 to 0 ml/min for 26 min.

Data were treated using Astra software. For the dn/dc (refractive index increment) a standard value of 0.185 cm^3/g for proteins was taken as input value in the Astra software.

2.7. Fluorescence spectroscopy

Protein samples (20 µg/ml) were first dialyzed at 5 °C with three

changes of phosphate buffered saline (PBS) 1X for 24 h using Slide-A-Lyser™ G2 dialysis cassettes (Fisher scientific, Illkirch, France) with a 3.5 kDa molecular mass cut off pore size. The protein spectra were recorded at 20 °C using a Horiba Jobin Yvon FluoroMax-4 spectrofluorometer. Quartz cells of 3.5 ml, 1 cm path-length were used. Emission spectra were collected after application of 280 and 295 nm excitation wavelengths. They were obtained using a range from 290 to 450 nm with a band pass of 4.25 nm for both excitation and emission monochromators. Fluorescence spectra were corrected by subtracting the buffer fluorescence spectra alone (PBS 1X) and emission maximum wavelength was obtained using FluoEssence software with an accuracy of ± 1 nm.

2.8. Circular Dichroism

Jasco J-1500 circular dischroism spectrometer equipped with a 150W Xenon lamp was used to collect protein spectra at 20 °C. CD spectra were taken from 250 to 180 nm in step of 0.5 nm and scan rate of 50 nm/min. A pass band of 1 nm was used for the monochromator. Quartz cells of 0.35 ml, 1 mm path-length were used. Protein samples (200 µg/ml) were dialyzed at 5 °C with three changes of Na_2HPO_4 8 mM, KH_2PO_4 2 mM and NaF 137 mM pH 7.4 buffer solution for 24 h. Slide-A-Lyser™ G2 dialysis cassettes (Fisher scientific, Illkirch, France) with a 3.5 kDa molecular mass cut off pore size were used for this step. CD signals were corrected by subtracting the buffer spectra and then converted to molar ellipticity. CD Multivariate SSE software was used to determine the content of α-helix, β-sheets, β-turns and random coils. The 2D structure was calculated using a multivariate algorithm partial least square (PLS) on the basis of a calibration data set of 26 different proteins.

3. Results

3.1. Characterization of formaldehyde-induced modifications on amino acids

Mass spectrometry was used to localize Schiff base and methylol adducts on the amino acids described in Table 1. MS total ions currents (TIC) spectra of TT and TTd after tryptic digestion are presented in Fig. 2. At least 13 main differences were observed between TT and TTd (diamonds on Fig. 2; a 25 relative abundance difference in threshold between TT and TTd peaks). The observed differences between TT and TTd digested peptides concerned the peaks abundance: some peaks were more intense in TT than in TTd (retention times of 23.73, 28.14, 34.99, 37.21, 40.77, 42.70, 53.27, 56.68, and 64.80 min in TT, red diamonds on Fig. 2). Other peaks were completely missing in the TTd batches (retention time of 37.58, 48.31 and 59.87 min in TT, blue diamonds on Fig. 2), and the presence of a peak at 31.85 min retention time in TTd was not visible in TT spectrum (green diamond on Fig. 2). As each digestion was conducted in triplicate, these observable differences between TT and TTd were not due to variability in the reaction. Moreover, the three different TTd batches presented similar profiles showing repeatability in the detoxification process.

Using Proteome Discoverer™ Software, peptide masses were compared to peptides containing (i) the theoretical mass, (ii) current post-translational modifications (deamidation and oxidation – used to identify the maximal number of peptides) [25] and (iii) known formaldehyde-induced modifications (Schiff base, methylol adducts). We began to look at methylol adducts as they are the first species obtained after formaldehyde detoxification. Then, we searched for Schiff bases as this modification was predictable. Other modifications (e.g. imine adducts) were too unstable to be characterized under the conditions used.

Sequence coverage of the TT and TTd peptides was similar for the peptides identified with a high confidence level (around 75%). However, some differences between the two proteins were visible in the sequence coverage representation (Fig. 3, red squares). It appears that at least three different peptides were missing; this was in agreement with the three missing peaks in TTd MS spectra (blue diamonds, Fig. 2, retention time 37.58, 48.31 and 59.87 min in TT). We compared all digested peptides masses using Proteome Discoverer™ software, and 41 modifications were identified on TTd (Table 2). Among these 41 modifications, 29 involved lysine residues (Schiff base modification was seen on 18 lysine residues, methylol adduct modification on three, and both modifications were seen on eight residues). This number of modified lysine residues was in agreement with Thaysen-Andersen's study [11]. Other identified methylol adducts were seen on six arginine, two tryptophan and four histidine residues. Not all these modifications were identified in each of the three TTd batches. The modification rate was calculated as follows: when a modification was identified x times on a peptide over all TTd batches, the modification rate was equal to (x/9)*100 (9 represented the 9 analyses of TTd: 3 TTd batches in triplicate). A 100% modification rate does not mean that the total population of a given peptide was 100% modified. Five modifications were found to be repeatable on all triplicates of the three TTd batches (modification rate of 100% in Table 2). These repeatable modifications were all located on lysine residues and they all contained the Schiff base modification; these five modified lysine residues may play a role in the neutralization of TT toxicity.

No modifications were found on asparagine or glutamine residues. In another study, less than 4% of these amino acids were found as modified by formaldehyde [12] but they were not observed in Thaysen-Andersen's study [11]. These modifications did not appear as essential in the TTd that we analyzed. These differences could be linked to the formaldehyde concentration used in the detoxification process (lab scale batches versus industrial batches).

MS and MS/MS analyzes also showed that at least three peptides were missing in the TTd batches (^{80}TDSDKDR^{86}, ^{94}LFNRIK^{99} and ^{938}AMDIEYNDMFNNFTVSFWLR^{957}). These peptides corresponded to the three missing peaks in Fig. 2 and to the three blanks in TTd sequence coverage in Fig. 3. These peptides were present in all TT triplicates but never seen in TTd. As one contained a tyrosine residue and the other two contained lysine residues, these missing peptides may have formed cross-links (Fig. 1). This modification was repeatable across the TTd batches and complete. Nevertheless, these peptides did not form a cross-link together as the different masses of the hypothetic cross-links were not found using MS analyses (e.g peptide ^{80}TDSDKDR^{86} cross-linked with peptide ^{938}AMDIEYNDMFNNFTVSFWLR^{957}).

Amino acid analysis (AAA) was conducted to further characterize the formaldehyde-induced modifications on TTd. The protein concentrations were estimated using Lowry assay prior to each assay [24]. Using AAA, we were able to compare almost all the TT and TTd amino acids in one run; methionine, cysteine, tryptophan and tyrosine residues excluded. Methionine residues underwent oxidation during the hydrolysis step and their corresponding retention times were changed. The same effect was observed on cysteine residues, in addition to the formation of cystine (cysteine dimer). As the proportion of cystine and oxidized cysteine fluctuated during each assay, it was difficult to correctly quantify this amino acid [26]. Different protections, alkylation [27] or cysteic acid formation [28], against oxidation were tested without success. The tryptophan residue peak was not sufficiently distinct from the background noise to enable quantification, due to destruction during acid hydrolysis [29]. Tyrosine residue peak was also indistinguishable from the background noise. In our study, we used the AQC reagent to derivate amino acids, contrary to Thaysen-Andersen's study which reported the use of ninhydrin [11]. This did not appear as optimal for tyrosine quantification as its coefficient response was low compared to other amino acids, even in the reference spectra (data not shown). Although it seemed that there was a difference between TT and TTd tyrosine peaks, this could not be concluded with certainty (Fig. 4B). Asparagine and glutamine residues were converted during hydrolysis to aspartic acid and glutamic acid, respectively, and

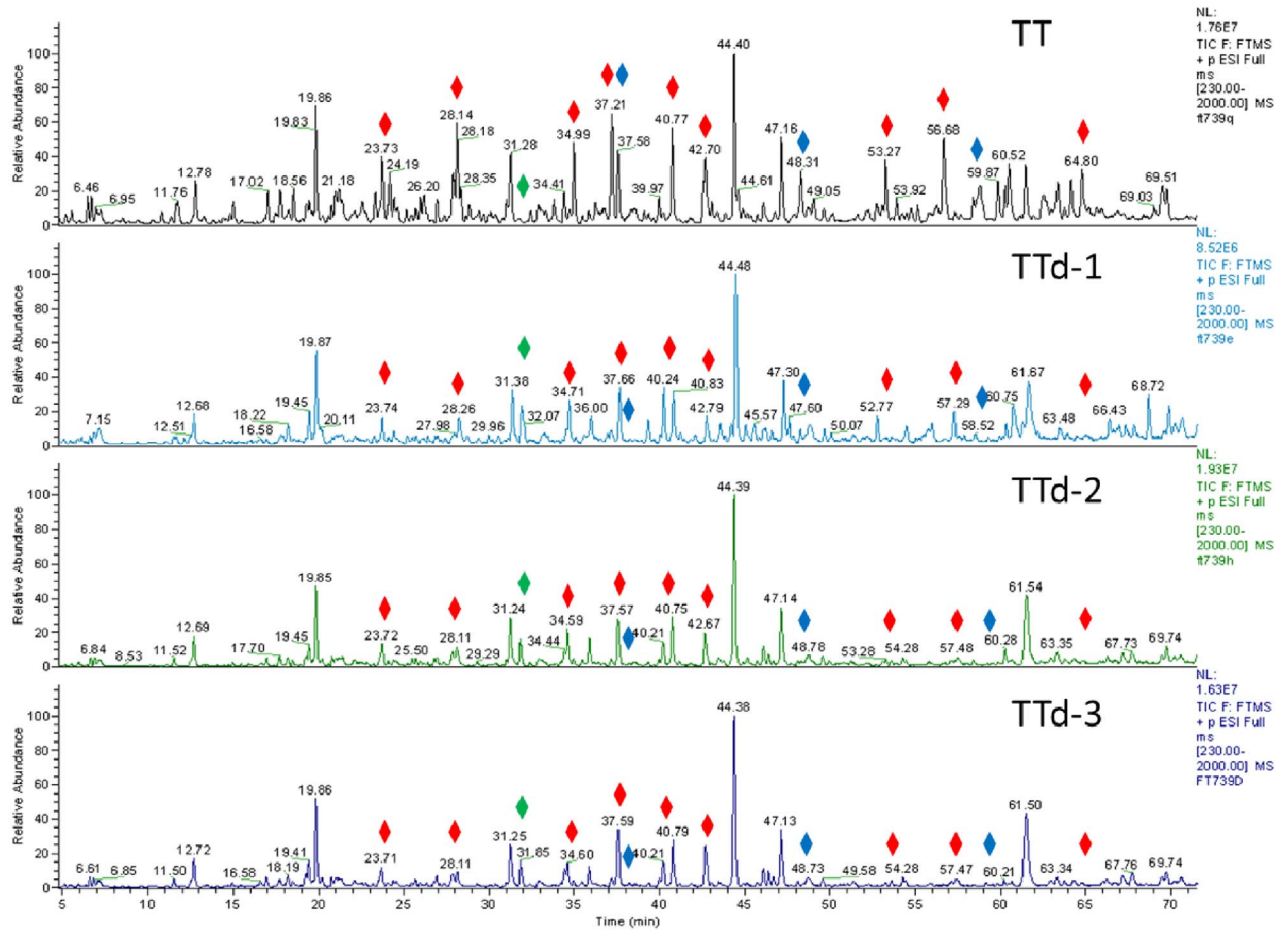


Fig. 2. MS total ions currents (TIC) spectra of TT and TTd after tryptic digestion. On top, profile of TT's digested peptides; below, 3 different batches of TTd. Diamonds point out main visible differences between TT and TTd (minimum threshold: difference of 25 relative abundance between TT and TTd peaks). Red diamonds correspond to a difference of peak intensity between TT and TTd; blue diamonds show missing peptides in TTd and the green diamonds indicate the presence of a peak in TTd profiles unseen in TT.

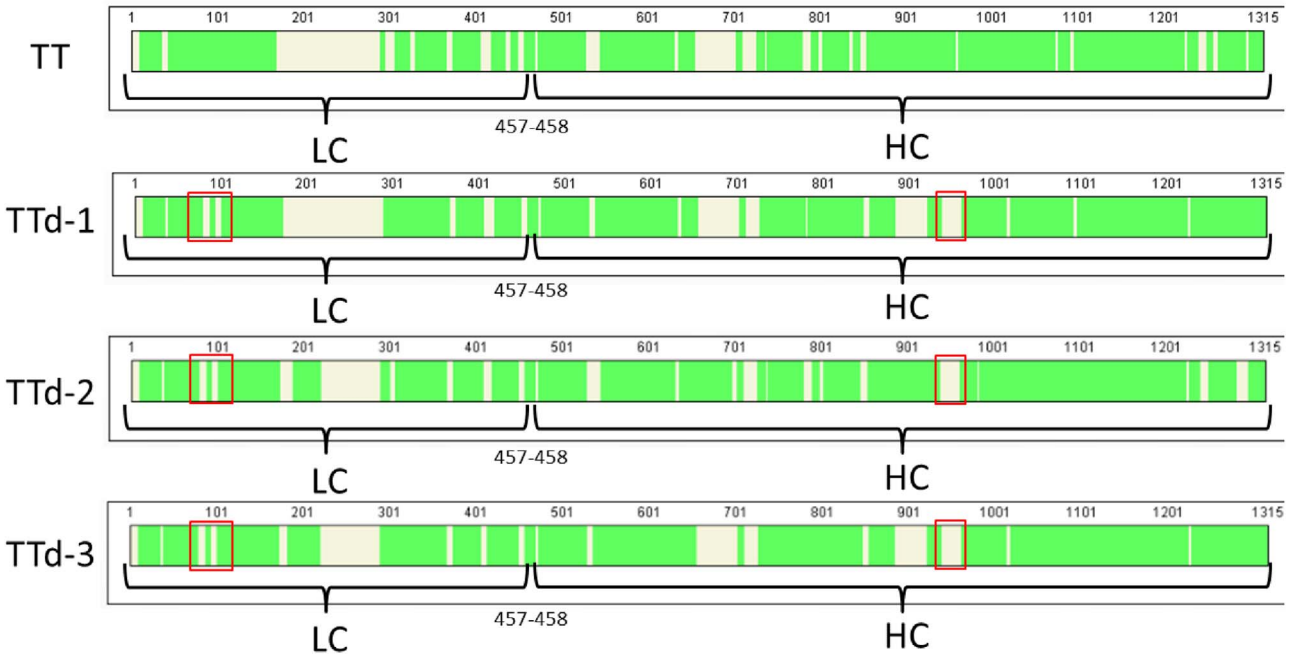


Fig. 3. TT and TTd high confidence peptide sequence coverage. Red squares highlight the main visible modifications.

Table 2
Modifications in TTd identified by MS. In peptide sequences, the letters in bold represent the modified amino acid. The amino acid column indicates its location on TT sequence. The modification types are Schiff base (+12.000 Da) and methylol adduct (+30.011 Da). The modification rate was calculated after analyzing 9 samples (3 different batches of TTd, each tested in triplicate).

Peptide sequence	Amino acid	Modification type	Modification rate (%)
^{10}YSDPVNNDTIIMMEPPYCKGLIDIYY**K**^{35}	K35	Schiff base/methylol	22
^{50}YEFGT**K**PEDFNPPSSLIEGASEYYDPNYLR^{79}	K55	Schiff base/methylol	77
^{50}YEFGTKPEDFNPPSSLIEGASEYYDPNYL**R**^{79}	R79	methylol	22
^{128}FDTNSNSVSFNLLEQDPSGATT**K**^{150}	K150	Schiff base	88
^{301}AIAN**K**LSQVTSCNDPNIDIDSYK^{323}	K305	Schiff base	33
^{301}AIANKLSQVTSCNDPNIDIDSY**K**^{323}	K323	Schiff base/methylol	44
^{330}YQFD**K**DSNGQYIVNEDK^{346}	K334	Schiff base/methylol	100
^{330}YQFDKDSNGQYIVNED**K**^{346}	K346	Schiff base	55
^{413}GQNM**R**VNTNAFRNVDGSGLVSKLIGLCK^{440}	R417	methylol	50
^{413}GQNMRVNTNAF**R**NVDGSGLVSKLIGLCK^{440}	R424	methylol	25
^{472}NEDLTFIAE**K**^{481}	K481	Schiff base/methylol	66
^{512}IIVDYNLQS**K**^{521}	K521	Schiff base	66
^{544}SNAASTIEIHNIDDNTIYQYLYAQ**K**^{568}	K568	Schiff base	88
^{603}VNQGAQGILFLQ**W**VR^{617}	W615	methylol	11
^{603}VNQGAQGILFLQWV**R**^{617}	R617	methylol	11
^{754}IIDYEY**K**^{761}	K761	Schiff base	88
^{762}IYSGPD**K**EQIADEINNLK^{779}	K768	Schiff base/methylol	100
^{780}N**K**LEEKANKAMININIFMRESSR^{852}	K781	Schiff base	50
^{780}NKLEE**K**ANKAMININIFMRESSR^{852}	K785	Schiff base/methylol	66
^{780}NKLEEKAN**K**AMININIFMRESSR^{852}	K788	methylol	66
^{789}AMININIFM**R**^{798}	R798	methylol	33
^{803}SFLVNQMINEA**K**K^{815}	K814	Schiff base	44
^{803}SFLVNQMINEAK**K**^{815}	K815	Schiff base	66
^{826}NILMQYI**K**^{833}	K833	Schiff base	88
^{826}NILMQYIKANS**K**^{837}	K837	Schiff base	100
^{922}AI**H**LVNNESSEVIVHK^{937}	H924	methylol	66
^{961}VSAS**H**LEQYGTNEYSIISSMKK^{982}	H965	methylol	66
^{961}VSASHLEQYGTNEYSIISSM**K**K^{982}	K981	Schiff base	33
^{961}VSASHLEQYGTNEYSIISSMK**K**^{982}	K982	Schiff base	33
^{983}HSLSIGSGWSVSLKGNNLI**W**TLK^{1005}	W1002	methylol	11
^{983}HSLSIGSGWSVSLKGNNLIWTL**K**^{1005}	K1005	Schiff base	44
^{1074}LD**R**CNNNNQYVSIDK^{1088}	R1076	methylol	29
^{1095}ALNP**K**EIEK^{1103}	K1099	Schiff base/methylol	57
^{1144}NITDYMYLTNAPSYTNG**K**^{1161}	K1161	methylol	22
^{1198}LYVSYNNNE**H**IVGYPK^{1213}	H1207	methylol	22
^{1198}LYVSYNNNEHIVGYP**K**^{1213}	K1213	methylol	22
^{1240}MEAV**K**LRDLK^{1249}	K1244	Schiff base	100
^{1245}LRDL**K**TYSVQLK^{1256}	K1249	Schiff base	80
^{1245}LRDLKTYSVQL**K**^{1256}	K1256	Schiff base	100
^{1250}TYSVQLKLYDD**K**^{1261}	K1261	Schiff base	77
^{1262}NASLGLVGT**H**NGQIGNDPNR^{1281}	H1271	methylol	22

were quantified using the peaks of their corresponding acid.

Other amino acids were correctly quantified: in our case, it was possible to quantify lysine residues as the industrial process did not use lysine as quencher, contrary to Thaysen-Andersen's study [11] (Fig. 4A). The recovery percentage was over 75%, except for isoleucine and valine residues (50%) as their peptide bonds are particularly resistant to acid hydrolysis [23,26]. We assumed that the difference of 25% of coverage was due to Lowry assay intrinsic variability when estimating protein concentration. As we compared two different kinds of proteins, we chose to keep the results as obtained, without normalization.

The main difference in the HPLC profile concerned the lysine amount; the peak area of lysine residues in TTd was lower than the corresponding TT peak (Fig. 4B). To confirm this difference, all data were analyzed by ANOVA and significant changes were found for lysine ($p < 0.001$, Fig. 5A). There was no significant difference for any other amino acid (Fig. 5B). The difference in the number of lysine residues (81 lysines in TT, 63 in TTd) was of the same order as that seen in the MS results (22% of the lysine residues are modified according to AAA compared to 27% by MS). These modifications were generally repeatable between the three TTd batches, they only differed by three modified lysine residues between batches.

3.2. Characterization of formaldehyde-induced modifications regarding protein structure

Some amino acids could not be characterized using the proteomic MS and AAA analyses. As such, we complemented the results by analyzing cysteine and tyrosine residues using colorimetric tests. Furthermore, using colorimetric tests on the entire proteins without a preliminary degradation step, allowed us to study amino acid accessibility (especially for lysine residues). Analysis of methionine [30] was not possible as the protein formed a complex with the nitroprusside in acidic medium (data not shown). Our hypothesis was that the accessibility of some amino acids might evolve because of the formation of cross-links which may modify the TT structure.

The average number of cysteine residues quantified in TT and TTd, were 3 and 2, respectively, and the number of tyrosine was 69 in TT compared to 51 in TTd, meaning that 18 tyrosine residues were less accessible or modified in TTd (Table 3). For the lysine residue, the difference observed between TT and TTd was higher than with the previous methods; 66 lysine residues were found for TT and only 16 for TTd (Table 3). Significant differences between TT and TTd were observed for all tested residues (assessed by ANOVA; $p < 0.001$). These results attested that not only the amino acids structure was modified, but also their accessibility. For example, for lysine residues, 61% were accessible in TT and only 15% in TTd, meaning that there

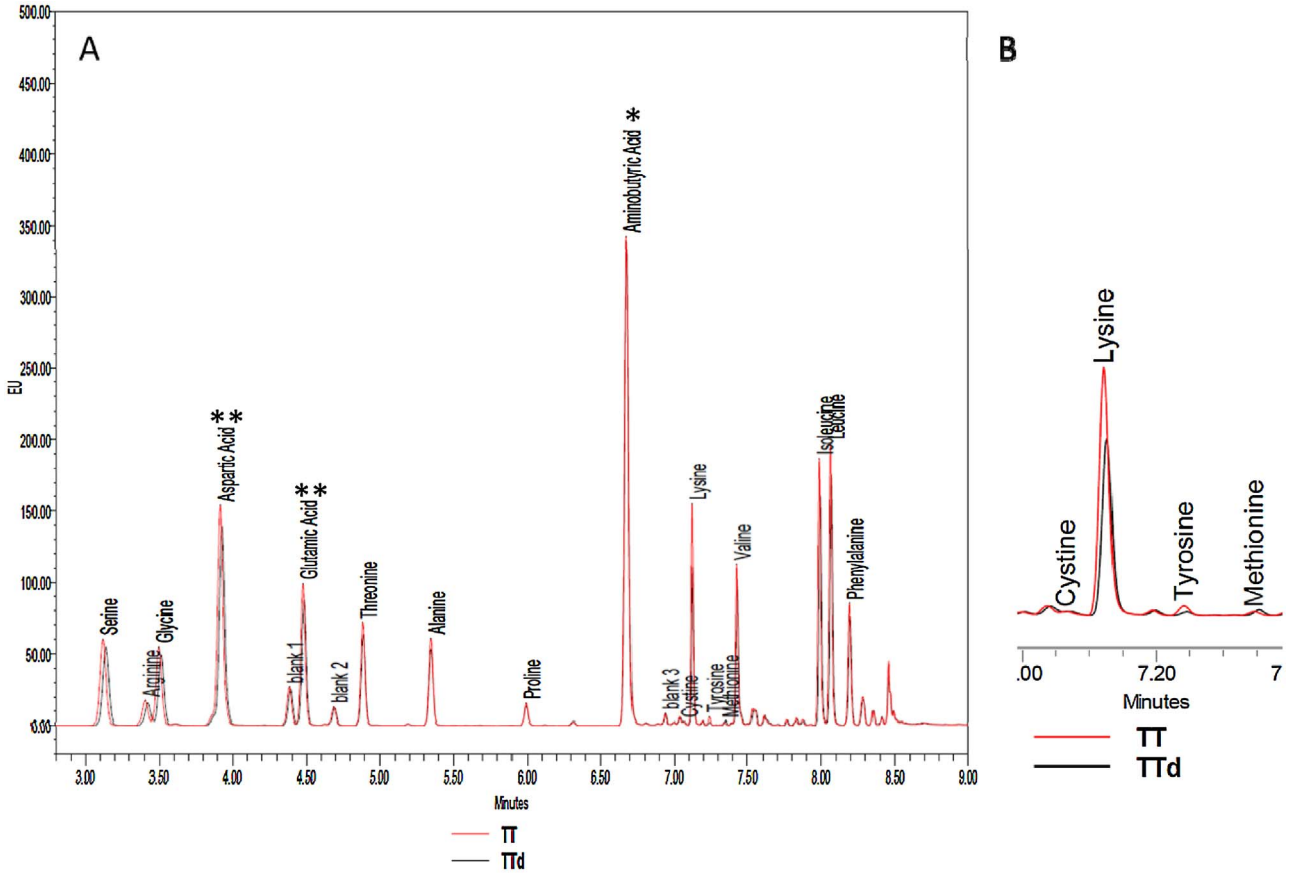


Fig. 4. AAA profile comparison between TT (in red) and TTd (in black). (A) Full view of TT and TTd amino acids. (B) Zoom on lysine residue; cystine, tyrosine and methionine residues are confused with the background noise.* Amino butyric acid is used as internal standard. ** These peaks represent asparagine and glutamine residues together with their corresponding acids.

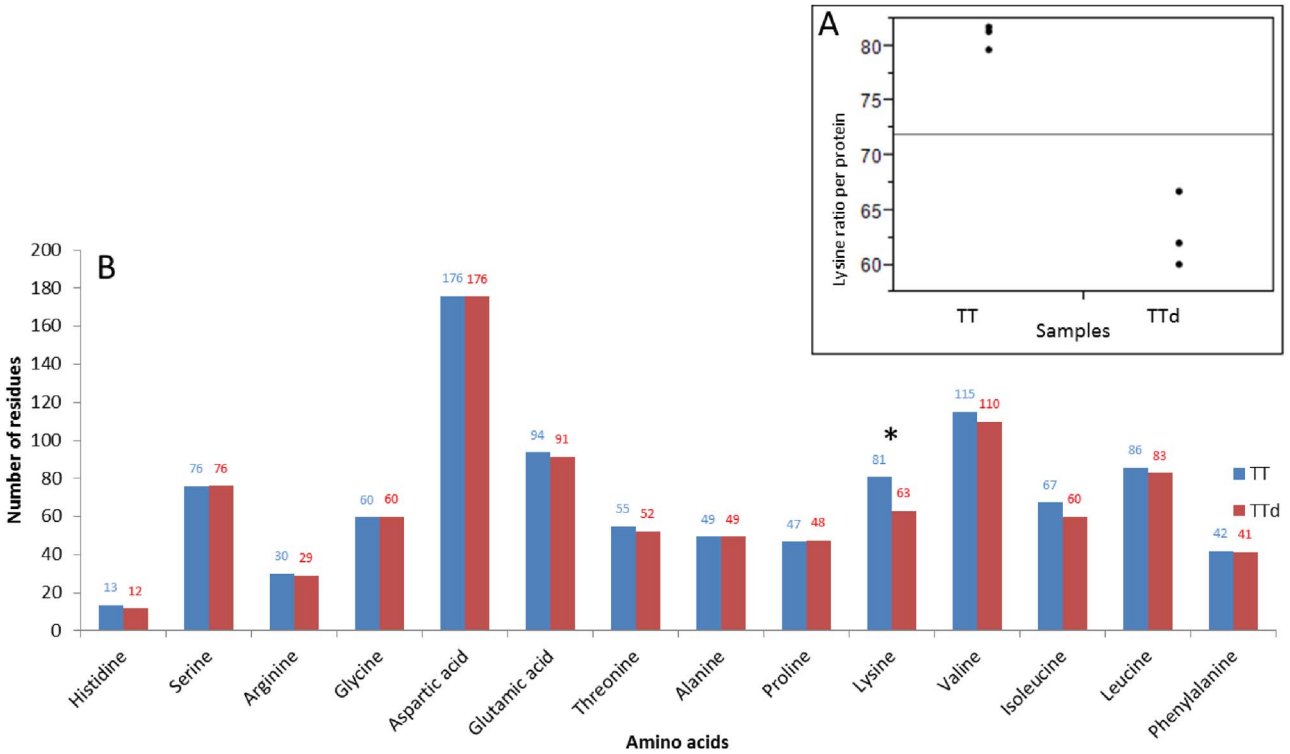


Fig. 5. (A) AAA ANOVA test graphical results for quantification of lysine residues. Amounts of lysine residues are significantly different in TT than in TTd, $p < 0.001$. (B) Amino acid quantification in TT (in blue) and TTd (in red). * represents a significant difference with $p < 0.001$. Other amino acid amounts were not significantly different between the two proteins.

Table 3
Amino acid quantification on TT, TTd and TTFC proteins using colorimetric tests. Number on left (-/) represents the number of quantified residues, and on the right (/-) the theoretical number of the same residue present on the protein.

Amino acid	TT	TTd batch 1	TTd batch 2	TTd batch 3
Lysine	66/107	14/107	18/107	16/107
Tyrosine	69/79	50/79	59/79	44/79
Cysteine	3/10	2/10	2/10	2/10

was a difference of 46% between TT and TTd while only 22 to 27% were found as structurally modified using MS and AAA. These results showed that the whole structure of the TT was modified by the detoxification process. The presence of cross-links may explain this change in structure.

To confirm the presence of cross-links on the TTd structure, SDS-PAGE analyses were conducted on TT and TTd under both reducing and non-reducing conditions. The most intense line in TT band was situated at 150 kDa which corresponds to TT theoretical size [4]. Other lines observed on this band may come from impurities or protein degradation during sample preparation (proteolytic degradation due to the presence of proteases in SDS-PAGE sample buffer). All TTd batches gave similar results; the main band was observed at 150 kDa, which corresponded to the theoretical TT size and another band was also seen at around 250 kDa (Fig. 6). The latter likely represents proteins formed when the 150 kDa TTd is linked to other TTd parts, and is suggestive of inter-protein cross-links formation not seen in the TT profile.

Under reducing conditions the TT 150 kDa line was distributed into a 100 kDa and a 50 kDa band (Fig. 6); in agreement with the theory that the TT light (50 kDa) and heavy (100 kDa) chains are linked by a disulfide bridge. Even under reducing conditions TTd still presented two lines, the most intense at 150 kDa and the other at around 250 kDa (Fig. 6). This indicates the formation of intra-protein cross-links as the heavy and the light chains of TTd remained linked even under reducing conditions. The three TTd batches gave similar results showing the repeatability of intra and inter-protein cross-links formation. The formaldehyde detoxification process induced the creation of these intra and inter links the protein.

Asymmetrical flow field-flow fractionation (AF4) analyses [31] were run to confirm the presence of cross-links using 'softer' analytical conditions, where the proteins were not denatured before the analysis. This tool also enabled us to characterize TT and TTd sizes and

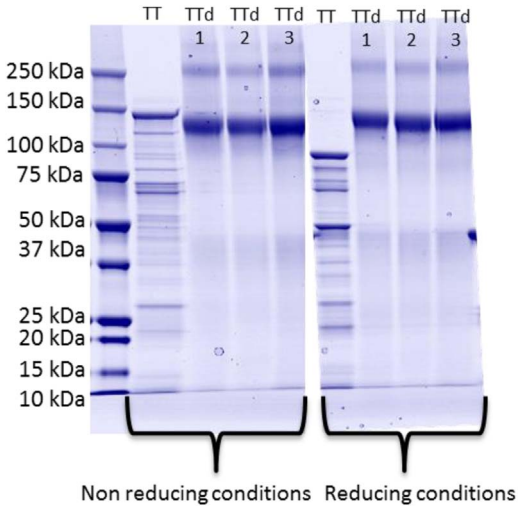


Fig. 6. SDS-PAGE gel: TT and TTd analyses. The first four samples were analyzed under non reducing conditions, the last four samples were analyzed under reduction buffer. TTd 1, 2 and 3 correspond to the three TTd batches.

molecular masses. The results of the AF4 are shown in Fig. 7, the peaks from left to right were of increasing size, contrary to classical size exclusion chromatography. Concentrations used as input data in the Astra software are presented in Table 4. Peak limits were manually defined by taking into account refractive index (RI) and light scattering (LS) signals. A mass recovery percentage (ratio of recovered mass over the total injected mass) was calculated using these peak limits and the sample concentrations. For TTd, only the batch 3 is displayed as the results were repeatable between batches.

For TT, the total mass recovery was 81.1%; this could be explained by either the overestimation of the concentration by Lowry assay or that the dn/dc (refractive index increment) value of the protein was different from the value we applied (0.185 cm^3/g) [32]. TTd had a total recovery mass of 100%. TT and TTd presented a main peak (peak 2, high RI and LS signals) between 12 and 15 min (Fig. 7), with corresponding mass recovery percentages of 62.1% for TT and 80.8% for TTd. The remaining mass recovery percentages were distributed over the two other peaks (peaks 1 and 3), showing the presence of other components in the proteins. TT and TTd had a small first eluted peak (peak 1) with a 50 kDa molecular weight. Some aggregates (peak 3, high molecular size and low concentration) were observed between 15 and 25 min. Data treatment was exclusively focused on peak 2 as it represented the main product.

The quasi electric light scattering (QELS) detector measures the translational diffusion coefficient of a particle in a given solvent. This detector measures the hydrodynamic radius (Rh) which is inversely proportional to the diffusion coefficient of the particle, the higher the diffusion coefficient, the lower the Rh [33]. The Rh of TT (4.9 nm) was found to be smaller than that of TTd (7.1 nm), meaning that TTd was slightly bigger than the TT.

Molecular weights (Mw) were determined using the multi angle light scattering (MALS) and the RI detectors. The dn/dc value was fixed at 0.185 cm^3/g as this is the average value for proteins in aqueous buffer [32]. A Zimm 1 model was applied to determine Mw values.

TT consists of three fragments: light chain (A), heavy chain N-terminal domain (B) and TTFC (C) and has an expected mass of 150.7 kDa. The A fragment has an expected mass of 52.4 kDa, the B fragment an expected mass of 46.7 kDa, and the TTFC an expected mass of 51.6 kDa [4,5]. The Mw found for TT, 142.7 kDa, was close to that expected. The calculated Mw value of TTd was 283.9 kDa, which may represent a mixture of unmodified and modified (e.g. dimers) proteins. TTd had a Mw almost double that of TT and its Rh is also higher confirming that TTd is bigger than TT.

There was a notable difference between TT and TTd in terms of profile and polydispersity (Figs. 7 and 8). TT was monodisperse (Mw/Mn values equal to 1, blue dotted line in Fig. 8). Conversely the TTd profile exhibited polydispersity (the Mw distribution represented by the green dotted line is upward, Fig. 8 and its Mw/Mn value equals 1.3). This polydispersity confirmed that the main peak present in the TTd profile (Fig. 7, peak 2) represented a mix of different species with an increasing Mw. The TT profile showed monodisperse proteins with a Mw consistent with TT theoretical size, while the TTd profile showed a mix of different species (polydisperse proteins) with an average Mw of 283.9 kDa. This indicates that the TTd protein was present in its monomeric form as well as in its dimeric and trimeric forms, due to inter-protein cross-links.

Fluorescence analyses were conducted on protein samples in order to obtain some indication of the secondary structural changes induced by formaldehyde. Emission spectra were collected after a 280 and a 295 nm excitation wavelength (λ_{ex}). Fluorescence emission wavelength maxima values (λ_{max}) were recorded and compared between protein samples [18,19]. At λ_{ex}280 nm, all samples gave λ_{max} values between 320 and 322 nm. At λ_{ex}295 nm, TTd samples gave λ_{max} values between 323 and 324 nm. TT λ_{max} was slightly higher at 325 nm. No significant differences were observed between the samples (Table 5, measurements accuracy of 1 nm). However, the 1 nm red-shift of TT protein at

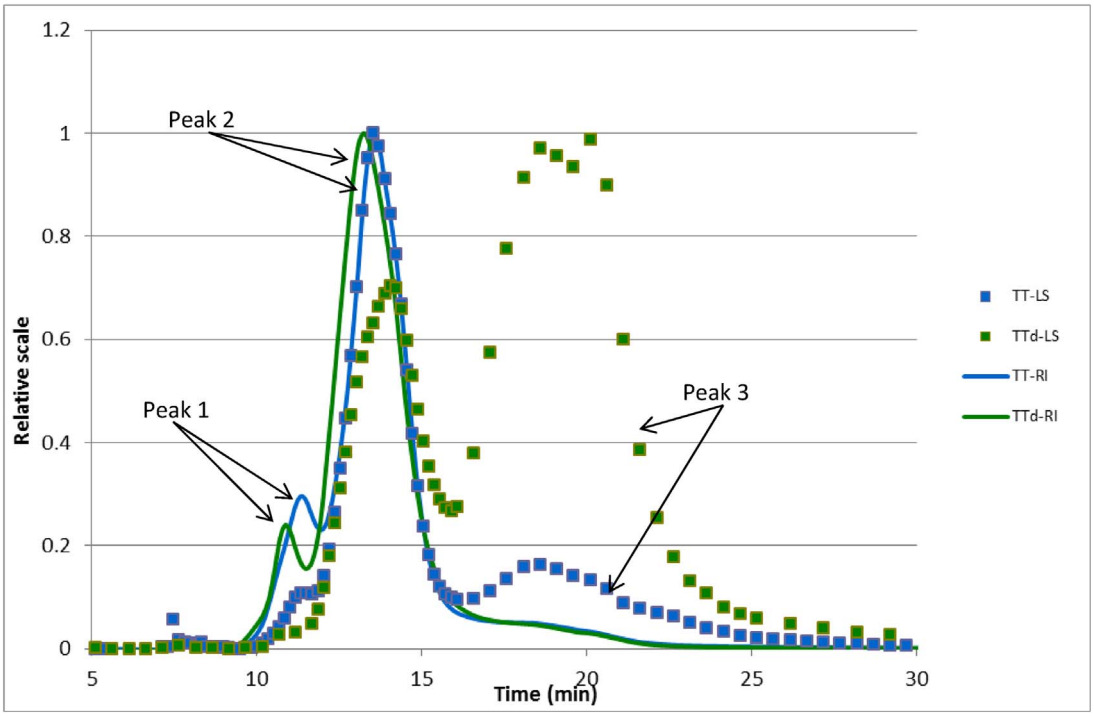


Fig. 7. TT (in blue) and TTd batch 3 (in green) AF4 fractogram. In full line, RI signal corresponding to the sample concentration, in dotted line, LS signal corresponding to sample component size.

Table 4
Concentration inputs in Astra software.

Samples	C (g/l)*
TT	2.6
TTd batch 1	4.1
TTd batch 2	3.9
TTd batch 3	2.9

* Concentrations were obtained using Lowry assay. Each sample was analyzed in triplicate.

Table 5
Intrinsic fluorescence maximum values of protein samples.

Samples	λ_{max} (nm) at λ_{ex}280 nm	λ_{max} (nm) at λ_{ex}295 nm
TT	322	325
TTd batch 1	321	324
TTd batch 2	322	323
TTd batch 3	320	324

λ_{ex}295 nm may reflect a change in the conformation of the protein on or in the vicinity of tryptophan residues (TT contains 13 tryptophan residues among its 1315 amino acids).

As the intra-protein cross-links may induce different structural constraints inside the protein, we decided to compare both TT and TTd secondary structures using circular dichroism (CD). The first differences between TT and TTd were observed when looking at the shape of the CD spectral curves. The spectral curves of the three TTd batches were essentially the same and clearly different from the TT spectra (Fig. 9), indicating differences in protein secondary structure. These differences were confirmed after CD spectra data processing using CD Multivariate SSE software (Table 6).

The TTd secondary structures were fairly similar across the three TTd batches (Table 6), thus indicating the repeatability of the formaldehyde-induced modifications. The TT structure contained less α-helix (17.9%), more β-sheets (30.7%) and slightly more β-turns (14.0%) than TTd (Table 6). A conformational transition from α-helix to β-sheets compositions was observed between the two proteins, which suggest that TT β-sheets were converted to α-helixes during the detoxification process. To our knowledge, this is the first time that such a difference between TT and TTd secondary structure has been reported [18]. This change in conformation may be due to intra-protein cross-links formation and may play an important role for the loss of toxicity.

4. Discussions

Understanding the action of formaldehyde on TT proteins during the detoxification process is necessary to elucidate the mandatory modifications required for detoxification. This investigation may allow the development of *in vitro* or analytical tests to control TTd production and so reduce the number of animals tested in order to determine safety before release of product batches [18]. Furthermore, it could also be the first step in the development of a TTd recombinant protein. Despite the different studies conducted on lab scale TTd batches, the exact mechanism of detoxification has not yet been resolved.

Our strategy was to combine seven analytical tools to extend TTd characterization on industrial batches in order to identify repeatable modifications. We first identified modifications on amino acids. As in Thaysen-Andersen's study [11], we found more than 40 modifications using MS (41 found here, 44 in Thaysen-Andersen's study). The most modified amino acid was lysine with 29 modifications; among these, seven modifications were identified in both studies (K35, K323, K521,

Table 6
Structural composition of proteins based on CD spectra.

Samples	α-helix (%)	β-sheets (%)	β-turns (%)	Random coils (%)
TT	17.9	30.7	14.0	37.4
TTd batch 1	36.4	21.4	10.6	31.6
TTd batch 2	37.8	19.8	10.7	31.7
TTd batch 3	39.6	18.7	10.7	31.0

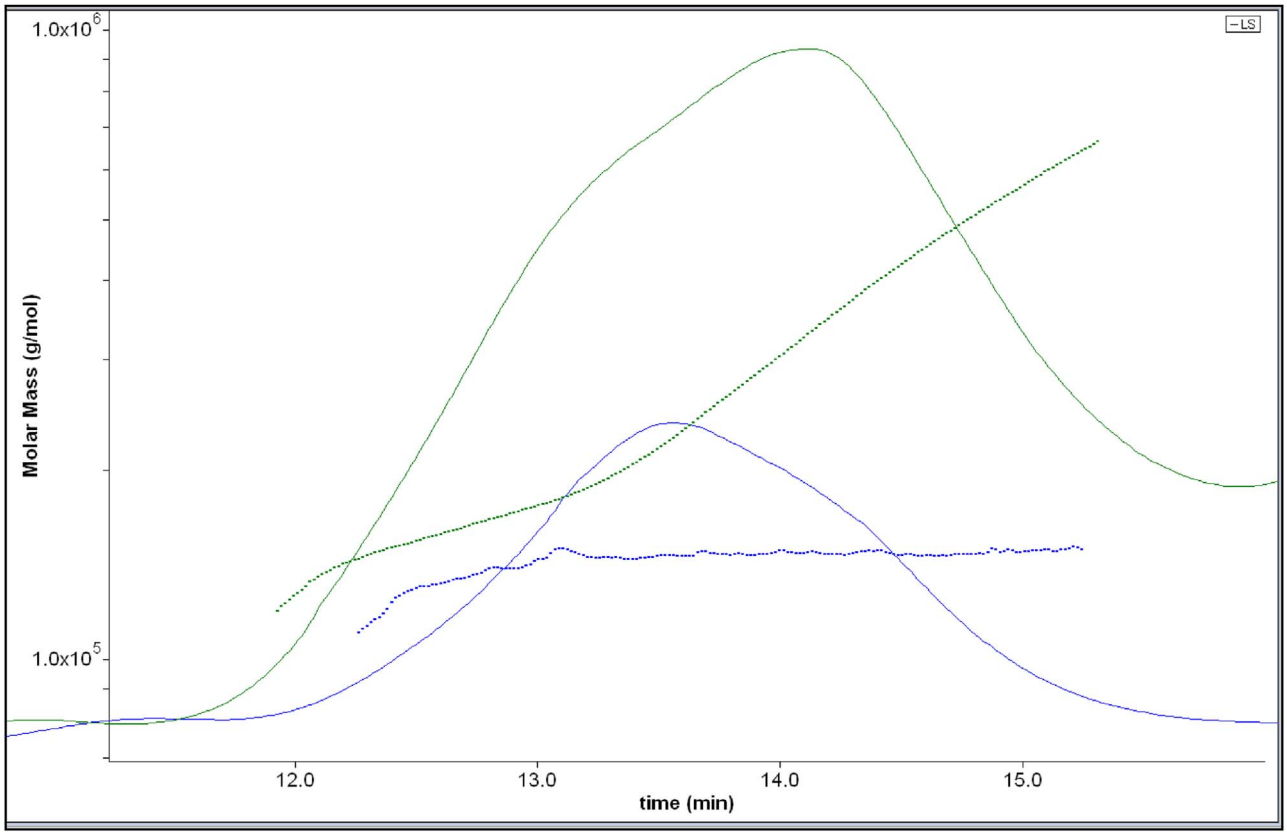


Fig. 8. TT (in blue) and TTd (in green) main peak 2. The dotted lines represent the molecular weight mass distribution in the peak. The more this line is horizontal, the more the product is monodisperse.

K815, K1005, K1249 and K1261). However, none of these corresponded to the five lysine residues we identified as 100% repeatable modifications, or situated on TT catalytic site [34]. Nevertheless, these modified lysine residues may play an important role in the detoxification process.

AAA results confirmed that more than 20% of the lysine residues were modified in the industrial batches of TTd assessed. The second most modified amino acid was tyrosine; as suspected previously [12],

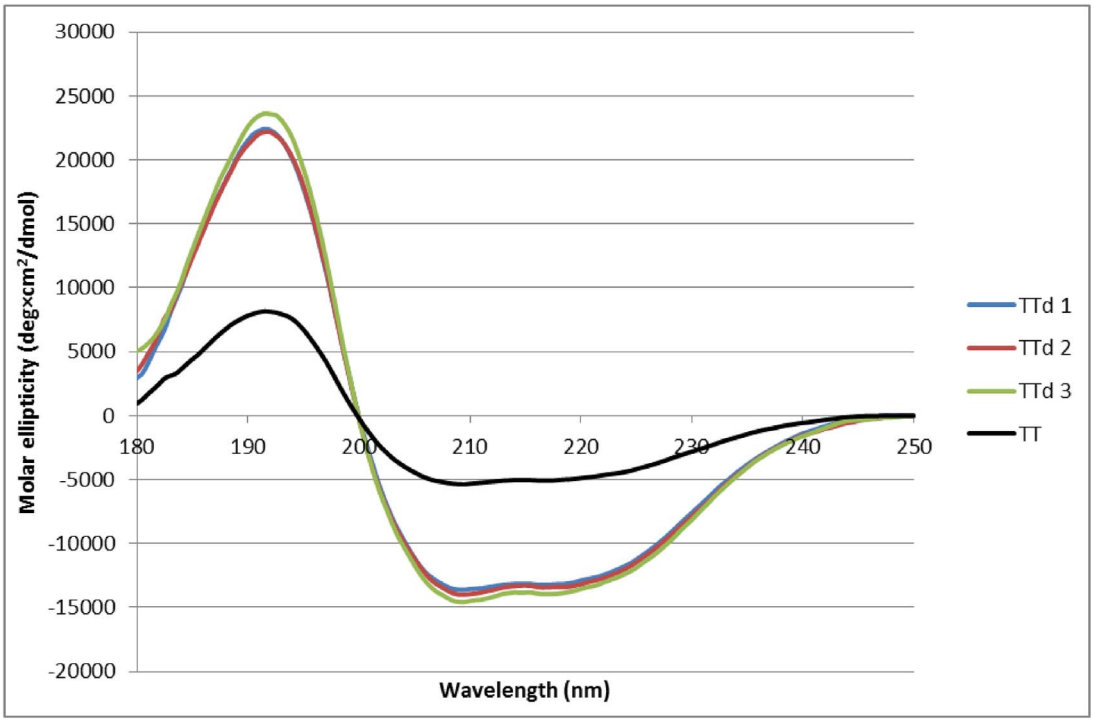


Fig. 9. TTd 1 (in blue), TTd 2 (in green), TTd 3 (in red) and TT (in black) CD spectra. Signals were converted into molar ellipicity signals. The three TTd batches presented the same profile whereas significant differences were observed between TTd and TT proteins.

Table 7
Main advantages and drawbacks of studied tools.

Tools	Advantages	Drawbacks
MS	Position of modified amino acids in TT sequence	Analysis after enzymatic digestion
	Known modifications (Schiff base or methylol) easily detected	Only seven amino acids studied
AAA	Almost all amino acids analyzed in one run (10 min)	Analysis after hydrolysis
	Quantification	Unknown nature of modifications
Colorimetric tests	Direct analysis (no hydrolysis or digestion)	Unknown nature of modifications
	Information about amino acids accessibility	One colorimetric test per amino acid
SDS-PAGE	Different conditions (reductive, non-reductive), several samples analyzed in the same gel	Use of standards for molecular weight estimation
	Quick view of protein integrity	Accuracy
AF4	Direct measurement of molecular weight	High product quantity needed
	Characterization of aggregates	Molecular weight calculated on peak fraction with sufficient dRI signal
Fluorescence spectroscopy	Fast and easy to handle	Accuracy
	Direct analysis	Less reliable interpretation
CD	Tertiary structure information	Buffer interferences
	Direct analysis	

tyrosine can form cross-links with formaldehyde-formed Schiff-base adducts on other amino acids. Colorimetric tests allowed us to determine that 26% of the tyrosine residues (18/69) were either modified or less accessible in TTd. In comparison, other investigators have reported that the amount of tyrosine was reduced to 4% (from 6% in TT, indicating that 33% of the tyrosine residues were modified) at high formaldehyde concentration [11]. Together, this information confirms that tyrosine is another amino acid widely affected by formaldehyde. Tyrosine residues are involved in cross-link formation; the identification of these intra and inter-protein cross-links using SDS-PAGE and AF4 analysis, were aligned with a change in the whole structure of TTd. This change in structure was consistent with the considerably reduced accessibility of lysine residues (reduction of 46%, compared to only 20% identified as structurally modified). CD analysis revealed a transition from TT β-sheets to TTd α-helix conformation, indicating cross-link induced changes in TT structure. This change in structure was repeatable in the three analyzed TTd batches. To our knowledge, this is the first time that such a difference has been observed between TT and TTd structures, but has previously been observed with other proteins (e.g. protein aggregation and SNARE proteins) [35,36]. The augmentation of β-sheets increased the accessibility of reactive groups involved in aggregation and so decreases the stability of the protein [35]. That statement applied to TT could mean that TT is less stable than TTd, the latter having a more stable 2D conformation due to the presence of cross-links. The β-sheets to α-helix transition was also observed in SNARE proteins, depending on the peptide/lipid ratio present in membranes, the conformation of these transmembrane proteins changed [36]. This change in TT conformation could then affect its capacity for crossing central nervous system membranes, and so be related to the non-toxicity of TTd.

In conclusion, determining which modification(s) is sufficient to detoxify TT protein remains to be established. We observed five Schiff base modifications on lysine residues that were completely repeatable between the three industrial TTd batches. Additionally, several cross-links were identified changing the whole TT structure; one of these cross-links may be formed between the lysine residues of the ^{80}TDSDKDR^{86} or ^{94}LFNRIK^{99} peptides; the tyrosine residue of the ^{938}AMDIEYNDMFNNFTVSFWLR^{957} peptide may have also reacted with another peptide containing a Schiff base modification. Again, these modifications were repeatable. To better understand if both Schiff base and cross-links are essential for TT loss of toxicity, further studies have to be conducted on the five repeatable lysine modifications, using protein mutation experiments for example. This study also allowed us to compile seven analytical tools to construct an efficient characterization strategy for the identification of formaldehyde-induced modifications. Each analytical method contributed to the information about protein structure and together enabled us to obtain a more comprehensive view of TTd structure. In Table 7, we list the main advantages and drawbacks of the seven tools. Each technique needed different sample preparations including sometimes protein degradation. While MS, AAA and SDS-PAGE analyses included a protein degradation step; colorimetric test, fluorescence spectroscopy, CD and AF4 directly analyzed the complete protein. MS precisely localized the modifications on the protein sequence and validated the nature of these modifications. AAA and colorimetric tests were used to compare unmodified and modified proteins. From these comparisons, hypotheses were established to identify supplementary modifications (e.g. cross-links on tyrosine). SDS-PAGE and AF4 analyses were used to estimate sample Mw, indirectly and directly, respectively. Finally, CD analyses brought the first 2D-structural information confirming that formaldehyde also modified TT structure. In the case of protein characterization and even more in the case of modified protein characterization, it is mandatory to master these analytical tools and to combine them before drawing definite conclusions. This efficient characterization strategy can be applied to other modified or detoxified proteins to map and identify structural modifications.

Conflict of interest

The authors declare that they have no conflicts of interest relevant to this article. Sébastien Peronin, Joseph Paladino and Philippe Talaga are employees of Sanofi Pasteur.

Funding

Funding for this assistance was provided by Sanofi Pasteur.

Author contributions

J.P. and M.L.B. directed the project, analyzed the results, wrote the paper, approved the final version to be published and are accountable for the accuracy and integrity of the results. C.B. designed the study, conducted most of the experiments, analyzed the data, wrote most of the paper, approved the final version to be published and is accountable for the accuracy and integrity of the results. S.P. oversaw proteomic aspects of the project, analyzed data related to mass spectrometry experiments, critically revised the manuscript, approved the final version to be published and is accountable for the accuracy and integrity of the results. E.J. performed AF4 analyses, critically revised the manuscript, approved the final version to be published and is

accountable for the accuracy and integrity of the results. P.T. analyzed the results, critically reviewed the manuscript, approved the final version to be published and is accountable for the accuracy and integrity of the results.

Acknowledgements

This study was sponsored by Sanofi Pasteur. We thank Loïc Petit and Soumia Zeghida for their assistance with the amino acid analyses, Catherine Girardon for assistance in fluorescence analysis and circular dichroism. We also thank Thierry Eynard, and Smaïl Boukenna for helpful discussions.

Manuscript editing services were provided by Nicola Truss, inScience Communications, Springer Healthcare.

References

[1] Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013, The Lancet. 385 (2015) 117-171. 10.1016/S0140-6736(14)61682-2.

[2] P. Emsley, C. Fotinou, I. Black, N.F. Fairweather, I.G. Charles, C. Watts, E. Hewitt, N.W. Isaacs, The Structures of the HC Fragment of Tetanus Toxin with Carbohydrate Subunit Complexes Provide Insight into Ganglioside Binding, J. Biol. Chem. 275 (2000) 8889–8894, http://dx.doi.org/10.1074/jbc.275.12.8889.

[3] L.C. Williamson, K.E. Bateman, J.C.M. Clifford, E.A. Neale, Neuronal Sensitivity to Tetanus Toxin Requires Gangliosides, J. Biol. Chem. 274 (1999) 25173–25180, http://dx.doi.org/10.1074/jbc.274.35.25173.

[4] U. Eisel, W. Jarausch, K. Goretzki, A. Henschen, J. Engels, U. Weller, M. Hudel, E. Habermann, H. Niemann, Tetanus toxin: primary structure, expression in E. coli, and homology with botulinum toxins, EMBO J. 5 (1986) 2495–2502.

[5] M. Matsuda, D.L. Lei, N. Sugimoto, K. Ozutsumi, T. Okabe, Isolation purification, and characterization of fragment B, the NH2-terminal half of the heavy chain of tetanus toxin, Infect. Immun. 57 (1989) 3588–3593.

[6] C. Montecucco, G. Schiavo, Mechanism of action of tetanus and botulinum neurotoxins, Mol. Microbiol. 13 (1994) 1–8.

[7] J.M. Toivonen, S. Oliván, R. Osta, Tetanus toxin C-fragment: the courier and the cure? Toxins. 2 (2010) 2622–2644, http://dx.doi.org/10.3390/toxins2112622.

[8] J. Guo, X. Pan, Y. Zhao, S. Chen, Engineering Clostridia Neurotoxins with elevated catalytic activity, Toxicon Off J. Int. Soc. Toxinol. 74 (2013) 158–166, http://dx.doi.org/10.1016/j.toxicon.2013.08.055.

[9] P. Costantino, R. Rappuoli, F. Berti, The design of semi-synthetic and synthetic glycoconjugate vaccines, Expert Opin, Drug Discov. 6 (2011) 1045–1066, http://dx.doi.org/10.1517/17460441.2011.609554.

[10] D. Goldblatt, Conjugate vaccines, Clin. Exp. Immunol. 119 (2000) 1–3, http://dx.doi.org/10.1046/j.1365-2249.2000.01109.x.

[11] M. Thaysen-Andersen, S.B. Jørgensen, E.S. Wilhelmsen, J.W. Petersen, P. Højrup, Investigation of the detoxification mechanism of formaldehyde-treated tetanus toxin, Vaccine 25 (2007) 2213–2227, http://dx.doi.org/10.1016/j.vaccine.2006.12.033.

[12] B. Metz, G.F.A. Kersten, P. Hoogerhout, H.F. Brugghe, H.A.M. Timmermans, A. de Jong, H. Meiring, J. ten Hove, W.E. Hennink, D.J.A. Crommelin, W. Jiskoot, Identification of formaldehyde-induced modifications in proteins: reactions with model peptides, J. Biol. Chem. 279 (2004) 6235–6243, http://dx.doi.org/10.1074/jbc.M310752200.

[13] B. Metz, G.F.A. Kersten, G.J.E. Baart, A. de Jong, H. Meiring, J. ten Hove, M.J. van Steenbergen, W.E. Hennink, D.J.A. Crommelin, W. Jiskoot, Identification of formaldehyde-induced modifications in proteins: reactions with insulin, Bioconjug. Chem. 17 (2006) 815–822, http://dx.doi.org/10.1021/bc050340f.

[14] L. Trézl, L. Hullán, Z.M. Jászay, T. Szarvas, I. Petneházy, B. Szende, J. Bocsi, Z. Takáts, K. Vékey, L. Töke, Antagonistic reactions of arginine and lysine against formaldehyde and their relation to cell proliferation, apoptosis, folate cycle and photosynthesis, in: J.F. Clark (Ed.), Guanidino Compd. Biol. Med. Springer, US, 2003, pp. 167–176.. Accessed 2 January 2017 http://link.springer.com/chapter/10.1007/978-1-4615-0247-0_25.

[15] A. Neuberger, The reaction between histidine and formaldehyde, Biochem. J. 38 (1944) 309–314.

[16] H. Fraenkel-Conrat, M. Cooper, H.S. Olcott, The Reaction of Formaldehyde with Proteins, J. Am. Chem. Soc. 67 (1945) 950–954, http://dx.doi.org/10.1021/ja01222a023.

[17] R.G. Kallen, W.P. Jencks, The Mechanism of the condensation of formaldehyde with tetrahydrofolic acid, J. Biol. Chem. 241 (1966) 5851–5863.

[18] B. Metz, W. Tilstra, R. van der Put, N. Spruit, J. van den Ijssel, J. Robert, C. Hendriksen, G. Kersten, Physicochemical and immunochemical assays for monitoring consistent production of tetanus toxoid, Biol. J. Int. Assoc. Biol. Stand. 41 (2013) 231–237, http://dx.doi.org/10.1016/j.biologicals.2013.05.001.

[19] K. Lockyer, F. Gao, J.P. Derrick, B. Bolgiano, Structural correlates of carrier protein recognition in tetanus toxoid-conjugated bacterial polysaccharide vaccines, Vaccine 33 (2015) 1345–1352, http://dx.doi.org/10.1016/j.vaccine.2015.01.046.

[20] M.M. Acharya, S.S. Katyare, An improved micromethod for tyrosine estimation, Z. Naturforschung C J. Biosci. 59 (2004) 897–900.

[21] P.W. Riddles, R.L. Blakeley, B. Zerner, Ellman's reagent: 5, 5′-dithiobis(2-nitrobenzoic acid)–a reexamination, Anal. Biochem. 94 (1979) 75–81.

[22] A.F.S.A. Habeeb, Determination of free amino groups in proteins by trinitrobenzenesulfonic acid, Anal. Biochem. 14 (1966) 328–336, http://dx.doi.org/10.1016/0003-2697(66)90275-2.

[23] S.M. Rutherfurd, G.S. Gilani, Amino Acid Analysis, Curr Protoc. Protein Sci. John Wiley & Sons, Inc., 2001 (Accessed 2 January 2017), http://onlinelibrary.wiley.com/doi/10.1002/0471140864.ps1109s58/abstract.

[24] O.H. Lowry, N.J. Rosebrough, A.L. Farr, R.J. Randall, Protein Measurement with the Folin Phenol Reagent, J. Biol. Chem. 193 (1951) 265–275.

[25] B. Yan, J. Valliere-Douglass, L. Brady, S. Steen, M. Han, D. Pace, S. Elliott, Z. Yates, Y. Han, A. Balland, W. Wang, D. Pettit, Analysis of post-translational modifications in recombinant monoclonal antibody IgG1 by reversed-phase liquid chromatography/mass spectrometry, J. Chromatogr. A (2007) 153–161, http://dx.doi.org/10.1016/j.chroma.2007.06.063.

[26] M. Fountoulakis, H.-W. Lahm, Hydrolysis and amino acid composition analysis of proteins, J. Chromatogr. A 826 (1998) 109–134, http://dx.doi.org/10.1016/S0021-9673(98)00721-3.

[27] A.S. Inglis, Single hydrolysis method for all amino acids, including cysteine and tryptophan, Methods Enzymol 91 (1983) 26–36.

[28] T.-Y. Liu, R.A. Boykins, Hydrolysis of proteins and peptides in a hermetically sealed microcapillary tube: High recovery of labile amino acids, Anal. Biochem. 182 (1989) 383–387, http://dx.doi.org/10.1016/0003-2697(89)90612-X.

[29] M. Friedman, J.W. Finley, Methods of tryptophan analysis, J. Agric. Food Chem. 19 (1971) 626–631, http://dx.doi.org/10.1021/jf60176a010.

[30] T.E. McCarthy, M.X. Sullivan, A New and Highly Specific Colorimetric Test for Methionine, J. Biol. Chem. 141 (1941) 871–876.

[31] P. Reschiglian, A. Zattoni, B. Roda, E. Michelini, A. Roda, Field-flow fractionation and biotechnology, Trends Biotechnol. 23 (2005) 475–483, http://dx.doi.org/10.1016/j.tibtech.2005.07.008.

[32] V. Ball, J.J. Ramsden, Buffer dependence of refractive index increments of protein solutions, Biopolymers 46 (1998) 489–492, http://dx.doi.org/10.1002/(SICI)1097-0282(199812)46:7<489:AID-BIP6>3.0.CO;2-E.

[33] P.J. Wyatt, Light scattering and the absolute characterization of macromolecules, Anal. Chim. Acta 272 (1993) 1–40, http://dx.doi.org/10.1016/0003-2670(93)80373-S.

[34] K.N. Rao, D. Kumaran, T. Binz, S. Swaminathan, Structural analysis of the catalytic domain of tetanus neurotoxin, Toxicon Off, J. Int. Soc. Toxinol. 45 (2005) 929–939, http://dx.doi.org/10.1016/j.toxicon.2005.02.032.

[35] H.R. Costantino, S.P. Schwendeman, K. Griebenow, A.M. Klibanov, R. Langer, The secondary structure and aggregation of lyophilized tetanus toxoid, J. Pharm. Sci. 85 (1996) 1290–1293, http://dx.doi.org/10.1021/js960148+.

[36] W. Yassine, N. Taib, S. Federman, A. Milochau, S. Castano, W. Sbi, C. Manigand, M. Laguerre, B. Desbat, R. Oda, J. Lang, Reversible transition between alpha-helix and beta-sheet conformation of a transmembrane domain, Biochim. Biophys. Acta 2009 (1788) 1722–1730, http://dx.doi.org/10.1016/j.bbamem.2009.05.014.