Detection of busulfan adducts on proteins

Acosta Martin, Adelina Elena; Antinori Malaspina, Paola; Uppugunduri Satyanarayana, Chakradhara Rao; Daali, Youssef; Ansari Djaberi, Marc Georges; Scherl, Alexander; Muller, Markus Johann; Lescuyer, Pierre

How to cite

ACOSTA MARTIN, Adelina Elena et al. Detection of busulfan adducts on proteins. In: RCM. Rapid communications in mass spectrometry, 2016, vol. 30, n° 23, p. 2517–2528. doi: 10.1002/rcm.7730

This publication URL: https://archive-ouverte.unige.ch/unige:88690
Publication DOI: 10.1002/rcm.7730

Detection of busulfan adducts on proteins

SHORT TITLE: busulfan adducts on proteins

Adelina E. Acosta-Martin [^,1, 2], Paola Antinori [^,2, 3], Chakradhara Rao S. Uppugunduri [4, 5], Youssef Daali [6], Marc Ansari [4, 5], Alexander Scherl [1, 2, 3], Markus Müller [7], Pierre Lescuyer [1, 2] *

[1] Department of Human Protein Science, Faculty of Medicine, University of Geneva, Geneva, Switzerland

[2] Division of Laboratory Medicine, Geneva University Hospitals, Geneva, Switzerland

[3] Swiss Centre of Applied Human Toxicology, Geneva, Switzerland

[4] Onco-Hematology Unit, Department of Pediatrics, Geneva University Hospitals, Geneva, Switzerland

[5] Cansearch Research Laboratory, Geneva Medical University, Geneva, Switzerland

[6] Division of Clinical Pharmacology & Toxicology, Geneva University Hospitals, Geneva, Switzerland

[7] SIB-Swiss Institute of Bioinformatics, University of Geneva, Switzerland

^ These authors contribute equally to this work.

*** Corresponding author:**

Pierre Lescuyer

SML/DMGL, Hôpitaux Universitaires de Genève, 1211 Genève 14, Switzerland

Tel.: 00 41 (0) 79 55 35 750 - Fax: 00 41 (0) 22 37 21 837

pierre.lescuyer@unige.ch

KEYWORDS: protein adducts, busulfan, mass spectrometry, open modification search.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/rcm.7730

ABSTRACT

RATIONALE

Busulfan is a bifunctional alkyl sulfonate antineoplastic drug. This alkylating agent was described as forming covalent adducts on proteins. However, only limited data are available regarding the interaction of busulfan with proteins. Mass spectrometry and bioinformatics were used to identify busulfan adducts on human serum albumin and hemoglobin.

METHODS

Albumin and hemoglobin were incubated with busulfan or control compounds, digested with trypsin and analyzed by LC-MS/MS on a Thermo Fisher LTQ Orbitrap Velos Pro. MS data were used to generate spectral libraries of non-modified peptides and an open modification search was performed to identify potential adduct mass shifts and possible modification sites. Results were confirmed by a second database search including identified mass shifts and by visual inspection of annotated tandem mass spectra of adduct-carrying peptides.

RESULTS

Five structures of busulfan adducts were detected and a chemical structure could be attributed to four of them. Two were primary adducts corresponding to busulfan monoalkylation and alkylation of two amino acid residues by a single busulfan molecule. Two others corresponded to secondary adducts generated during sample processing. Adducts were mainly detected on Asp, Glu, and His residues. These findings were confirmed by subsequent database searches and experiments with synthetic peptides.

CONCLUSIONS

The combination of *in vitro* incubation of proteins with the drug of interest or control compounds, high-resolution mass spectrometry, and open modification search allowed confirming direct interaction of busulfan with proteins and characterizing resulting adducts. Our results also showed that careful analysis of the data is required to detect experimental artifacts.

INTRODUCTION

Human body is exposed to exogenous chemicals during the entire lifetime.[1] Some of these compounds are electrophiles or become electrophiles after bioactivation, thus being able to covalently bind to nucleophilic residues on biomolecules to form adducts. Unlike DNA adducts, protein adducts are not removed by enzymatic repair systems and can consequently induce permanent protein damage, leading to the manifestation of pathological states.[2,3] Precise information about the reactivity towards proteins to form adducts is however missing for many pharmaceutical drugs currently used for disease treatment. An example of such pharmaceutical compounds is 1,4-butanediol dimethanesulfonate (busulfan), one of the oldest chemotherapeutic agents, which was introduced in clinical practice in the 1950s. Busulfan is a bi-functional alkylating drug currently used as a component of conditioning regimens before hematopoietic stem cell transplantation in adult and pediatric patients.[4,5] Being an electrophile, busulfan is supposed to covalently interact with blood and cellular proteins during these treatments. A study using ^{14}C-busulfan suggested indeed that busulfan was irreversibly binding to plasma and blood cell proteins.[6] However, no information was provided in this work on the adducts structure and on the target aminoacids. More recently, it was shown that γ-glutamyldehydroalanylglycine (EdAG), an electrophile compound formed during busulfan metabolism, was able to bind to different proteins: the glutathione S-transferase GSTA1-1[7] and glutaredoxins-1 and -2.[8] EdAG reacts with sulfhydryl groups of proteins to form a stable lanthionine linkage. Nevertheless, reaction with EdAG cannot account for the results obtained from the interaction of ^{14}C-busulfan with blood proteins[6] since this metabolite is formed from glutathione by a β-elimination reaction involving the loss of hydrogen sulfide. Accordingly, busulfan probably forms adducts by direct interaction with proteins. In addition, due to the bi-functional nature of busulfan, internal rearrangement or cross-links are expected. Investigations remain therefore to be done to confirm the formation of busulfan-protein adducts, to determine the structure of the resulting adduct(s) and to identify the target residues on proteins. Such data would be helpful for better understanding the mechanisms of cellular and toxic effects of busulfan.

Historically, the characterization of drug-adducts on proteins relied on the complete hydrolysis of the protein followed by the analysis of isolated aminoacids by liquid (LC) or gas chromatography and mass spectrometry (MS).[9] Since then, developments in proteomics and MS have enabled the characterization of chemical modifications on peptides or even intact proteins.[9, 10] In this context, the standard process for searching protein modifications

from LC-MS/MS data has some limitations since it requires a pre-established knowledge of the modification, i.e. its chemical composition and the target modification site. Thus, finding unknown or unexpected protein modifications needs a different approach than standard peptide sequence searches. Trial and error searches become prohibitive both in terms of search time and number of false positives when including too many modifications. Open modification searches (OMS) should be considered as an alternative.[11] OMS does not require prior definition of the modifications but extract the information directly from the analyzed tandem mass data.[12-14] This is possible because OMS is based on the alignment of modified query spectra to unmodified database spectra, where some fragment peaks are shifted in accordance with the mass of the modification.[15-17] In this way, information about both mass shifts and modification sites is provided. In most practical applications, OMS is performed after a reduction of the search space to limit search time and to increase sensitivity. Thus, query spectra can be searched in either a small set of peptides likely to be present in the sample[18] or in a library of previously identified consensus tandem mass spectra.[19-21] Furthermore, the use of consensus tandem mass spectra increases accuracy of OMS alignments compared to theoretical spectra.[20] Detected modifications can be summarized in mass shift histograms (MSH), counting in how many spectrum alignments a certain mass shift was observed.[18, 22] Hence, the application of OMS for the characterization of unknown drug-adducts on proteins seems to be more advantageous than classical protein modification searches.

The present study reports the use of an approach combining MS-based proteomics and OMS to identify adducts formed by busulfan on proteins. The workflow consisted in the following steps: *in vitro* incubation of human serum albumin (HSA) and hemoglobin (Hb) with busulfan or control compounds; trypsin digestion and LC-MS/MS analysis in a high resolution tandem mass spectrometer; database search and generation of spectral libraries; OMS to identify unknown mass shifts and possible modification sites of generated adducts; confirmation of the OMS results by a second database search including the potential drug-adduct modification masses provided by the OMS; and finally, annotation and visual inspection of tandem mass spectra of adduct-carrying peptides.

EXPERIMENTAL PROCEDURES

Materials

HSA, Hb, triethyl ammonium bicarbonate (TEAB), tris(2-carboxyethyl)phosphine (TCEP), iodoacetamide, α-cyano-4-hydroxycinnamic acid (CHCA), dimethyl sulfoxide (DMSO), and HPLC grade acetonitrile (ACN) were purchased from Sigma-Aldrich (St. Louis, Missouri, MO). Busulfan, sulfolane, trifluoroacetic acid (TFA) and methanol were obtained from Fluka Chemika. Urea and ammonium phosphate were obtained from Merck (Darmstadt, Germany), HPLC grade formic acid (FA) from Biosolve Chemical (Valkenswaard, The Nederlands), HPLC grade water from Romil (Cambridge, UK), phosphate-buffered saline (PBS) Gibco from Life Technologies (Zug, Switzerland), and butyl methanesulfonate (BMS) from Ark Pharm, Inc. (Illinois, USA). C18 macro-spin columns were purchased from Harvard apparatus (Holliston, MA, USA), ZipTip C18 tips and ultrafiltration Millipore AMICON 0.5 (cut-off 3 kDa) from Merck-Millipore (Billerica, MA, USA), and sequencing grade modified trypsin from Promega. The peptide AQYLQQCPFEDHVK was synthetized by Thermo Fisher Scientific (Ulm, Germany).

In vitro generation of drug-adducts and sample preparation for MS analysis

Purified Hb and HSA were incubated separately with different chemical agents at a molar ratio of protein to chemical agent of 1:150, mixing 50 µL of protein solution with sulfolane, busulfan, or BMS. As a control condition, 50 µL of protein were incubated with DMSO. For all conditions, triplicate incubations were performed at 37°C during 6 h in a PCR thermocycler (Biorad, Hercules, CA, USA). After the incubation, samples were filtered with Amicon Ultra-0.5 3K filters according to manufacturer's instructions. After evaporation, 6 M urea in 0.1 M TEAB pH=8 was added to 50 µg of incubated proteins. Disulfide bonds were reduced with 30 mM TCEP at 37°C for 1 h and alkylated with 130 mM iodoacetamide at 37°C for 1 h in the dark with shaking. Samples were then diluted before addition of trypsin (enzyme to protein ratio of 1:50). Protein digestion was carried out overnight at 37°C.

A synthetic peptide (AQYLQQCPFEDHVK), corresponding to a C-terminal portion of the tryptic peptide containing Cys 34 of HSA (ALVLIAFAQYLQQCPFEDHVK), was incubated with busulfan at a peptide to reagent molar ratio of 1:150, mixing peptide solution with busulfan. As a control condition, peptide was incubated with DMSO. Duplicate incubations were performed at 37°C during 6 h in a PCR thermocycler. Peptides were then desalted and immediately analyzed by MALDI-TOF MS.

Detailed description can be found as supplementary experimental procedures.

LC-MS/MS analysis

Each of the 24 digested samples was run in triplicate. HPLC was performed on a Waters NanoAcquity LC system (Waters, Manchester, UK) using a home-made 0.1 × 20 mm trap column (5 µm, 200 Å) packed with Magic C18 AQ (Bruker-Michrom, Auburn, CA, USA) stationary phase and a commercial 0.075 x 150 mm, C18, 5 µm, 100 Å Nikkyo analytical nano-column (Nikkyo Technology, Japan). More details can be found as supplementary experimental procedures.

MS analysis was performed on a LTQ Orbitrap Velos Pro (Thermo Electron, San Jose, CA, USA) hybrid ion trap-orbitrap mass spectrometer. MS survey scans in positive ion mode were acquired in the FT-orbitrap analyzer using a *m/z* window from 400 to 2000, a resolution of 60´000, and an automatic gain control target setting of $5x10^5$. The three most intense precursor ions were selected for the acquisition of tandem mass spectra (7'500 resolution) using higher-energy collisional dissociation (HCD) prior to FT-orbitrap detection. Charge states 2+ and higher were included for precursor selection. Monoisotopic precursor selection was activated Normalized collision energy was set to 30%, activation time to 0.1 ms, isolation width to 2.5 *m/z*, and automatic gain control target value to $2x10^5$. Total cycle time was 85 min. Dynamic exclusion for precursor ions was applied for 45 s.

Mass measurement error calculations and peptide identification searches

LC-MS/MS analysis RAW files were converted to mgf peak lists using EasyProtConv v1.6.[23] Each file was analyzed with Preview software to calculate mass measurement errors for both precursor and fragment ions.[24] Protein searches were performed against databases created in-house from FASTA files containing exclusively the sequence of either Hb (UniProtKB accession: P69905 and P68871) or HSA (UniProtKB accession: P02768).

For the generation of spectral libraries, protein identification was performed from mgf files the EasyProt search engine v2.3.[23] The same protein database was used for the calculation of mass measurement errors. Search parameters are described in the supplementary experimental procedures. These EasyProt results were exported to pepXML files.

MALDI-TOF MS analysis

MALDI-TOF MS spectra of the synthetic peptide (AQYLQQCPFEDHVK) were recorded in reflectron positive ion mode using a 4800 MALDI-TOF/TOF mass spectrometer (Applied Biosystems Inc., Framingham, MA, USA) equipped with a Nd:YAG laser operating at 355 nm and 200 Hz. For each extracted and desalted peptide sample, 0.8 μL of the peptide solution were mixed with 1 μL of matrix solution (5 mg of CHCA dissolved in 1 mL of 0.1% TFA/50% ACN, and addition of 10 μL of 10 mM ammonium phosphate) on the MALDI-TOF MS plate. An accelerating voltage of 20 kV was applied and reflectron voltages were 14.1 kV and 20.4 kV. MS spectra were acquired from 800 to 4000, from 1750 to 1960, and from 3350 to 3660 *m/z*. Each spectrum represents the cumulative average of 1200 laser shots. The mass spectra were then inspected and calibrated using DataExplorer software (Applied Biosystems Inc., Framingham, MA, USA).

Generation of tandem mass spectral libraries

To build the tandem mass spectral libraries (HSA-Lib and Hb-Lib), all HSA and Hb peptides confidently identified by EasyProt in the 72 LC-MS/MS analyses were exported as pepXML files. Consensus spectra were calculated and exported to sptxt format (SpectraST) using the in-house Liberator tool v2.0.[25] Fragment peaks were merged into a consensus peak if their *m/z* values were within 0.3 units and if the peak was present in at least 20% of the spectra. Decoy spectra were generated using the Deliberator tool (version 2.0) to calculate the FDR.[26] Deliberator implements a variation of the peptide shuffling and peak repositioning algorithm, generating realistic but randomized spectra.[27] For each original library spectrum, the algorithm appended a decoy spectrum to the library file. Both Liberator and Deliberator tools are based on the MzJava class library (mzjava.expasy.org).

OMS for mass shifts

The OMS for mass shifts of generated adducts was performed using QuickMod software v1.03 (http://javaprotlib.sourceforge.net/packages/tools/).[20] For both HSA and Hb, QuickMod aligned every query spectrum to every candidate spectrum in the corresponding spectral library (HSA-Lib or Hb-Lib), which had the same charge and a neutral precursor mass within the modification mass tolerance of the query spectrum precursor. Query spectra for the four experimental conditions (PBS/DMSO, Sulfolane, Busulfan, and BMS) were treated separately. After alignment, spectra were rank transformed and spectrum-spectrum matches (SSM) were scored using the normalized dot product score. For every match to a

library spectrum, a positioning score (S_{pos}) was generated for all amino acid positions in the peptide sequence and the modification was assigned to the position(s) with the best score. QuickMod parameters were set as follows: modification mass tolerance = 320 Da; fragment ion mass tolerance = 0.02 Da; charge states for query precursor ions = +2 to +5; and library spectrum candidates for SSM for each query spectrum = 1. The FDR was calculated separately for modified and unmodified spectra and for each precursor charge. Score thresholds were set to a value that corresponded to a FDR of 1%.

Integration of OMS results and differential analysis of mass shifts

In order to perform differential statistical analyses between the experimental conditions (PBS/DMSO, busulfan, sulfolane, and BMS), OMS results were integrated using R scripts as described in the supplementary experimental procedures.

Amino acid positioning of mass shifts

For each condition (PBS/DMSO, busulfan, sulfolane, BMS), the tendency $W(aa,\Delta M)$ of a given modification mass shift ΔM to be located on a particular amino acid residue *aa* was assessed by the implementation of a R script as described in the supplementary experimental procedures.

In order to compare the amino acid positioning tendency of the different busulfan modifications, the tendency $W(aa, \Delta M)$ of each amino acid was divided by the amount of that amino acid in each protein sequence (corrected tendency). These results were then normalized by the highest value to compare the amino acid positioning tendency between HSA and Hb.

Confirmation of OMS results via database searches

The database searches to confirm mass shifts provided by QuickMod applied the same parameters as described above except for the following: (i) The validation of BMS included BMS modification (+56.0625968 Da for $+C_4H_9(-H)$) on Asp, Glu, and His as variable; (ii) The validation of the primary adduct of busulfan included carbamidomethylation of Cys as well as busulfan modification (+150.035063 Da for $+C_5H_{11}O_3S(-H)$) on Cys, Asp, Glu, His, Lys and Arg as variable; (iii) Additionally, searches including the mass of the busulfan intra-

peptide cross-link (busulfan loop; +54.0469476 Da for $+C_4H_8(-2H)$) on Asp, Glu, and His as a variable modification were performed to aid the validation process.

Annotation and visual inspection of tandem mass spectra

Amino acid residues carrying both BMS and busulfan +150.035 Da adducts in EasyProt searches were selected for further validation through visual inspection of corresponding tandem mass spectra with the aid of the peak labeling software pLabel (http://pfind.ict.ac.cn). Immonium ions, as well as single, double and triple charged fragment ions of the series *y, y-H_2O, b, b-H_2O, a*, and *a-H_2O* were selected for peak matching using the following parameters: error tolerance, 15 ppm; intensity threshold, 2.5%; mass measurement, monoisotopic; and match type, highest. Busulfan modification was defined as +150.035063 Da for $+C_5H_{11}O_3S(-H)$ and BMS modification as +56.0625968 Da for $+C_4H_9(-H)$. Adducted residues were considered as valid when more than 3 consecutive ions of the same series including the modification site were detected. The same parameters of peak matching were used to visualize QuickMod aligned spectra and validate the existence of the following modifications: +72.0575118 Da for $+C_4H_9O(-H)$ and +100.0524268 Da for $+C_5H_9O_2(-H)$. In order to evaluate the formation of busulfan loops, tandem mass spectra matching peptides with the corresponding mass shift (+54.0469476 for $+C_4H_8(-2H)$) were examined. Fragment ion patterns were compared against tandem mass spectra of the same unmodified peptide fragmented at the same charge state.

RESULTS

Generation of spectral libraries

In the present study, we developed a workflow combining high accuracy MS analysis and bioinformatics to identify busulfan adducts generated *in vitro* on HSA and Hb (Figure 1). In addition to busulfan, proteins were incubated with several chemicals used as control conditions: (1) BMS, an alkylating agent with the same chemical structure as busulfan but containing only one methanesulfonate group; (2) sulfolane, a non-alkylating metabolite of busulfan;[28] (3) DMSO, the solvent used to prepare busulfan and BMS solutions. After incubation, proteins were digested with trypsin and analyzed by LC-MS/MS in a data-dependent acquisition manner. Median values of mass measurement accuracy were calculated for each of the 72 MS acquisitions. For precursor ions, median accuracy values were between 1 and 2 ppm, and between 1.5 and 3.2 ppm for fragments. Database search provided identification of non-adducted peptides with a protein sequence coverage of 93.10% for HSA,

99.30% for Hb alpha chain, and 95.92% for Hb beta chain (Suppl. Information 1). Spectra corresponding to these identifications were used to generate high coverage, high mass accuracy spectral libraries of non-adducted peptides. HSA library included 610 consensus spectra generated from 30,697 identified spectra and Hb library included 469 consensus spectra generated from 15,638 identified spectra.

Modification mass shifts produced during incubations

The OMS for mass shifts of generated adducts was performed using QuickMod software. Tandem mass spectra of HSA and Hb peptides corresponding to the four incubation conditions (PBS/DMSO, busulfan, sulfolane, BMS) were matched against HSA and Hb spectral libraries. Precursor mass shifts of adducted peptides in each condition were determined and statistical analyses between incubation conditions allowed distinguishing which adduct modifications were specifically generated in each condition. When comparing busulfan incubations against PBS/DMSO controls (Figure 2A), five mass shifts were found specific for busulfan: +54.045, +72.059, +90.025, +100.054, and +150.037 Da (these mass values correspond to the mean mass shifts in theirs MSH bins). All these mass shifts were also highlighted when comparing busulfan against sulfolane incubations (Figure 2B). These results suggested that busulfan generated multiple adduct structures on HSA and Hb. After incubation with BMS, only the mass shift of +56.060 Da was found to be specific when compared to PBS/DMSO controls (Figure 2C). Sulfolane incubations were also compared against PBS/DMSO controls and no significantly enriched mass shift was found (Figure 2D).

Hypothetical chemical structures of busulfan and BMS adducts

The chemical structure of adducts corresponding to the mass shifts detected by the OMS was hypothesized and monoisotopic molecular weights were calculated. For busulfan, two of the OMS mass shifts corresponded to primary adducts (Figure 3A). The mass shift of +150.037 Da was attributed to busulfan monoalkylation (+150.035063 Da for $+C_5H_{11}O_3S(-H)$), and the mass shift of +54.045 Da to the alkylation of two amino acid residues on the same peptide by a single busulfan molecule (+54.0469476 Da for $+C_4H_8(-2H)$), thus generating an intra-peptide crosslink (busulfan loop). For the mass shifts of +72.059 Da and +100.054 Da, we hypothesized the generation of secondary busulfan adducts (Figure 3B, 3C), resulting from additional chemical modifications of primary adducts (+72.0575118 Da for $+C_4H_9O(-H)$ and +100.0524268 Da for $C_5H_9O_2(-H)$). Finally, we were not able to formulate any hypothesis on chemical structure for the modification at +90.025 Da. This busulfan mass shift and the

corresponding potential adduct were thus excluded for further analyses. Regarding BMS, for the mass shift of +56.060 Da detected during the OMS, we hypothesized a chemical structure (+56.0625968 Da for $+C_4H_9(-H)$) corresponding to the primary alkylation by this compound (Figure 3D).

Amino acid specificity of busulfan adducts

Statistical analyses of QuickMod results allowed evaluating the amino acid positioning of busulfan adducts (Equation 6). The distribution of adducted residues in HSA and Hb were consistent for the primary (+150.035 Da) and secondary (+100.052 Da and +72.057 Da) busulfan adducts (Suppl. Figure 2). Furthermore, since these calculations were based on the number of spectral alignments of each mass shift, the higher number of modifications for the busulfan adduct of +150.035 Da suggested that this modification was the most abundant adduct generated by busulfan. For both HSA and Hb, the distribution of amino acid positioning tendency of the busulfan primary adduct (+150.035 Da) was normalized to the highest value. Figure 4A shows that amino acid residues with highest specificity for busulfan alkylation were Cys, Asp, Glu, His, Lys and Arg. Moreover, busulfan alkylation occurred with similar specificity in both proteins, suggesting that alkylation specificity does not depend on the protein itself. Because this mass shift corresponded to the primary adduct structure and to the most abundant modification, it was selected for further validation of the OMS results with conventional database search including the modification.

Visual inspection of tandem mass spectra for busulfan and BMS adducted peptides

In order to confirm OMS results and validate the presence of primary busulfan adducts, database searches for the identification of HSA and Hb were repeated while including the busulfan modification (+150.035 Da) on Cys, Asp, Glu, His, Lys and Arg. First, identifications of busulfan adducted Lys and Arg were evaluated. The use of trypsin for digestion involved the cleavage of proteins at the carboxyl side of Lys and Arg. Rationally, if the alkylation of Lys and Arg is produced before digestion, trypsin should not cleave efficiently and very few tryptic peptides would be identified. Hence, the identification of tryptic peptides with adducted Lys or Arg suggests that busulfan alkylation of the C-terminal amino acid of the peptide took place after digestion. Additionally, tandem mass spectra of peptides with missed cleavages were annotated and visually inspected (Suppl. Information 2), and only two out of 190 spectra showed that busulfan reacted with the side chain of an uncleaved Lys residue. This may be explained by the fact that Lys and Arg residues are less

nucleophilic when protonated as is the case at the incubation pH. Additionally, this indicated that busulfan alkylation of Lys and Arg mostly occurred post-digestion at the carboxyl group of the C-terminal side of tryptic peptides. We hypothesized that due to the highly hydrophobic nature of the molecule and its low solubility in aqueous solutions, some busulfan was retained by the Amicon cellulose membrane and did not get filtered prior to digestion. Second, regarding busulfan alkylation of Cys residues, database searches for Hb did not return any peptide with modified Cys. In the case of HSA, the sulfhydryl group of Cys34 was the only available to be modified during busulfan incubations. Busulfan adducted Cys34 was identified in only two tandem mass spectra that were visually examined. Despite the fact that mass errors on the precursor ions were lower than 8 ppm, fragment ions did clearly not match the peptide sequence (Suppl. Information 3). Finally, database searches confirmed the identification of busulfan modifications on diverse Glu, Asp and His residues of both HSA and Hb. Since both busulfan and BMS contain the same alkylating group, BMS experiments verified the alkylated residues were Glu, Asp and His. Thus, a database search for protein identification including BMS (+56.063 Da) on Glu, Asp and His as variable modification was performed, and only adducted sites in common to both chemicals were kept for annotation and visual inspection of corresponding tandem mass spectra. For both busulfan (Suppl. Information 4) and BMS (Suppl. Information 5), in addition to the matching of *y* or *b* series of fragments surrounding the modification, immonium ions of modified residues were detected. They corresponded to *m/z* = 238.074, 252.090 and 260.106 for busulfan adducted Asp, Glu and His, respectively, and to *m/z* = 144.102, 158.118 and 166.134 for BMS adducted Asp, Glu and His, respectively. In total, 35 residues in HSA (Figure 4B) and 21 in Hb (Figure 4C) were successfully validated as being alkylated by busulfan. Figure 5 shows the annotated spectrum of modified HSA peptide LVNEVT**E**FAK, in which the entire *y* series of fragment ions was detected as well as the immonium ions of modified Glu.

The generation of busulfan loops on HSA and Hb was carefully evaluated by annotation and visual inspection of spectra matching for peptides carrying a modification corresponding to the added mass (+54.047 Da) and by comparing the fragmentation pattern of looped peptides to the pattern of unmodified peptides (Suppl. Information 6). Altogether, eight pairs of amino acid residues in HSA and Hb were consistent with a busulfan loop modification (Figure 4B and 4C). In all modified peptides, fragment ions within the cross-linked amino acid domain were not observed whereas they were clearly detected on unmodified peptides. Taking the example of the Hb peptide GTFATLSELHCDKLHV**D**P**E**NFR (Figure 6), the confirmation of the busulfan loop was not only based on the mass shift observed in fragment ions

containing looped residues, but also on the fact that the strong "proline effect" (preference for fragmentation at the N-terminal of proline residues) observed for the unmodified peptide (intense y_5^+ ion fragment) was lost in the looped peptide due to the positioning of the proline between the two cross-linked residues.

The presence of secondary busulfan adducts (+72.057 Da and +100.052 Da) was also confirmed (Suppl. Information 7 and Suppl. Information 8). Tandem mass spectra from the OMS results were manually annotated and visually inspected using the molecular weight of proposed chemical structures and the same criteria as in the validation of primary busulfan adducts.

Discrimination of primary and secondary busulfan adducts by MALDI-TOF MS

The synthetic peptide AQYLQQCPFEDHVK was incubated with busulfan or with PBS/DMSO and analyzed by MALDI-TOF MS to validate that the presence of secondary adducts was a consequence of sample treatment. For both the linear and the disulfide-linked structures of the peptide, ions corresponding to busulfan adducts were exclusively detected in samples incubated with busulfan (Figure 7), which is in agreement with OMS results obtained with HSA and Hb. Additionally, only peaks of primary adducts (+150.035 Da and +54.047 Da) were detected in the MALDI spectra of the busulfan incubated peptide, suggesting that the formation of secondary adducts (+72.057 Da and +100.052 Da) in HSA and Hb did not occur during the incubation but in a subsequent step of sample processing.

DISCUSSION

The aim of the present study was to identify adducts on proteins resulting from the covalent interaction with busulfan. The experimental workflow combined the generation of busulfan adducts on purified proteins *in vitro* with MS-based proteomics and OMS. This set-up allowed working with highly standardized and reproducible assay conditions and to reduce system complexity. As shown by our results, these aspects were important for a reliable and accurate interpretation of OMS data. Indeed, LC-MS/MS analysis of purified HSA and Hb favored high sequence coverage identifications, resulting in the compilation of high sequence coverage spectral libraries. This increased the confidence in OMS results and the possibility to identify low abundance peptides carrying adducts. In addition, the incubation of a synthetic peptide and its analysis by MALDI-TOF allowed the discrimination of primary busulfan adducts from secondary busulfan adducts and led to the conclusion that secondary adducts were artifacts generated during post-incubation sample treatment. We hypothesize that the

primary busulfan adduct of +150.035 Da underwent basic hydrolysis during trypsin digestion (pH = 8), resulting in the replacement of the second methanesulfonate group by a hydroxyl group (secondary adduct of +72.057 Da). For the secondary adduct of +100.052 Da, two hypotheses involving formic acid, which was used to stop digestion and into sample and LC buffers for LC-MS/MS analyses, seem possible: (i) the hydroxyl group of the +72.057 adduct reacted with formic acid to form the corresponding ester; (ii) the second methanesulfonate group of the busulfan primary adduct (+150.035) reacted with formate as a nucleophile. For the analysis of the synthetic peptide, trypsin digestion (basic conditions), as well as the use of formic acid, was avoided, as opposed to the analysis of purified proteins. These results highlighted the fact that adducted groups may undergo additional chemical modifications during sample processing. Hence, careful considerations in sample processing steps are required for the particular study of unknown adducts.

The use of tandem mass spectra acquired with high measured mass accuracy is favorable for OMS. High mass accuracy measurement of fragment ions was shown to be the acquisition mode of choice for the identification of peptides containing post-translational modifications.[29] Accurate alignments of spectra carrying busulfan adducts with the corresponding unmodified spectra from the spectral library resulted in highly specific peptide-spectrum matches. Additionally, high mass accuracy measurements offered the possibility to confirm peptide modifications in tandem mass spectra with high confidence. Moreover, confirming the position of a presumed modification on a particular amino acid residue implies the detection of peptide fragments surrounding the modification site. Regarding residues in the two or three first positions from the N- or C-termini of peptides, fragment ions needed for validation are usually in the low *m/z* region, difficult to access with traditional ion-trap collision-induced dissociation. Activation of peptide ions by HCD prior to Orbitrap detection avoids this low-mass cut-off, and allows more precise position determination of the modification. Indeed, modifications of C-terminal Lys and Arg were clearly identified due to the presence of low mass *m/z* ions in the tandem mass spectra, specifically y_1^+ ions. Furthermore, immunium ions of modified internal residues were also detected in the low *m/z* range and provided a key element to validate busulfan and BMS modification sites.

The presented strategy included the use of OMS for the identification of mass shifts resulting from busulfan adducts. OMS with spectral libraries followed by statistical analyses offered the possibility to find modifications that did not need to be included in a database, which is of particular interest for the identification of unknown modifications. The combination of the OMS and the statistical analysis served as a screening and filtering process for the detection of the chemical adducts. The *in vitro* incubation of purified proteins and the high number of technical replicates used to compare different conditions strengthened the statistical analysis and the OMS returned a few mass shift candidates with high significance for further validation. In the case of BMS, which only allowed a single alkylation, the OMS returned only a single significant mass shift. For busulfan, which contains two alkylating groups, five mass shifts were found statistically specific. These few mass shifts were relatively easy to handle in terms of structural determination and further validation of tandem mass spectra. These results confirmed that OMS tools and statistical analysis are highly efficient bioinformatics tools for the study of unknown modifications on proteins.

The OMS identified mass shifts resulting from busulfan and BMS adducts by the spectral alignment of modified against unmodified spectra. Mass shifts to which we could assign a chemical structure were validated through manual annotation and visual inspection of tandem mass spectra of modified peptides using strictly defined criteria. In the case of busulfan primary adducts and BMS, inspected spectra were those identified by database search, and for secondary busulfan adducts, inspected spectra were those identified by OMS. Database searches were performed using theoretical values of adduct chemical structures and more restricted tolerances for precursor and fragment ion matches, confirming the identification of peptides carrying adducts. Database searches did not only validate OMS identifications, but also provided additional modified peptides. This was due to the fact that OMS spectral alignment identified a maximum of one mass shift per peptide, which was not the case of identifications resulting from database search. From the point of view of a general application of the workflow, a modified peptide won't be identified during the OMS if the corresponding non-modified peptide is not included in the spectral library. In contrast, database searches of tandem mass spectra were not restricted by this limitation.

Lastly, visual inspection of annotated tandem mass spectra represents an essential step of validation of modified peptides that cannot be entirely replaced by automatic software analysis. However, the application of adapted software solutions to visualize and annotate spectra can dramatically reduce verification time. Using dedicated software, we could visualize hundreds of tandem mass spectra and apply rigorous criteria to validate

modification sites of adducted peptides. For example, busulfan adducted Cys residues were highlighted with a high frequency in the results generated by the OMS, but this modification was confirmed neither on HSA nor on Hb. The OMS modification site assignment is unrestricted, i.e. all sites within a peptide are possible, and the same positioning score can be attributed to amino acids in the vicinity of the true modification site due to noise peaks or missing peaks. Cys residues in HSA and Hb are close to Asp, Glu or His residues and were affected by such miss-assignments. Additionally, in the particular case of Cys modifications, thiol alkylation by busulfan may lead to a dehydroalanine from the loss of hydrogen sulfide by way of a tetrahydrothiophenium intermediate, which corresponds to a loss of 34.0809 Da. Dehydroalanine formation resulting from the direct interaction of busulfan with Cys residues of proteins has been described on glutaredoxin and thioredoxin.[30] Such a modification could not be observed in the present work since the OMS performed only screened for positive mass shifts, i.e. adduct formation leading to an increase in peptide mass. In these situations, validation by database searches and visual inspection of tandem mass spectra becomes mandatory to guarantee an accurate interpretation of the OMS results.

Using the presented workflow, we were able to characterize busulfan-derived adducts on HSA and Hb, providing information about the reactivity of busulfan towards these proteins. Moreover, the workflow allowed the identification of primary and secondary adducts, the formulation of their hypothetical structures and the identification of the modified aminoacids. These data could represent the basis for further studies aiming at investigating busulfan mechanisms of action and toxicity in cellular or animal models. Also, targeting the identified modifications in patient samples with high sensitivity quantitative MS assays would be of interest to determine whether bulsulfan-protein adducts could represent useful biomarker for monitoring the treatment. Finally, the approach used in this study represents a generic methodology for the identification of protein adducts resulting from the interaction with drugs and other chemicals compounds.

ACKNOWLEDGMENT

This study was supported by Swiss National Science Foundation (grant No 32003B_143809) and the Swiss Centre for Applied Human Toxicology.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's website.

REFERENCES

[1] S. M. Rappaport, H. Li, H. Grigoryan, W. E. Funk, E. R. Williams. Adductomics: characterizing exposures to reactive electrophiles. Toxicol. Lett. **2012**, *213*, 83-90.

[2] M. Tornqvist, C. Fred, J. Haglund, H. Helleberg, B. Paulsson, P. Rydberg. Protein adducts: quantitative and qualitative aspects of their formation, analysis and applications. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. **2002**, *778*, 279-308.

[3] D. C. Liebler. Protein damage by reactive electrophiles: targets and consequences. Chem. Res. Toxicol. **2008**, *21*, 117-128.

[4] M. Hassan, B. S. Andersson. Role of pharmacogenetics in busulfan/cyclophosphamide conditioning therapy prior to hematopoietic stem cell transplantation. Pharmacogenomics. **2013**, *14*, 75-87.

[5] S. O. Ciurea, B. S. Andersson. Busulfan in hematopoietic stem cell transplantation. Biol. Blood Marrow Transplant. **2009**, *15*, 523-536.

[6] H. Ehrsson, M. Hassan. Binding of busulfan to plasma proteins and blood cells. J. Pharm. Pharmacol. **1984**, *36*, 694-696.

[7] I. R. Younis, M. Elliott, C. J. Peer, A. J. Cooper, J. T. Pinto, G. W. Konat, M. Kraszpulski, W. P. Petros, P. S. Callery. Dehydroalanine analog of glutathione: an electrophilic busulfan metabolite that binds to human glutathione S-transferase A1-1. J. Pharmacol. Exp. Ther. **2008**, *327*, 770-776.

[8] M. Scian, W. M. Atkins. The busulfan metabolite EdAG irreversibly glutathionylates glutaredoxins. Arch. Biochem. Biophys. **2015**, *583*, 96-104.

[9] F. M. Rubino, M. Pitton, D. Di Fabio, A. Colombi. Toward an "omic" physiopathology of reactive chemicals: thirty years of mass spectrometric study of the protein adducts with endogenous and xenobiotic compounds. Mass Spectrom. Rev. **2009**, *28*, 725-784

[10] X. Yang, M. G. Bartlett. Identification of protein adduction using mass spectrometry: Protein adducts as biomarkers and predictors of toxicity mechanisms. Rapid Commun. Mass Spectrom. **2016**, *30*, 652-64.

[11] E. Ahrne, M. Muller, F. Lisacek. Unrestricted identification of modified proteins using MS/MS. Proteomics. **2010**, *10*, 671-686.

[12] P. Hernandez, R. Gras, J. Frey, R. D. Appel. Popitam: towards new heuristic strategies to improve protein identification from tandem mass spectrometry data. Proteomics. **2003**, *3*, 870-878.
[13] R. Matthiesen, M. B. Trelle, P. Hojrup, J. Bunkenborg, O. N. Jensen. VEMS 3.0: algorithms and computational tools for tandem mass spectrometry based identification of post-translational modifications in proteins. J. Proteome Res. **2005**, *4*, 2338-2347.
[14] D. Tsur, S. Tanner, E. Zandi, V. Bafna, P. A. Pevzner. (2005) Identification of post-translational modifications by blind search of mass spectra. Nat. Biotechnol 23, 1562-1567.
[15] P. A. Pevzner, V. Dancik, C. L. Tang. Mutation-tolerant protein identification by mass spectrometry. J. Comput. Biol. **2000**, *7*, 777-787.
[16] B. C. Searle, S. Dasari, M. Turner, A. P. Reddy, D. Choi, P. A. Wilmarth, A. L. McCormack, L. L. David, S. R. Nagalla. High-throughput identification of proteins and unanticipated sequence modifications using a mass-based alignment algorithm for MS/MS de novo sequencing results. Anal. Chem. **2004**, *76*, 2220-2230.
[17] S. Na, N. Bandeira, E. Paek. Fast multi-blind modification search through tandem mass spectrometry. Mol. Cell. Proteomics **2012**, *11*, M111 010199.
[18] M. M. Savitski, M. L. Nielsen, R. A. Zubarev. ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures. Mol. Cell. Proteomics. **2006**, *5*, 935-948.
[19] J. A. Falkner, J. W. Falkner, A. K. Yocum, P. C. Andrews. A spectral clustering approach to MS/MS identification of post-translational modifications. J. Proteome Res. **2008**, *7*, 4614-4622.
[20] E. Ahrne, F., Nikitin, F. Lisacek, M. Muller. QuickMod: A tool for open modification spectrum library searches. J. Proteome Res. **2011**, *10*, 2913-2921.
[21] D. Ye, Y. Fu, R. X. Sun, H. P. Wang, Z. F. Yuan, H. Chi, S. M. He. Open MS/MS spectral library search to identify unanticipated post-translational modifications and increase spectral identification rate. Bioinformatics. **2010**, *26*, i399-406.
[22] F. Potthast, B. Gerrits, J. Hakkinen, D. Rutishauser, C. H. Ahrens, B. Roschitzki, K. Baerenfaller, R. P. Munton, P. Walther, P. Gehrig, P. Seif, P. H. Seeberger, R. Schlapbach. The Mass Distance Fingerprint: a statistical framework for de novo detection of predominant modifications using high-accuracy mass spectrometry. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. **2007**, *854*, 173-182.

[23] F. Gluck, C. Hoogland, P. Antinori, X. Robin, F. Nikitin, A. Zufferey, C. Pasquarello, V. Fetaud, L. Dayon, M. Muller, F. Lisacek, L. Geiser, D. Hochstrasser, J. C. Sanchez, A. Scherl. EasyProt--an easy-to-use graphical platform for proteomics data analysis. J. Proteomics. **2013**. *79*, 146-160.

[24] Y. J. Kil, C. Becker, W. Sandoval, D. Goldberg, M. Bern. Preview: a program for surveying shotgun proteomics tandem mass spectrometry data. Anal. Chem. **2011**, *83*, 5259-5267.

[25] H. Lam, E. W. Deutsch, J. S. Eddes, J. K. Eng, N. King, S. E. Stein, R. Aebersold. Development and validation of a spectral library searching method for peptide identification from MS/MS. Proteomics. **2007**, *7*, 655-667.

[26] E. Ahrne, Y. Ohta, F. Nikitin, A. Scherl, F. Lisacek, M. Muller. An improved method for the construction of decoy peptide MS/MS spectra suitable for the accurate estimation of false discovery rates. Proteomics. **2011**, *11*, 4085-4095.

[27] H. Lam, E. W. Deutsch, R. Aebersold. Artificial decoy spectral libraries for false discovery rate estimation in spectral library searching in proteomics. J. Proteome Res. **2010**, *9*, 605-610.

[28] M. Hassan, G. Oberg, H Ehrsson, M. Ehrnebo, I. Wallin, B. Smedmyr, T. Tötterman, S. Eksborg, B. Simonsson. Pharmacokinetic and metabolic studies of high-dose busulphan in adults. Eur. J. Clin. Pharmacol. **1989**, *36*, 525-530.

[29] A. Scherl, S. A. Shaffer, G. K. Taylor, P. Hernandez, R. D. Appel, P. A. Binz, D. R. Goodlett. On the benefits of acquiring peptide fragment ions at high measured mass accuracy. J. Am. Soc. Mass Spectrom. **2008**, *19*, 891-901.

[30] M. Scian, M. Guttman, S.D. Bouldin, C.E. Outten, W.M. Atkins. The Myeloablative Drug Busulfan Converts Cysteine to Dehydroalanine and Lanthionine in Redoxins. Biochemistry. **2016**, *55,* 4720-4730.

ABBREVIATIONS

ACN, acetonitrile; BMS, butyl methanesulfonate; CHCA, α-cyano-4-hydroxycinnamic acid; DMSO, dimethyl sulfoxide; FA, formic acid; FDR, false discovery rate; Hb, human hemoglobin; HCD, higher-energy collisional dissociation; HSA, human serum albumin; LC, liquid chromatography; MS, mass spectrometry; MSH, mass shift histogram; OMS, open modification search; PBS, phosphate-buffered saline; SSM, spectrum-spectrum match; TCEP, tris(2-carboxyethyl)phosphine; TEAB, triethyl ammonium bicarbonate; TFA, trifluoroacetic acid.

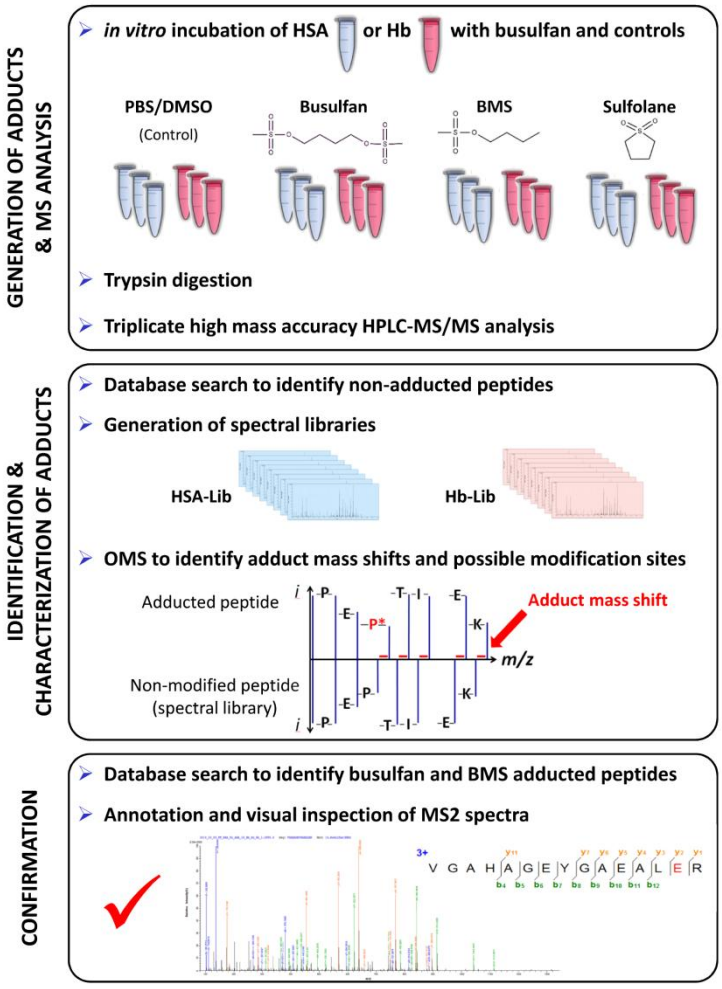

Figure 1

Figure 1. Experimental workflow for the detection of busulfan protein adducts. HSA: human serum albumin, Hb: human hemoglobin, OMS: Open modification search.

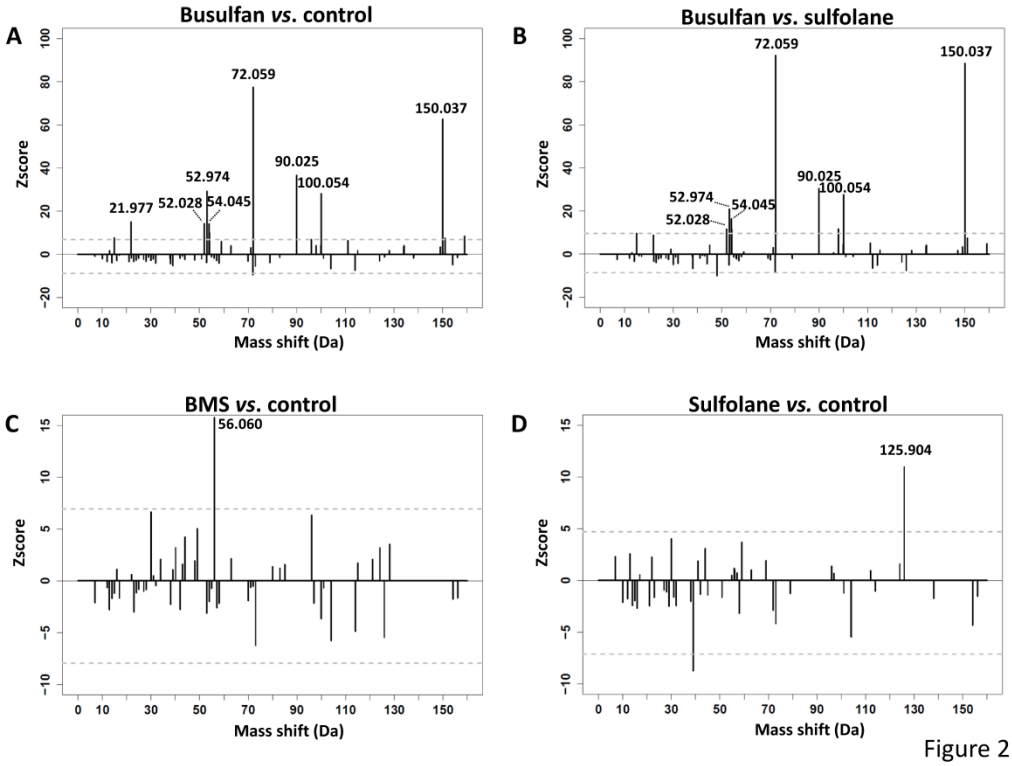

Figure 2

Figure 2. Open modification search of busulfan-derived adducts. Tandem mass spectra from HSA and Hb incubated with busulfan, sulfolane, BMS or PBS/DMSO (control condition) were searched for unrestricted modifications against spectral libraries. Precursor mass shifts from 0 to +160 Da are shown for the combined analysis of HSA and Hb incubated samples. ZScores histograms (Equation 1) for peptide count differences between **A)** busulfan and control incubations, **B)** busulfan and sulfolane incubations, **C)** sulfolane and control incubations, and **D)** BMS and control incubations. Horizontal dashed lines designate ZScore thresholds for statistical significant difference.

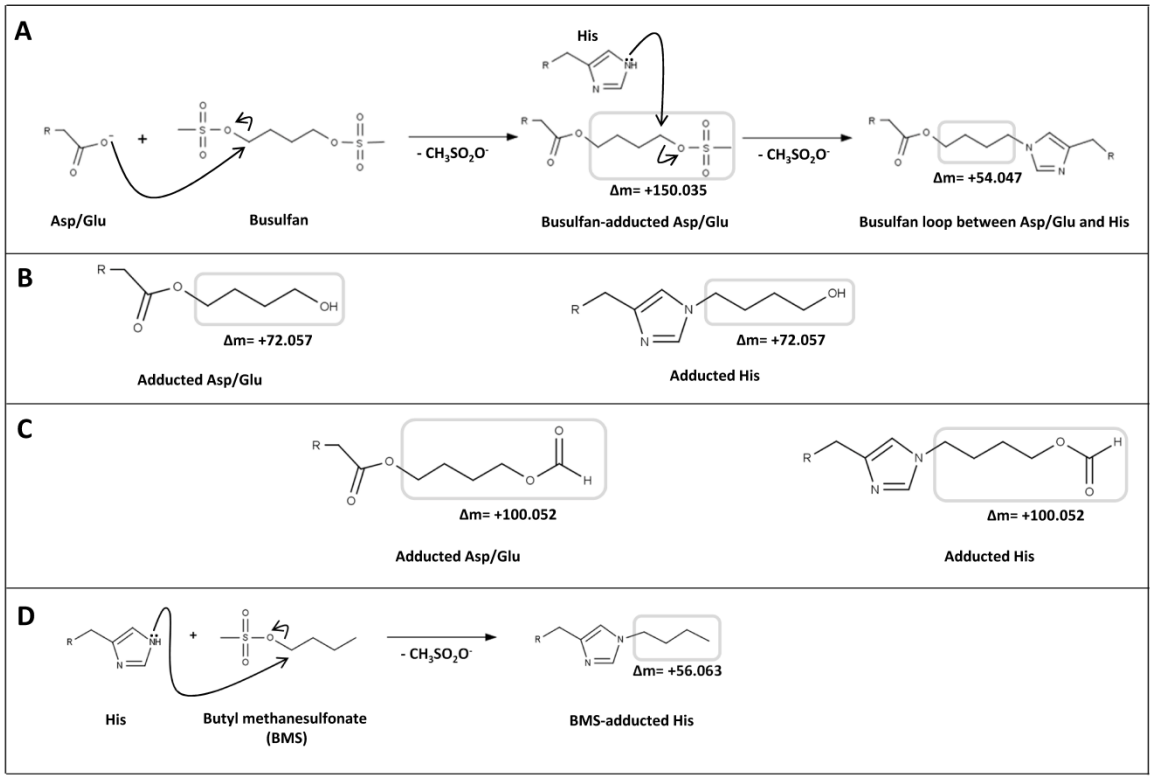

Figure 3

Figure 3. Hypothesis on busulfan and BMS adduct chemical structure. Mass values correspond to the monoisotopic molecular weight for the proposed chemical structure. **A)** Proposed alkylation reactions of busulfan to form primary adducts with a mass shift of +150.035 and intra-peptide cross links (loops) between two amino acids, producing a mass shift of +54.047. **B)** Proposed chemical structure for the busulfan secondary adduct with a mass shift of +72.057. **C)** Proposed chemical structure for the busulfan secondary adduct with a mass shift of +100.052. **D)** Propose alkylation reaction of BMS to form adducts with a mass shift of +56.063.

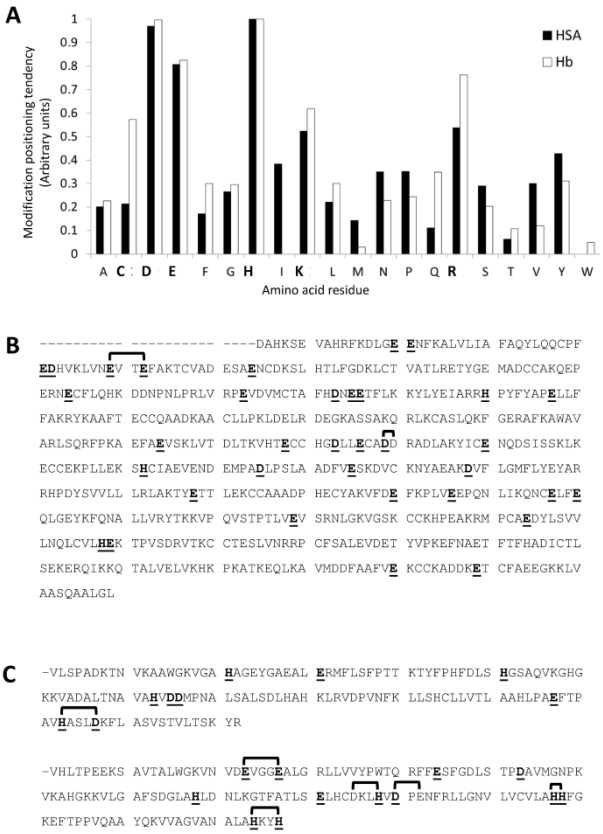

Figure 4

Figure 4. Primary busulfan +150.037 Da adduct positioning on HSA and Hb. **A)** Busulfan adduct (+150.037 Da) positioning tendency for each amino acid residue on HSA and Hb. Residues presenting higher tendency are highlighted in bold. Normalized values of corrected tendencies W(aa, ΔM) (Equation 6) are shown. **B)** Busulfan adducted (+150.035 Da) sites that were successfully validated after manual inspection of tandem mass spectra (highlighted in bold and underlined) are shown for HSA and **C)** Hb protein sequences. Possible positioning of busulfan loops (+54.047 Da) after evaluation of tandem mass spectra are indicated with horizontal brackets.

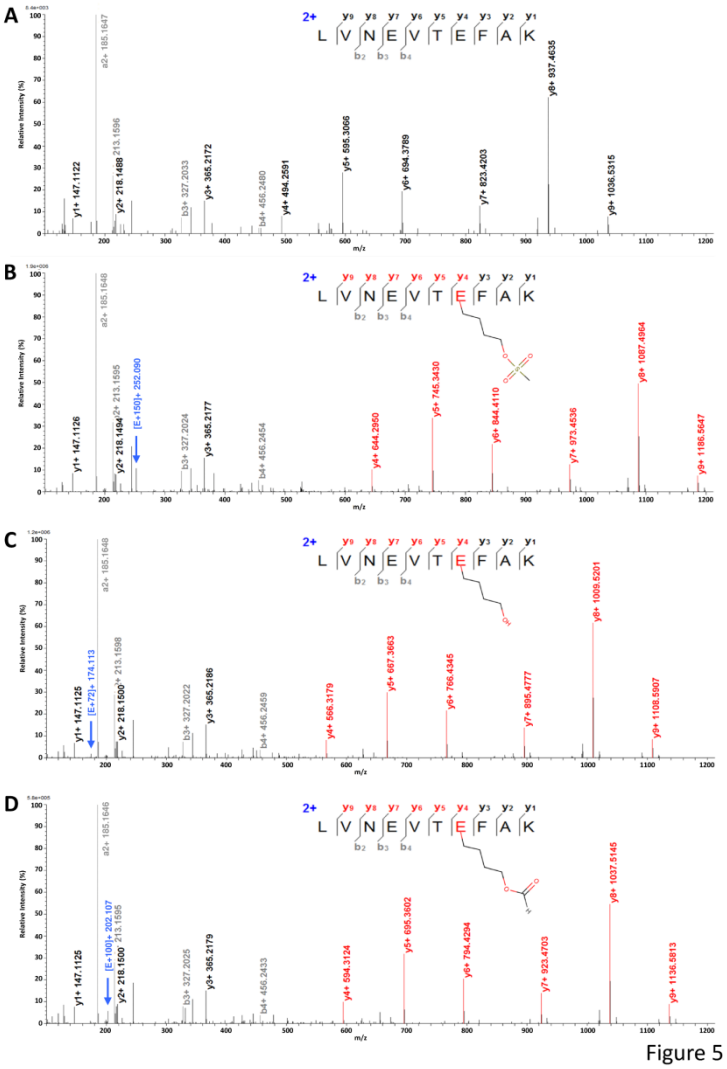

Figure 5

Figure 5. High measured mass accuracy identification of a modified peptide. Manual annotation and visual inspection of spectra of HSA peptide LVNEVTEFAK. Fragment ions are annotated for the peptide when carrying **A)** no modification, **B)** the +150.035 busulfan primary adduct, **C)** the +72.057 busulfan secondary adduct, and **D)** the +100.052 busulfan secondary adduct. From the fragment y_4^+ (in red), mass shifts are observed when Asp-7 carries the modification. Immonium ions of modified Asp are indicated with a blue arrow.

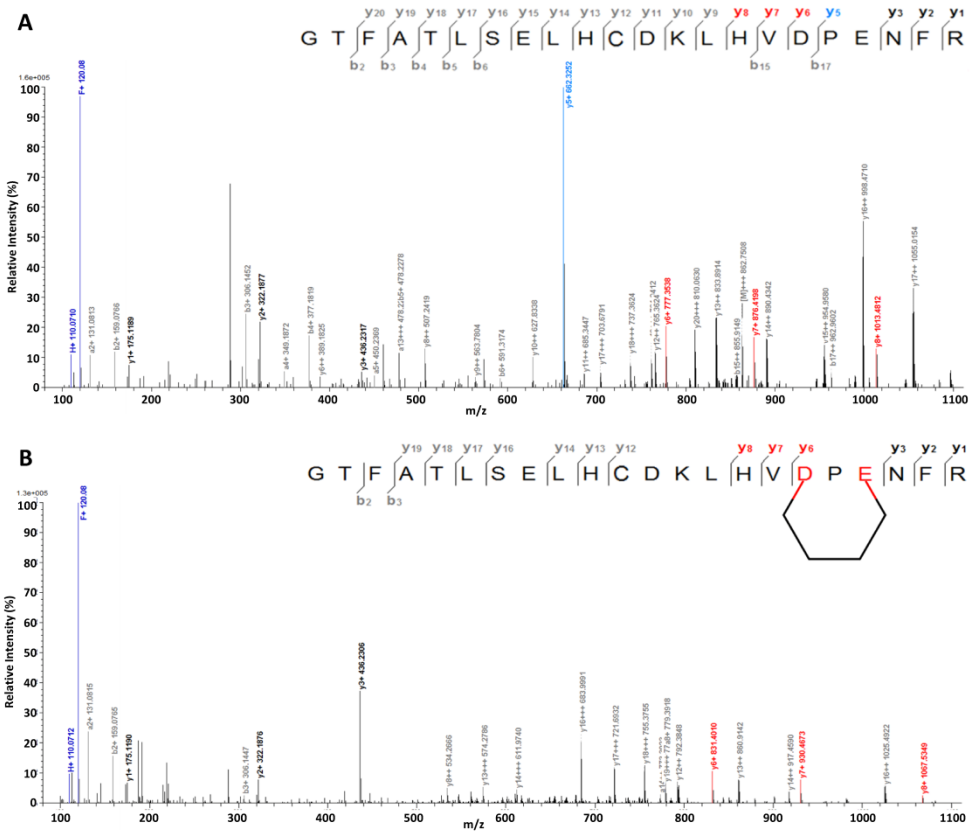

Figure 6

Figure 6. Evaluation of intra-peptide cross-links. Tandem mass spectra of Hb beta chain peptide GTFATLSELHC$_{CAM}$DKLHVDPENFR when carrying **A)** no modification, and **B)** the +54.047 busulfan loop. Fragments y_1^+, y_2^+ and y_3^+ (in black) are common in both spectra. The prominent y_5^+ ion (in blue) observed for the non-modified peptide, is not detected in the presence of the loop. The mass shift of +54.047 Da is observed from the fragment y_6^+ (in red).

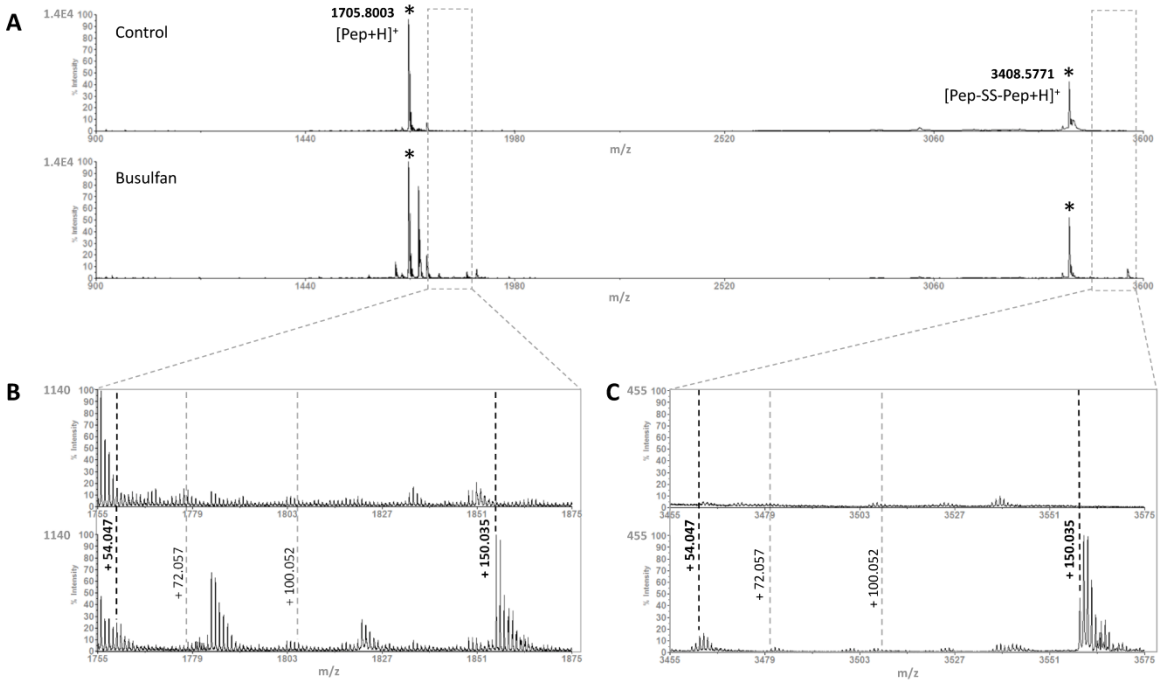

Figure 7

Figure 7. Modification of a synthetic peptide by busulfan. Positive ion mode MALDI-TOF MS spectra of the synthetic peptide AQYLQQCPFEDHVK incubated in PBS/DMSO (control condition; upper spectrum), and with busulfan (lower spectrum). **A)** *m/z* values are given for the non-modified and the disulfide-linked peptide (peaks indicated with an asterisk). **B)** Spectra acquired at a narrower *m/z* region to observe peaks of the busulfan-adducted linear peptide (1759.8472 for +54.047; 1777.8578 for +72.057; 1805.8527 for +100.052, and 1855.8354 for +150.035) and **C)** busulfan-adducted disulfide-linked peptide (3462.6241 for +54.047; 3480.6346 for +72.057; 3508.6296 for +100.052, and 3558.6122 for +150.035).