

# SCIENTIFIC REPORTS



OPEN

## Cross Reactive Material 197 glycoconjugate vaccines contain privileged conjugation sites

Uwe Möglinger<sup>1,2</sup>, Anja Resemann<sup>3</sup>, Christopher E. Martin<sup>1,2</sup>, Sharavathi Parameswarappa<sup>1</sup>, Subramanian Govindan<sup>1,†</sup>, Eike-Christian Wamhoff<sup>1,2</sup>, Felix Broecker<sup>1,2</sup>, Detlev Suckau<sup>3</sup>, Clancy Lebev Pereira<sup>1</sup>, Chakkumkal Anish<sup>1,‡</sup>, Peter H. Seeberger<sup>1,2</sup> & Daniel Kolarich<sup>1</sup>

Received: 12 August 2015  
Accepted: 05 January 2016  
Published: 04 February 2016

Production of glycoconjugate vaccines involves the chemical conjugation of glycans to an immunogenic carrier protein such as Cross-Reactive-Material-197 (CRM<sub>197</sub>). Instead of using glycans from natural sources recent vaccine development has been focusing on the use of synthetically defined minimal epitopes. While the glycan is structurally defined, the attachment sites on the protein are not. Fully characterized conjugates and batch-to-batch comparisons are the key to eventually create completely defined conjugates. A variety of glycoconjugates consisting of CRM<sub>197</sub> and synthetic oligosaccharide epitopes was characterised using mass spectrometry techniques. The primary structure was assessed by combining intact protein MALDI-TOF-MS, LC-MALDI-TOF-MS middle-down and LC-ESI-MS bottom-up approaches. The middle-down approach on CNBr cleaved glycopeptides provided almost complete sequence coverage, facilitating rapid batch-to-batch comparisons, resolving glycan loading and identification of side products. Regions close to the N- and C-termini were most efficiently conjugated.

Pathogens such as bacteria and parasites exhibit specific glycan structures on their surface<sup>1</sup>. These accessible biomolecules are targets for pathogen specific vaccination. Glycoconjugate vaccines that combine a glycan antigen with a carrier protein have for the most part replaced polysaccharide vaccines that fail to induce a strong immune response<sup>2</sup> and do not result in B-cell memory in infants under the age of two<sup>3,4</sup>.

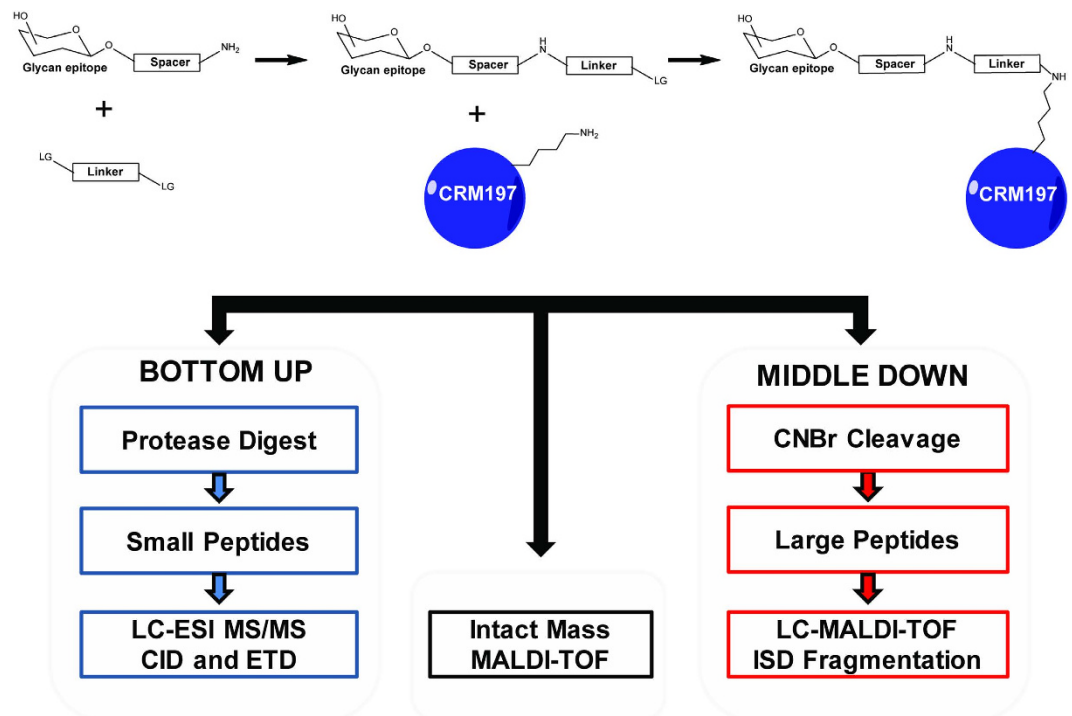
Successful and broadly marketed carbohydrate conjugate vaccines are based on just a few FDA-approved carrier proteins. First-generation carrier proteins such as diphtheria toxin and tetanus toxin require detoxification with formaldehyde eliminating part of the lysine residues needed for glycan attachment, thereby limiting conjugation efficacy<sup>5</sup>. One of the most widely used and highly effective carrier protein is Cross-Reactive-Material-197 (CRM<sub>197</sub>), a mutant version of the diphtheria toxin, where the single amino acid exchange of a glycine in position 52 to a glutamic acid renders the protein non-toxic<sup>6</sup>. Conjugate vaccines such as HibTITER<sup>®</sup> (*Haemophilus influenzae* type b associated diseases), Prevnar<sup>™</sup> (pneumococcal diseases), and Menveo<sup>®</sup> (meningococcal diseases)<sup>7</sup> use CRM<sub>197</sub> as carrier protein.

Almost all currently marketed glycoconjugate vaccines contain isolated, large and heterogeneous glycans<sup>8,9</sup>. These vaccines are very successful, but technically, the isolation from natural sources is challenging. Not all pathogenic organisms can be grown to produce antigenic carbohydrates *in vitro*. Culturing and processing large amounts of human pathogens safely is complex and cost intensive. Downstream processing inevitably produces heterogeneous glycan mixtures of various lengths and modifications that are difficult to characterise before and after conjugation. In some cases, glycan processing can affect epitope integrity<sup>10,11</sup>. Regular in-depth molecular level characterisation of the final product that will be used for immunisation is thus not possible for these biologics.

The chemical synthesis of the immunogenic glyco-epitopes<sup>12</sup> holds many advantages but only one semi-synthetic glycoconjugate vaccine (Quimi Hib) is currently on the market<sup>13</sup>. Synthetic, minimal

<sup>1</sup>Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, 14424 Potsdam, Germany.

<sup>2</sup>Institute of Chemistry and Biochemistry, Freie Universität Berlin, Germany. <sup>3</sup>Bruker, Bremen, Germany. <sup>†</sup>Present address: a) Department of chemistry, Indian Institute of Technology Tirupati, Tirupati-Renigunta Road, Tirupati, Andhra Pradesh, India 517506 and b) Department of Civil Engineering, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India 600036. <sup>‡</sup>Present address: Bacterial Vaccine Discovery Team, Janssen Pharmaceuticals of Johnson & Johnson, Bioscience Park Leiden, Zernikedreef 9, 2333 CK Leiden, Netherlands. Correspondence and requests for materials should be addressed to D.K. (email: daniel.kolarich@mpikg.mpg.de)



**Figure 1. Schematic overview of the three orthogonal approaches employed for the in-depth characterisation of CRM<sub>197</sub> conjugated with defined synthetic minimal glycoepitope vaccine candidates.**

Step 1: The glycan part is synthesised with a spacer carrying an amine group. Step 2: addition of the linker molecule with two leaving groups (LG). Step 3: addition of the glycan linker construct to the carrier protein and subsequent conjugation to a primary amine (e.g. lysine residue).

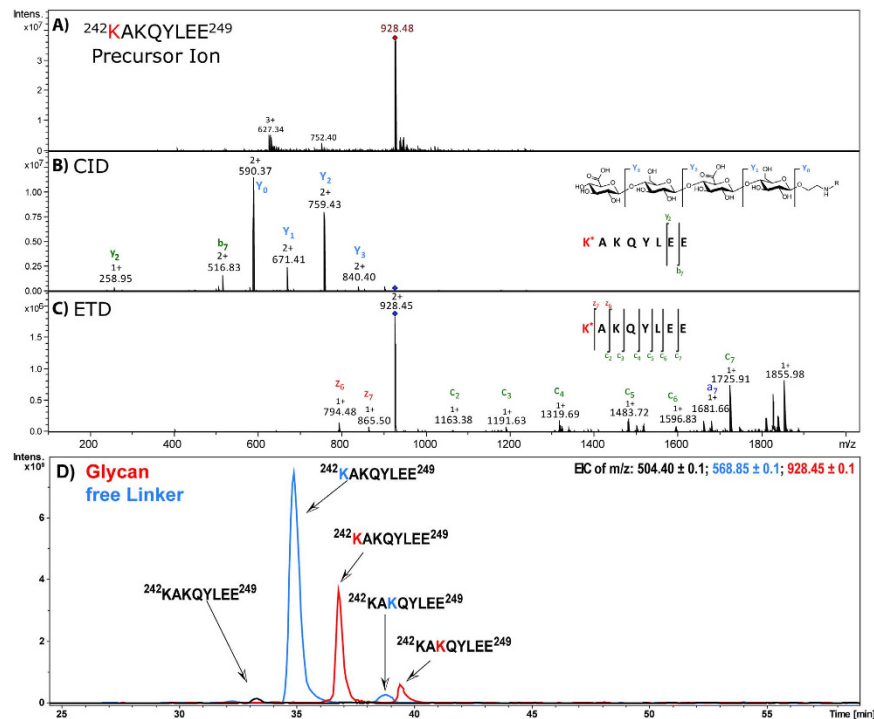
immunogenic glyco-epitopes are structurally well-defined and homogenous<sup>14</sup>. The synthetic approach enables the placement of a spacer carrying a unique functional group such as an amine to facilitate protein attachment at a single site without further activation.

Typically, glycans are coupled to the primary amine side chains of lysine residues and to the protein N-terminus via a linker molecule. Alternatively, sulfhydryl groups in cysteine or carboxyl side chains can serve as sites for attachment. Prior to coupling, non-synthetic glycans need to be activated<sup>15</sup> either arbitrarily on hydroxyl or carboxyl groups<sup>16</sup>, leading to attachment at various points or via a single amine group, introduced by reductive amination, at the reducing end<sup>17</sup>. The choice of the linker is crucial since it influences vaccine properties<sup>18</sup> and may be immunogenic itself<sup>19</sup>. In addition, coupling efficiencies have to be considered. Conjugation of a synthetic glyco-epitope to a carrier protein such as CRM<sub>197</sub> that contains 40 primary amine side chains will result in heterogeneous glycoconjugates since some amino acid side chains more readily engage in conjugation reactions than others<sup>20</sup>. Since variations in epitope loading influence the effectiveness of vaccines<sup>21,22</sup> understanding which side chains preferentially react with glycans represents an important step towards producing better defined, more effective and safer conjugate vaccines. Quality control and batch-to-batch comparison relies on a combination of physicochemical, spectroscopic and spectrometric methods such as nuclear magnetic resonance (NMR) or mass spectrometry (MS)<sup>23–25</sup>. Glycan attachment to the carrier protein is typically monitored by SDS-PAGE and MALDI-TOF-MS<sup>26</sup>. These methods provide information about average loading and cannot detect any side reactions or alterations occurring on the glycan epitopes. New methods are necessary for quality assurance of conjugate vaccines and to define conjugation sites on the protein.

We established a mass spectrometry-based assay for the in-depth characterisation of glycoconjugates obtained by coupling synthetic glycans to CRM<sub>197</sub> (Fig. 1). An integrated new middle-down LC-MALDI-TOF-MS approach simplified batch-to-batch comparison, and enabled relative quantification of regional conjugation efficacy. In this study we evaluated the influence of glycan size and amino acid microenvironment on conjugation efficiencies, assessed glycan integrity and determined which amino acid sites were preferably conjugated. Two regions close to the N- and C-termini of the protein engage primarily in the chemical conjugation process and exhibited the highest loading ratios. In addition to accessibility, our data clearly illustrates that the local amino acid environment significantly influences which lysine residues will be modified.

## Results and Discussion

**Protease based bottom up strategies for the evaluation of CRM<sub>197</sub> glycoconjugate vaccine candidates.** Assuring glycan integrity of the glycoconjugate is important after the chemical conjugation step and to evaluate the storage stability. We studied this property by a classical (glyco)proteomics bottom up approach, which provided qualitative data on individual, repeatedly conjugated sites in CRM<sub>197</sub>.



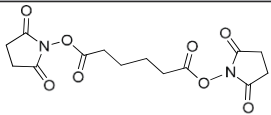
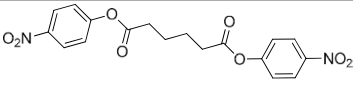
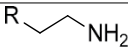
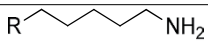
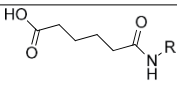
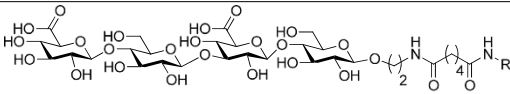
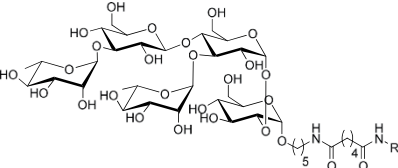
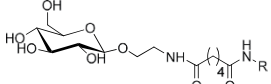
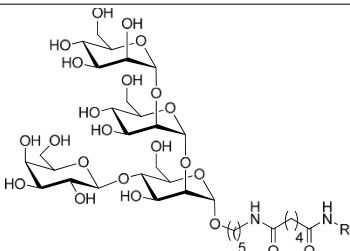
**Figure 2. Conjugation site determination by tandem mass spectrometry.** (A) After Glu-C digestion a doubly charged precursor ion corresponding to glycopeptide  $^{242}\text{KAKQYLEE}^{249}$  carrying an ST3 epitope was detected and selected for fragmentation by CID and ETD. The red labeled K indicates the site of conjugation identified by tandem MS. (B) CID fragmentation resulted in a prominent Y-ion series indicating the loss of glucuronic acid and glucose, but no significant peptide backbone fragments. (C) ETD fragmentation of the peptide backbone confirmed peptide identity as well as K242 as the site of conjugation. (D) Extracted ion chromatogram of doubly charged ions of unconjugated and conjugated peptide 242–249 showing the separation of isobaric conjugate products by C18 LC [overlay of  $m/z = 504.40$  (unconjugated peptide, black line),  $568.85$  (peptide with conjugated linker, blue line) and  $928.45$  (peptide with conjugated glycan, red line)]. Despite the proximity of the two potential conjugation sites, K242 was more effectively conjugated.

For unmodified CRM<sub>197</sub> the use of trypsin resulted in an average sequence coverage of 50–67%. Using Glu-C usually 50% of the entire sequence could be identified (see Supplementary Fig. S2). The highest coverage ( $\approx 70$ –85%) was obtained using a sequential digestion employing Glu-C followed by trypsin or Asp-N followed by Glu-C. Nevertheless, the results varied considerably from digest to digest indicating the limitation of such a bottom-up approach for routine analyses of chemically glycosylated neoglycoproteins.

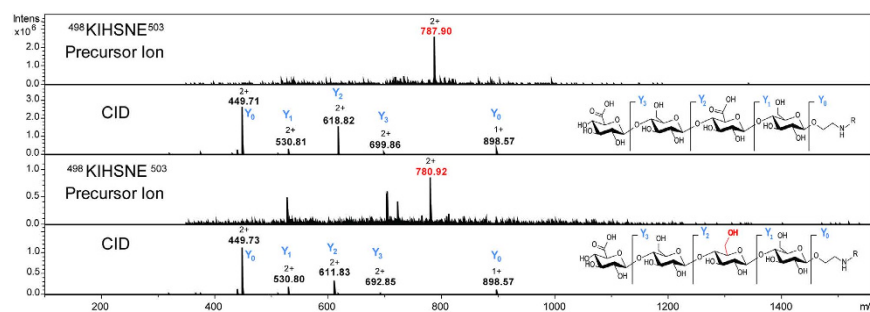
Trypsin did not cleave at modified lysine residues resulting in conjugated peptides carrying at least one missed cleavage site<sup>27</sup>. As a result, the C-terminal lysine in any tryptic peptide had to be unmodified in order to be cleaved. This facilitated the assignment of the actual conjugation site in particular as for most glycopeptides only one lysine residue remained as a possible site of conjugation. On the other hand, the modification induced missed cleavage sites lead to an increased heterogeneity of peptide fragments and consequently reduced signal intensities.

The problem could be avoided by the use of alternative proteases not cleaving at lysine residues (Glu-C, Asp-N). However, these proteases produced peptides with multiple lysine residues. In order to elucidate the conjugation sites it was crucial to acquire sufficient sequence information from the MS/MS experiment. Besides the qualitative data an in-depth conjugation site analysis also required quantitative information on site occupancy. The protease assisted bottom-up approach, however, primarily provided qualitative data since a direct mass spectrometric quantitative comparison between different peptide backbones is impossible. Despite the fact that in some cases the conjugation reactivity of individual lysine residues located on the same peptide could be determined (Fig. 2D), a protease based approach appeared to be unsuitable for any broad quantitative comparison of region/site specific conjugation efficiencies between different conjugate vaccines.

**CRM<sub>197</sub> neoglycopeptides exhibit unusual CID and ETD fragmentation patterns.** Naturally occurring glycopeptides usually exhibit strong and specific oxonium ions when subjected to collision induced dissociation (CID)<sup>28,29</sup>. This feature is frequently applied to quickly filter glycopeptide spectra from the bulk of MS/MS data. However, CRM<sub>197</sub> neoglycopeptides carrying the glyco-epitope constructs used in this study (Table 1) produced a different fragmentation pattern which exhibited an extremely prominent Y-ion series when subjected to CID while showing almost no or very low intensity oxonium ions (see Fig. 2B and Supplementary Fig. S3). Assuming that the majority of protons are associated with the peptide rather than with the glycan moiety, it appears that the linker construct prevents the protons from effectively migrating to the glycan fragments,

LINKER AND SPACER MOLECULES				
Structure	Abbreviation	Composition	Description	
	DSA	C <sub>14</sub> H <sub>16</sub> N <sub>2</sub> O <sub>8</sub>	di-succinimido-adipate linker <i>Leaving group: succinimide</i>	
	PNP	C <sub>18</sub> H <sub>16</sub> N <sub>2</sub> O <sub>8</sub>	Adipic acid <i>p</i> -nitro-phenol linker <i>Leaving group: para-nitrophenol</i>	
	C2	C <sub>2</sub> H <sub>6</sub> N	2 carbon unit spacer	
	C5	C <sub>5</sub> H <sub>12</sub> N	5 carbon unit spacer	
Linker side product (depicted as present after conjugation)				
Structure	Abbreviation	Composition	$\Delta m/z$	Description
	L	C <sub>6</sub> H <sub>10</sub> O <sub>3</sub>	128.06	Adipic acid/free linker #
CONJUGATES (depicted as present after conjugation)				
	ST3	(GlcA) <sub>2</sub> (Glc) <sub>2</sub> -C2- DSA	847.80	Epitope from <i>Streptococcus pneumoniae</i> serotype 3
	PS1	(Rha) <sub>2</sub> (Glc) <sub>3</sub> -C5-DSA	991.42	Epitope from <i>Clostridium difficile</i>
	GLC	Glc- C2- PNP	333.14	Glucose test glycan
	LPG	(Man) <sub>3</sub> (Gal)-C5-DSA	861.36	Lipophosphoglycan capping oligosaccharide <i>Leishmania infantum</i>

**Table 1. Overview on linker, spacer and glycan epitope constructs used in this study.**  $\Delta m/z$ : monoisotopic mass addition upon conjugation to a primary amine. Glc = glucose, Rha = rhamnose, GlcA = glucuronic acid, Hep = heptose, Man = mannose, Gal = galactose; # Conjugation of linker molecule and subsequent hydrolysis of the leaving groups results in the same product for DSA linker and PNP linker.



**Figure 3.** MS spectra of the doubly charged precursor ions  $m/z = 787.90$  and  $m/z = 780.92$  as well as their respective CID MS/MS spectra. Using CID fragmentation even minor amounts of synthesis byproducts containing glucose instead of a glucuronic acid at the second position could be identified after being conjugated to CRM<sub>197</sub>.

resulting mostly in the detection of neutral loss fragments. This observed phenomenon appeared to be independent of the length of the carbon spacer. Despite the lack of oxonium ions the observed Y-ions showed a characteristic serial neutral loss of each monosaccharide in the glycan chain. This feature was used to distinguish MS/MS spectra of conjugated peptides from unconjugated ones. CID analyses also enabled confirmation of the glyco-epitope integrity after conjugation to the protein.

Electron transfer dissociation (ETD) of the peptide backbone<sup>30</sup> enabled the acquisition of sufficient data to confirm peptide identities and map the site(s) of conjugation on the majority of the detected glycopeptides (Fig. 2C). Nevertheless, the ETD MS/MS spectra derived from the CRM<sub>197</sub> neoglycopeptides showed unusual fragmentation patterns. Z-ions were highly underrepresented, especially when obtained from tryptic glycopeptides. The c-ions series continued at most until the amino acid neighbouring the conjugated lysine, and z-ions containing this residue could generally not be detected.

**Glyco-Epitope screening by nanoLC-ESI-MS/MS.** Liquid chromatography enabled separation of isobaric modified peptides carrying the conjugate at different lysine residues. In addition to the glycoconjugates we detected conjugates of adipic acid formed by residual linker reacting with the lysine residues of CRM<sub>197</sub>. This is a common problem since the di-succinimido-adipate (DSA) linker hydrolyses very easily hampering an extensive purification. These adipic acid conjugation artefacts are referred to as "free linker" in this study. Addition of the glycan construct as well as of the free linker resulted in an increased retention time with the glycan-linker construct showing the biggest impact (Fig. 2D).

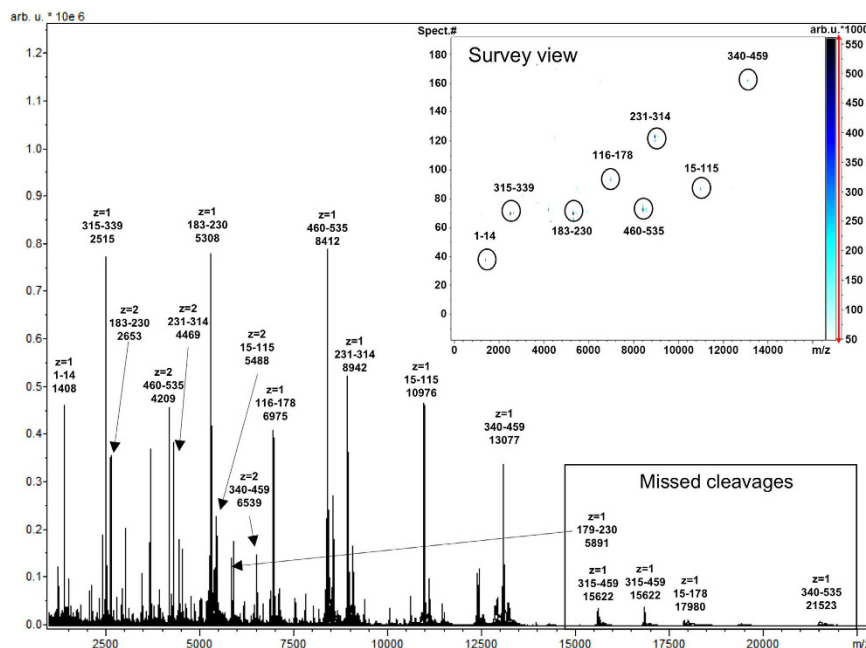
Minor unintended structural features or contaminations might not always be picked up prior to conjugation. The multiple dimensions of nano LC-ESI-MS/MS analysis detected low abundant features such as a peptide carrying a minor side product of the chemical synthesis of ST3. CID fragmentation revealed that one of the two glucuronic acids of the tetrasaccharide was substituted by a glucose resulting in a mass shift of  $-14$  Da from the expected value (Fig. 3). This technique also allows detection and monitoring of any mass modification due to degradation.

Overall, the bottom up approach provided useful data to confirm conjugate integrity, to enable conjugation site mapping and to monitor site-specific conjugation preferences. Nevertheless, issues such as incomplete sequence coverage or the uncontrollable heterogeneity of proteolytic (glyco)peptide products<sup>31</sup> made it less suited for ensuring batch-to-batch consistency of chemically glycosylated CRM<sub>197</sub> in a pharmaceutical context. Therefore, an orthogonal LC-MALDI-TOF MS based approach was developed with the intention to overcome these shortcomings and enable fast batch-to-batch comparisons while still providing more detailed data on glycosylation occupancy compared to intact protein analyses.

**Improved CRM<sub>197</sub>-glycoconjugate sequence coverage using chemical cleavage combined with LC-MALDI-TOF-MS.** We developed and applied an LC-MALDI-TOF-MS middle-down proteomics approach based on CNBr mediated cleavage to increase sequence coverage and simplify batch-to-batch comparisons while achieving a better general overview on the entire protein sequence detecting any intended and unintended modifications. CNBr mediated protein cleavage results in larger glycopeptides that should compensate or significantly reduce ionisation suppression effects. Glycosylation was shown to reduce ionisation efficiency of tryptic N-glycopeptides, however ionisation efficiency appeared to increase when the peptide to glycan mass ratio was shifted towards the peptides<sup>32</sup>.

CNBr cleavage occurs on the C-terminal side of methionine residues<sup>33</sup>. In the case of CRM<sub>197</sub> CNBr cleavage results in nine distinct (glyco)peptides for subsequent LC-MALDI-TOF-MS analysis (see Supplementary Table S4). Eight out of nine peptides could be reproducibly detected, corresponding to a sequence coverage of  $> 99\%$  (Fig. 4). The missing four amino acid peptide (<sup>179</sup>YEYM<sup>182</sup>) exhibited a mass  $< 600$  Da, which is below the optimal operation range of MALDI-TOF-MS. This particular peptide is one of two that do not carry any potential conjugation sites such as lysine residues or the protein's N-terminus, ultimately leaving only seven peptides that needed to be considered in a simplified and comprehensive batch-to-batch comparison.

Due to the large size of the peptides C8- as well as C4-reversed phase LC were tested for (glyco)peptide separation. The C4-chromatography provided better peak shapes and separation reproducibility and was therefore



**Figure 4.** Overlay of all LC-MALDI-MS spectra acquired during the LC run of unconjugated CRM<sub>197</sub>. For each identified peak charge, peptide position and average mass are indicated. The major peaks are all corresponding to the singly charged CNBr peptides without any missed cleavages. All of the eight expected peptides were detected. Doubly charged species were also found for larger peptides as well as low abundant signals corresponding to peptides with missed cleavage sites. The survey view illustrates retention time differences of the various peptides.

used for the majority of analyses. As expected, the most intense  $m/z$  signals corresponded to singly charged ions while larger peptides ( $m/z > 5000$  Da) were detected as minor signals of doubly charged species as well (Fig. 4).

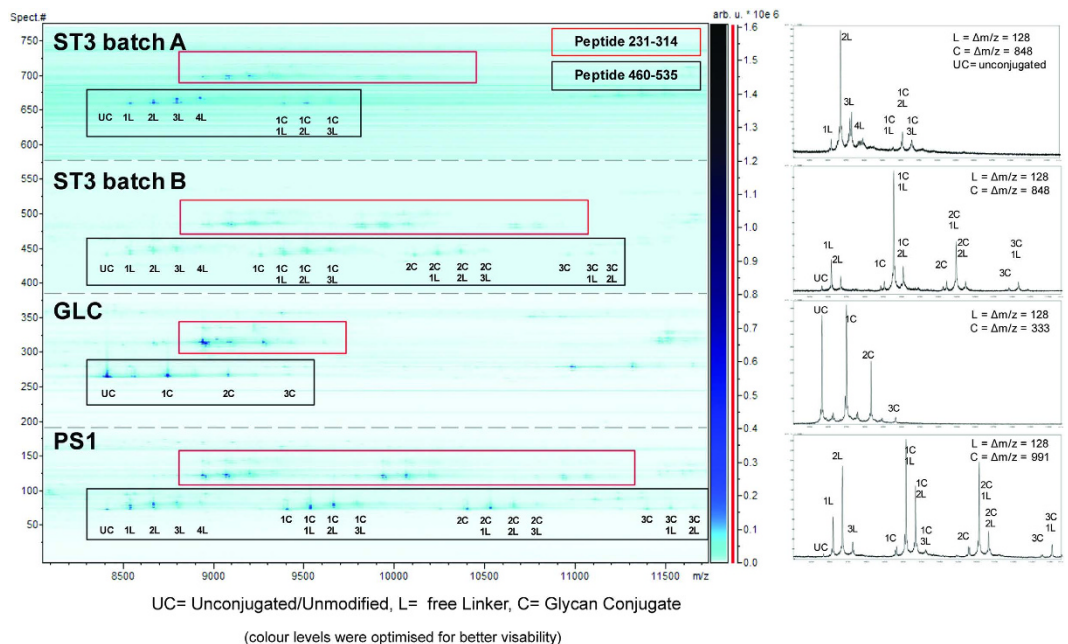
Two batches of ST3 conjugate were analysed in a first run. Depending on their size, the peptides usually eluted during approx. 1–2 minutes. ST3 conjugated glycopeptides were detected by their specific mass shift of + 847.80 Da, corresponding to the average mass of one glycoconjugate unit. A mass shift of + 128.06 Da indicated the presence of free linker.

The presence of conjugated adipic acid on a CNBr peptide resulted in a retention time shift of around + 30 seconds whereas the presence of a glycoconjugate barely had any effect. In contrast, in the bottom up approach the separation of (glyco)peptides using C18 chromatography, free linker and especially glycoconjugates showed a much larger influence on retention time (Fig. 5). The survey view of the WARP-LC software allows for a simultaneous assay like representation of retention time,  $m/z$  and intensity. This provided a clear overview of the conjugation status of each CNBr cleaved (glyco)peptide. Increasing amounts of glycoconjugates as well as the addition of several adipic acid units could be easily depicted as characteristic peak patterns (Fig. 5).

**SurveyViewer assisted comparison of batch-to-batch variances.** Analyses of several batches of conjugates demonstrated that differences in loading and the extent of side reactions could be quickly detected in a protein region specific manner and evaluated using this approach. Major differences in free linker loading and consequently in glycan loading could be observed between the different conjugates (Fig. 5). Batch B showed one additional glycan conjugation in each of the peptides whereas batch A exhibited an increased free linker loading. More detailed information was obtained when relative quantification of the respective (glyco)peptides was applied (Fig. 6). The degree of conjugation with the free linker differed considerably in the two batches. Batch A generally displayed an elevated loading. This increased occupation of lysine residues with free linker molecules resulted in a reduced number of available sites for glycan conjugation, explaining the observed lower glycan loading.

To address the issue of adipic acid conjugation onto the protein and further benchmark the assay a new set of CRM<sub>197</sub> glycoconjugate was prepared using the *p*-nitrophenol (PNP) activation method<sup>34</sup> with GLC as a test epitope (Table 1, Fig. 5). Compared to DSA, the PNP-linker is less prone to hydrolysis allowing an extensive purification after activation. The analysis showed that glycan loading was generally lower, however only conjugated glycans and no free linker molecules were detected in both batches. Our findings demonstrate the potential of the assay to reliably detect and differentiate intended and unintended modifications occurring on CRM<sub>197</sub>.

**Evaluation of CNBr cleavage conditions on glycan conjugate integrity.** CNBr cleavage requires highly acidic conditions (50% TFA). Even though the sample preparation is performed at low temperature (4 °C) the acidity could potentially affect glycan integrity. Therefore, we evaluated the stability of the glycan epitopes used in this study (Table 1). As exemplified on the ST3, PS1 and GLC-CRM<sub>197</sub> conjugates no degradation



**Figure 5. Comparison of conjugation efficiency on peptides 231–314 and 460–535 shown for four samples carrying different glycoepitopes.** In the SurveyViewer (retention time vs.  $m/z$ ) of a (C4) LC-MALDI-MS experiment (left) intensities of the detected signals are plotted vs. their retention times and signal intensities in a gel-like representation, providing a quick overview on the number of conjugates present on each CNBr peptide (right). Spectra obtained for peptide 460–535 at a given retention time. Samples conjugated with the DSA-linker appeared in clusters with varying amounts of linker and glycoconjugates. Linker addition induced a shift of + 128 Da and minor shifts in retention time. A single glycoconjugate addition increased the mass by 848 Da (ST3), 333 Da (GLC), or 991 Da (PS1), respectively. Major differences in loading of free linker and glycans were easily detectable by this approach. ST3 batch-to-batch comparison revealed batch A containing up to one glycan conjugated to peptide 231–314 and up to two conjugated to peptide 460–535, whereas the loading was increased in batch B with up to two glycoconjugates attached to peptide 231–314 and up to three to peptide 460–535, respectively. The lower glycan loading in batch A correlated with stronger signals for linker conjugation. GLC was conjugated via a PNP-linker and showed no detectable free linker additions and therefore no clusters as seen for the DSA-conjugated peptides.

**Batch to Batch Comparison of ST3**

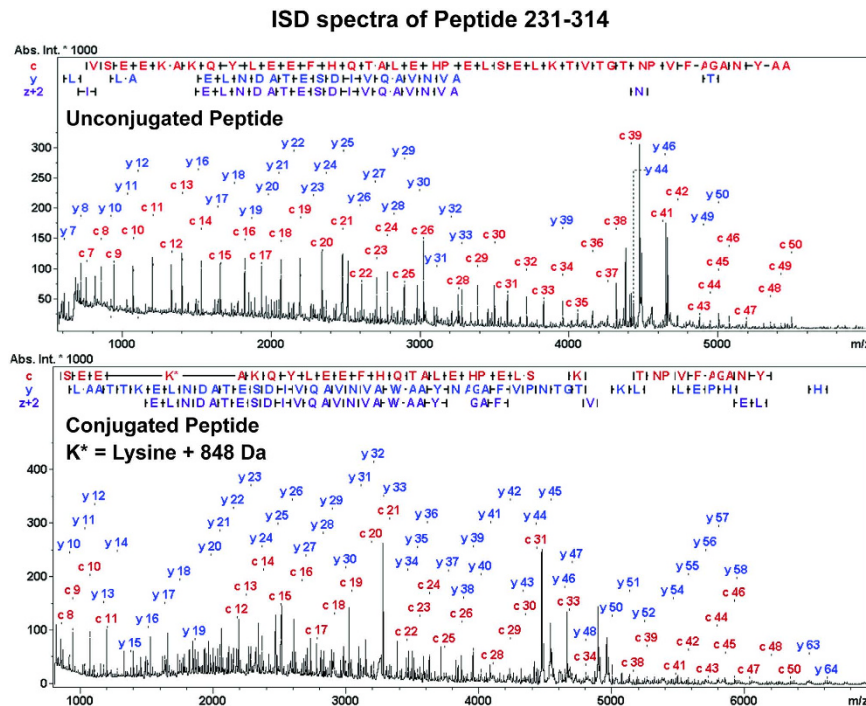
Batch	Peptide	UC	1L	2L	3L	4L	1C	1C 1L	1C 2L	1C 3L	2C	2C 1L	2C 2L	2C 3L	3C	3C 1L	4C	Relative Abundance
A	1-14	53	47															%
B		54	31	2			13											
A	15-115	1	7	19	22	14	1	3	4		8	8	4	5	2	2	3	%
B		4	8	6	5	4	2	10	13	8	4	8	6	2	4	6	3	
A	116-178	46	38	7			3	2	2	3								%
B		41	26	8			18	5			1							
A	183-230	14	25	28	14		9	8	1									%
B		11	24	8			23	22	6		6							
A	231-314	7	24	25	25	14	2	2	1									%
B		6	27	17	5		15	17	6		5	3						
A	340-459	19	49	32			3											%
B		38	30	14			11	7										
A	460-535	1	10	24	30	16	1	8	10	6								%
B		1	6	13	8	3	4	19	17	7	4	9	5		1	1		

UC= Unconjugated/Unmodified, L= free Linker, C= Glycan Conjugate

**Figure 6. Relative quantitation of the seven lysine containing CNBr peptides derived from two ST3 batches.** Each peptide section was quantified individually. The sum of the area under the curve from all detected signals obtained for a given peptide (with and without modifications) was set to 100%. The percentages presented in the graph are rounded to the full digit. Differences in glycan and free linker loading could be directly compared for each peptide. The results clearly demonstrated that glycan loading was more effective in batch B over the entire protein sequence. Inversely batch A exhibited higher loading of free linker molecules.

products could be detected (Fig. 5). Our data indicate that most glycan structures appear to tolerate the applied cleavage conditions.

**Conjugation site identification on CNBr glycopeptides using ISD.** To determine site-specific attachment information from the conjugated glycopeptides we tested fragmentation by in source dissociation



**Figure 7.** Exemplary MALDI-MS/MS spectra of CRM<sub>197</sub> peptide 231–314 acquired from the unconjugated control (top) and an ST3 conjugated batch (bottom) after separation by C8 chromatography. An  $m/z + 848$  shift of the  $c_{12}$ -ion observed between the top and bottom spectra indicated K242 to be the major modified lysine residue in this peptide. The ISD-spectra also provided extensive amino acid sequence data of the peptide.

(ISD) using this LC-MALDI-TOF-MS approach. This technique has been frequently used for N- and C-terminal protein sequencing of medium size and larger proteins<sup>35–37</sup>, and in the case of smaller proteins even top-down *de novo* sequencing of the entire protein was achieved<sup>38</sup>.

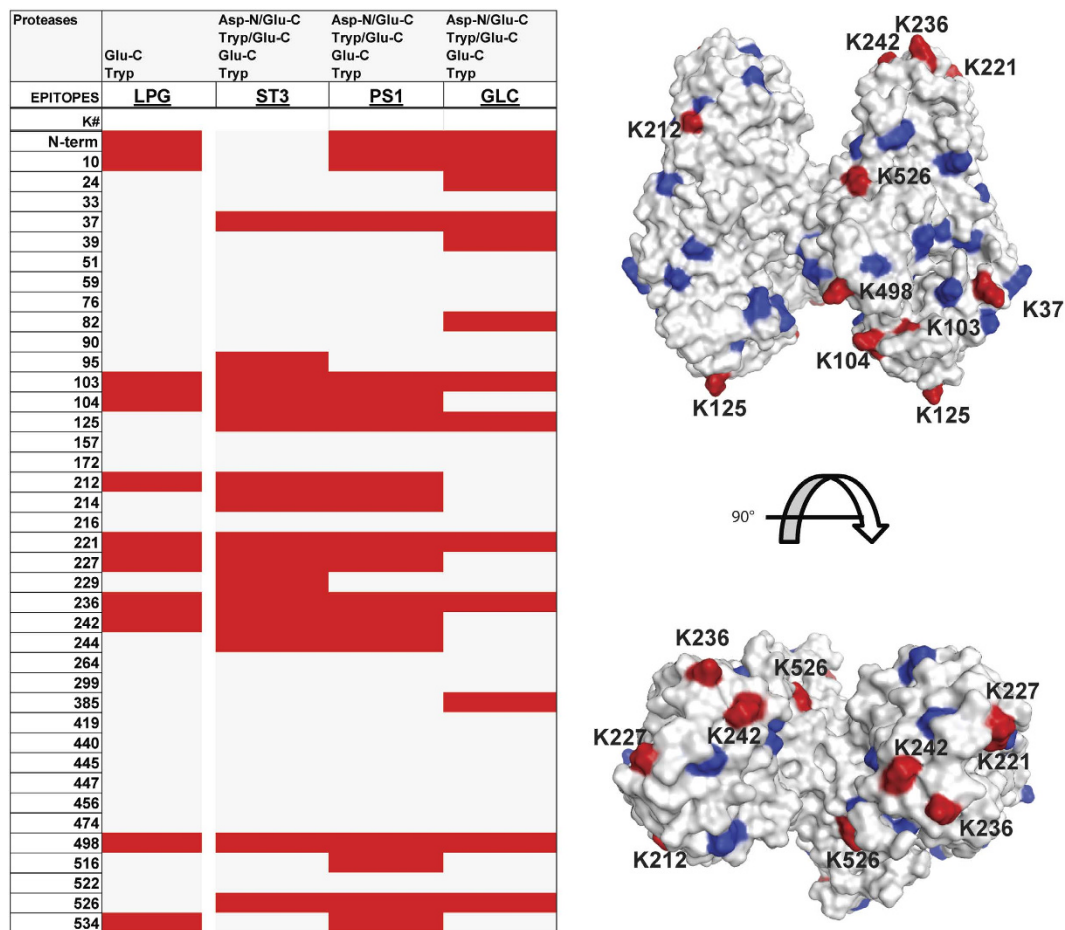
In order to assign the conjugated glycan to a specific lysine residue, the unconjugated peptide should ideally be LC baseline separated from the different conjugated forms. As mentioned previously in the text just minor retention time differences were observed and the necessary separation required for ISD was not achieved when using the C4 column. In contrast, C8 chromatography resulted in sufficient separation in the case of two singly conjugated (glyco)peptides. ISD analysis of these compounds allowed sequence confirmation of the individual peptides and assignment of the modification to the respective amino acid (Fig. 7).

Despite the LC separation it must be assumed that the selected  $m/z$  signals still contained a mixture of conjugation site isomers. Therefore, fragment ions specific for less abundant isomers are likely suppressed by the most intense glycopeptide isomer(s), providing information on the site(s) most efficiently modified within a glycopeptide. In peptide 231–314, K242 was identified as the most prominent conjugation site for both, conjugations with the free linker as well as with the glycan. A similar observation was made for K212 on peptide 183–230 (see Supplementary Fig. S5), indicating that these lysine residues are more frequently conjugated. These results also correlated well with the data obtained for K242 within peptide 242–249 after digestion with Glu-C (Fig. 2D) and demonstrate the principal applicability of ISD fragmentation to obtain in-depth attachment site information on chemically conjugated glycopeptides.

**Evaluation of conjugation site occupancy.** Even though CRM<sub>197</sub> contains 39 lysine residues and the N-terminal amine that can serve as sites of conjugation, not all of them seem to be equally reactive. Steric accessibility, the individual pKa of the respective lysine residues as well as reaction conditions and conjugate size are likely to influence the conjugation efficacy of individual sites. To evaluate whether particular lysine residues represent favoured sites of conjugation, various CRM<sub>197</sub> conjugates (ST3, PS1, LPG and GLC, Table 1) were digested either with trypsin or orthogonal proteases and protease combinations. Thereby, a qualitative conjugation site map of all modified primary amines was established based on the presence of conjugation sites in the different samples (Fig. 8).

Three areas within CRM<sub>197</sub> seem to be preferred for the attachment of glycan epitopes: In addition to lysines K37, K103, K104 and K125, a region including residues K212, K221, K227, K236 and K242 as well as the C-terminal part (in particular K498 and K526) were repeatedly found to be conjugated (Fig. 8). Some peptides such as the peptide <sup>519</sup>DHTKVNKLSLFFE<sup>532</sup> containing K526 could only be detected using certain protease combinations. In terms of attachment sites the results are in good agreement with those from Crotti *et al.*<sup>20</sup>, who evaluated site reactivity based on the presence of a linker molecule without the glycan. *In silico* mapping of the lysine residues on a CRM<sub>197</sub> structure model<sup>39</sup> showed that most conjugation sites identified in the course of this work are located on outer edges (Fig. 8).



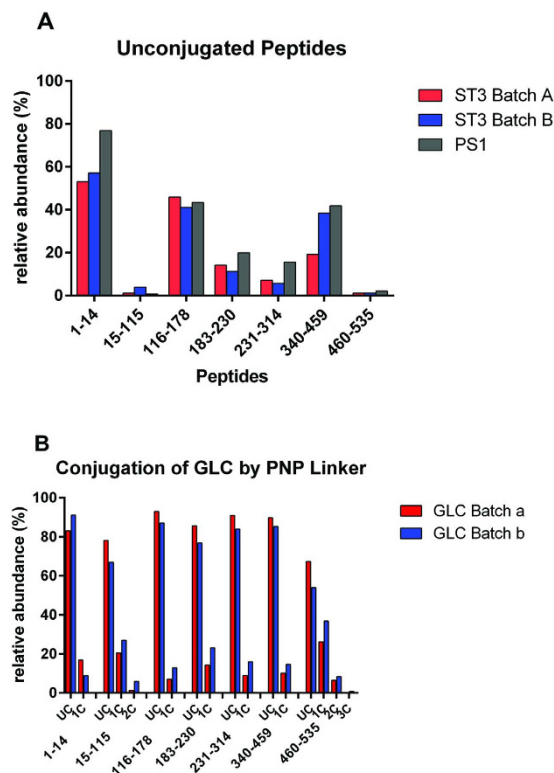


**Figure 8. (Left):** Heatmap showing the lysine residues detected to be conjugated (red label) after proteolytic digestion with trypsin, Glu-C or combinations of trypsin/Glu-C or Glu-C/Asp-N. Grey labelled lysine residues were not observed as being conjugated. Data obtained for four independent conjugates carrying a variety of glycoepitopes (LPG, ST3, PS1 and GLC as represented in each column) showed that several lysine residues were commonly found conjugated. (Right) 3D crystal structure of CRM<sub>197</sub> dimer (PDB entry: 4AE0). Lysine residues are labelled blue. Lysine residues that were frequently found conjugated with a glycan in all the samples are labelled in red.

The conjugation of mono- to pentasaccharides as performed in this study as well as the conjugation of just a linker molecule as described by Crotti and colleagues did not show marked differences in conjugation site selectivity. Based on these data we conclude that smaller variations in conjugation efficiency observed within the different epitopes are in majority due to unavoidable batch-to-batch variations resulting from the chemical glycosylation protocol rather than induced by structural differences of the epitopes.

In contrast, the microenvironment of the conjugation site apparently influenced the conjugation efficiency significantly, as demonstrated in particular by glycopeptide <sup>242</sup>KAKQYLEE<sup>249</sup> obtained after Glu-C-digestion. The two lysine residues K242 and K244 are located within an alpha helical part in close proximity to each other but differed significantly in their respective conjugation reactivity. An exclusive steric effect is unlikely to account for the reactivity differences, however variations in the regional pKa of the respective lysine residues might be involved. Studies on helix stabilisation by glutamic acid and lysine residues showed that these amino acids, when spaced three to four amino acids apart, were efficiently stabilizing alpha helices by the formation of salt bridges<sup>40</sup>. Unlike K242, K244 is located in sufficient proximity to three glutamic acid residues (E240, E241, E248, Fig. S6) and could be involved in forming a salt bridge, stabilizing the positive charge of K244 and hence decreasing its nucleophilicity, rendering K244 less reactive.

**Semi-quantitative regio-specific occupancy determination by LC-MALDI-TOF MS.** The signal intensities of the unconjugated peptide and all of its modified forms detected in the LC-MALDI-TOF MS experiments were summed up and the relative amounts for each peptide/glycopeptide group were determined. In the case of ST3 and PS1, peptides 15–115 and 460–535 were most efficiently conjugated, followed by peptides 183–230 and 231–314 (Fig. 9A). Peptides 1–14, 116–178 and 340–469 were found to be least efficiently conjugated. A similar result was obtained for GLC-CRM<sub>197</sub> conjugates that were produced via the PNP method, where no free linker interfered with the conjugation. Again, peptides 15–115 and 460–535 showed the lowest amount of



UC= unconjugated/unmodified, L= free Linker, C= Glycan Conjugate

**Figure 9. Quantitation results from LC-MALDI-MS measurements.** (A) Quantitation of unconjugated peptides from of ST3 batch A, ST3 batch B and PS1 (only peptides containing lysine residues were taken into account). Peptides 15–115 and 460–535 show the lowest amounts of unconjugated peptide, consequently carrying larger amounts of glycoconjugates and/or free linker, indicating that conjugation was more efficient at lysine residues present in these peptides. (B) Quantitation of two batches of GLC (batch a and batch b) conjugated via the PNP-linker. Only glycan conjugates were found to be present on the peptides. Again, peptides 15–115 and 460–535 show the highest loading.

unconjugated peptide and the highest conjugate loading (Fig. 9B). Apart from peptide 1–15, the LC-MALDI-MS results for the least conjugated region are qualitatively correlating with the LC-ESI-MS data.

The data acquired using various orthogonal approaches in the course of this study allowed us to semi-quantitatively evaluate conjugation efficiency on the individual lysine residues. We concluded that amino acid residues K103, K498, and K526 were among the most reactive ones in the peptides 15–115 and 460–535, since these peptides showed the highest overall conjugation. In the slightly less reactive peptides 183–230 and 231–314 amino acids K212, K221, K242 and K236 were among the frequently conjugated sites.

## Conclusions

Characterisation of glycoconjugate vaccines requires sophisticated and orthogonal approaches. In the course of this study we applied well-established glycoproteomic techniques and developed a novel approach for the in-depth characterisation of the widely used immunogenic carrier protein CRM<sub>197</sub> that was modified with a variety of defined synthetic oligosaccharide epitopes. Chemical cleavage using CNBr followed by a "middle-down" LC-MALDI-MS detection strategy provided several advantages towards any protease based assays for comprehensive and in-depth semi-quantitative evaluation of region specific conjugation efficiency for CRM<sub>197</sub> glycoconjugate vaccine candidates. This technique largely facilitated conjugation product screening by providing virtually complete sequence coverage, good overview on intended as well as unintended modifications and allowed a reproducible, protease independent opportunity for batch-to-batch evaluations. From the 40 possible primary amines (39 lysines + N-terminus) present in CRM<sub>197</sub>, just a few participate in quantitatively relevant conjugation reactions. Steric accessibility, the local amino acid environment and protein secondary structure are likely the most relevant parameters influencing the conjugation reaction under non-denaturing conditions. Our findings can ensure quality control of effective and safe vaccines as well as improving rational vaccine design in the future. It remains to be seen how conjugation variations on these specific sites possibly influence CRM<sub>197</sub> conjugate vaccines effectiveness, nevertheless the presented assay provides an important step towards functional studies applying CRM<sub>197</sub> glycoconjugate vaccine candidates carrying defined synthetic glyco-epitopes.

## Material & Methods

**Reagents.** Sequencing grade trypsin was purchased from Roche (Mannheim, Germany), Glu-C and Asp-N from Protea Bioscience (Morgantown, WV). Acetonitrile (ACN), formic acid (FA), trifluoroacetic acid (TFA), cyanogen bromide (CNBr), iodoacetamide (IAA), dithiothreitol (DTT), ammonium bicarbonate and triethylamine amine were purchased from Sigma Aldrich (Munich, Germany). Recombinant CRM<sub>197</sub> was obtained from Pfenex (San Diego, CA). The BCA assay was purchased from Pierce (Rockford, IL), C<sub>18</sub> ZipTips<sup>®</sup> and Cetricon<sup>®</sup> diafilters were purchased from Millipore (Tullagreen, Ireland) and hydrophilic lipophilic balanced solid phase extraction cartridges 30 MG (HLB SPE) were obtained from Supelco/Sigma Aldrich (Bellefonte, PA). Generally, the MIAPE and MIRAGE reporting guidelines are followed throughout this work<sup>41,42</sup>.

**Glycan epitope synthesis.** The glycan epitopes ST3, PS1 and LPG (Table 1) were synthesised from mono-saccharide building blocks as described previously<sup>43–45</sup>.

**Synthesis of p-nitrophenol activated ester (GLC-PNP).** The details on the synthesis of GLC-PNP are described in Supplementary Material S7.

**DSA conjugation of synthetic glycans to CRM<sub>197</sub>.** Oligosaccharide (3.46 μmol) was dissolved in 100 μL DMSO and added drop wise to a solution of 0.33 mmol of di-succinimido adipate in 100 μL DMSO. Before the addition of glycan, a catalytic amount of triethylamine was added to the linker solution. After 2 hours 0.5 mL 0.1 M phosphate buffer pH 7.4 was added to the reaction mixture and residual unreacted linker was extracted with 14 mL of Chloroform. The extraction procedure was repeated three times and the resultant aqueous layer was centrifuged (300 g, 5 min) to separate traces of chloroform. The resultant aqueous layer was added to 1 mL of protein solution (CRM<sub>197</sub> 1 mg/mL in 0.1 M phosphate buffer pH 7.4) and the reaction was allowed to continue for 5–6 hours with gentle stirring. The final reaction mixture was purified by ultrafiltration and the final protein concentration was determined by micro BCA assay following the manufacturer's recommendations.

**p-Nitrophenol (PNP) conjugation of synthetic glycans to CRM<sub>197</sub>.** PNP half ester (172.4 nmol, 4 μg/μL in DMSO) was added slowly to 17.24 nmol of CRM<sub>197</sub> in 500 μL of 0.1 M phosphate buffer pH 7.9 and allowed to react for 24 h at room temperature. After the reaction was complete, the protein was dialysed by diafiltration (30 kDa Centricon Diafilters). Conjugation efficiency was monitored by SDS-PAGE and MALDI-TOF MS.

**Proteolytic Digestion.** Protein (5 μg) was added to 4 μL of a 4x SDS-PAGE sample buffer (250 mM Tris-HCl, 40% glycerine, 4% SDS, 0.015% Bromophenol blue containing 50 mM DTT) and denatured at 96 °C for 5 minutes before 0.5 μL of 500 mM IAA were added to achieve a final concentration of 50 mM IAA. The solution was incubated 30 min in the dark before loading the protein onto a 10% SDS-PAGE gel for electrophoretic separation. Proteolytic digestion and peptide extraction were performed as described earlier<sup>28</sup>. Briefly, excised bands were cut into pieces destained in 50% ACN and dried. Proteases (Trypsin, Glu-C, Asp-N) were used in a protease to protein ratio of 1:50 respectively in the single digests or successively in double digests (Tryp/Glu-C; Glu-C/Asp-N). After peptide extraction with 2x 50% ACN and 2x 5% FA the samples were dried in a Speed Vac concentrator and reconstituted in 0.1% FA for further MS analysis.

**Cyanogen bromide cleavage.** One crystal of cyanogen bromide (CNBr) was dissolved in 200 μL 50% TFA. 10 μg (30 μg for larger amounts) of protein were dissolved in 190 μL 50% TFA. Subsequently, 10 μL of the CNBr stock solution were added. The solution was topped with N<sub>2</sub>, sealed with Parafilm and incubated for 48 hours at 4 °C in the dark. The reaction was quenched by the addition of 1 mL of water with subsequent removal of the liquid by Speed Vac drying. The sample was reconstituted in 10 μL 0.1% FA and purified by C<sub>18</sub> ZipTips eluting in 10 μL 80% FA. Larger amounts of initial protein were reconstituted in 50 μL and purified by HLB SPE. The sample was washed three times with 400 μL of 0.1% FA and peptides were eluted using three times 200 μL 80% ACN containing 0.1% FA. The sample was dried and reconstituted in 20 μL of 0.1% TFA for LC-MALDI-TOF-MS analysis.

**LC-ESI-MS/MS.** NanoLC-ESI-MS analysis was carried out on an Ultimate 3000 RSLC-nano system (Dionex/Thermo Scientific, Sunnyvale, CA) coupled to an amaZon speed ETD ion trap (Bruker, Bremen, Germany). In each run peptides corresponding to 0.15 μg of a CRM<sub>197</sub> digest were injected. The peptides were concentrated on a C18 precolumn (Acclaim PepMap100<sup>™</sup>, Thermo, 100 μm × 20 mm, 5 μm particle size) and separated by reversed phase chromatography on a C18 analytical column (Acclaim PepMap<sup>™</sup>, Thermo, 75 μm × 15 cm, particle size 3 μm). The samples were loaded in 99% buffer A (0.1% FA) for 6 min on the precolumn at a flow rate of 6 μL/min before the captured peptides were subjected to nanoLC at a flowrate of 350 nL/min using a gradient of increasing ACN as follows: buffer B (ACN containing 0.1% FA) from 1% to 20% (7–57 min) followed by a further increase to 50% (57–75 min) and a steep increase to 90% (75–86 min) before returning to the starting conditions. The mass spectrometer was set-up to perform CID as well as ETD fragmentation on the three most intense signals in every MS scan. An *m/z* range from 380–1800 Da was used for data dependent precursor scanning. The MS data was recorded using the instrument's "enhanced resolution mode". MS/MS data was acquired in "ultra-mode" over an *m/z* range from 100–2000. Details on the MS CID and ETD settings can be found as Supplementary Table S1.

**LC-MALDI-TOF-MS analysis.** LC-MALDI-TOF-MS experiments were performed on an ultrafleXtreme MALDI-TOF/TOF (Bruker, Bremen, Germany). Peptides were separated via an Agilent 1200 by C8 (Zorbax 300SB-C8, 3.5 μm, 100 × 0.3 mm from Agilent) and C4 (Phenomenex, Jupiter C4 300A, 150 × 0.3 mm) capillary reversed phase separation. Typically, 4–8 μg of samples were injected. Buffer A was 0.1% TFA. Buffer B was ACN with 0.1% TFA. The 40 min gradient was as followed: Buffer B: 5–20% in 5 min and 20–55% over 30 min. Fractions of 3 μL were spotted every 15 seconds by a PROTEINEER fc II on a MTP BigAnchor 384 MALDI-target (both

Bruker, Bremen, Germany). Super dihydroxybenzoic acid (sDHB) was used as matrix. Aliquots of 0.5  $\mu$ L of a 50 mg/mL matrix solution in 50% ACN/water/0.1% TFA were spotted to each dried fraction.

The fractions were then analysed in an automated way by WARP-LC 1.3 and Compass 1.4 in positive linear ion mode and spectra were displayed according to their retention time in the SurveyViewer of the software. In-source decay (ISD) spectra were then manually acquired on selected fractions using a positive reflector method optimised for ISD. Typically, 5000–10000 laser shots were accumulated for each ISD spectrum. External calibration of ISD spectra was performed with ISD fragments of bovine ubiquitin spotted on the calibration chips of the MTP 384 BigAnchor target.

**MS Data Processing.** LC-MS data was analysed with Compass 4.1 analysis software (Bruker, Bremen, Germany). Compound spectra were created with a retention time window of 0.5 min and an intensity threshold of 100000. EIC for neutral loss of respective monosaccharide glycan masses were screened for conjugated peptides as well as all compound spectra were checked manually for presence of conjugated peptides. Protein coverage was evaluated using ProteinScape 3.1 (Bruker, Bremen, Germany) and the MS data was searched against a custom-made database using Mascot server 2.3 including all available entries for CRM and diphtheria toxin. The cleavage sites were defined as non-specific since some unspecific cleavage products were identified.

LC-MS data was analysed and the area under the curve (AUC) was quantified on SurveyViewer (Bruker). For quantitation AUC of the EICs of the respective peptide/glycopeptide signals was integrated. For each peptide the sum of all AUC of the modified forms and the non-modified peptides was set as 100%. The ISD spectra were processed in Compass 4.1 using SNAP as peak picking algorithm for monoisotopic peak annotation and further analysed in BioTools 3.2 SR4 (Bruker). Peptide sequences were modified with the conjugates and matched with the ISD spectra to determine the modification sites. Each annotation was further validated manually. The mapping of the reactive lysine residues in the 3D model was performed using Molecular Operating Environment (MOE), 2014.09 (Chemical Computing Group Inc., 1010 Sherbooke St. West, Montreal, QC, Canada).

## References

- Bazaka, K., Crawford, R. J., Nazarenko, E. L. & Ivanova, E. P. Bacterial Extracellular Polysaccharides. In: *Bacterial Adhesion* (eds Linke D., Goldman A.). 715, 213–226. Springer Netherlands (2011).
- Stein, K. E. Thymus-independent and thymus-dependent responses to polysaccharide antigens. *J of infec Dis* **165** Suppl 1, S49–52 (1992).
- Avery, O. T. & Goebel, W. F. Chemo-immunological studies on conjugated carbohydrate-proteins: II. Immunological specificity of synthetic sugar-protein antigens. *J Exp Med* **50**, 533 (1929).
- Peltola, H., Kayhty, H., Sivonen, A. & Makela, H. Haemophilus influenzae type b capsular polysaccharide vaccine in children: a double-blind field study of 100,000 vaccinees 3 months to 5 years of age in Finland. *Pediatrics* **60**, 730–737 (1977).
- Pon, R. A. Exploiting the Bacterial Surface: The Successful Application of Glycoconjugate Vaccines In: *Bacterial glycomics: current research, technology and applications* (eds Reid C. W., Reid A. N., Twine S. M.), 243–262. Horizon Scientific Press (2012).
- Giannini, G., Rappuoli, R. & Ratti, G. The amino-acid sequence of two non-toxic mutants of diphtheria toxin: CRM45 and CRM<sub>197</sub>. *Nucleic Acids Res* **12**, 4063–4069 (1984).
- Shinefield, H. R. Overview of the development and current use of CRM<sub>197</sub> conjugate vaccines for pediatric use. *Vaccine* **28**, 4335–4339 (2010).
- Jin, S. D. *et al.* Optimization of capsular polysaccharide production by Streptococcus pneumoniae type 3. *J Microbiol Biotechnol* **19**, 1374–1378 (2009).
- Bertolo, L. *et al.* The design of a capsule polysaccharide conjugate vaccine against Campylobacter jejuni serotype HS15. *Carbohydr Res* **366**, 45–49 (2013).
- Pujar, N. S. *et al.* Base hydrolysis of phosphodiester bonds in pneumococcal polysaccharides. *Biopolymers* **75**, 71–84 (2004).
- Frasch, C. E. Preparation of bacterial polysaccharide-protein conjugates: analytical and manufacturing challenges. *Vaccine* **27**, 6468–6470 (2009).
- Anish, C., Schumann, B., Pereira Claney, L. & Seeberger Peter, H. Chemical Biology Approaches to Designing Defined Carbohydrate Vaccines. *Chem Biol* **21**, 38–50 (2014).
- Verez-Bencomo, V. A Synthetic Conjugate Polysaccharide Vaccine Against Haemophilus influenzae Type b. *Science* **305**, 522–525 (2004).
- Seeberger, P. H. & Werz, D. B. Synthesis and medical applications of oligosaccharides. *Nature* **446**, 1046–1051 (2007).
- Costantino, P., Rappuoli, R. & Berti, F. The design of semi-synthetic and synthetic glycoconjugate vaccines. *Expert Opin Drug Discov* **6**, 1045–1066 (2011).
- Chu, C. Y. *et al.* Preparation, characterization, and immunogenicity of conjugates composed of the O-specific polysaccharide of Shigella dysenteriae type 1 (Shiga's bacillus) bound to tetanus toxoid. *Infect Immun* **59**, 4450–4458 (1991).
- Broker, M., Dull, P. M., Rappuoli, R. & Costantino, P. Chemistry of a new investigational quadrivalent meningococcal conjugate vaccine that is immunogenic at all ages. *Vaccine* **27**, 5574–5580 (2009).
- Peeters, J. M., Hazendonk, T. G., Beuvery, E. C. & Tesser, G. I. Comparison of four bifunctional reagents for coupling peptides to proteins and the effect of the three moieties on the immunogenicity of the conjugates. *J Immunol Methods* **120**, 133–143 (1989).
- Buskas, T., Li, Y. & Boons, G. J. The immunogenicity of the tumor-associated antigen Lewis(y) may be suppressed by a bifunctional cross-linker required for coupling to a carrier protein. *Chemistry* **10**, 3517–3524 (2004).
- Crotti, S. *et al.* Defined conjugation of glycans to the lysines of CRM<sub>197</sub> guided by their reactivity mapping. *ChemBioChem* **15**, 836–843 (2014).
- Li, Q., Rodriguez, L. G., Farnsworth, D. F. & Gildersleeve, J. C. Effects of hapten density on the induced antibody repertoire. *ChemBioChem* **11**, 1686–1691 (2010).
- Desaymard, C. & Howard, J. G. Role of epitope density in the induction of immunity and tolerance with thymus-independent antigens. II. Studies with 2,4-dinitrophenyl conjugates *in vivo*. *Eur J of Immunol* **5**, 541–545 (1975).
- Jones, C. The Regulatory Framework for Glycoconjugate Vaccines. In: *Carbohydrate-Based Vaccines* (eds Roy, R.). 989, 21–35. American Chemical Society (2008).
- Jones, C. Vaccines based on the cell surface carbohydrates of pathogenic bacteria. *An Acad Bras Cienc* **77**, 293–324 (2005).
- Jones, C. & Currie, F. Control of components of bacterial polysaccharide vaccines by physical methods. *Biologicals* **19**, 41–47 (1991).
- Monteiro, M. A. *et al.* Capsule polysaccharide conjugate vaccine against diarrheal disease caused by Campylobacter jejuni. *Infect Immun* **77**, 1128–1136 (2009).
- Allen, G. Chapter 3 Specific cleavage of the protein. In: *Laboratory Techniques in Biochemistry and Molecular Biology* (eds Allen G.). 9, 43–71. Elsevier (1981).

28. Kolarich, D., Jensen, P. H., Altmann, F. & Packer, N. H. Determination of site-specific glycan heterogeneity on glycoproteins. *Nat Protoc* **7**, 1285–1298 (2012).
29. Wuhrer, M., Catalina, M. I., Deelder, A. M. & Hokke, C. H. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J Chromatogr B Analyt Technol Biomed Life Sci* **849**, 115–128 (2007).
30. Alley, W. R., Jr., Mechref, Y. & Novotny, M. V. Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. *Rapid Commun Mass Spectrom* : RCM **23**, 161–170 (2009).
31. Leymarie, N. *et al.* Interlaboratory Study on Differential Analysis of Protein Glycosylation by Mass Spectrometry: the ABRF Glycoprotein Research Multi-Institutional Study 2012. *Mol Cell Proteomics*, (2013).
32. Stavenhagen, K. *et al.* Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *J Mass Spectrom* : JMS **48**, 627–639 (2013).
33. Gross, E. & Witkop, B. Selective Cleavage of the Methionyl Peptide Bonds in Ribonuclease with Cyanogen Bromide. *J Am Chem Soc* **83**, 1510–1511 (1961).
34. Wu, X., Ling, C. C. & Bundle, D. R. A new homobifunctional p-nitro phenyl ester coupling reagent for the preparation of neoglycoproteins. *Org Lett* **6**, 4407–4410 (2004).
35. Suckau, D. & Resemann, A. MALDI top-down sequencing: calling N- and C-terminal protein sequences with high confidence and speed. *J Biomol Tech* **20**, 258 (2009).
36. Hardouin, J. Protein sequence information by matrix-assisted laser desorption/ionization in-source decay mass spectrometry. *Mass Spectrom Rev* **26**, 672–682 (2007).
37. Suckau, D. & Resemann, A. T3-sequencing: targeted characterization of the N- and C-termini of undigested proteins by mass spectrometry. *Anal Chem* **75**, 5817–5824 (2003).
38. Resemann, A. *et al.* Top-Down *de Novo* Protein Sequencing of a 13.6 kDa Camelid Single Heavy Chain Antibody by Matrix-Assisted Laser Desorption Ionization-Time-of-Flight/Time-of-Flight Mass Spectrometry. *Anal Chem* **82**, 3283–3292 (2010).
39. Malito, E. *et al.* Structural basis for lack of toxicity of the diphtheria toxin mutant CRM<sub>197</sub>. *Proc Natl Acad Sci USA* **109**, 5229–5234 (2012).
40. Marqusee, S. & Baldwin, R. L. Helix stabilization by Glu...Lys+ salt bridges in short peptides of *de novo* design. *Proc Natl Acad Sci USA* **84**, 8898–8902 (1987).
41. Taylor, C. F. *et al.* The minimum information about a proteomics experiment (MIAPE). *Nat Biotechnol* **25**, 887–893 (2007).
42. Kolarich, D. *et al.* The minimum information required for a glycomics experiment (MIRAGE) project: improving the standards for reporting mass-spectrometry-based glycoanalytic data. *Mol Cell Proteomics* **12**, 991–995 (2013).
43. Martin, C. E., Weishaupt, M. W. & Seeberger, P. H. Progress toward developing a carbohydrate-conjugate vaccine against *Clostridium difficile* ribotype 027: synthesis of the cell-surface polysaccharide PS-I repeating unit. *Chem Commun (Camb)* **47**, 10260–10262 (2011).
44. Lefeber, D. J., Kamerling, J. P. & Vliegthart, J. F. Synthesis of *Streptococcus pneumoniae* type 3 neoglycoproteins varying in oligosaccharide chain length, loading and carrier protein. *Chemistry* **7**, 4411–4421 (2001).
45. Liu, X. *et al.* Enhancement of the immunogenicity of synthetic carbohydrates by conjugation to virosomes: a leishmaniasis vaccine candidate. *ACS Chem Biol* **1**, 161–164 (2006).

## Acknowledgements

We thank the Max Planck Society, European Union (Seventh Framework Programme "Glycoproteomics", grant number PCIG09-GA-2011-293847) and BMBF (03IS2201G "Taschentuchlabor") for generous financial support. We would like to thank Andreas Geißner, Anika Reinhardt and Annette Wahlbrink for support during the protein conjugation process, Dr. Christopher Jones, Prof. Neil Ravenscroft and Kathrin Stavenhagen for fruitful discussion and input on the topic. We acknowledge Pfenex Inc. (reagent proteins) for providing CRM<sub>197</sub> at a discounted price. C.A. and F.B. acknowledge support from the German Federal Ministry of Education and Research (Grant 0315447).

## Author Contributions

U.M., A.R. and D.S. performed the MS experiments and analyses; C.E.M., S.P., S.G. and C.L.P. synthesised the glycans. C.A. designed the epitopes. C.A. and F.B. prepared the CRM<sub>197</sub> glycoconjugates. E.W. performed molecular modelling, D.K., C.A. and U.M. designed the study. U.M., D.K., P.H.S. and C.L.P. wrote the manuscript with discussion and input from all the authors.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** A.R. and D.S. are employees of Bruker, Bremen. The other authors declare no competing financial interests.

**How to cite this article:** Möglinger, U. *et al.* Cross Reactive Material 197 glycoconjugate vaccines contain privileged conjugation sites. *Sci. Rep.* **6**, 20488; doi: 10.1038/srep20488 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>