



# Molecular modeling provides insights into the loading of sialic acid-containing antigens onto CRM<sub>197</sub>: the role of chain flexibility in conjugation efficiency and glycoconjugate architecture

Michelle M. Kuttel<sup>1</sup> · Francesco Berti<sup>2</sup> · Neil Ravenscroft<sup>3</sup>

Received: 25 January 2021 / Revised: 26 February 2021 / Accepted: 10 March 2021 / Published online: 15 March 2021  
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

Vaccination is the most cost-effective way to control disease caused by encapsulated bacteria; the capsular polysaccharide (CPS) is the primary virulence factor and vaccine target. *Neisseria meningitidis* (Nm) serogroups B, C, Y and W all contain sialic acid, a common surface feature of human pathogens. Two protein-based vaccines against serogroup B infection are available for human use while four tetravalent conjugate vaccines including serogroups C, W and Y have been licensed. The tetravalent Menveo® conjugate vaccine is well-defined: a simple monomeric structure of oligosaccharides terminally conjugated to amino groups of the carrier protein CRM<sub>197</sub>. However, not only is there a surprisingly low limit for antigen chain attachment to CRM<sub>197</sub>, but different serogroup saccharides have consistently different CRM<sub>197</sub> loading, the reasons for which are unclear. Understanding this phenomenon is important for the long-term goal of controlling conjugation to prepare conjugate vaccines of optimal immunogenicity. Here we use molecular modeling to explore whether antigen flexibility can explain the varying antigen loading of the conjugates. Because flexibility is difficult to separate from other structural factors, we focus on sialic-acid containing CPS present in current glycoconjugate vaccines: serogroups NmC, NmW and NmY. Our simulations reveal a correlation between Nm antigen flexibility (NmW > NmC > NmY) and the number of chains attached to CRM<sub>197</sub>, suggesting that increased flexibility enables accommodation of additional chains on the protein surface. Further, *in silico* models of the glycoconjugates confirm the relatively large hydrodynamic size of the saccharide chains and indicate steric constraints to further conjugation.

**Keywords** *Neisseria meningitidis* · Capsular polysaccharide · Molecular modeling · Antigen conformation · Conjugation efficiency · Glycoconjugate structure

## Introduction

Infections by *Neisseria meningitidis* (Nm) cause life-threatening illnesses such as meningitis, bacteremia and pneumonia. There are twelve known Nm serogroups, six of which — NmA, NmB, NmC, NmW, NmX, and NmY — are

associated with invasive meningococcal disease (IMD). Vaccination is the most cost-effective way to control meningococcal disease; the meningococcal capsular polysaccharide (CPS) is the primary virulence factor and vaccine target. Four tetravalent conjugate vaccines against serogroups NmA, NmC, NmW and NmY (Menactra®, Menveo®, Nimenrix®, MenQuadfi®), as well as two NmB protein-based vaccines are currently licensed. In addition, the monovalent MenAfriVac® has successfully reduced the incidence of NmA disease in the Meningitis Belt of sub-Saharan Africa [1]. Higher valency vaccines against serogroups NmACWYX and NmABCWY are in development [2].

The six primary meningococcal serogroups can be split into two families on the basis of the CPS repeat unit (Table 1): the NmA and NmX homopolymers have phosphodiester glycosidic linkages in common, whereas the family comprising NmB, NmC, NmW and NmY all contain sialic acid residues. In this work we focus on conformational

This article belongs to the Topical Collection: Glycoconjugate vaccines: classic and novel approaches.

✉ Neil Ravenscroft  
neil.ravenscroft@uct.ac.za

<sup>1</sup> Department of Computer Science, University of Cape Town, 7701 Cape Town, South Africa

<sup>2</sup> Technical Research & Development, GSK Vaccines S.r.l, Via Fiorentina 1, 53100 Siena, Italy

<sup>3</sup> Department of Chemistry, University of Cape Town, 7701 Cape Town, South Africa

**Table 1** Chemical structures of the CPS repeating unit backbone for the six main meningococcal serogroups; the pattern of O-acetylation is indicated

Serogroup	CPS Chemical structure
NmA	->6)- $\alpha$ -D-ManpNAc3/4Ac-1-P-(O->
NmB	->8)- $\alpha$ -D-NeupNAc-(2->
NmC	->9)- $\alpha$ -D-NeupNAc7/8Ac-(2->
NmW	->6)- $\alpha$ -D-Galp-(1->4)- $\alpha$ -D-NeupNAc7/9Ac-(2->
NmX	->4)- $\alpha$ -D-GlcpNAc-1-P-(O->
NmY	->6)- $\alpha$ -D-Glcp-(1->4)- $\alpha$ -D-NeupNAc7/9Ac-(2->

analysis of the sialic-acid containing CPS. Sialic acid is exposed on the surface of many human pathogens — e.g. *Campylobacter jejuni*, *Escherichia coli*, *Neisseria meningitidis*, and *Streptococcus agalactiae* — where it is thought to function as a molecular mimic and thus facilitate evasion of the host's immune response [3]. Indeed, the NmB CPS is not an appropriate conjugate vaccine target as it is poorly immunogenic and can potentially cause autoimmune reactions because of the similarity of the polysaccharide to sialic-acid epitopes expressed on mammalian tissues [2]. For this reason, in this study we restrict our conformational analysis to the sialic acid-containing meningococcal CPS that are present in current glycoconjugate vaccines: NmC, NmW and NmY.

The licensed glycoconjugate vaccines are structurally diverse. They are prepared with a variety of conjugation strategies and protein carriers, producing either a complex cross-linked lattice or a well-defined monomeric structure [4]. Manufacturing consistency of the conjugates is ensured by control of the saccharide size, derivatization (if applicable) and the stoichiometry and conditions of the conjugation reaction. The simplest monomeric conjugate is prepared by terminal activation of a pool of oligosaccharides of defined average degree of polymerization (avDP), to introduce an adipic acid linker bearing a terminal active ester, which is conjugated to primary amino groups of the carrier protein CRM<sub>197</sub> [5]. The final conjugate structure produced by this process is defined by physicochemical methods, including the molecular size distribution, the saccharide and protein content, and the ratio of saccharide to protein (degree of glycosylation), as described in the WHO and Pharmacopoeia guidelines. The monomeric Nm oligosaccharide CRM<sub>197</sub> conjugate approach was first applied to the monovalent NmC vaccine licensed in 2000 (Menjugate<sup>®</sup>) and then extended to the tetravalent Menveo<sup>®</sup> conjugate vaccine against NmACWY licensed in 2010 [2]. This well-defined product is a suitable model for understanding the parameters that determine the conjugation efficiency, with the long-term goal of controlling conjugation so as to prepare conjugate vaccines of optimal immunogenicity. Firstly, the terminal conjugation of oligosaccharides of a

defined size using the same conjugation chemistry for each serogroup results in conjugate vaccines that are amenable to detailed physicochemical characterization. Secondly, the differences between the conjugates formed using the same conjugation chemistry can be attributed to structural differences between the serogroup antigens. Structural characterization of the tetravalent Menveo<sup>®</sup> conjugate vaccine is reported in Table 2 [5], where the average degree of glycosylation and molecular size is determined by the number and length of oligosaccharide chains attached.

A first observation is that there is a relatively low limit for oligosaccharide chain attachment: over a certain limit, increasing the ratio of oligosaccharide-active ester to CRM<sub>197</sub> has little effect on the degree of glycosylation. For Menveo<sup>®</sup> the highest number of 18 RU chains loaded onto CRM<sub>197</sub> is 6.9 for NmW. (Note that, for the homopolymers NmA and NmC, a chain of 18 RU contains 18 monosaccharides, whereas for NmW and NmY with disaccharide RUs an avDP of 18 corresponds to 36 monosaccharides). Studies on synthetic oligosaccharide-based conjugates show that this limit is higher for loading of short oligosaccharide chains onto CRM<sub>197</sub> than for longer strands [8–10]. For example, in a *Candida albicans* vaccine candidate, a 30:1 ratio of a  $\beta$ -glucan hexasaccharide-active ester to CRM<sub>197</sub> achieved a loading of 16.8 chains, in contrast to 7.5 chains for the corresponding 15-mer glucan [11].

A second observation is that, across each of the four serogroups, there is a marked difference in the loading of the saccharide chains onto the carrier protein CRM<sub>197</sub>: the average number of chains per protein ranges from 4.5 to 6.9 (Table 2, right column). The reasons for this are not clear, as for each CPS the conjugation reaction conditions for oligosaccharide-active ester coupling to CRM<sub>197</sub> were optimized to reach the maximum loading of saccharide chains per mol of protein (i.e. the highest saccharide:protein ratio) [5]. One explanation is that the conjugation efficiency may be affected by the charge density of the Nm antigens i.e. serogroup C has a higher charge density in comparison to serogroup W and Y. However, it remains puzzling that the almost identical polymers of hexose-sialic acid NmY and NmW (Table 1) have significant differences in loading — NmY has 4.5 chains and NmW has 6.9 chains — and NmC (the conjugate of the homopolymer of sialic acid) falls between them, with an intermediate loading of 5.6 chains per CRM<sub>197</sub>. Further, if length was the main factor in limiting chain attachment, we would expect more chains of NmC (with 18 monosaccharides in the chain) to be attached than the longer NmY and NmW chains (36 monosaccharides), which is not the case.

Molecular flexibility may play a role in determining the glycosylation efficiency and degree of loading, as several groups have reported improvements in conjugation efficiency by the introduction of adipic spacers [12] or flexible

**Table 2** Physicochemical characteristics of Nm oligosaccharide CRM<sub>197</sub> conjugates [5], listed in order of saccharide loading

Serogroup	avDP	Degree of glycosylation (wt/wt)	Mw, weight-average molar mass (kDa)	Approximate number of chains per CRM <sub>197</sub> <sup>1</sup>
NmW	18 <sup>2</sup>	1.00	110.1	6.9
NmA	18 <sup>2</sup>	0.55 <sup>3</sup>	88.5	5.6
NmC	18 <sup>2</sup>	0.55 <sup>3</sup>	85.2	5.6
NmY	18 <sup>2</sup>	0.65 <sup>4</sup>	84.6	4.5

<sup>1</sup> Calculated from the reported average degree of glycosylation using the molecular weights of the Nm repeating unit: W 470 g/mol including a percentage of O-acetylation of 41.5 % (average of lots reported in [5]); A 317 g/mol including a percentage of O-acetylation of 79.7 % (average of lots reported in [5]); C 320 g/mol including a percentage of O-acetylation of 69.2 % (average of lots reported in [5]); Y 468 g/mol including a percentage of O-acetylation of 36.0 % (average of lots reported in [5]); CRM<sub>197</sub> 58,413 g/mol. <sup>2</sup> Defined as averaged value for calculation according to information published in literature [6, 7], <sup>3</sup> as an average of a range of 0.5–0.6, <sup>4</sup> as an average of a range of 0.6–0.7

poly(ethylene glycol) linkers [13, 14]. For Menveo® the same adipic acid linker is used for each serogroup, however, different conjugation efficiencies are observed. Here we use molecular modeling to compare the conformations and molecular flexibility of the antigens to correlate flexibility with saccharide loading onto CRM<sub>197</sub>. Molecular dynamics has proven to be a useful tool for investigation of the physical properties of carbohydrates, such as flexibility and conformation [15]. Because flexibility is difficult to separate from other structural factors (such as charge density, glycosidic versus phosphodiester linkages), we focus on sialic-acid containing CPS: serogroups NmC, NmW and NmY. As part of our on-going investigation into cross-protection between similar Nm serogroups, we have previously performed molecular dynamics simulations of 3 RU of the CPS antigens NmW and NmY [16]; and 2 to 10 RU of NmA and NmX [17]. These simulations revealed conformational and flexibility differences between the related pairs. Here we expand these prior simulations, adding simulations of the NmC CPS, as well as extended simulations of longer 6 RU strands of NmW and NmY, to allow for a complete comparison of the saccharide conformations and relative flexibilities of the CPS backbone. In addition, we build models of the three glycoconjugate vaccines, performing “*in silico*” conjugations to favored lysine residues on the surface of the crystal structure of CRM<sub>197</sub> to antigen polysaccharides (18 RU) in their dominant simulation conformations, to give some insight into the ability of the conjugated protein to accommodate additional chains.

## Material and methods

We performed longer simulations (both increased molecular length and simulation time) of NmW and NmY than we did previously [16]; and as well as new simulations of NmC. The shorter 3 RU strands do not adequately reveal the molecular flexibility of the antigens. To allow for a fair comparison, we

ran separate 1 μs simulations of CPS chains containing approximately the same number of residues – 10 RU for the homopolymer NmC (10 sugar residues) and 6 RU for the heteropolymers NmW and NmY (12 sugar residues).

## Simulation protocol

All simulations were performed using the NAMD molecular dynamics simulation software [18] with CUDA extensions for accelerated calculation of long-range electrostatic potentials and non-bonded forces on graphics processing units [19]. All structures were built with our in-house CarbBuilder software [20, 21] and modeled with the CHARMM36 additive force field for carbohydrates [22, 23]. The TIP3P model [24] was used to simulate water.

All starting structures underwent 10,000 steps of standard NAMD minimization in vacuum and were then solvated in a cubic water box with the *solvate* command in the Visual Molecular Dynamics (VMD) program [25]. Sodium ions were randomly distributed in each simulation using VMD, to neutralize the negative charge from each sialic acid. Periodic boundary conditions were employed for the solvated simulations. All solvated MD simulations first underwent a minimization-and-heating protocol of 122,000 steps, consisting of 5 K incremental temperature reassignments beginning at 10 K up to 310 K, with 1,000 steps of NAMD minimization and 1,000 steps of MD at each temperature reassignment.

A Leap-Frog Verlet integrator was used to integrate the MD equations of motion using a step size of 1 fs. Simulations were sampled under the isothermal-isobaric (nPT) ensemble at 310 K, which was achieved with the use of a Langevin piston barostat [18] and a Nose-Hoover thermostat (a hybrid of the Nose-Hoover constant pressure method [26] with piston fluctuation controlled using Langevin dynamics [27] as implemented in NAMD. Long-range electrostatics were implemented with particle mesh Ewald (PME)

summation [28] with  $k = 0.20 \text{ \AA}^{-1}$  and PME grid dimensions that were set to  $90 \text{ \AA}$  for the 6 RU simulations. Non-bonded forces were truncated at  $15.0 \text{ \AA}$  with a switching function applied from  $12.0 \text{ \AA}$  to  $15.0 \text{ \AA}$ . The 1–4 interactions were not scaled, in accordance with CHARMM force field recommendations.

## Simulation data analysis

### Chain extension

We define  $r$ , the end-to-end distance of each chain, as follows. NmC: C6 of RU2 and C2 of RU9; and for the heteropolymers NmW and NmY: C4 of galactose/glucose in RU2 and C1 of galactose/glucose in RU5.

### Conformational families

Conformations from both MD simulation trajectories were clustered using VMD's internal *measure cluster* command. Clustering analysis used time series frames 250 ps apart, discarding the first 100 ns as equilibration. First, all simulation conformations were aligned on the saccharide rings for the middle four residues in the chain. Then all conformations were clustered into families according to a rmsd fit (cutoff of  $3.5 \text{ \AA}$ ) to the chain backbone (the ring and glycosidic linkage atoms), ignoring the first two and last two residues in the chain. Only clusters comprising 5 % or more of simulation time are reported; more minor clusters are ignored. The percentage frequency of each cluster was calculated as fraction of the number of frames in the cluster relative to the total number of frames in the simulation data set (excluding the first 100 ns as equilibration).

### The effect of O-acetylation

Using our CarbBuilder software [20], we built a 10 RU NmC chain with O-acetylation at C7 and dihedral angles at the average values from the de-O-acetylated NmC molecular dynamics simulation. This structure was then relaxed with 10,000 steps of standard NAMD minimization in vacuum.

### Glycoconjugate models

To build the *in silico* conjugate models, the CRM<sub>197</sub> crystal structure was obtained from the Protein Data Bank (PDB id: 4ae0). The 18 RU chains were attached to privileged surface lysine residues in the order K103, K221, K242, K236, K498, K526 and K95, which have been determined to be privileged conjugation sites in this order [29, 30]. In the crystal structure file, the K103, K242, K236 and K498 lysine amino acids are truncated, missing most of the lysyl side chain ( $(\text{CH}_2)_4\text{NH}_2$ ). These residues were repaired by adding the missing C<sub>3</sub>N

backbone moiety to these side chains (note that the protein crystal structure does not contain hydrogens). Other truncated lysines appear in the structure, but were not repaired.

The conjugates were built using an in-house extension to our CarbBuilder software [20] that generates dihedral angle values according to a Gaussian distribution around a given angle (with  $\sigma$  specified). Each 18 RU carbohydrate chain was attached through an N-linkage to the terminal nitrogen of the specified lysine (atom name "NZ"). Saccharide chains were built with glycosidic linkage orientations chosen randomly from a Gaussian distribution ( $\sigma$  obtained from the simulation distribution) around the most favored orientation(s) in the MD simulations.

## Results

In analyzing our MD simulations of the three sialic acid-containing Nm serogroup antigens, we first contrast the extension and flexibility of the saccharide chains, and then progress to a comparison of the dominant conformational families found for each antigen, concluding with a comparison of representative models of the three glycoconjugates.

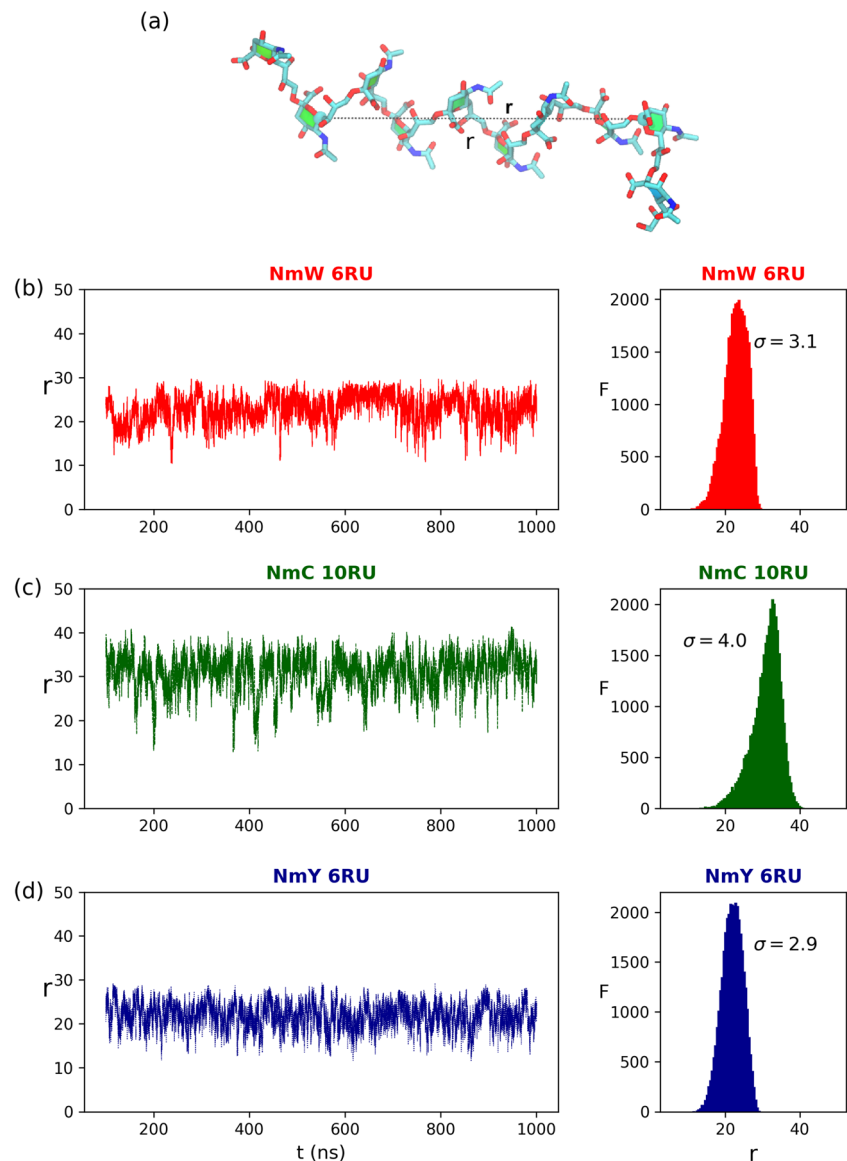
### Chain extension and flexibility

As a first approximation, the flexibility of carbohydrate antigens can be compared according to the relative extension of the molecular chains throughout the simulations. The extension of a carbohydrate chain may be measured simply by the distance between the ends of the strand, usually termed the "end-to-end distance" or  $r$ . Here we define  $r$  as the distance between the penultimate terminal residues of the meningococcal carbohydrate backbone (Fig. 1a). The terminal residues in the chains are excluded, as these are more flexible and hence not good representatives of the behavior of sugar residues in a polysaccharide.

Figure 1 lists the plots of  $r$  versus time and the corresponding histograms for each of the three sialic acid-containing Nm conjugate vaccine antigens in order of saccharide loading on the CRM<sub>197</sub> protein (as per Table 2). A broad comparison of the graphs down the rows indicates a general correlation between the distribution of  $r$  values and increased saccharide loading, as follows. All the  $r$  distributions are roughly Gaussian, but the antigens show different ranges and standard deviations (Table 3). NmC (Fig. 1c) shows the broadest distribution of  $r$  values ( $\sigma = 4.0 \text{ \AA}$ ), with similar values reported for NmW ( $3.1 \text{ \AA}$ , Fig. 1b) and NmY ( $2.9 \text{ \AA}$ , Fig. 1d).

We have previously reported the chain dynamics for shorter chains (3 RU) of the NmW and NmY pairs, finding that NmW is markedly more flexible than NmY [16]. However, the distribution of end-to-end values in our 6 RU simulations only show a small difference in flexibility: NmW has a slightly broader

**Fig. 1** Comparison of the end-to-end distance ( $r$ ) time series and distribution for the three meningococcal vaccine antigens, excluding the first 100 ns of each simulation. **a** A model of the 10 RU NmC chain with  $r$  shown. The line graphs of the  $r$  time series (left column) and the corresponding histograms (right column) are shown in order of decreasing saccharide loading in the conjugates: **(b)**, red lines) NmW, **(c)**, green lines) NmC and **(d)**, blue lines) NmY.  $F$  is the frequency of each value of  $r$



range of  $r$  than NmY ( $\sigma = 3.1$  Å versus 2.9 Å). Further, the NmC antigen is an anomaly, with a broader range of  $r$  than NmW (4.0 versus 3.1), but a lower saccharide loading (5.6 versus 6.9 chains). Overall, the end-to-end distance is a useful, but imperfect, measure of flexibility and an analysis of the conformations and rotations of the saccharide chains provides more support for our primary hypothesis that chain loading is directly correlated with chain flexibility.

### Nm saccharide chain conformations

The flexibility of a carbohydrate molecule comprises not only the ability to bend the chain (which is roughly measured by  $r$ ), but also how easy it is to rotate the backbone linkages, which will alter the conformation but may not produce a significant change in the end-to-end distance. A useful measure of flexibility is therefore the number of different conformations

**Table 3** Statistics for the molecular end-to-end ( $r$ ) distance for the simulations of the three Nm oligosaccharides, listed in order of saccharide loading in the CRM<sub>197</sub> conjugates

Serogroup	Average	Minimum	Maximum	Range	Standard deviation( $\sigma$ )
NmW	22.8	10	30	20	3.1
NmC	30.6	13	41	28	4.0
NmY	21.9	11	29	18	2.9

All values in Å



explored by an antigen, which can be estimated by grouping the simulation frames (post-equilibration) into families of similar conformations. The frequency (number of simulation frames) of each conformational family can then also be estimated, to allow for identification of the dominant conformational families for each serogroup. Comparison of the number of conformational families and their associated frequencies for each serogroup gives a more nuanced estimate of relative molecular flexibility. Such clustering analysis for the Nm antigens reveals a clearer correlation between molecular flexibility (the number of dominant conformational families) and saccharide loading, as shown in Fig. 2.

Conformational analysis reveals a larger difference between the two structurally most similar antigens, NmW and NmY, than is indicated by the  $r$  time series. NmW has six conformational families (Fig. 2a) while NmY has only one (Fig. 2c). The dominant curved NmW conformation is equivalent to the sole NmY conformation (85 %) but is occupied for only 36 % of the time. NmW transitions between the different conformational families throughout the course of the simulation, whereas NmY is almost entirely in one conformation, as is shown by the  $r$  time series colored according to conformation in the right column of Fig. 2. This marked conformational difference between very similar antigens is in agreement with our earlier study (3 RU) corroborated by NMR NOESY data [16]. The variety of conformations for NmW involve rotation about the backbone but have similar extensions – hence the small variation seen in  $r$  for this antigen. It arises from the three distinct orientations (*gg*, *gt*, and *tg*) for the primary hydroxyl linkage of Gal (NmW) compared to a single dominant orientation for Glc (NmY). These differences result in a much more conformationally diverse and flexible NmW, which is correlated with its higher loading of 6.9 chains, whereas NmY has a lower loading of 4.5 chains on CRM<sub>197</sub> (Table 2). This conformational analysis thus provides a compelling rationale for the otherwise confusingly different saccharide loading in these two closely related antigens, which differ only in the configuration at C4 of the hexose component.

In contrast to NmW, analysis of conformational families for NmC (Fig. 2c) indicates that this antigen is less flexible than is indicated by the variation in end-to-end distance. The broad range of end-to-end distances seen for NmC resolves into five quite similar conformational families. All conformations show a zig-zag ribbon, with 2 RU per turn; variations in  $r$  are a result of the ribbon “breathing” (stretching and compressing), without an overall change in conformation. Support for this breathing motion is that transitions between conformations are relatively rapid (as compared to NmW) – compare the  $r$  time series colored according to conformation in the right column of Fig. 2.

Therefore, using the number and frequency of conformational families as a proxy for flexibility for the Nm antigens produces a flexibility order of NmW > NmC > NmY that is directly correlated with the saccharide loading on CRM<sub>197</sub>.

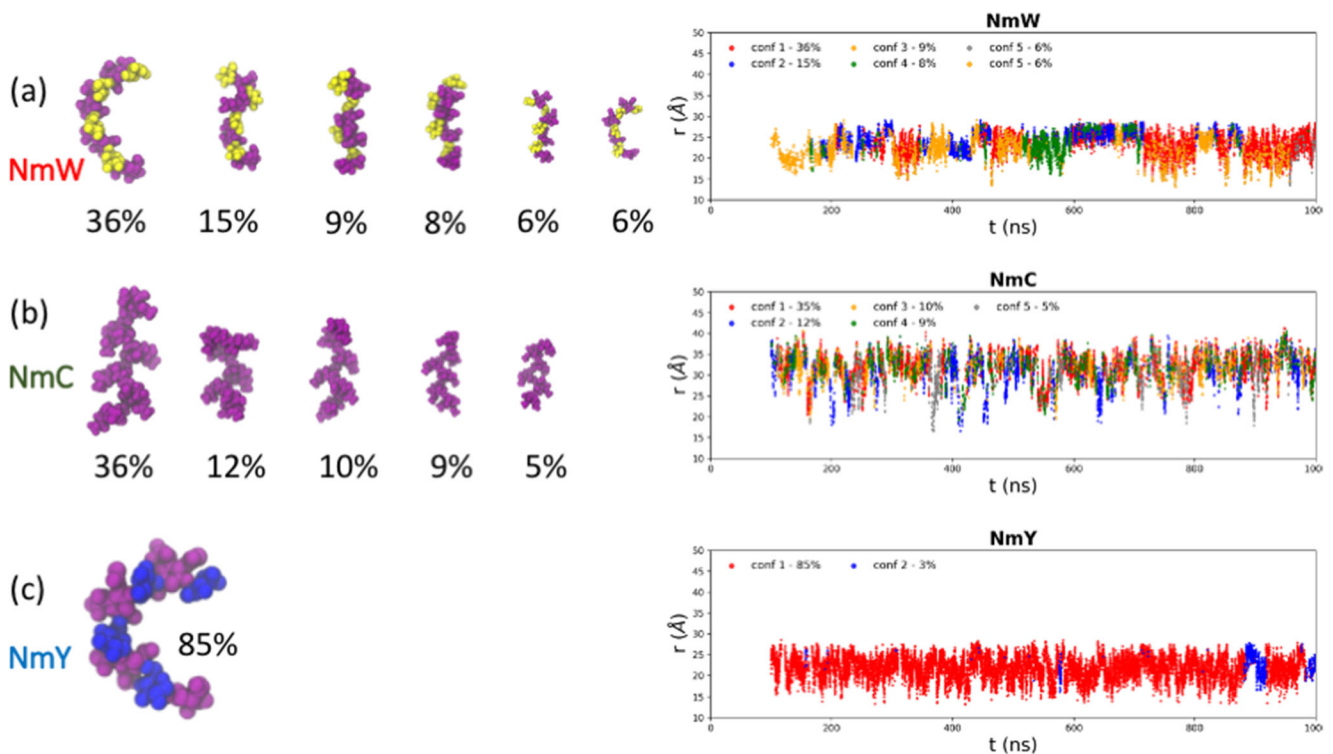
### The effect of O-acetylation

We do not consider the effects of O-acetylation on conformation in this investigation. The NmY and NmW polysaccharides are partly O-acetylated at C7/C9 and NmC O-acetylation is initially O-acetylated on C8, but it migrates to C7 during the amination and activation steps [31]. O-acetylation is not thought to be important for antigenicity of the NmY and NmW as pre-clinical and clinical studies indicate that the de-O-acetylated epitope is the primary target for bactericidal antibodies [32]. However, although the monovalent NmC conjugate vaccine NeisVac-C, licensed for human use, is not O-acetylated, the four licenced MCV4 vaccines are [4]. To address the possible impact of O-acetylation on the Nm C conformation described above, we built and relaxed an NmC chain bearing O-acetylation at C7 and compared this static model to the primary conformational family of the de-O-acetylated NmC (Fig. 3). It is clear that the conformation of the de-O-acetylated NmC (Fig. 3a) is not significantly distorted in the 7-O-acetylated chain (Fig. 3b), likely because the location of the 7-O-acetyl substitution (colored pink in Fig. 3b) is such that it does not cause a direct steric clash with the rest of the chain. This is in agreement with analysis of <sup>1</sup>H-NMR spectra by Michon et al. [33], who concluded that the 7-O-acetyl substitution does not change the conformation of the glycosidic linkages relative to the de-O-acetylated chain, whereas the 8-O-acetyl substitution does. Deeper investigation of the effects of O-acetylation on the NmC polysaccharide is left as a topic for future work.

### Glycoconjugate models

Figure 4 shows models of the CRM<sub>197</sub> glycoconjugates for each of the three sialic acid-containing Nm vaccine antigens in the tetravalent vaccine, with the number of carbohydrate chains commensurate with the calculated saccharide loading for each antigen. All chains are 18 RU, are attached to privileged conjugation sites and were built with random dihedral angle conformations assigned in accordance with the distributions obtained from the MD simulations. The NmW conjugate has seven chains (Fig. 4a), NmC six chains (Fig. 4b) and NmY four chains (Fig. 4c). A model of the unconjugated CRM<sub>197</sub> is shown in Fig. 4d, with surface lysines colored pink.

Note that these are static models, not dynamic simulations, so the care must be taken in drawing conclusions from them. However, the models are to scale and highlight the large size of the 18 RU saccharide chains relative to the CRM<sub>197</sub> protein,

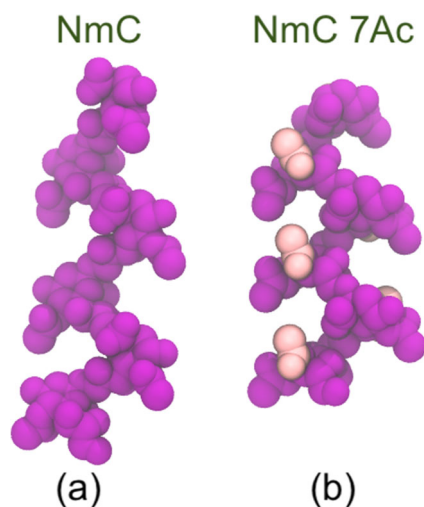


**Fig. 2** Conformational families calculated for each of the Nm antigens, listed in order of saccharide loading. **a** NmW, **b** NmC and **c** NmY. The percentage frequency of each conformational family is shown, with families comprising less than 5% of the simulation (post-equilibration) excluded. The percentage frequency of each cluster was calculated as fraction of the number of frames in the cluster relative to the total number

of frames in the simulation data set (excluding the first 100 ns as equilibration). The left column shows the corresponding time series of the end-to-end distance ( $r$ ) for each antigen, colored according to the conformational family adopted at each time step. Sugar residues are colored according to identity:  $\alpha$ -D-NeupNAc purple,  $\alpha$ -D-Galp yellow and  $\alpha$ -D-Glcp blue

giving a rationale for the low limit for oligosaccharide chain attachment. Further, given the relative size, it is clear why it is that the carbohydrate component has been found to dictate the hydrodynamic properties of the conjugates [34]. Furthermore, for all the conjugate models, there is considerable scope for

some degree of wrapping of the flexible polysaccharide chains around the protein core. This has previously been suggested to occur on the basis of persistence lengths and other hydrodynamic behavior, including intrinsic viscosity of the conjugates compared with the native or activated polysaccharides [34].

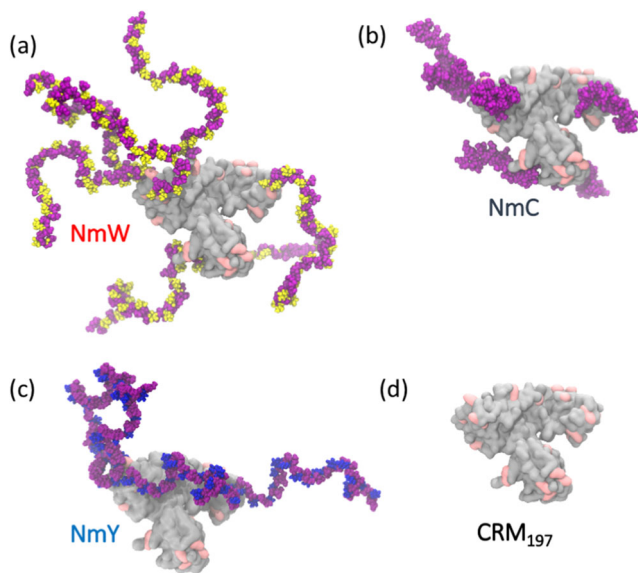


**Fig. 3** Comparison of the primary conformation of **(a)** NmC with a static, relaxed model of **(b)** NmC 7-O-acetylated. Hydrogen atoms are not shown and O-acetyl substitutions are colored pink

The radius of gyration ( $R_g$ ) is a commonly used parameter to describe the distribution of molecule's mass. For the four conjugate models in Fig. 4,  $R_g$  increases in the order NmC (43 Å) < NmY (55 Å) < NmW (62 Å). NmC therefore has the most compact conjugate and NmW the least compact structure. The increased size of the heteropolysaccharide conjugates (NmY and NmW) relative to the homopolysaccharide conjugates (NmC) is partially a result of the longer chains attached (36 residues versus 18 residues). Aside from this aspect, greater flexibility in a chain results in a larger surface area for the carbohydrate and hence a larger size for a conjugate (NmW > NmY).

## Discussion

We have identified a correlation between conjugation efficiency (determined by the number of chains attached to CRM<sub>197</sub> for the monomeric conjugates for each serogroup) and the



**Fig. 4** *In silico* models of CRM<sub>197</sub> conjugated to 18 RU polysaccharide antigens, listed in order of number of chains attached. **a** Seven chains of NmW. **b** Six chains of NmC. **c** Four chains of NmY. **d** Unconjugated CRM<sub>197</sub>, colored grey with surface lysines highlighted in pink. Sugar residues are colored according to identity:  $\alpha$ -D-NeupNAc purple,  $\alpha$ -D-Galp yellow and  $\alpha$ -D-GlcP blue

flexibility of the Nm serogroups calculated from molecular simulations, which both decrease in the order NmW > NmC > NmY. The conjugation chemistry for all the Nm saccharides involves the same reactive species: nucleophilic attack of the saccharide-linker-active ester terminus by a surface amino group of CRM<sub>197</sub> to form a stable amide bond. Moreover, as the conjugation is performed using the same stoichiometry and reaction conditions (buffer, temperature and pH), the difference in conjugation efficiency must arise from the nature of the antigens involved. The antigen structure affects the access/availability of the active ester group to the surface amino groups. We postulate that the correlation arises from a combination of two effects. First, the ability of the active ester at the carbohydrate chain terminus to access the surface lysine on CRM<sub>197</sub> for conjugation and, second, the capacity of the glycosylated CRM<sub>197</sub> to sterically accommodate the attachment of additional chains. More flexible antigens such as NmW may be better able to locate the surface amino groups for conjugation, as compared to the more rigid antigens. The existence of steric constraints to chain attachment is supported by the fact that there is a relatively low limit for oligosaccharide chain attachment to CRM<sub>197</sub> and that shorter oligosaccharides achieve higher saccharide loading [8–10]. More flexible antigens have a greater ability to shift to remove steric clashes; the introduction of long flexible linkers has been shown to increase the degree of glycosylation [13, 14]. Further, the raising of the loading limit when loading short oligosaccharide chains [10] can be explained by the reduction of steric clashes between the chains.

We expect this relationship between antigen chain flexibility and conjugation efficiency to hold for other conjugation chemistries that involve attachment of antigen chains to a carrier protein via adipic acid dihydrazide, maleimide or other terminal groups. While the flexibility of the chain may influence the random activation of antigens at multiple sites - such as hydroxyls by use of periodate oxidation, cyanylation or active carbonyls [35] - the effects would not be readily discernible after formation of complex cross-linked lattice conjugates.

Our models of the glycoconjugate vaccines calculated using the predominant antigen conformational distributions with the chains attached to the preferred surface lysines of CRM<sub>197</sub> identified experimentally confirm the relatively large hydrodynamic size of the saccharide chains, highlighting the high saccharide-to-protein size ratio and thus providing an explanation for the limited number of antigen chains that can be attached to CRM. Simulations can thus provide information on saccharide conformation and dynamics can provide understanding of antigen properties and behavior important for vaccine development. However, it is important to bear in mind these results are preliminary and that this *in silico* study requires future validation by experiment.

**Acknowledgements** N.R. and F.B. would like to thank our friend and colleague Dr Paolo Constantino for his mentorship in the field of conjugate vaccine development. Computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: .

**Author contributions** Conceptualization, N.R. and F.B.; methodology, M.K.; software, M.K.; validation, M.K.; resources, M.K. and N.R.; data curation, M.K.; writing—original draft preparation, M.K. and N.R.; writing—review and editing, M.K., F.B. and N.R.; visualization, M.K.; funding acquisition, M.K. and N.R.

**Funding** This research was funded by the South African National Research Foundation [NRF Grant 103,805, (M.K.)]. This work was sponsored by GlaxoSmithKline Biologicals SA.

**Data Availability** All data has been disclosed.

**Code Availability** Not applicable.

## Declarations

**Conflicts of interest** F.B. is an employee of the GSK companies.

**Ethics approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** All authors have read and agreed to the published version of the manuscript.



## References

- Trotter, C.L., et al.: Impact of MenAfriVac in nine countries of the African meningitis belt, 2010–15: an analysis of surveillance data. *Lancet. Infect. Dis* **17**(8), 867–872 (2017)
- Pizza, M., Bekkat-Berkani, R., Rappuoli, R.: Vaccines against meningococcal diseases. *Microorganisms* **8**(10), 1521 (2020). <https://doi.org/10.3390/microorganisms8101521>
- McDonald, N.D., Boyd, E.F.: Structural and biosynthetic diversity of nonulosonic acids (NulOs) that decorate surface structures in bacteria. *Trends Microbiol.* **29**(2), 142–157 (2021)
- Berti, F.: Multivalent meningococcal conjugate vaccines: chemical conjugation strategies used for the preparation of vaccines licensed or in clinical trials. In: Prasad K. (ed.) *Carbohydrate-based vaccines: from concept to clinic*, p. 123–137. American Chemical Society, Washington, D.C. (2018)
- Bardotti, A., et al.: Physicochemical characterisation of glycoconjugate vaccines for prevention of meningococcal diseases. *Vaccine* **26**(18), 2284–2296 (2008)
- Ravenscroft, N., et al.: Size determination of bacterial capsular oligosaccharides used to prepare conjugate vaccines. *Vaccine* **17**(22), 2802–2816 (1999)
- Bardotti, A., Berti, A.G., Berti, F., Galli, S., Giannini, C., Fabbri, S., Proietti, B., Ravenscroft, D., Ricci, N. S.: Size determination of bacterial capsular oligosaccharides used to prepare conjugate vaccines against *Neisseria meningitidis* groups Y and W135. *Vaccine* **23**(16), 1887–1899 (2005)
- Lefebvre, D.J., Kamerling, J.P., Vliegthart, J.F.: Synthesis of *Streptococcus pneumoniae* type 3 neoglycoproteins varying in oligosaccharide chain length, loading and carrier protein. *Chem. Eur. J.* **7**(20), 4411–4421 (2001)
- Carboni, F., et al.: Evaluation of immune responses to group B streptococcus type III oligosaccharides containing a minimal protective epitope. *J. Infect. Dis.* **221**(6), 943–947 (2020)
- Schumann, B., et al.: A semisynthetic *Streptococcus pneumoniae* serotype 8 glycoconjugate vaccine. *Sci. Transl. Med.* **9**(380), eaaf5347 (2017)
- Adamo, R., et al.: Synthesis of laminarin fragments and evaluation of a  $\beta$ -(1, 3) glucan hexasaccharide-CRM197 conjugate as vaccine candidate against *Candida albicans*. *J. Carbohydr. Chem.* **30**(4–6), 249–280 (2011)
- Micoli, F., et al.: Development of a glycoconjugate vaccine to prevent meningitis in Africa caused by meningococcal serogroup X. *Proc. Natl. Acad. Sci. U.S.A.* **110**(47), 19077–19082 (2013)
- Stefanetti, G., et al.: Sugar–protein connectivity impacts on the immunogenicity of site-selective *Salmonella* O-Antigen glycoconjugate vaccines. *Angew. Chem. Int. Ed. Engl.* **54**(45), 13198–13203 (2015)
- Xu, L., et al.: Development of meningococcal polysaccharide conjugate vaccine that can elicit long-lasting and strong cellular immune response with hepatitis B core antigen virus-like particles as a novel carrier protein. *Vaccine* **37**(7), 956–964 (2019)
- Kuttel, M.M., Ravenscroft, N.: The role of molecular modeling in predicting carbohydrate antigen conformation and understanding vaccine immunogenicity. In: *Carbohydrate-based vaccines: from concept to clinic*, p. 139–173. ACS Publications, Washington, D.C. (2018)
- Kuttel, M.M., Timol, Z., Ravenscroft, N.: Cross-protection in *Neisseria meningitidis* serogroups Y and W polysaccharides: A comparative conformational analysis. *Carbohydr. Res.* **446**, 40–47 (2017)
- Hlozek, J., Kuttel, M.M., Ravenscroft, N.: Conformations of *Neisseria meningitidis* serogroup A and X polysaccharides: The effects of chain length and O-acetylation. *Carbohydr. Res.* **465**, 44–51 (2018)
- Phillips, J.C., et al.: Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**(16), 1781–1802 (2005)
- Stone, J.E., et al.: Accelerating molecular modeling applications with graphics processors. *J. Comput. Chem.* **28**(16), 2618–2640 (2007)
- Kuttel, M.M., Stähle, J., Widmalm, G.: CarbBuilder: Software for building molecular models of complex oligo- and polysaccharide structures. *J. Comput. Chem.* **37**(22), 2098–2105 (2016)
- Kuttel, M., et al.: (2011) CarbBuilder: an adjustable tool for building 3D molecular structures of carbohydrates for molecular simulation. In: 2011 IEEE Seventh International Conference on eScience. IEEE, New York (2011)
- Guvench, O., et al.: CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. *J. Chem. Theory Comput.* **5**(9), 2353–2370 (2009)
- Guvench, O., et al.: Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.* **29**(15), 2543–2564 (2008)
- Jorgensen, W.L., et al.: Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**(2), 926–935 (1983)
- Humphrey, W., Dalke, A., Schulten, K.: VMD: visual molecular dynamics. *J. Mol. Graph.* **14**(1), 33–38 (1996)
- Martyna, G.J., Tobias, D.J., Klein, M.L.: Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**(5), 4177–4189 (1994)
- Feller, S.E., et al.: Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* **103**(11), 4613–4621 (1995)
- Darden, T., York, D., Pedersen, L.: Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**(12), 10089–10092 (1993)
- Crotti, S., et al.: Defined conjugation of glycans to the lysines of CRM197 guided by their reactivity mapping. *ChemBioChem* **15**, 836–843 (2014)
- Möglinger, U., Resemann, A., Martin, C.E., Parameswarappa, S., Govindan, S., Wamhoff, E.C., et al.: Cross reactive material 197 glycoconjugate vaccines contain privileged conjugation sites. *Sci. Rep.* **6**(1), 1–13 (2016)
- Ravenscroft, N.: The application of NMR spectroscopy to track the industrial preparation of polysaccharide and derived glycoconjugate vaccines, pp. 131–144. *Pharmeuropa, Special Edition* (2000)
- Fusco, P.C., Farley, E.K., Huang, C.H., Moore, S., Michon, F.: Protective meningococcal capsular polysaccharide epitopes and the role of O acetylation. *Clin. Vaccine Immunol.* **14**(5), 577–584 (2007)
- Michon, F., et al.: Structure activity studies on group C meningococcal polysaccharideprotein conjugate vaccines: effect of O-acetylation on the nature of the protective epitope. *Dev. Biol. Basel* **103**, 151–160 (2000)
- Abdelhameed, A.S., et al.: Solution conformation and flexibility of capsular polysaccharides from *Neisseria meningitidis* and glycoconjugates with the tetanus toxoid protein. *Sci. Rep.* **6**, 35588 (2016)
- Prasad, A.K., Kim, J., Gu, J.: Design and development of glycoconjugate vaccines. In: Prasad, K. (ed.) *Carbohydrate-based vaccines: from concept to clinic*, p. 75–100. American Chemical Society, Washington, D.C. (2018)

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.