

Comparative simulation of pneumococcal serogroup 19 polysaccharide repeating units with two carbohydrate force fields.

Michelle Kuttel^{a,*}, Marc Gordon^a, Neil Ravenscroft^b,

^a*Department of Computer Science, University of Cape Town, Cape Town, South Africa*

^b*Department of Chemistry, University of Cape Town, Cape Town, South Africa*

Abstract

Streptococcus pneumoniae causes meningitis, pneumonia, and severe invasive disease (IPD) in young children. Although widespread infant immunisation with the PCV7 seven-valent pneumococcal conjugate vaccine has led to a dramatic decrease in IPD, infections due to non-vaccine serotypes, particularly serotype 19A, have increased. As the 19F polysaccharide differs from 19A at a single linkage position, it was assumed that PCV7 (containing 19F) would cross-protect against 19A disease. However, vaccination with PCV7 results in only 26% effectiveness against IPD caused by 19A. We explored the conformations and dynamics of the polysaccharide repeating units from serotypes 19F and 19A, comparing free energy surfaces for glycosidic linkages with 100 ns aqueous Molecular Dynamics simulations of the di- and trisaccharide components. All calculations were performed with both the CHARMM and the GLYCAM carbohydrate force fields to establish whether the choice of model affects the predicted molecular behaviour. Although we

*Tel:+27 21 6505107, email: mkuttel@cs.uct.ac.za

identified key differences between the force fields, overall they were in agreement in predicting a 19F repeating unit with a wider range of conformation families than the more restricted 19A trisaccharide. This suggests a probable conformational difference between the 19F and 19A polysaccharides, which may explain the low cross-protection of 19F vaccines against 19A disease.

Keywords: Streptococcus pneumoniae, vaccine, pneumococcal, force fields, Molecular Modelling

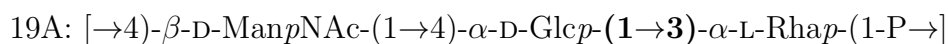
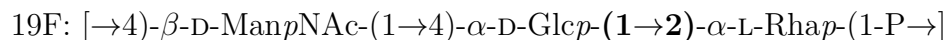
1. Introduction

The bacterium *Streptococcus pneumoniae* is a leading cause of meningitis, pneumonia, febrile bacteremia, osteomyelitis and, less seriously, otitis media, sinusitis and bronchitis in humans. The greatest impact of pneumococcal invasive disease is on young children: the World Health Organisation estimates an annual global mortality of 700 000 to 1 000 000 for children under the age of five and the highest disease burden is in the developing world.^{1,2} Therefore, increasing the coverage of vaccination of infants and young children against pneumococcal disease is a current global health priority.³

The polysaccharide capsule is the main virulence factor of *S. pneumoniae* and the target of all pneumococcal vaccines developed thus far. While over 90 pneumococcal serotypes have been described, only a small subset is responsible for the majority of severe disease.⁴ Vaccine manufacturers in the developing world need to select the serotypes to be included in their vaccine carefully, aiming for a vaccine that covers at least 60% of the invasive disease isolates in the target region⁵ and yet has sufficiently low valency (number of serotypes) to be affordable.

Pfizer’s 7-valent conjugate vaccine (PCV7) comprises the pneumococcal serotypes responsible for the majority of disease in the United States prior to widespread vaccination: serotypes 4, 6B, 9V, 14, 18C, 19F and 23F.⁶ Since introduction in 2000, PCV7 has dramatically reduced the incidence of invasive pneumococcal disease^{7,8}, especially for disease due to the vaccine serotypes. However, the efficacy of PCV7 varies considerably worldwide: it is highest in the USA and lowest in the developing world.²

Serogroup 19 is currently responsible for the bulk of pneumococcal disease. This serogroup comprises serotypes 19F, 19A, 19B and 19C, with disease mainly caused by 19A and 19F.⁹ When the more stable 19F¹⁰ was selected for PCV7, it was thought that the antibodies elicited would be cross-protective within the structurally similar serogroup. However, although the 19F and 19A capsular polysaccharides comprise very similar trisaccharide repeating units —



— cross-reactivity has proven to be very limited. Indeed, post PCV7 the proportion of infections caused by serotype 19A has increased: 19A currently appears to be the most prevalent and resistant pneumococcal serotype.^{8,11,12} Vaccination with PCV7 has been estimated to result in only 26% effectiveness against IPD caused by 19A.¹³

The new higher valency conjugate vaccines aim to provide greater coverage of the problematic pneumococcal strains. GSK’s PCV10 adds serotypes 1, 5 and 7F; whereas Pfizer’s PCV13 contains additional serotypes 1, 3, 5, 6A, 7F and 19A. 19A was omitted from PCV10, despite the high levels of

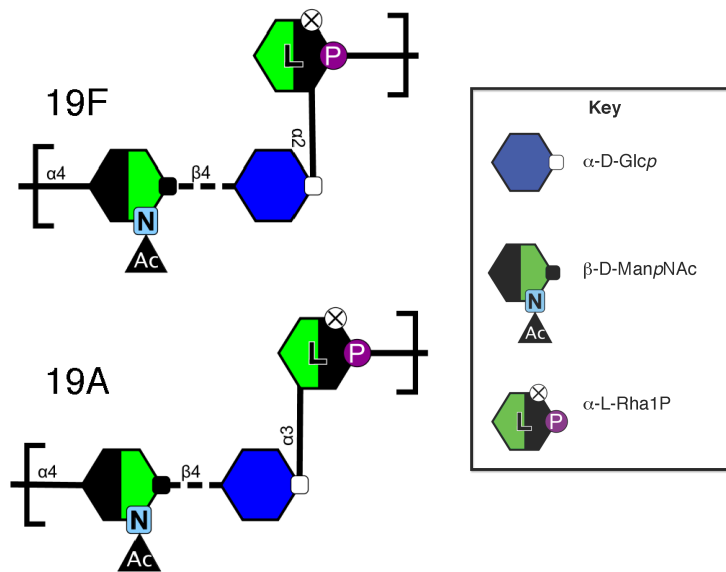


Figure 1:

disease due to this serotype, because analysis of sera that showed that the coupling chemistry employed for serotype 19F in PCV10 may provide better cross protection against 19A disease.¹⁴ Given their recent licensure, the efficacy of these higher valency vaccines in children is yet to be established.

When considering the molecular basis for cross protection in *S. pneumoniae* serogroup 19, a key question is whether the $\alpha(1\rightarrow3)$ linkage configuration in serotype 19A (Fig. 1) alters the molecular conformation sufficiently to affect the binding of 19F antibodies to 19A antigens and hence allow 19A to evade the host immune response in subjects vaccinated with 19F. Ciuffreda *et al.* performed an early computational investigation into the conformation of serotypes 19A and 19F, using both the hard-sphere HSEA and the MM3 force fields.¹⁵ This study found no key differences between the linkages that

would account for a lack of cross protection between the serotypes. Much later, Legani et al. employed short Molecular Dynamics simulations (500ps) with the MM+ force field to indicate whether a carba-analogue of the 19F repeating unit would have similar conformation and dynamics to the original repeating unit.¹⁶ The two chief determinators of the reliability of MD simulations are conformational sampling and force field accuracy. As both of these studies employed very short simulations and non-carbohydrate specific force fields, there is considerable scope for expanding the calculations with modern force fields with parameters optimised for the treatment of carbohydrates.¹⁷

Here we investigate the molecular basis of cross-protection by simulation of the conformation of the repeating units in the carbohydrate antigens for pneumococcal serotypes 19A and 19F. We follow a systematic procedure of investigation, starting with analysis of the preferred conformations of the individual linkages in the polysaccharides and proceeding to extended simulations of the two trisaccharides in aqueous solution. In the absence of experimental data, the reliability of the force field becomes a more pressing question. Therefore, we performed our calculations both the CHARMM¹⁸ and the GLYCAM¹⁹ carbohydrate force fields to highlight similarities and differences between the force fields and thus give a guide to the reliability of our predictions.

2. Methods

2.1. General Approach

We undertook a systematic, incremental approach to computational analysis of the repeat units in *S. pneumoniae* capsular polysaccharide serotypes

19A and 19F. Calculation of the preferred conformation of the isolated glycosidic linkages in vacuum (*Phase One*) progressed to aqueous solution simulations of the disaccharide components (*Phase Two*) and then, finally, to solution simulations of the trisaccharides (*Phase Three*). All calculations were performed twice, firstly with the new CHARMM¹⁸ force field and secondly with the GLYCAM¹⁹ force field, to enable comparison of the predictions of these two carbohydrate models.

2.1.1. Phase One: comparative disaccharide PMF calculations

As saccharide rings are relatively rigid, oligosaccharide conformation is determined chiefly by the orientations of the glycosidic linkages. Therefore, in the first phase of analysis, we identified the preferred conformations of the each of the disaccharide linkages in isolation by calculation of the potential of mean force (PMF) for rotation about the glycosidic linkage dihedral angles, ϕ and ψ . To do this, we divided the 19A and 19F repeat units (Fig. 1) into representative disaccharide fragments:

- β -D-Man p NAc-(1 \rightarrow 4)- α -D-Glcp (19A and 19F - termed M14G herein)
- α -D-Glcp-(**1** \rightarrow **2**)- α -L-Rhap (19F - termed G12R herein)
- α -D-Glcp-(**1** \rightarrow **3**)- α -L-Rhap (19A - termed G13R herein)
- α -L-Rhap-(1-P \rightarrow 4)- β -D-Man p NAc (19A and 19F - not analysed, force field parameters not available).

2.1.2. Phase Two: comparative disaccharide solution simulations

The next phase comprised simulations of the key G12R and G13R disaccharides in aqueous solution. All simulations were started from conforma-

tions representing the minimum energy orientations of the individual glycosidic linkages, as identified in Phase One.

2.1.3. Phase Three: comparative trisaccharide PMF calculations

In the final phase, we performed solution simulations of the complete trisaccharide repeating units for 19F and 19A. Again, simulations were started from conformations representing the minimum energy orientations of the individual glycosidic linkages, as identified in Phase One. In addition, the trisaccharide conformations from the solution molecular dynamics trajectories were clustered to reveal the principal conformational families explored during the aqueous simulations.

Neither the CHARMM nor the GLYCAM carbohydrate force fields parameters currently possess parameters for the highly flexible anomeric phosphodiester linkage. While parameters for phosphate substituents were recently added to the CHARMM force field²¹, these do not extend to flexible anomeric phosphodiester linkages. Therefore, simulation of more than a single repeating unit was not attempted in this study.

2.2. Molecular Dynamics Simulations

All simulations were performed with version 2.9 of the NAMD molecular dynamics program.²² The force field simulations employed the the CHARMM carbohydrate parameter set version 1.98 or the GLYCAM06 force field version h.

The disaccharide and trisaccharide models were built using psfgen²³ structure building tool for the CHARMM force field simulations and the online GLYCAM Biomolecular Builder²⁴ for the GLYCAM simulations. To build

the trisaccharides in CHARMM, it was necessary to first add the ManNAc residue to CHARMM topology file, using the GlcNAc as a template.²⁵ In each case, the initial configurations were optimised through 1000 steps of standard NAMD minimization.

All simulations were preceded by a 10 000 step minimisation phase with a temperature control and equilibration regime involving 5 K temperature re-assignments from 0 K culminating in a maximum temperature of 310 K. For the Molecular Dynamics simulations, equations of motion were integrated using a Leap-Frog Verlet integrator with a step size of 1 fs. Energy calculations in the gas phase were performed with an infinite non-bonded cut-off, while simulations in aqueous-phase employed periodic boundary conditions. A dielectric constant of unity was employed in all calculations. Long-range electrostatic interactions were treated using particle mesh Ewald (PME) summation. Non-bonded interactions were truncated using a switching function applied between 12.0 and 15.0 Å on a neutral group basis (groups correspond to electrically neutral collections of atoms). For all CHARMM and GLYCAM simulations, 1-4 interactions were not scaled, according to the force field recommendations. Simulations were performed under isothermal-isobaric (nPT) conditions at 310 K maintained using a Langevin piston barostat²⁶ and a Nose-Hoover²⁷ thermostat.

All solution simulations were carried out with the explicit inclusion of TIP3P²⁰ water solvent molecules in a periodic cubic unit cell. Carbohydrate molecules were solvated with either the GLYCAM Biomolecular Builder (GLYCAM) or the *solvate* plugin to the VMD analysis package (CHARMM). For all the water simulations, box dimensions were adjusted to give a mass

density of 0.92 g/cm³. The CHARMM disaccharide simulations for G12R and G13R comprised a single solute disaccharide in a solution box of 464 TIP3P waters (dimensions 25 × 25 × 25 Å) and a PME grid size of 26 in each dimension. For GLYCAM, G12R was placed in a 29.3 × 29.3 × 29.3 Å box of 775 TIP3P waters using a PME grid size of 30 in each dimension; while G13R was placed in a 29.3 × 29.3 × 29.3 Å box of 757 TIP3P waters with a PME grid size of 30 in each dimension. The simulations of trisaccharide M14G12R and M14G13R molecules in CHARMM and GLYCAM all comprised a single saccharide in a box of 4298 TIP3P water molecules, with 36 Å for each dimension of the box and a PME grid size of 36 in each dimension.

Solution simulations for the disaccharides (G12R and G13R) and corresponding trisaccharides (M14G12R and M14G13R) using the CHARMM and GLYCAM force fields (8 simulations in total) were run for 100 ns each, where the first 10 ns comprised equilibration and the last 90 ns were used for data collection.

2.3. Vacuum ϕ, ψ Potential of Mean Force calculations

Potential of mean force free energy surfaces were calculated with the metadynamics²⁸ routine incorporated into NAMD²² version 2.9, with the ϕ, ψ glycosidic linkage torsion angles used as collective variables. In metadynamics, the system is simulated by a standard molecular dynamics simulation to which a special bias potential is progressively added to enhance sampling. In each case, the (1→X) glycosidic linkage is defined by the torsion angles, $\phi = H1'-C1'-O1'-CX$ and $\psi = C1'-O1'-CX-HX$. These definitions for ϕ and ψ are analogous to ϕ_H and ψ_H in IUPAC convention.

Metadynamics simulations in vacuum were performed on the M14G, G12R and G13R disaccharides with both the CHARMM and the GLYCAM carbohydrate force fields using the MD protocol described above. Each metadynamics simulation comprised a 1500 ns MD simulation, with a Gaussian hill height of 0.05 kcal/mol and a hill width of 2.5 degrees.

2.4. Post-Simulation Analysis

The 2D PMF files output from NAMD were graphed and contoured using the gnuplot²⁹ utility. Molecular conformations extracted from the MD simulations were depicted using the VMD visualization package³⁰ incorporating the Twister visualization algorithm for carbohydrates³¹ to highlight the conformation of the glycan rings. For the di- and trisaccharide solution time series plots, frames 2.5 ps apart in the last 90 ns of the run were extracted and the ϕ, ψ angles extracted using VMD and plotted on the appropriate PMF surface. The trisaccharide conformations from the solution molecular dynamics trajectories were clustered using VMD's internal *measure cluster* command that calculates clusters according to the quality threshold algorithm.³² Clustering was performed with a cut-off of 1.0 on an RMSD fit to the sugar residue ring atoms and the glycosidic linkage oxygens. Only clusters with prevalence greater than 1% of the simulation time were analysed.

3. Results and Discussion

3.1. Force Field Comparisons: M14G

We start our analysis of pneumococcal capsular polysaccharide conformation with the β -D-Man p NAc-(1 \rightarrow 4)- α -D-Glcp disaccharide (M14G) that

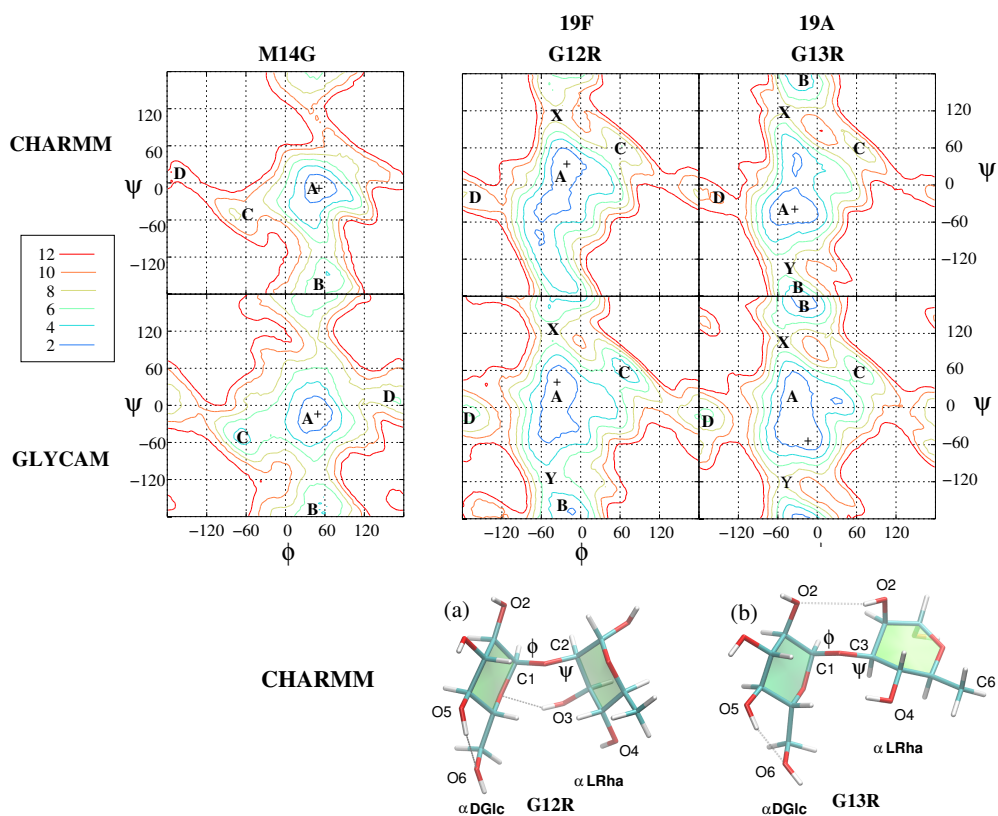


Figure 2:

is common to 19A and 19F. The contoured ϕ, ψ PMF surfaces in vacuum for these linkages (Fig. 2, left column) allow for comparison of the force field conformational predictions for this linkage. In both the CHARMM (top) and GLYCAM (bottom) force fields, a central syn-syn well (A) contains the M14G global minimum, which is at $\phi, \psi = 51^\circ, -6^\circ$ in CHARMM and $\phi, \psi = 49^\circ, -14^\circ$ in GLYCAM (indicated by black crosses in Fig. 2). Both force fields also have a secondary low-energy anti- ψ conformation (B-well) of similar relative energy, at $\phi, \psi = 66^\circ, -151^\circ$ ($4\Delta G = \text{kcal/mol}$) in CHARMM and $\phi, \psi = 54^\circ, -166^\circ$ ($\Delta G = 4 \text{ kcal/mol}$) in GLYCAM, although the well is

broader in GLYCAM.

However, although the tertiary minimum (C) has a similar location across the force fields — at $\phi, \psi = -71^\circ, -59^\circ$ ($\Delta G = 7$ kcal/mol) in CHARMM and $\phi, \psi = -71^\circ, -56^\circ$ ($\Delta G = 3$ kcal/mol) in GLYCAM — the energy is 4 kcal/mol higher with CHARMM. Force fields have fairly complex fitting routines and reasons for differences are typically multifactorial (arising from both bonded and non-bonded interactions) and difficult to pinpoint. However, a contributing factor for energy differences in well C are interactions of the aliphatic ring hydrogens on the glucose and mannose residues: the GLYCAM force field has no partial charges on aliphatic hydrogen atoms, while in CHARMM aliphatic hydrogens are assigned small partial charges of 0.09 electrons.¹⁷ C-well conformations bring the mannose H2 and the glucose H4 into close proximity - the resultant repulsive Coulombic interactions in CHARMM would contribute to the raising of the energy of these conformations. Similar H-H interactions would account for the differences in the high-energy anti- ϕ minimum conformation (well D) in this disaccharide. In GLYCAM the D-well is at $\phi, \psi = 171^\circ, 6^\circ$ ($\Delta G = 6$ kcal/mol), but it is almost entirely absent in CHARMM. D-well conformations bring the aliphatic methyl hydrogens on the N-Acetyl substituent of mannose into close proximity with the glucose ring aliphatic hydrogens.

Despite these differences, the surfaces are very similar, with close agreement on the global and secondary minimum conformations. However, the broader wells and lowered energy of the C-well in GLYCAM means that simulations with this force field will show a more flexible M14G disaccharide, with a broader range of allowed conformations than in CHARMM. Solution

dynamics of this linkage in the trisaccharide repeating units are discussed below.

3.2. Force Field Comparisons: G12R and G13R

We continue our analysis with comparisons of the structure and dynamics for the disaccharides representing the sole point of difference between 19A and 19F : the α -D-Glcp-(**1**→**2**)- α -L-Rhap (G13R) versus the α -D-Glcp-(**1**→**3**)- α -L-Rhap (G12R) dimer. Contoured ϕ, ψ PMF surfaces for these linkages (Fig. 2) enable both comparison of the force field results for a specific linkage (CHARMM surfaces on the top row and GLYCAM on the bottom) and of the $\alpha(1\rightarrow2)$ and $\alpha(1\rightarrow3)$ linkages (G12R left column, G13R right column). We first compare the results for the force fields and then progress to a comparison of the linkages.

3.2.1. G12R (19F)

Vacuum PMFs for the G12R disaccharide, (Fig. 2, left column) produced with the CHARMM (top) and GLYCAM (bottom) carbohydrate force fields both reveal that the (1→2) linkage to the axial O2 in L-Rhamnose creates a relatively flexible dimer: in both surfaces the 4 kcal/mol contour encompasses a wide area. The force fields agree on the general location of the global minimum conformation in the central syn-syn well (A): at $\phi, \psi = -21^\circ, 34^\circ$ for CHARMM and $\phi, \psi = -36^\circ, 41^\circ$ for GLYCAM. In both CHARMM and GLYCAM, the minimum energy conformation combines a lack of steric strain with a stabilising inter-residue O3-O5 hydrogen bonding interaction (shown for CHARMM in Fig. 2a). The energetic advantage of inter-residue hydrogen bonds is expected to be reduced in aqueous solution.³³ The minimum

Table 1:

	Conformational region									
	<i>syn-φ/syn-ψ</i>			<i>syn-φ/anti-ψ</i>			saddle points			
	A-well		C-well	B-well		X	Y			
	ϕ, ψ	(ΔG)	ϕ, ψ	(ΔG)	ϕ, ψ	(ΔG)	ϕ, ψ	(ΔG)		
G12R										
CHARMM	-21, 34	(0)	64 64	(5.9)	-	-34, 114	(8.5)	-		
	-41, -24	(0.7)								
GLYCAM	-36, 41	(0)	74 56	(2.6)	-14, -161	(1.5)	-51, 127	(6.9)	-39, -119	(4.7)
	-41, -19	(0.6)								
MM3 ¹⁵	-12, 49	(1.3)	-							
	-34, -35	(0)								
MM+ ¹⁶	-35, -38	(0)	-							
G13R										
CHARMM	-34, -39	(0)	69, 56	(6.7)	-29, 171	(2.2)	-59 119	(8.2)	-51 -134	(6.7)
	-28, 43	(1.3)								
GLYCAM	-14, -54	(0)	66, 51	(4.2)	-11, 171	(0.8)	-54, 109	(7.5)	-56 -119	(7.8)
	-36, 16	(0.3)								
	29, 6	(1.4)								
MM3 ¹⁵	-36, -40	(0)	-		-28, 175	(4.3)				
	-23, 54	(2.0)								

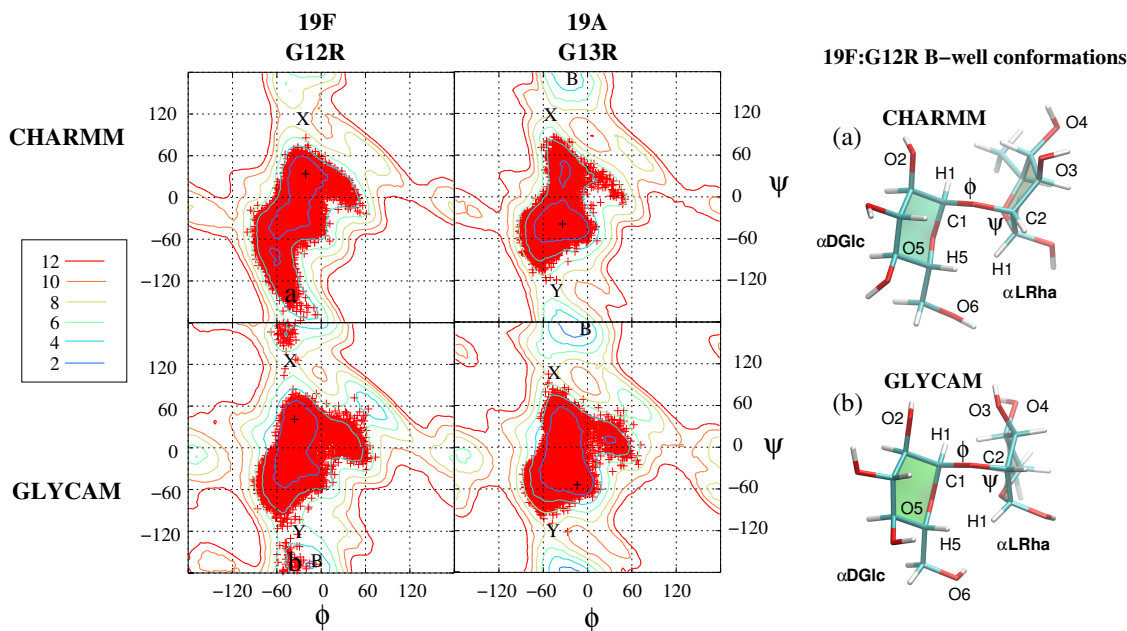


Figure 3:

energy conformation is not a close match with either the global minimum previously calculated for the MM3 force field ($\phi, \psi = -34^\circ, -35^\circ$)¹⁵ or with that estimated for the MM+ ($\phi, \psi = -35^\circ, -38^\circ$)¹⁶ force field (Table 1): in both these cases the negative ψ region of the A-well was favoured.

In contrast, there is a clear difference for the *anti- ψ* well in the G12R PMF's for the two force fields (labelled B in Fig. 2). GLYCAM has a distinct secondary minimum energy well ($\phi, \psi = -14^\circ, -161^\circ$) and a $\Delta G = 4.2$ kcal/mol barrier (Y) separates the A-well and B-well (Table 1). However, in the CHARMM force field the A- and B-wells coalesce: the A-well extends in a long broad valley to the B-region, with no barrier at Y and no clear minimum at B (the point of lowest energy in this region is at $\phi, \psi = -41^\circ, -136^\circ$).

In addition, both force fields predict a small tertiary energy minimum

(C) — at $\phi, \psi = 64^\circ, 64^\circ$ with the CHARMM model and $\phi, \psi = 74^\circ, 56^\circ$ with GLYCAM — but it is considerably more favoured in GLYCAM ($\Delta G = 2.6$ kcal/mol versus 5.9 kcal/mol, Table 1). The quaternary syn- ϕ , anti- ψ minimum (D) is also much more favoured in GLYCAM (5.2 kcal/mol) than CHARMM (7.6 kcal/mol). However, in both force fields this minimum is too high in energy to be extensively explored at standard temperature and pressure, as is the barrier at X ($\Delta G = 8.5$ kcal/mol in CHARMM and 6.9 kcal/mol in GLYCAM). This is demonstrated in the ϕ, ψ time series of 100 ns unbiased runs in aqueous solution superimposed on the PMF surfaces (Fig. 3, left column). While the GLYCAM simulation (bottom) shows slight occupation of the C-well and X-barrier, they are not occupied at all in the CHARMM simulation.

The time series also shows the effect of the differences in the A- and B-well’s between the two force fields on G12R disaccharide dynamics. While in the CHARMM run G12R occupies the extended A-well, including the valley stretching to the B-region, the GLYCAM simulation shows only a small, distinct B-well population, with little presence on either the X- or Y-saddle region. Snapshots of anti- ψ conformations in aqueous solution for CHARMM ($\phi, \psi = -51^\circ, -140^\circ$) and GLYCAM ($\phi, \psi = -178^\circ, -52^\circ$) are shown in Fig. 3. The low-energy anti- ψ region has few steric clashes and the “flipped” conformation brings the equatorial Glc H1 into close proximity with the Rha O5 ring oxygen and, conversely, the Rha H1 into close proximity (≈ 3 Å) with the Glc O5 ring oxygen. This is another instance where Coulombic interactions involving aliphatic hydrogens play a role. With the small partial charges assigned to these groups in CHARMM, pairwise attractive H1-O5 Coulombic

interactions will contribute to lowering the the energy of the Y-region conformations and, conversely, repulsive H1-H5 and H4-H1 interactions would raise the energy of tighter B-well conformations, which bring these aliphatic hydrogens atoms into closer proximity. These interactions are absent in the GLYCAM force field. Charges on the aliphatic hydrogens also contribute to the higher energy of the D-well in CHARMM, which brings the Rha H2 and H1 into close contact with the glucose H3 and H5 atoms. Note that, although aliphatic charges also affect the C6 methyl group in rhamnose, this does not influence conformations of the G12R disaccharide as the methyl group is located far away from the glycosidic linkage and is not in contact with the glucose residue.

3.2.2. G13R (19A)

The ϕ, ψ free energy maps for the G13R disaccharide (Fig. 2, right column) produced for the CHARMM (top) and GLYCAM (bottom) force fields show a closer agreement than is the case for the G12R disaccharide. The central syn-syn well (A) contains the global minimum, at $\phi, \psi = -34^\circ, -39^\circ$ for CHARMM and $\phi, \psi = -14^\circ, -54^\circ$ for GLYCAM. In both models, this conformation in vacuum is stabilised by an inter-residue hydrogen bond between the equatorial O2 on the glucose and the axial O2 on the rhamnose residues (Fig. 2b). Both force fields show two conformational regions within the A-well: the favoured region corresponding to negative ψ values; and a positive ψ region, corresponding to the global minimum for the $\alpha(1 \rightarrow 2)$ linkage. The A-well is again broader and shallower in GLYCAM, with the secondary local minimum in the well lower in energy in the GLYCAM model ($\phi, \psi = -36^\circ, 16^\circ, \Delta G = 0.3$ kcal/mol, Table 1) than in CHARMM ($\phi, \psi = -28^\circ, 43^\circ, \Delta G = 1.3$

kcal/mol). Prior calculations with the HSEA and MM2 force fields estimated the global minimum conformation for G13R at $\phi, \psi = -36^\circ, -40^\circ$ ¹⁵, which is close to the CHARMM minimum for this disaccharide.

In contrast to G12R, for G13R the distinct secondary *anti- ψ* minimum energy well B appears in both models at approximately the same location: CHARMM $\phi, \psi = -29^\circ, 171^\circ$ and GLYCAM $\phi, \psi = -11^\circ, 171^\circ$. However, the B-well is 1 kcal/mol lower, and hence slightly preferred, in GLYCAM (Fig. 1). Prior molecular mechanics calculations with the MM2 force fields also identified this B minimum, albeit at a high enthalpy, at $\phi, \psi = -28^\circ, -175^\circ$ ¹⁵ again closest to the CHARMM prediction.

The higher Y-barrier in the G13R disaccharide (6.7 kcal/mol in CHARMM model and 7.8 kcal/mol in GLYCAM) makes transition to *anti- ψ* conformations unlikely at standard temperature and pressure, as is clearly seen in ϕ, ψ time series of 100 ns unbiased runs in aqueous solution superimposed on the PMF surfaces (Fig. 3, right column). Over this time period, no transitions to the B-well occurring for either force field. The GLYCAM model also shows increased occupation of the third, tertiary energy minimum at C.

In general, for all the PMF's calculated here, CHARMM and GLYCAM show close agreement in the minimum energy conformations identified. However, there is a general pattern of GLYCAM having broader minimum wells and secondary minima lower in energy in relative to the global minima than is the case for CHARMM. This results in more flexible linkages in the GLYCAM than in the CHARMM model.

3.3. Comparison of the 1→2 (19F) and 1→3 (19A) linkages

The early 1992 Molecular Mechanics study with the MM3 force field did not find a significant difference between the energy surfaces for G12R and G13R, and predicted that Glc-Rha linkages in general showed rather rigid conformations.¹⁵ The subsequent development of carbohydrate-specific force fields has altered this picture. Our calculations in both CHARMM and GLYCAM indicate that the linkage to the Rhamnose O2 (axial) in G12R results in fewer steric clashes and a more flexible linkage, with a higher probability of anti- ψ conformations than is the case for G13R. Both force fields predict a significant rise of the barriers to rotation to *anti- ψ* configurations on mutation of an $\alpha(1\rightarrow2)$ to an $\alpha(1\rightarrow3)$ glycosidic linkage. This effect is more pronounced in the CHARMM model, which lacks a clear Y barrier for G12R and has 6.7 kcal/mol barrier for G13R. The GLYCAM model predicts an approximate 3 kcal/mol rise in the Y saddle point, to 7.8 kcal/mol in G13R. In contrast to G12R, for G13R the distinct secondary *anti- ψ* minimum energy B-well appears in both models.

The key difference for both force fields is the prevalence of B-well conformations in G12R and opposed to G13R: *anti-* conformations much more likely in G12R (19A) than G13R (19F). While our vacuum PMF calculations with both force fields predict a ψ shift of the global energy minimum on mutation of an $\alpha(1\rightarrow2)$ to an $\alpha(1\rightarrow3)$ glycosidic linkage into the negative ψ region, in aqueous solution this shift has little impact: the entire 4 kcal/mol regions is well occupied for the 90 ns simulations at standard temperature and pressure.

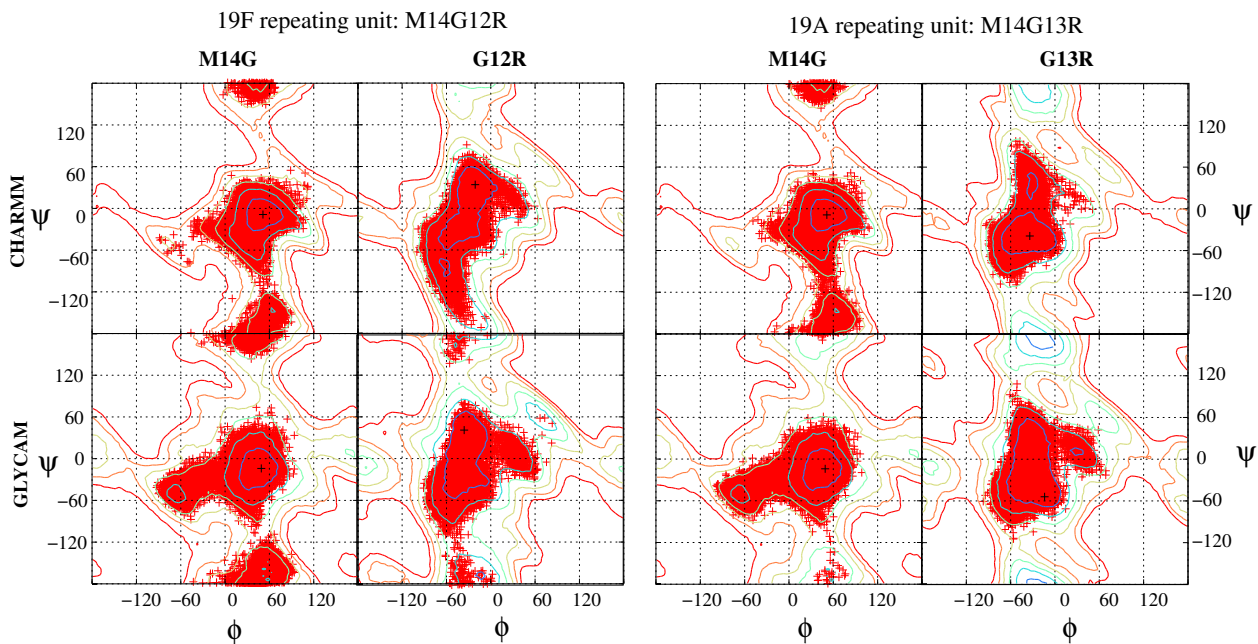


Figure 4:

3.4. Extension to repeating units: $M14G12R$ and $M14G13R$

We now explore the aqueous dynamics of the pneumococcal capsular polysaccharide repeat units in 19F and 19A. Fig. 4 compares the ϕ, ψ time series for 90 ns aqueous simulations of the trisaccharide repeating units in 19F ($M14G12R$, left column) and 19A ($M14G13R$, right column) with the CHARMM (top row) and GLYCAM (bottom row) force fields.

As is expected from inspection of the M14G vacuum PMFs, the M14G C-well (labelled in Fig. 2) is extensively visited in both of the GLYCAM trisaccharide simulations, but only briefly in the $M14G12R$ CHARMM simulation. In common with GLYCAM (but not CHARMM), a prior 500 ps MD simulation of the 19F trisaccharide repeating unit in vacuum with the MM+ force field reported two major conformations of the M14G linkage, at $\phi, \psi = 19^\circ, -$

51° and 43°,5°.¹⁶ In contrast, M14G B-well anti- ψ conformations are prevalent in both CHARMM trisaccharide simulations, but, while the GLYCAM M14G12R simulation shows anti- ψ populations equal to CHARMM, anti- ψ M14G conformations only occur briefly in the GLYCAM M14G13R simulation.

From a comparison of Fig. 3 and Fig. 4, it is clear that the addition of the ManNAc residue to the G12R disaccharide does not affect the dynamics around the G12R and G13R linkages significantly: the time series are qualitatively very similar. Once again, there is good agreement in the behaviour of the G12R and G13R linkages across the force fields. In both simulations, the G12R linkage shows a wider range of conformations than the G13R linkage; neither simulation exhibits anti- ψ conformations for the G13R linkage (19A). However, there are some differences in the force field simulations for the G12R linkage in M14G12R. As was the case for the G12R disaccharide, the trisaccharide in CHARMM explores the long narrow valley extending to anti- ψ conformations, but does not show the distinct population of anti- ψ conformations apparent for GLYCAM.

Clustering of the conformations from the trisaccharide aqueous simulations produced the conformational families shown in Fig. 5. For both force fields, the M14G12R trisaccharide shows a broader range of conformations. The CHARMM M14G12R simulations (top left) have four conformational families: a major conformational population representing central A-well conformations for both the M14G and G12R linkages (92%, blue), a small secondary population distinguished by an anti- ψ B-well bend in the M14G linkage (4%, red), and two minor conformers representing positive (2%, cyan)

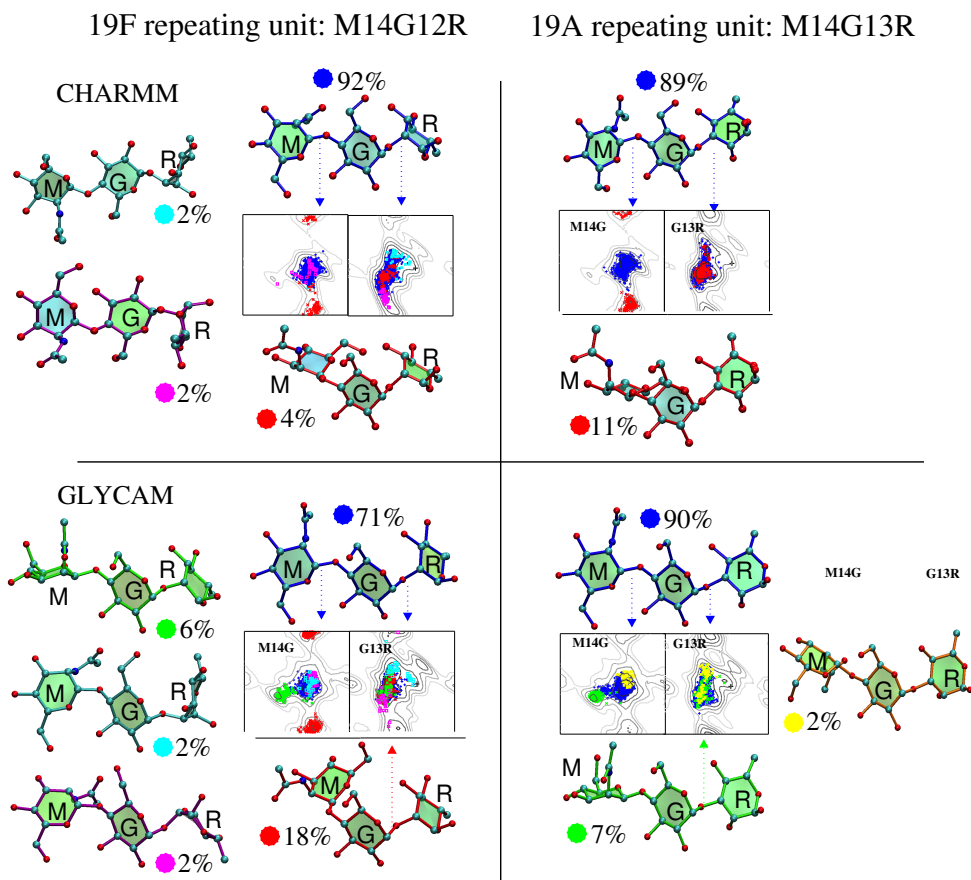


Figure 5:

and negative (2%, purple) ψ -bends in the G12R linkage. GLYCAM has five M14G12R conformational families. Four of these correspond to those seen for the CHARMM simulations: the major conformational population represents A-well conformations for both the M14G and G12R linkages (71%, blue), a secondary population is distinguished by an anti- ψ B-well bend in the M14G linkage (18%, red) and two minor conformers represent positive (2%, cyan) and negative (2%, mauve) ψ -bends in the G12R linkage. In addition, the GLYCAM simulations for M14G12R show a conformational family

associated with M14G C-well conformations (6%, green), the consequence of this well being considerably lower in energy in GLYCAM than in CHARMM. Note that, although the range of conformations is similar for the two force fields, GLYCAM predicts a lower prevalence of the primary conformation and a higher prevalence of conformations with anti- ψ B-well bend in the M14G linkage. Clustering also reveals that, for 19F, large ψ rotations are never associated with bent “flipped” anti- ψ conformations of the M14G linkage.

For both force fields, the M14G13R trisaccharide is less conformationally varied than the corresponding (1 \rightarrow 2)-linked trisaccharide: there are no significant populations showing anti- ψ conformations for the G13R linkage. For the M14G12R CHARMM simulation, only two principle populations are seen: a major conformational population representing central A-well conformations for both the M14G and G12R linkages (89%, blue) and a small secondary population distinguished by an anti- ψ B-well bend in the M14G linkage (11%, red). In the GLYCAM simulation, none of the principle populations show “flipped” anti- ψ conformations of the M14G linkage. There is the usual major conformational population representing A-well conformations for both the M14G and G12R linkages (90%, blue), a conformational family associated with M14G C-well conformations as seen for M14G12R (7%, green), and an additional conformational family associated with M14G conformations in the upper-right corner of the A-well (2%, yellow).

Thus, the primary difference between the force fields is a distinct population in both the GLYCAM simulations associated with C-well conformations of the M14G linkage. Such conformations do not occur in the CHARMM trisaccharide simulations. In general, the GLYCAM force field shows more

flexible linkages with a wider range of conformations than CHARMM. However, despite this, the conformation families predicted by the CHARMM and GLYCAM force fields are remarkably similar. The principle difference between the M14G12R and M14G13R trisaccharides in both force fields is that M14G12R has a much wider range of conformations than M14G13R. M14G12R has two families of “bent” conformations of the G12R linkage (the cyan and purple conformations in the left column of Fig. 5), which do not occur for the G13R linkage. These conformations, if prevalent in a polysaccharide, would result in a bend at the point of the phosphodiester linkage. However, because of the great flexibility of the phosphodiester linkage, the affect of such a bend on polysaccharide conformation is difficult to predict. Factors to be considered are the potential for intra-chain hydrophobic interactions with the aliphatic methyl group and potential counter-ion interactions with the charged phosphate group.

4. Conclusions and Future Work

We performed conformational analysis of the repeating units in pneumococcal capsular polysaccharides serotypes 19A and 19F. Our calculations reveal clear differences in the allowed conformations of serotypes 19A and 19F trisaccharide repeating units, with the $\text{Glc}_p\text{-}(\mathbf{1}\rightarrow\mathbf{2})\text{-}\alpha\text{-L-Rhap}$ linkage showing “bent” conformational families that do not occur for the $\text{Glc}_p\text{-}(\mathbf{1}\rightarrow\mathbf{3})\text{-}\alpha\text{-L-Rhap}$ linkage. This is in contrast to previous modelling studies, which predicted no significant difference between the linkages. This difference in conformational preference could result a change in molecular conformation significant enough to affect binding of the antibody to the polysaccharide

antigen and thus account for the lack of cross-protection of serotype 19F vaccination against serotype 19A.

It is encouraging that the CHARMM and GLYCAM carbohydrate force fields showed close agreement in the minimum energy conformations identified. However, generally GLYCAM exhibited more flexible glycosidic linkages, with broader minimum wells and secondary minima lower in energy in relative to the global minima than was the case for CHARMM. Small charges on the aliphatic hydrogens present in CHARMM, but not in GLYCAM, can account for some of the differences seen for the linkages examined here.

This work represents the first stage in analysis of a complex system and the effect of the flexible phosphodiester linkages and counter-ions, inter-residue interactions and increasing chain length have still to be determined, once suitable force field parameters for the phosphodiester linkage are available in CHARMM or GLYCAM carbohydrate force fields. In addition, depending on availability of suitable samples, predicted conformations could be validated by use of NMR spectroscopy.

5. Acknowledgements

Computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: <http://hpc.uct.ac.za>. We thank the South African National Research Foundation funding agency for financial support for this work.

References

- [1] World Health Organisation, *Wkly Epidemiol Rec.* 82 (12) (2007) 93–104.

- [2] K. L. O'Brien, L. J. Wolfson, J. Watt, et al., *Lancet* 374 (9693) (2009) 893–902.
- [3] A. Ginsburg, M. Alderson, *Drugs Today* 47 (3) (2011) 207–212.
- [4] H. L. Johnson, M. Deloria-Knoll, O. S. Levine, S. K. Stoszec, L. F. Hance, R. Reithinger, L. Muenz, K. L. O'Brien, *PLoS Med.* 7 (10) (2010) e1000348.
- [5] WHO Meeting report., 26–27 October 2006, Geneva, *Vaccine* 25 (2007) 6557–6564.
- [6] Center for Disease Control, *MMWR* 49 (RR-9) (2000) 1–38.
- [7] J. Schranz, *Procedia Vaccinol.* 1 (2009) 189–205.
- [8] D. J. Isaacman, E. D. McIntosh, R. R. Reinert, *Int J Infect Dis* 14 (2010) e197–e209.
- [9] K. Elberse, S. Witteveen, H. van der Heide, I. van de Pol, C. Schot, A. van der Ende, G. Berbers, L. Schouls, *PLoS ONE* 6 (9) (2011) e25018.
- [10] W. Egan, R. Schneerson, K. E. Werner, G. Zon, *J. Am. Chem. Soc.* 104 (1982) 2898–2910.
- [11] M. E. Pichichero, J. R. Casey, *JAMA* 298 (15) (2007) 1772–1778.
- [12] R. R. Reinert, M. R. Jacobs, S. L. Kaplan, *Vaccine* 28 (2010) 4249–4259.
- [13] C. G. Whitney, T. Pilishvili, M. M. Farley, W. Schaffner, A. Craig, et al., *Lancet* 368 (2006) 1495–1502.

- [14] J. Poolman, C. Frasch, A. Nurkka, H. Käyhty, R. Biemans, L. Schuerman, *Clin. Vaccine Immunol.* 18 (2) (2011) 327–336.
- [15] P. Ciuffreda, D. Colombo, F. Ronchetti, L. Toma, *Carbohydr. Res.* 232 (2) (1992) 327–339.
- [16] L. Legnani, S. Ronch, S. Fallarini et al., *Synthesis, Org. Biomol. Chem.* 7 (2009) 4428–4463.
- [17] E. Fadda, R. J. Woods, *Drug Discov Today* 15 (15/16) (2010) 596–609.
- [18] O. Guvench, S. S. Mallajosyula, E. P. Raman, E. Hatcher, K. Vanommeslaeghe, T. J. Foster, F. W. Jamison, II, A. D. MacKerell, Jr., *J. Chem. Theory Comput.* 7 (2011) 3162–3180.
- [19] K. Kirschner, A. Yongye, S. Tschampel, J. Gonzalez-Outeirino, C. Daniels, B. Foley, R. Woods, *J. Comput. Chem.* 29 (2008) 622–655.
- [20] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys* 79 (2) (1983) 926–935.
- [21] S. S. Mallajosyula, O. Guvench, E. Hatcher, A. D. MacKerell, Jr., *J. Chem. Theory Comput.* 8 (2012) 759–776.
- [22] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten., *J. Comput. Chem.* 26 (2005) 1781–1802.
- [23] <http://www.ks.uiuc.edu/Research/vmd/plugins/psfgen/> [last accessed November 2013].

- [24] http://glycam.ccruc.uga.edu/ccrc/biombuilder/biomb_index.jsp [last accessed November 2013].
- [25] A. D. MacKerell, Jr., personal communication (October 2013).
- [26] S. E. Feller, Y. Zhang, R. W. Pastor, B. R. Brooks, *J.Chem. Phys.* 103 (11) (1995) 4613–4621.
- [27] G. J. Martyna, D. J. Tobias, M. L. Klein, *J. Chem. Phys* 101 (5) (1994) 4177–4189.
- [28] A. Laio, M. Parrinello, *Proc Natl Acad Sci USA* 99 (2002) 12562–12565.
- [29] <http://www.gnuplot.info> [last accessed November 2013].
- [30] W. Humphrey, A. Dalke, K. Schulten, *J. Molec. Graphics* 14 (1996) 33–38.
- [31] M. Kuttel, J. Gain, A. Burger, I. Eborn, *J. Mol. Graphics Modell.* 25 (2006) 380–388.
- [32] L. J. Heyer, S. Kruglyak, S. Yooseph, *Genome Res.* 9 (1999) 1106–1115.
- [33] M. M. Kuttel, *Carbohydr. Res.* 343 (6) (2008) 1091–1098.

Figure and Table Captions

Table 1. Values of the ϕ, ψ dihedral angles (degrees) for local minima and saddle points on the PMF surfaces for the G12R and G13R disaccharides. Approximate free energy values (in kcal/mol) relative to the global energy minimum are quoted in parenthesis. Values from prior Molecular Mechanics studies are included, but here energies denote enthalpies, not free energy.

Figure 1. Schematic diagram of the trisaccharide repeating units of the capsular polysaccharides in pneumococcal serotypes 19F and 19A.

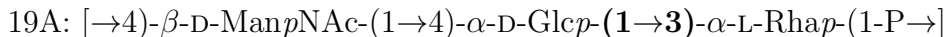
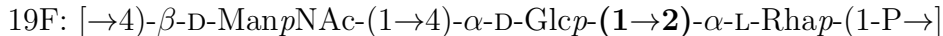


Figure 2. Contoured disaccharide ϕ, ψ PMF surfaces in vacuum produced with the CHARMM (top) and GLYCAM (bottom) carbohydrate force fields: M14G (common to 19A/19F, left column), G12R (middle column) and G13R (right column). Contours are drawn at 2 kcal/mol intervals, to a maximum cut-off of 12 kcal/mol. Locations of the global energy minima are indicated by black crosses. Minimum energy regions are labeled A through D and the X- and Y-saddle points are also indicated for G12R and G13R. Structures shown are snapshots of the G12R and G13R disaccharides in A-well minimum energy conformations taken from a vacuum simulation with the CHARMM force field. The ϕ and ψ labels on the glycosidic linkage identify the central bond in these dihedral angles. (a) G12R (19F): $\phi, \psi = -21^\circ, 34^\circ$. (b) G13R

(19A): $\phi, \psi = -34^\circ, -39^\circ$. Stabilising hydrogen bonding interactions are indicated by dashed lines.

Figure 3. Scatterplots of the ϕ, ψ time series for 90 ns unbiased runs in aqueous solution with the G12R (left column) and G13R (right column) disaccharides using the CHARMM (top row) and GLYCAM (bottom row) force fields. Scatter plots are superimposed on the corresponding vacuum PMF surfaces. Molecular structures shown are snapshots of the G12R disaccharide in B-well anti conformations taken from aqueous solution simulations. Stabilising hydrogen bonding interactions are indicated by dashed lines. The ϕ and ψ labels on the glycosidic linkage identify the central bond in these dihedral angles. (a) CHARMM conformation with $\phi, \psi = -51^\circ, -140^\circ$. (b) GLYCAM conformation with $\phi, \psi = -178^\circ, -52^\circ$.

Figure 4. Scatterplots for 90 ns unbiased runs in aqueous solution of the ϕ, ψ time series in both linkages of the of the capsular polysaccharide trisaccharide repeating units in pneumococcal serotypes 19F (M14G12R, left column) and 19A (M14G13R, right column) using the CHARMM (top row) and GLYCAM (bottom row) force fields. Scatter plots are superimposed on the corresponding vacuum PMF surfaces.

Figure 5. Comparison of representative structures from the principle conformational families formed in solutions simulations of the trisaccharide

repeating units in 19F (left column) and 19A (right column) using the CHARMM (top row) and GLYCAM (bottom row) carbohydrate force fields. The ϕ, ψ time series in the M14G and G12R/G13R linkages are indicated on the corresponding PMF surfaces. The principle cluster is shown in blue, followed by secondary clusters in descending order in red, green, cyan and purple. Population percentages of each of the clusters during the simulation are shown.