Identification of rare Lewis oligosaccharide conformers in aqueous solution using enhanced sampling molecular dynamics

Irfan Alibay[†], *Kepa K. Burusco*[†], *Neil J. Bruce*[‡] *and Richard A. Bryce*[*†]

† Division of Pharmacy and Optometry, School of Health Sciences, Manchester Academic Health Sciences Centre, University of Manchester, Oxford Road, M13 9PL, UK

‡ Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnenweg 35, 69118, Heidelberg, Germany

KEYWORDS: Enhanced sampling, carbohydrates, molecular dynamics, swarm-enhanced sampling, accelerated molecular dynamics, umbrella sampling, oligosaccharide conformation, hydrogen mass repartitioning (HMR), sialyl Lewis X, sialyl Lewis A

Corresponding Author

*Richard Bryce, Division of Pharmacy and Optometry, School of Health Sciences, University of Manchester, Manchester, M13 9PT, U.K. Email: R.A.Bryce@manchester.ac.uk, Tel: (0)161-275-8345, Fax: (0)161-275-2481; ORCID 0000-0002-8145-2345

Abstract

Determining the conformations accessible to carbohydrate ligands in aqueous solution is important for understanding their biological action. In this work, we evaluate the conformational free energy surfaces of Lewis oligosaccharides in explicit aqueous solvent using a multidimensional variant of the swarm-enhanced sampling molecular dynamics (msesMD) method; we compare with multi-microsecond unbiased MD simulations, umbrella sampling and accelerated MD approaches. For the sialyl Lewis A tetrasaccharide, msesMD simulations in aqueous solution predict conformer landscapes in general agreement with the other biased methods and with triplicate unbiased 10 μs trajectories; these simulations find a predominance of closed conformer and a range of low occupancy open forms. The msesMD simulations also suggest closed-to-open transitions in the tetrasaccharide are facilitated by changes in ring puckering of its GlcNAc residue away from the $^4C_1$ form, in line with previous work. For sialyl Lewis X tetrasaccharide, msesMD simulations predict a minor population of an open form in solution, corresponding to a rare lectin-bound pose observed crystallographically. Overall, from comparison with biased MD calculations, we find that triplicate 10 μs unbiased MD simulations may not be enough to fully sample glycan conformations in aqueous solution. However, the computational efficiency and intuitive approach of the msesMD method suggest potential for its application in glycomics as a tool for analysis of oligosaccharide conformation.

1. Introduction

Glycoscience is an important emerging area with potential applications from personalized medicines and food security, to biofuels and advanced biomaterials.[2] Key to the rational design of carbohydrate-based compounds is the ability to accurately understand their conformational behavior.[3] Achieving this, however, can be challenging as carbohydrates can exhibit subtle conformational preferences. In this regard, the use of computational methods have proven useful; in particular, the application of molecular dynamics (MD) simulations is capable of providing conformational landscapes in atomistic detail not readily accessible *via* experimental approaches. For example, conformational analysis of the core pentasaccharide derived from the lipooligosaccharide of pathogenic bacterium *Moraxella catarrhalis* was recently performed using molecular simulation and NMR. It was shown that the addition of one glucosyl unit to each of the α-(1$\rightarrow$4) and α-(1$\rightarrow$2) branches of this core pentasaccharide led to a significant change in conformation, into a more folded compact structure,[4] with significant potential implications for the minimum epitope to target in vaccine design.

In another recent study, Blaum et al.[5] examined the pentasaccharide glycan component of ganglioside GM1, a membrane-bound ligand for a range of proteins, including those of viruses and bacteria. In their work, they identified an alternative conformation of GM1, exhibited in its complex with a 155 kDa protein, Factor H. Subsequent long time scale MD simulations (10 μs) of GM1 in explicit solvent identified this structure as a secondary low-lying minimum on the free energy surface. Interestingly, this alternative conformation of GM1, although consistent with NMR data from a prior combined NMR/modelling study, did not correspond to a good set of NOE restraints and thus had not been reported.[6]

The Lewis blood group oligosaccharides constitute key mediators of biomolecular recognition.[7] This blood group mainly consists of six oligosaccharides (Figure 1), all of which share a common trisaccharide core comprising galactose (Gal), N-acetylglucosamine (GlcNAc) and fucose (Fuc) residues. The six Lewis sugars are divided into two types, I and II, based on the linkage configuration of this core. For type I Lewises, the core trisaccharide is Galβ(1→3)(Fucα(1→4))GlcNAcβ, denoted Le^a (Figure 1a); this can be further sialylated (sLe^a, Figure 1b) or fucosylated (Le^b, Figure 1c). For type II sugars, the glycosidic linkage connectivity of the core trisaccharide is interchanged to give Galβ(1→4)(Fucα(1→3))GlcNAcβ, known as Le^x (Figure 1d); as for Type I Lewises, this can be further sialylated (sLe^x, Figure 1e) or fucosylated (Le^y, Figure 1f). While important contributors to normal physiological function, the Lewis oligosaccharides have also been associated with several disease states due to their role in the immune system, including pathogenic inflammatory response and carcinomas.[8-10] Characterising their conformational behavior is therefore useful in further understanding their role in these disease states and potentially assisting in the development of novel therapeutics.

The Lewis oligosaccharides have generally been viewed as somewhat rigid structures, adopting a compact closed conformation at equilibrium, where a stable stacking interaction is formed between the fucose and galactose rings (Figure 2a).[11] Although the presence of more open non-stacked structures has previously been suggested by both NMR and computational studies,[12-13] this could not be confirmed.[14] Recently, however, unusual conformations adopted by Lewis antigens Le^x and sLe^x were reported by Topin and co-workers.[1] In this case, crystal structures of Le^x and sLe^x bound to pathogenic *Ralstonia solanacearum* lectin (RSL) were found

to adopt poses distinct from the closed shape. Subsequent long time scale MD simulation of Le^x in explicit solvent by Topin et al.[1] detected transitions from closed to open conformations on the μs timescale. These transitions however were rather infrequent: of the 30 MD trajectories acquired, five showed transitions from the closed to open forms of Le^x; of the combined 75 μs of simulation, open forms comprised 1% of the total solution ensemble.

These Le^x simulations highlight the need for long timescale trajectories to characterize as fully as possible the ensemble of glycan structures.[14-16] However, despite recent advances in hardware development and utilization,[13,14,17-20] exhaustively characterizing the multi-microsecond landscapes to identify rare conformational states,[1] remains a time consuming task. Recently, enhanced sampling MD techniques[21,22-23,24,25-27] have been applied to the challenge of exploring carbohydrate conformation.[23, 28-34] Here, we map out the conformational landscape of Lewis oligosaccharides in aqueous solution, introducing and applying a computationally efficient variant of our earlier biased MD method.[35-36] We term this variant multi-dimensional swarm-enhanced sampling molecular dynamics (msesMD). We first assess the ability of msesMD to explore sLe^a conformations in explicit aqueous solvent, comparing with multi-microsecond unbiased MD simulations, umbrella sampling and accelerated molecular dynamics (aMD). We then apply both msesMD and multi-microsecond unbiased MD simulations to investigate the ability of sLe^x and the unmodified Lewis cores, Le^a and Le^x, to adopt non-closed conformations in aqueous solution.

2. Computational methods

2.1 msesMD methodology

Previously we have introduced a multicopy molecular dynamics method for enhanced conformational sampling, based on the coupling of a swarm of simulation replicas via attractive and repulsive pair potentials acting on dihedral angles of interest.[35-36] Specifically this potential is a function of the root-mean-square (RMS) dihedral angle distance over all *K* dihedrals *j* between swarm replicas α and β, ie. $(M^{-1}\sum_j^K\left(\varphi_j^\alpha - \varphi_j^\beta\right)^2)^{1/2}$. Here we introduce a variant of this swarm methodology, which we term the multi-dimensional swarm enhanced sampling method (msesMD). This approach decomposes the biasing swarm potential U^{ses} into a series of sub-swarm energy terms:

$$U^{ses}(\{\varphi^\alpha\}) = \sum_\alpha^M U_\alpha^{ses}$$

$$= \frac{1}{2}\sum_\alpha^M\sum_{\beta\neq\alpha}^M\sum_j^K\left(Aexp\left[-Bd_j^{\alpha\beta}(\varphi^\alpha,\varphi^\beta)\right] - Cexp\left[-Dd_j^{\alpha\beta}(\varphi^\alpha,\varphi^\beta)\right]\right) \quad (1)$$

where $d_j^{\alpha\beta}(\varphi^\alpha,\varphi^\beta)$ is given by $\left|\varphi_j^\alpha - \varphi_j^\beta\right|$ and A - D are parameters for the repulsive (A, B) and attractive components (C, D). The dihedral-specific swarm energy evaluations entailed by Eq (1) are inherently suited to distribution with the domain decomposition algorithm as implemented in the molecular dynamics framework of the *pmemd* module of AMBER14.[37]

2.2. Computational details

Solvated oligosaccharides **a - d** (Figure 1) were modelled using the Glycam06j-1[38] and TIP3P[39] force fields for the sugars and water respectively. Each oligosaccharide was built and then solvated in an octahedral water box with solvent molecules placed up to a minimum of 12 Å away using the *leap* module in AmberTools14.[37] Sodium ions were added to neutralize the solute charge where appropriate.

Molecular dynamics simulations used the *pmemd* and *pmemd.cuda* modules of AMBER14,[37] with the former having been modified to support the msesMD methodology. A hydrogen mass repartitioning scheme (HMR)[40] was used for all unbiased MD simulations, scaling solute hydrogen masses to 3 amu;[41] all msesMD, aMD and umbrella sampling simulations used standard masses. Unless otherwise indicated, a time step of 4 fs was used for mass repartitioned systems and 2 fs for standard mass systems. In all simulations, hydrogen atom motions were constrained using the SHAKE[42] and SETTLE[43] algorithms for the solutes and waters respectively. Long range electrostatic interactions were calculated using the particle mesh Ewald method and a 9 Å cutoff for short range nonbonded interactions. Thermal control was achieved using the Langevin thermostat[44] and a collision frequency of 3.0 ps^{-1}.

Equilibration. After energy minimization, all systems were heated under NVT conditions, where the temperature was raised from 0 K to 298 K over 500 ps. This was then followed by 1 ns of NPT simulation to equilibrate box densities at a pressure of 1 bar using the Monte Carlo barostat,[44] with volume exchange attempts every 100 steps. This was then followed by a final 1 ns of NVT equilibration at 298 K. At the end of the equilibration phase, all Lewis solutes were observed to occupy the closed conformational state.

Unbiased simulations. For all oligosaccharide systems, three replicate unbiased NVT simulations at 298 K were carried out with a coordinate sampling frequency of 5 ps. Each replicate simulation was performed for a total of 10 μs.

Umbrella sampling calculations. One-dimensional potential of mean force (PMF) profiles were computed using umbrella sampling along the ψ angles of the Fucα(1→4)GlcNAc and Galβ(1→3)GlcNAc glycosidic linkages of sLe^a, denoted ψ_F and ψ_G respectively (Figure 1b). A total of 120 window simulations along each ψ torsion were computed, subjected to a harmonic biasing force of 200 kcal mol^{-1} rad^{-2}. Distinct from the other enhanced sampling simulations in this work, a 1 fs time step was used to improve stability. The systems were equilibrated using the equilibration protocol above, and then each window, under the influence of its respective biasing potential, was subject to an additional 500000 steps of energy minimization and 1 ns of NVT equilibration prior to 20 ns of NVT production dynamics, recording dihedral values every 50 steps.

Swarm-enhanced sampling simulations. Swarm-enhanced sampling simulations were conducted for all systems using the msesMD potential detailed above. After system equilibration, eight independent and unbiased NVT trajectories were propagated from the equilibrated structure at 298 K for 1 ns; and then coupled via msesMD in the NVT ensemble by gradually introducing the swarm potential over a period of 600 ps, followed by a further 5.4 ns of equilibration under the full influence of the biasing potential. This was then followed by 250 ns per replica of production simulation. Parameters for the msesMD boost potential were taken as A = 3.195 kcal

mol^{-1}, B = 2.625 rad^{-1}, C = 0.75 kcal mol^{-1} and D = 0.5 rad^{-1}, based on short (5 ns) trial simulations of Le^a, mannotriose and alanine dipeptide (data not shown) and guided by the extent of torsional sampling, range of boost energies and normality of their distribution. The swarm boost potential was applied to every φ/ψ glycosidic torsion in the systems of interest.

Accelerated molecular dynamics simulations. Similar to the msesMD simulations, eight independent trajectories were spawned from an equilibrated system and simulated for an extra 1 ns each in the NVT ensemble. An aMD dual boost potential was then applied using boost parameters E_{dih} = 1.7 kcal mol^{-1} and α_{dih} = 3.2 kcal mol^{-1} for the dihedral potential energies; and E_{tot} = -15000.3 kcal mol^{-1} and α_{tot} = 822.7 kcal mol^{-1} for the total potential energies. It is noted that the aMD parameters were derived from previous empirical parameter results for peptides and proteins.[45] Each independent aMD simulation was carried out for a total of 500 ns, with the first 6 ns being discarded as equilibration.

Analysis. Simulation analyses were primarily carried out using *cpptraj*[46] in AmberTools16[47] and in-house python scripts. Additionally, the WHAM code by the Grossfield group[48] and the PyReweighting scripts from the McCammon group[45] were used to analyze the umbrella sampling and aMD simulations respectively. Glycosidic torsions were defined as $\varphi = O_5\text{-}C_1\text{-}O\text{-}C_n$, $\psi = C_1\text{-}O\text{-}C_n\text{-}C_{(n-1)}$ for $\alpha/\beta(1\rightarrow n)$ linkages; and $\varphi = O_6\text{-}C_2\text{-}O\text{-}C_3$, $\psi = C_2\text{-}O\text{-}C_3\text{-}C_4$ for the $\alpha(2\rightarrow 3)$ linkages of sLe^a and sLe^x. Free energy surfaces for the glycosidic linkages were generated using and the expression $\Delta A = -k_B T ln(\rho_x/\rho_{max})$, where ρ_x is the bin density, ρ_{max} is the highest occupied bin density, k_B is the Boltzmann constant and T the temperature. Bin densities ρ_x were recovered from biased estimates ρ'_x via msesMD simulations using the approach of Torrie and

Valleau[18] where replica contributions are reweighted according to the biasing energy term, U_α^{ses}, using the expression:

$$\langle \rho_x \rangle = \sum_{\alpha}^{M} \langle \rho'_{x_x}(r_\alpha) \exp(\beta U_\alpha^{ses}) \rangle \Big/ \langle \exp(\beta U_\alpha^{ses}) \rangle \qquad (2)$$

For the computation of the free energy profile associated with the glycosidic linkages and the Cremer and Pople θ pucker angle, bin sizes of 8° and 6° were used, respectively. Errors in the reweighted estimates of the φψ free energy surfaces were obtained by bootstrap resampling the surfaces 25000 times and calculating the standard deviation in the bin estimates across all bootstrap samples. In order to further analyze the conformational diversity of the Lewis oligosaccharides, the unbiased and msesMD trajectories of systems **a-d** (Figure 1) were clustered based on the RMS distance of core Gal-GlcNAc-Fuc ring atoms using the DBSCAN algorithm[49] using an ε distance cutoff of 0.1725 $Å^2$ and minimum criteria of four frames to create a cluster. Due to high memory costs associated with clustering, the clusters were first created using $1/130^{th}$ and $1/8^{th}$ of the frames for unbiased MD and msesMD respectively. The remaining frames were then assigned to the generated clusters. For msesMD, cluster densities were then reweighted according to the swarm contributions of the configurations in each cluster. For the umbrella sampling simulations, an implementation of the weighted histogram analysis method (WHAM)[48] by the Grossfield Lab was used to recover a one-dimensional PMF from the biased windows. The aMD simulations were reweighted according to the total boost energy in each frame, using a 10^{th} order Maclaurin series expansion to reduce noise from the biasing potential.[45] All three dimensional pictures were generated using the VMD molecular graphics viewer version 1.9.[50]

4. Results and discussion

4.1. Conformational sampling of sLe^a

The free energy surface for the three φψ glycosidic linkages of sLe^a in explicit solvent, starting from its closed conformation, are first computed from triplicate 10 μs unbiased MD simulations and compared with the initial 1 μs of these three trajectories (Figure 3). The three linkages of sLe^a are Neu5Acα(2-3)Gal, Fucα(1→4)GlcNAc and Galβ(1→3)GlcNAc (Figures 1b and 2b). We first note that free energy surface predicted for the Neu5Acα(2-3)Gal linkage via the triplicate 1 μs and 10 μs simulations are very similar (Figure 3), suggesting good sampling on both timescales, due to a lack of intramolecular hindrance. We next consider the free energy surfaces of the Fucα(1→4)GlcNAc and Galβ(1→3)GlcNAc glycosidic linkages, labelling minima found on these surfaces as F and G regions respectively. Based on the first 1 μs of the trajectories (referred to as the triplicate 1 μs trajectories from this point), sLe^a samples mainly around its initial closed conformational state, occupying the regions F_1 and G_1 (Figure 3). This closed conformer is denoted from this point as the C state. Although no X-ray structure of sLe^a could be found in the RCSB Protein Data Bank[51], we note that this C state appears to correspond to all observed Le^a conformations in protein-bound crystal structures (Supporting Information, Table S1). From the triplicate 1 μs simulations, less frequently occupied conformations in non-closed geometries are also predicted, denoted $F_{2\text{-}3}$ and $G_{2\text{-}3}$ respectively (Figure 3). These open conformers tend to maintain partial interactions between the fucose and galactose rings; in certain cases these residues can orient themselves into a near perpendicular arrangement (Figure 2b), distinct from the parallel stacking of the closed state (Figure 2a).

From the more extensive sampling of sLe^a of the three replicate 10 μs simulations, we observe a greater population of F_3, in addition to the detection of new open conformers which occupy F_4 and G_4 regions (Figure 3); these structures have ψ angle values differing by ~150° from the native C positions, F_1 and G_1, in their respective linkages (Figure 3). Inspection of the trajectories indicates that for the Fucα(1→4)GlcNAc linkage, access to the F_4 regions occurs primarily by transitions through the F_2 and F_3 regions, in a $F_1 \rightarrow F_{2/3} \rightarrow F_4$ sequence. As for the Galβ(1→3)GlcNAc linkage, access to the G_4 well can occur as a mixture of either directly from G_1 or in a $G_1 \rightarrow G_2 \rightarrow G_3 \rightarrow G_4$ route. Transitions to the G_4 and F_4 regions can occur simultaneously, resulting in a conformational state which is stabilized by the formation of a hydrogen bond between the hydroxyl groups at the C2 positions of both the fucose and galactose rings (Figure 2c). Nevertheless, sampling of the F_4 and G_4 wells is highly infrequent, as can be seen from the time profiles for the ψ angle of both glycosidic linkages (Figure S1a,b). Aside from one replicate trajectory (black, Figure S1), access to the F_4 and G_4 regions, with ψ values of 100° and -60° respectively, is infrequent with lifetimes of up to a few tens of nanoseconds (Figure S1a,b). In fact, for one of the three 10 μs trajectories (red), this is only achieved once for a duration 10.8 ns and 270 ps for F_4 and G_4 respectively (Figure S1a,b). We can therefore conclude that a 10 μs simulation could easily fail to identify either well during the lifetime of its time evolution. This observation accords with an earlier 10 μs study of sLe^a by Sattelle et al., where no open conformers were detected.[14]

msesMD simulation. We compare these unbiased 1 μs and 10 μs MD simulations with a 245 ns msesMD simulation, using a swarm of eight coupled replicas of sLe^a in explicit solvent, interacting via their proximity in the values of the six glycosidic torsion angles. Again we

observe that the flexible Neu5Acα(2→3)Gal linkage is readily sampled, with similar free energy landscapes to both the unbiased 1 μs and 10 μs simulations (Figure 3). However, the msesMD free energy surfaces of the core Fucα(1→4)GlcNAc and Galβ(1→3)GlcNAc linkages of sLe^a correspond considerably more closely to the full 10 μs simulations than the first 1 μs of MD, with broad sampling of the $F_{1\text{-}4}$ and $G_{1\text{-}4}$ low energy regions. The enhanced coverage of open forms by the swarm replicas is clearly shown by the sampling of ψ values for both glycosidic linkages (Figure S1c,d). Interestingly, on the Galβ(1→3)GlcNAc free energy surface obtained via msesMD, an additional low energy region is identified, which we denote G_5 (Figure 3).

Overall, the free energy surface computed via msesMD appears to predict similar stabilities to the profile obtained from 10 μs simulations (Figure 3); the exception is the G_4 well, where the free energy is predicted as 2 kcal mol^{-1} less stable than from unbiased MD. We note from the dihedral time series of the 10 μs simulations, the lifetime of the G_4 state occupancy is disproportionally longer in the first replicate trajectory (black in Figure S1a,b) compared to the other two unbiased trajectories (blue, red). As the G_4 state is sampled only once briefly in each unbiased trajectory, it is not possible to discern if this constitutes an outlier; however it does highlight the fact that the free energy estimates for the G_4 region via unbiased MD are limited due to inadequate sampling.

The msesMD simulation also appears to reproduce more subtle features of the unbiased simulations, in particular sampling of GlcNAc ring puckering. This can be seen from the free energy profiles as a function of Cremer-Pople pucker angle θ (Figure 4a). The msesMD simulations, as well as unbiased triplicate 1 μs and 10 μs MD trajectories sample the full

spectrum of puckers, from $^4C_1$ through skew-boat intermediates to $^1C_4$ conformations, tracing similar profiles (Figure 4a). For the remaining sugar rings, the pucker free energy profiles obtained from msesMD broadly reflect those from unbiased MD albeit with some differences regarding the highest energy regions (Figure 4b-d); the greatest discrepancy is found for the high energy $^4C_1$ state of Neu5Ac where biased and unbiased estimates differ by ~2.5 kcal mol^{-1} (Figure 4b).

Potentials of mean force. We next compare the conformational free energy profiles obtained from msesMD simulation of sLe^a with those computed via other biasing methods. Firstly, in order to observe the rotational barrier along the $F_1 \rightarrow F_4$ transition, a 1-dimensional potential of mean force was computed using umbrella sampling around the ψ_F torsion angle of sLe^a, ie. ψ of the Fucα(1→4)GlcNAc. This involved computing 120 windows of 20 ns production dynamics each. As expected, for this free energy profile, the closed form C is predicted as the lowest energy conformation, occupying the F_1 well (Figure 5a). On this profile we also observe other local minima: $F_{2/3}$ (we group F_2 and F_3 as they occupy similar positions along ψ) and F_4. The $F_{2/3}$ and F_4 wells have calculated free energies relative to F_1 of 3.2 and 2.8 kcal mol^{-1} (Figure 5a). The profile also indicates that direct transition of F_1 conformer to F_4 would encounter an energetic barrier of 9.8 kcal mol^{-1}; this compares with barriers of 4.6 and 7.9 kcal mol^{-1} for the indirect route via $F_{2/3}$ intermediates; this indeed corresponds to the $F_1 \rightarrow F_4$ route observed from the triplicate 10 μs unbiased MD simulations.

The free energy profile constructed from the triplicate unbiased 10 μs simulations provides a reasonable approximation of the ψ_F torsional PMF (Figure 5a), although the relative stability of

the $F_{2/3}$ and F_4 wells is predicted as close to degenerate. From the unbiased simulations, the $F_1 \rightarrow F_{2/3}$ transition also appears to have a near negligible barrier height, approximately 2 kcal mol^{-1} lower than the one predicted via umbrella sampling. The triplicate 1 μs simulations unfortunately fail to adequately sample the complete range of the ψ_F coordinate, but otherwise agree with the 10 μs simulations. Using msesMD simulation, the estimated free energy profile is similar to the unbiased simulations, although the msesMD barrier heights for the $F_{2/3} \rightarrow F_4$ and $F_4 \rightarrow F_1$ transitions are in closer agreement with the PMF from umbrella sampling. Nevertheless, it is noted that msesMD predicts the stability of the F_4 region to be 0.7 kcal mol^{-1} lower than the umbrella sampling and unbiased MD values. As found in the unbiased simulations, the free energy barrier for $F_1 \rightarrow F_{2/3}$ is predicted by msesMD to be minimal, much lower than found via umbrella sampling. This discrepancy may be due to inadequate sampling of the φ_F rotation in the 1-D PMF, using windowing samples of only 20 ns that lead to an inaccurate description of the $F_1 \rightarrow F_{2/3}$ transition path. In fact, inspection of the φψ normalized average occupation density of visited φψ states reconstructed from all umbrella sampling windows shows that a lack of sampling of the F_2 well is evident, with the path seeming to hop discontinuously from the F_1 to F_3 wells (Figure S2).

A second PMF profile was also calculated by umbrella sampling about the ψ rotation of the Galβ(1→3)GlcNAc glycosidic linkage (ψ_G), in order to examine transitions between G wells. The closed conformer C, occupying G_1, is again predicted as the lowest energy state (Figure 5b). The G_4 well is computed as 5.7 kcal mol^{-1} above G_1 via the potential of mean force calculation. Additionally, the profile indicates an energetic barrier of 10.3 kcal mol^{-1} going from the G_1 to G_4 wells directly, while the route which should involve the $G_{2/3}$ region shows a barrier of 12.6 kcal

mol^{-1} (Figure 5b). This would seem to indicate that the most appropriate route is a direct transition from G_1 to G_4, rather than the alternative path via G_3.

The triplicate 10 μs MD profile finds a defined $G_{2/3}$ well 2.5 kcal mol^{-1} above G_1. The G_4 well stability is also increased relative to the umbrella sampling surface, becoming equienergetic with $G_{2/3}$. The energetic barriers are also significantly lowered, with a value of 7.2 kcal mol^{-1} for the direct $G_1 \rightarrow G_4$ transition, compared to 2.5 and 5.3 kcal mol^{-1} via the indirect $G_1 \rightarrow G_{2/3} \rightarrow G_4$ path. The msesMD free energy profile for ψ_G shows agreement in overall shape with the unbiased MD prediction, but differs in finding a decreased stability of G_4 by ~2 kcal mol^{-1} and an increase in energetic barrier for the direct and indirect $G_1 \rightarrow G_4$ routes, by 2.0 and 2.7 kcal mol^{-1} respectively. As discussed above, this discrepancy between the msesMD and unbiased MD predictions is potentially due to a poor estimate of the G_4 population in the unbiased MD simulation, arising from one of the replicate trajectories becoming kinetically trapped in the higher energy G_4 well for a disproportionate amount of time. The much larger overall differences between msesMD and the umbrella sampling profile appear to be caused, as for ψ_F, by a lack of sampling of the φ_G rotation. Although not as obvious as in the ψ_F profile, there appears to be incomplete sampling of the G_3 region by the PMF, with the path consisting primarily of a transition from $G_2 \rightarrow G_4$ (Figure S3).

The discrepancies seen in the umbrella sampling profiles of both ψ_F and ψ_G angles highlights the limitation of using one-dimensional coordinates to describe complex systems such as carbohydrates. The use of two-dimensional PMF profiles for carbohydrates would be preferred, but are very computationally expensive, even for relatively small systems such as sLe^a.

Accelerated molecular dynamics. A third biasing approach, accelerated molecular dynamics (aMD), was also employed for comparison. We report the free energy surface for sLe^a in explicitly modelled water, obtained using dual boost aMD. In this approach, a boost potential is applied based on the total and dihedral potential energies of the system. To compare with the eight msesMD replicas, we combine the results of eight independent 494 ns aMD simulations to obtain the final free energy surfaces (Figure 3). The aMD simulations provide good coverage of the Fucα(1→4)GlcNAc and Galβ(1-3)GlcNAc glycosidic linkages, identifying all the main wells, $F_{1\text{-}4}$ and $G_{1\text{-}5}$, in addition to an high energy F_5 well not seen in either unbiased MD or msesMD simulations. However, as demonstrated by the ψ_F and ψ_G time series (Figure S1), the F_4 and G_4 wells appear to be infrequently sampled.

In general, we note the somewhat flatter topology features of the reweighted aMD free energy surfaces relative to msesMD and unbiased MD, which are particularly evident around the $F_{1\text{-}3}$ and $G_{1\text{-}3}$ regions (Figure 3). This may be a reflection of the magnitude of the resultant aMD boost energy, which has a mean and range of 4.9 and 24.2 kcal mol^{-1} respectively; this compares to an applied swarm energy for msesMD of mean 4.1 kcal mol^{-1} and range 18.4 kcal mol^{-1} (Figure S4). The use of softer boost parameters for aMD did not result in sufficient transitions to the F_4 and G_4 regions (data not shown). Nonetheless, the free energy topology of the aMD simulations broadly reflects that obtained from the unbiased triplicate 10 μs simulations and to a greater degree, the msesMD simulations.

4.2. Evaluating closed and open conformational of sLe^x

From the above comparison of biased and unbiased MD simulations, we conclude that msesMD reproduces the key topological features of the sLe^a conformational landscape; we next apply msesMD alongside unbiased simulations to investigate the conformational flexibility of sLe^x and subsequently Le^x and Le^a. In particular, we examine the ability of the sLe^x tetrasaccharide to adopt open conformational states in aqueous solution that may resemble the previously reported lectin-bound pose.[1]

Unbiased MD simulations. Firstly, we consider unbiased simulations of sLe^x in solution. Unlike sLe^a, triplicate 1 μs trajectories only sample closed conformations (F_1,G_1), with no sampling of neighboring $F_{2\text{-}3}$ and $G_{2\text{-}3}$ regions (Figure 6). The longer 10 μs simulations explore non-closed minima of sLe^x, in the $F_{2\text{-}3}$ and $G_{2\text{-}3}$ regions, in addition to some infrequent sampling of the G_5 region (Figure 6). However, the F_4 and G_4 minima, which tend to be associated with the open protein-bound conformations of the tetrasaccharide,[1,13] are not sampled, perhaps suggesting that sLe^x may be more rigid in solution than sLe^a. This potential rigidity is consistent with previous 25 μs simulations of sLe^x, which similarly did not appear to detect such conformations.[14] As for sLe^a, the Neu5Acα(2→3)Gal rotation in sLe^x is readily sampled (Figure 6).

On computing pucker free energy profiles, we find that for the GlcNAc ring in sLe^x, the triplicate 10 μs simulations sample the complete $^4C_1$-to-$^1C_4$ pucker transition (Figure 7a); by contrast, the triplicate 1 μs simulations fail to sample the $^1C_4$ pucker. Given that the 1 μs simulations do not sample any non-closed conformations, this difference in sampling of pucker appears to support the hypothesis of Topin et al.[1] that GlcNAc ring puckering is important to the

exploration of open conformational states. This is further indicated by a shift in the ring pucker states when moving along the structural path from closed (F_1,G_1) to the F_2 and G_2 wells (Figure S5). The GlcNAc ring primarily occupies a $^4C_1$ pucker ($\theta \sim 10°$) when in the closed (F_1,G_1) state; this changes to a mixture of boat/skew-boat and $^1C_4$ puckers in the F_2 and/or G_2 states (Figure S5). As for sLe^a simulations, access to the F_2 and G_2 wells is the primary route to open conformations, thus supporting the theory that changes in the GlcNAc ring pucker are necessary to achieve such events. For the remaining residues, in contrast to the 10 μs simulations, the triplicate 1 μs MD trajectories do not sample the boat/skew-boat puckers of the fucose and galactose rings, or the $^4C_1$ pucker of Neu5Ac (Figure 7b-d).

msesMD simulation. As for sLe^a, the 245 ns msesMD simulations of sLe^x in solution display greater similarity to the triplicate 10 μs rather than 1 μs unbiased MD simulations, with extensive sampling of the $F_{1\text{-}3}$ and $G_{1\text{-}3,5}$ regions (Figure 6). Significantly, F_4 and G_4 minima are also identified by msesMD. Interestingly the F_4 and G_4 torsional values from msesMD agree well with those of the RSL-bound sLe^x X-ray conformer. For the latter, φ,ψ values of (-70°,150°) and (-70°,-120°) are determined for the Fucα(1→3)GlcNAc and Galβ(1→4)GlcNAc linkages respectively.[1] By clustering the msesMD trajectories, a structure closely matching the RSL-bound conformer was identified (Figure 2d). It should be noted that whilst this lectin-like conformer was identified within the top 10 most occupied clusters upon reweighting, the cluster was low in population, with an estimated relative free energy of 5.5 kcal mol^{-1} above the closed state. This rare conformation could have been overlooked if fewer clusters had been examined and thus highlights the importance of considering relatively low occupation conformations in identifying important glycan shapes.

Regarding the ring puckering in sLe^x, we find that for its GlcNAc residue, the complete $^4C_1$-to-$^1C_4$ landscape is sampled by msesMD simulation (Figure 7a), which is broadly similar to the triplicate 10 μs unbiased MD simulations. For the other pyranose rings, msesMD generally traces more complete free energy profiles than the triplicate 1 μs MD simulations, but differs from the triplicate 10 μs simulations in estimates of the higher energy conformers (Figure 7b-d).

4.3. Effect of sialylation on conformation

Finally, we calculate the effect of sialylation of Le^a and Le^x on their closed/open equilibria in solution. Unbiased triplicate 10 μs molecular dynamics in aqueous solution of Le^a (Figure 8a) indicate a broadly similar pattern of sampling to the corresponding triplicate 10 μs trajectories of sLe^a (Figure 3). We note two small differences in the topologies of the free energy surface for glycosidic linkage Galβ(1-3)GlcNAc: first the detection of an additional well, termed G_5, which is not observed in the unbiased sLe^a simulations, although found in the enhanced sampling simulations of sLe^a; secondly, the G_3 region appears to be more favored in simulations of Le^a. Furthermore, the F_4 and G_4 well stabilities differ slightly with the relative free energy estimates in Le^a being 0.9 and 0.6 kcal mol^{-1} less stable than for sLe^a (Figures 3 and 8).

Using msesMD, the Fucα(1→4)GlcNAc free energy surfaces for both Le^a (Figure 8a) and sLe^a (Figure 3) are near identical with similar well depths for all major regions. However, differences of over 1 kcal mol^{-1} are seen in the reweighted Galβ(1-3)GlcNAc surfaces. Interestingly, unlike the unbiased MD simulations, the msesMD results estimate the F_4 and G_4 wells to be more stable in Le^a rather than sLe^a by an average of 0.5 and 1.2 kcal mol^{-1} respectively. We note that errors

in msesMD free energy estimates, the errors calculated via bootstrap sampling are generally within ± 0.25 kcal mol^{-1} (Supplementary Figure S7).

Turning to consider simulation of Le^x in solution, similarly to Le^a, analysis of the unbiased MD simulations appears to show that sialylation has a small impact on the conformer ensemble. The triplicate 10 μs MD simulations of Le^x (Figure 8b) sample the same $F_{1\text{-}3}$ and $G_{1\text{-}3,5}$ torsional wells identified in the sLe^x simulations (Figure 6), albeit with a slightly lower occupancy of the G_5 region. Whilst our unbiased MD simulations did not sample F_4 and G_4 regions, crucial to adopting the observed RSL-bound conformations[1] of Le^x and sLe^x, msesMD simulations were able to find several such conformers (Figure 8b). Cluster analysis of the msesMD trajectories identified conformations which correspond to the bound crystal open poses (Figure 2e-g). Comparing the relative free energy profiles for Le^x and sLe^x generated by msesMD (Figure 8b), sialylation appears to impact the stability of the F_4 and G_4 conformers, with an estimated 0.9 and 0.7 kcal mol^{-1} average decrease in stability respectively going from Le^x to sLe^x (Figure 6 and 8b). Interestingly, as discussed above, the msesMD simulations for Le^a and sLe^a also predict a decrease in stability for the F_4 and G_4 wells upon sialylation. This may indicate that sialylation stiffens the Fuc-GlcNAc and Gal-GlcNAc linkages, preventing the large ~ 140-200° flips in the ψ angle associated with these two wells.

5. Conclusions

In this study, we have detailed the propensity of Lewis blood sugar antigens to adopt closed and open conformations using unbiased and enhanced sampling molecular dynamics methods. We find that 245 ns msesMD simulations of sLe^a, using a swarm of eight simulation replicas,

reproduces key features of triplicate 10 μs unbiased trajectories: the topological features of the glycosidic linkage free energy surfaces in general agree well; the extent of ring puckering of the key central GlcNAc ring is also in good agreement.

The msesMD trajectories yield free energy profiles along ψ_F, the torsion angle most diagnostic of open and closed conformations, in closer accord with the profile from triplicate 10 μs simulations than that calculated via umbrella sampling. The latter appears to experience limited sampling due to the 20 ns windows employed. Furthermore, msesMD simulations identify a new G_5 well for sLe^a which is absent from the triplicate 10 μs simulations, but features in aMD simulations of sLe^a and unbiased MD calculations of Le^a (Figures 3 and 8a). Differences in the stability of sLe^a conformers as calculated by unbiased MD and msesMD are found, particularly for the G_4 well, and appear to arise from limited sampling in the unbiased MD simulations. Overall, these outcomes highlight the need for even longer unbiased MD simulation lengths in order to exhaustively sample the conformational space of sLe^a.

Application of msesMD simulation to four Lewis oligosaccharides in solution predicts that the closed form is the lowest energy conformer, comprising populations of 93%, 89%, 92% and 97% for Le^a, Le^x, sLe^a and sLe^x respectively. This agrees with the ubiquity of closed forms in both free and bound structures observed from X-ray spectroscopy (Figures 6 and 8a,b). Several clearly defined energy minima were identified on the torsional landscapes of the Gal-GlcNAc and Fuc-GlcNAc glycosidic linkages, corresponding to a large ensemble of non-closed conformational states. These minima have a range of stabilities, with those closer to the main closed (F_1 and G_1) wells, e.g. $G_{2/3}$ and $F_{2/3}$, lying within 1-2 kcal mol^{-1} of the closed form; and

then more distant wells, e.g. G_5 and F_5, potentially reaching over 5 kcal mol^{-1}. Access to the different open forms is achieved via transition paths which frequently require initial access to the G_2 and F_2 states. As demonstrated for sLe^x, this appears to be mediated by a shift in the GlcNAc pucker, from $^4C_1$ to either boat/skew-boat or $^1C_4$. This agrees with previous proposals[1] that puckering of the GlcNAc ring can facilitate closed-to-open transitions.

Access to the G_4 and F_4 regions, via a >150° transition in glycosidic torsion space from the closed regions, is crucial to formation of the RSL-bound geometries of Le^x and sLe^x. Analysis of the msesMD simulations appears to show these open conformer regions as slightly more stable when sialylation is absent. The msesMD simulations of Le^a/sLe^a suggest a similar behavior may also be true for these two systems too, albeit possibly to a lesser extent, especially considering the lack of agreement with the unbiased MD simulations. Interestingly, the K_d has been measured for the binding of a Le^x tetrasaccharide and sLe^x pentasaccharide to RSL by Topin et al.[1] These sugars have the primary structure Galβ(1→4)(Fucα(1→3))GlcNAcβ(1→3)Gal and Neu5Acα(2→3)Galβ(1→4) (Fucα(1→3))GlcNAcβ(1→3)Gal respectively, differing from Le^x and sLe^x in possessing an extra galactose residue distal to the sialylation site. The K_d of Le^x tetrasaccharide was found to be approximately half that of sLe^x pentasaccharide, at 26 and 58 μM respectively, and could support the proposed higher energy of the bound pose on sialylation. From msesMD simulations, we also note that higher energy conformers closely matching the RSL-bound crystal poses were found (Figure 2e-g). This reflects the induced fit nature of protein-ligand binding, where rare higher energy conformational states are stabilized by interactions with protein side chains.

In this study, we find that msesMD constitutes a potentially useful tool in overcoming energy barriers to sample rare minima that are usually accessed at very long timescales. The msesMD method reasonably reproduces free energy surfaces obtained by multi-microsecond unbiased MD simulations and is approximately five times faster in achieving this, acquiring 245 ns in 5 days, relative to the 25 days required to generate 10 μs MD simulations using hydrogen mass repartitioning; this is obtained using four Infiniband connected 24 core Haswell E5-2680 nodes (msesMD) and three Nvidia K20 GPU nodes (unbiased MD). It is noted that without the use of HMR, the unbiased simulations would have doubled the calculation time. Indeed one could envisage the use of msesMD in tandem with HMR in order to further lower simulation times. While *a priori* knowledge of possible boost coordinates is required when using msesMD, the use of glycosidic torsions, as previously noted by similar locally biased methods,[52] seems to be appropriate. The form of the biasing potential used here appears to be effective for these carbohydrates, only introducing a slight amount of noise as seen in free energy profiles sLe^a and sLe^x when the number of boosted dihedrals increases (Figures 3 and 6). In conclusion, the msesMD method appears well suited as an intuitive efficient tool for the exploration of oligosaccharides and other similarly sized systems, providing important physical insights into solution conformations for understanding of function and potential design of ligands.

Supporting Information. Structural analyses of Lewis oligosaccharides from crystallographic data; structural, energetic and error analyses of MD simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

Acknowledgements

This project made use of time on ARCHER granted via the UK High-End Computing Consortium for Biomolecular Simulation, HECBioSim (http://hecbiosim.ac.uk), supported by EPSRC (grant no. EP/L000253/1). This work made use of the facilities of N8 HPC Center of Excellence, provided and funded by the N8 consortium and EPSRC (grant no. EP/K000225/1). The authors would like to acknowledge the use of the Computational Shared Facility at the University of Manchester.

References

1. Topin, J.; Lelimousin, M.; Arnaud, J.; Audfray, A.; Pérez, S.; Varrot, A.; Imberty, A., The Hidden Conformation of Lewis x, a Human Histo-Blood Group Antigen, Is a Determinant for Recognition by Pathogen Lectins. *ACS Chem Biol* **2016,** *11*, 2011-2020.
2. de la Fuente, J. M.; Penadés, S., Glyconanoparticles: Types, synthesis and applications in glycoscience, biomedicine and material science. *Biochim Biophys Acta* **2006,** *1760*, 636-651.
3. Woods, R. J.; Tessier, M. B., Computational glycoscience: characterizing the spatial and temporal properties of glycans and glycan–protein complexes. *Curr Opin Struct Biol* **2010,** *20*, 575-583.
4. Frank, M.; Collins, P.; Peak, I.; Grice, I.; Wilson, J., An Unusual Carbohydrate Conformation is Evident in Moraxella catarrhalis Oligosaccharides. *Molecules* **2015,** *20*, 14234.
5. Blaum, B. S.; Frank, M.; Walker, R. C.; Neu, U.; Stehle, T., Complement Factor H and Simian Virus 40 bind the GM1 ganglioside in distinct conformations. *Glycobiology* **2016,** *26*, 532-539.
6. DeMarco, M. L.; Woods, R. J.; Prestegard, J. H.; Tian, F., Presentation of Membrane-Anchored Glycosphingolipids Determined from Molecular Dynamics Simulations and NMR Paramagnetic Relaxation Rate Enhancement. *J Am Chem Soc* **2010,** *132*, 1334-1338.
7. Watkins, W. M., The ABO blood group system: historical background. *Transfus Med* **2001,** *11*, 243-265.
8. Yuriev, E.; Farrugia, W.; Scott, A. M.; Ramsland, P. A., Three-dimensional structures of carbohydrate determinants of Lewis system antigens: Implications for effective antibody targeting of cancer. *Immunol Cell Biol* **2005,** *83*, 709-717.
9. Lorant, D. E.; Topham, M. K.; Whatley, R. E.; McEver, R. P.; McIntyre, T. M.; Prescott, S. M.; Zimmerman, G. A., Inflammatory roles of P-selectin. *J Clin Invest* **1993,** *92*, 559-570.
10. Sugasaki, A.; Sugiyasu, K.; Ikeda, M.; Takeuchi, M.; Shinkai, S., First Successful Molecular Design of an Artificial Lewis Oligosaccharide Binding System Utilizing Positive Homotropic Allosterism. *J Am Chem Soc* **2001,** *123*, 10239-10244.
11. Lemieux, R. U.; Bock, K.; Delbaere, L. T. J.; Koto, S.; Rao, V. S., The conformations of oligosaccharides related to the ABH and Lewis human blood group determinants. *Can J Chem* **1980,** *58*, 631-653.

12. Mishra, S. K.; Kara, M.; Zacharias, M.; Koča, J., Enhanced conformational sampling of carbohydrates by Hamiltonian replica-exchange simulation. *Glycobiology* **2014,** *24*, 70-84.

13. Haselhorst, T.; Weimar, T.; Peters, T., Molecular Recognition of Sialyl Lewis X and Related Saccharides by Two Lectins. *J Am Chem Soc* **2001,** *123*, 10705-10714.

14. Sattelle, B. M.; Almond, A., Shaping up for structural glycomics: a predictive protocol for oligosaccharide conformational analysis applied to N-linked glycans. *Carbohydr Res* **2014,** *383*, 34-42.

15. Sattelle, B. M.; Hansen, S. U.; Gardiner, J.; Almond, A., Free Energy Landscapes of Iduronic Acid and Related Monosaccharides. *J Am Chem Soc* **2010,** *132*, 13132-13134.

16. Sattelle, B. M.; Shakeri, J.; Almond, A., Does Microsecond Sugar Ring Flexing Encode 3D-Shape and Bioactivity in the Heparanome? *Biomacromolecules* **2013,** *14*, 1149-1159.

17. Harvey, M. J.; Giupponi, G.; Fabritiis, G. D., ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J Chem Theory Comput* **2009,** *5*, 1632-1639.

18. Friedrichs, M. S.; Eastman, P.; Vaidyanathan, V.; Houston, M.; Legrand, S.; Beberg, A. L.; Ensign, D. L.; Bruns, C. M.; Pande, V. S., Accelerating Molecular Dynamic Simulation on Graphics Processing Units. *J Comput Chem* **2009,** *30*, 864-872.

19. Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J Chem Theory Comput* **2012,** *8*, 1542-1555.

20. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J Chem Theory Comput* **2013,** *9*, 3878-3888.

21. Hamelberg, D.; Mongan, J.; McCammon, J. A., Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J Chem Phys* **2004,** *120*, 11919-11929.

22. Torrie, G. M.; Valleau, J. P., Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J Comput Phys* **1977,** *23*, 187-199.

23. Hansen, H. S.; Hünenberger, P. H., Using the local elevation method to construct optimized umbrella sampling potentials: Calculation of the relative free energies and interconversion barriers of glucopyranose ring conformers in water. *J Comput Chem* **2010,** *31*, 1-23.

24. Laio, A.; Parrinello, M., Escaping free-energy minima. *Proc Natl Acad Sci USA* **2002,** *99*, 12562-12566.

25. Swendsen, R. H.; Wang, J.-S., Replica Monte Carlo Simulation of Spin-Glasses. *Phys Rev Lett* **1986,** *57*, 2607-2609.

26. Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* **1999,** *314*, 141-151.

27. Hansmann, U. H. E., Parallel tempering algorithm for conformational studies of biological molecules. *Chem Phys Lett* **1997,** *281*, 140-150.

28. Islam, S. M.; Richards, M. R.; Taha, H. A.; Byrns, S. C.; Lowary, T. L.; Roy, P.-N., Conformational Analysis of Oligoarabinofuranosides: Overcoming Torsional Barriers with Umbrella Sampling. *J Chem Theory Comput* **2011,** *7*, 2989-3000.

29. Perić-Hassler, L.; Hansen, H. S.; Baron, R.; Hünenberger, P. H., Conformational properties of glucose-based disaccharides investigated using molecular dynamics simulations with local elevation umbrella sampling. *Carbohydr Res* **2010,** *345*, 1781-1801.

30. Spiwok, V.; Tvaroška, I., Conformational Free Energy Surface of α-N-Acetylneuraminic Acid: An Interplay Between Hydrogen Bonding and Solvation. *J Phys Chem B* **2009,** *113*, 9589-9594.

31. Yang, M.; MacKerell, A. D., Conformational Sampling of Oligosaccharides Using Hamiltonian Replica Exchange with Two-Dimensional Dihedral Biasing Potentials and the Weighted Histogram Analysis Method (WHAM). *J Chem Theory Comput* **2015,** *11*, 788-799.

32. Mallajosyula, S. S.; MacKerell, A. D., Influence of Solvent and Intramolecular Hydrogen Bonding on the Conformational Properties of O-Linked Glycopeptides. *J Phys Chem B* **2011,** *115*, 11215-11229.

33. Plazinski, W.; Drach, M., The influence of the hexopyranose ring geometry on the conformation of glycosidic linkages investigated using molecular dynamics simulations. *Carbohydr Res* **2015,** *415*, 17-27.

34. Biarnés, X.; Ardèvol, A.; Planas, A.; Rovira, C.; Laio, A.; Parrinello, M., The Conformational Free Energy Landscape of β-d-Glucopyranose. Implications for Substrate Preactivation in β-Glucoside Hydrolases. *J Am Chem Soc* **2007,** *129*, 10686-10693.

35. Atzori, A.; Bruce, N. J.; Burusco, K. K.; Wroblowski, B.; Bonnet, P.; Bryce, R. A., Exploring Protein Kinase Conformation Using Swarm-Enhanced Sampling Molecular Dynamics. *J Chem Inf Model* **2014,** *54*, 2764-2775.

36. Burusco, K. K.; Bruce, N. J.; Alibay, I.; Bryce, R. A., Free Energy Calculations using a Swarm-Enhanced Sampling Molecular Dynamics Approach. *ChemPhysChem* **2015,** *16*, 3233-3241.

37. Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossvary, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. *AMBER 14*, University of California, San Francisco, 2014.

38. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J Comput Chem* **2008,** *29*, 622-655.

39. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **1983,** *79*, 926-935.

40. Feenstra, K. A.; Hess, B.; Berendsen, H. J. C., Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J Comput Chem* **1999,** *20*, 786-798.

41. Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E., Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J Chem Theory Comput* **2015,** *11*, 1864-1874.

42. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* **1977,** *23*, 327-341.

43. Miyamoto, S.; Kollman, P. A., Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem* **1992,** *13*, 952-962.

44. Allen, M. P.; Tildesley, D. J., *Computer simulation of liquids*. Clarendon Press: Oxford England, 1989; p 385.

45. Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R. C.; McCammon, J. A., Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *J Chem Theory Comput* **2014,** *10*, 2677-2689.

46. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **2013,** *9*, 3084-3095.

47. Case, D.; Betz, R.; Cerutti, D.; Cheateham III, T.; Darden, T.; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K.; Monard, G.; Nguyen, H.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Roe, D.; Roitberg, A.; Sagui, C.; Simmerling, C.; Botello-Smith, W.; Swails, J.; Walker, R.; Wang, J.; Wolf, R.; Wu, X.; Xiao, L.; Kollman, P. *AMBER 16*, University of California: San Francisco, 2016.

48. Grossfield, A. *WHAM: the weighted histogram analysis method*, 2.0.9.1; 2013.

49. Ester, M.; Kriegel, H.-P.; Jorg, S.; Xu, X., A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, AAAI Press: Portland, Oregon, 1996; pp 226-231.

50. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996,** *14*, 33-8, 27-8.

51. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res* **2000,** *28*, 235-242.

52. Yang, M.; Huang, J.; MacKerell, A. D., Enhanced Conformational Sampling Using Replica Exchange with Concurrent Solute Scaling and Hamiltonian Biasing Realized in One Dimension. *J Chem Theory Comput* **2015,** *11*, 2855-2867.

Figure captions

Figure 1. The six major Lewis oligosaccharides (**a**) Le^a, (**b**) sLe^a, (**c**) Le^b, (**d**) Le^x, (**e**) sLe^x and (**f**) Le^y. For sLe^a, ψ_F and ψ_G angles in sLe^a, indicated in red and blue respectively.

Figure 2. (**a**) Closed conformation of Le^a; (**b**) example T-shaped open conformer of sLe^a; **(c)** hydrogen bond stabilized F_4/G_4 conformer of sLe^a (distance in Å). Note: for **a-c** Fuc is labelled in red, Gal in blue and GlcNAc by atom type. (**d**) Overlap of 5AJC sLe^x RSL bound pose (colored by atom type) with clustered conformation from msesMD (blue). (**e-g**) Overlaps of the 5AJC RSL bound Le^x conformations,[1] superimposing the core ring atoms of the crystal (colored by atom type), with selected cluster centroid conformations generated via msesMD (blue). These crystal structures correspond to the notation of Topin et al.[1] as (**e**) Open I (**f**) Open II and (**g**) Open III.

Figure 3. Free energy surfaces (in kcal mol^{-1}) of the $\phi\psi$ glycosidic torsions (degrees) of sLe^a computed via MD, msesMD and aMD

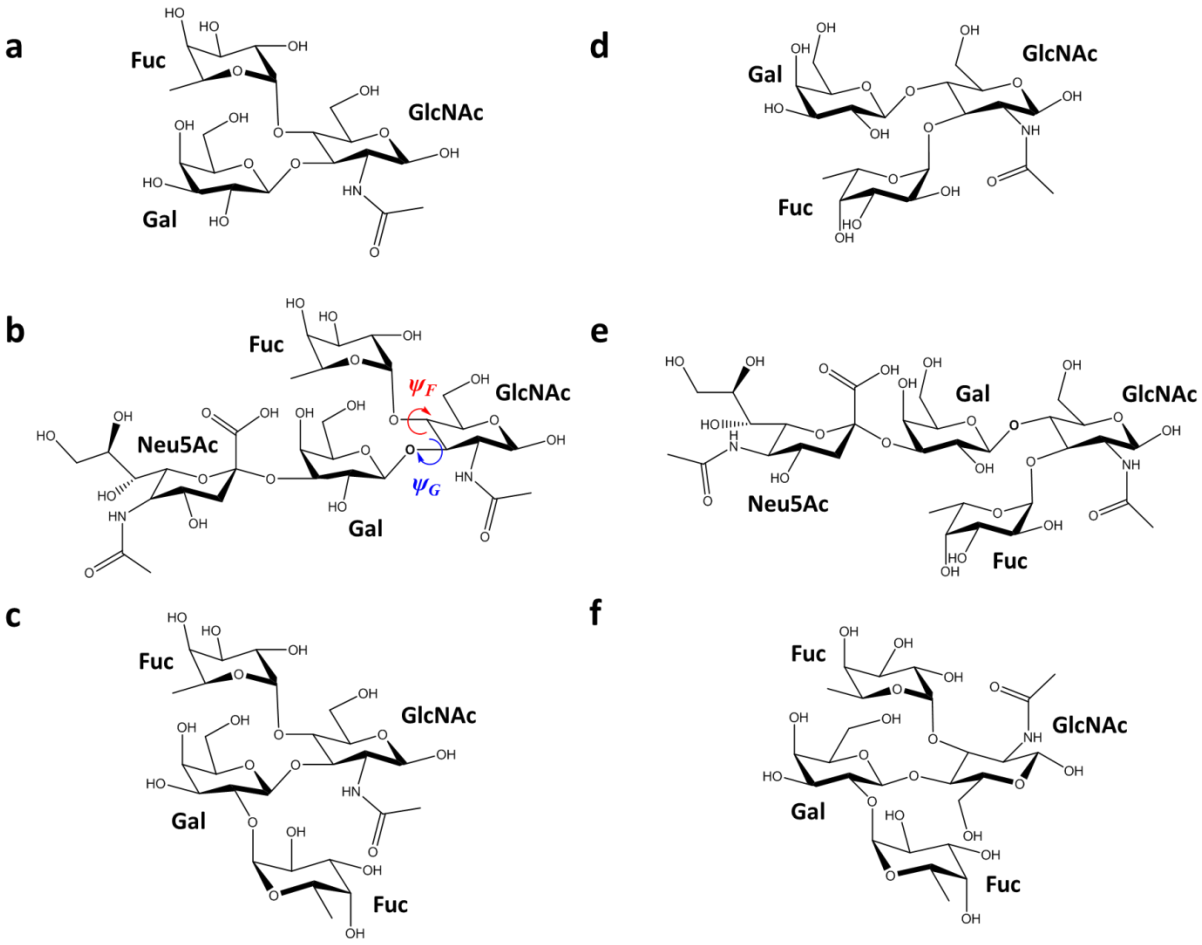
Figure 4. Cremer-Pople θ puckering free energy profiles of sLe^a (in kcal mol^{-1}) as calculated by triplicate 1 and 10 μs MD and msesMD for the four saccharide rings (**a**) GlcNAc, (**b**) Neu5Ac, (**c**) Gal, and (**d**) Fuc.

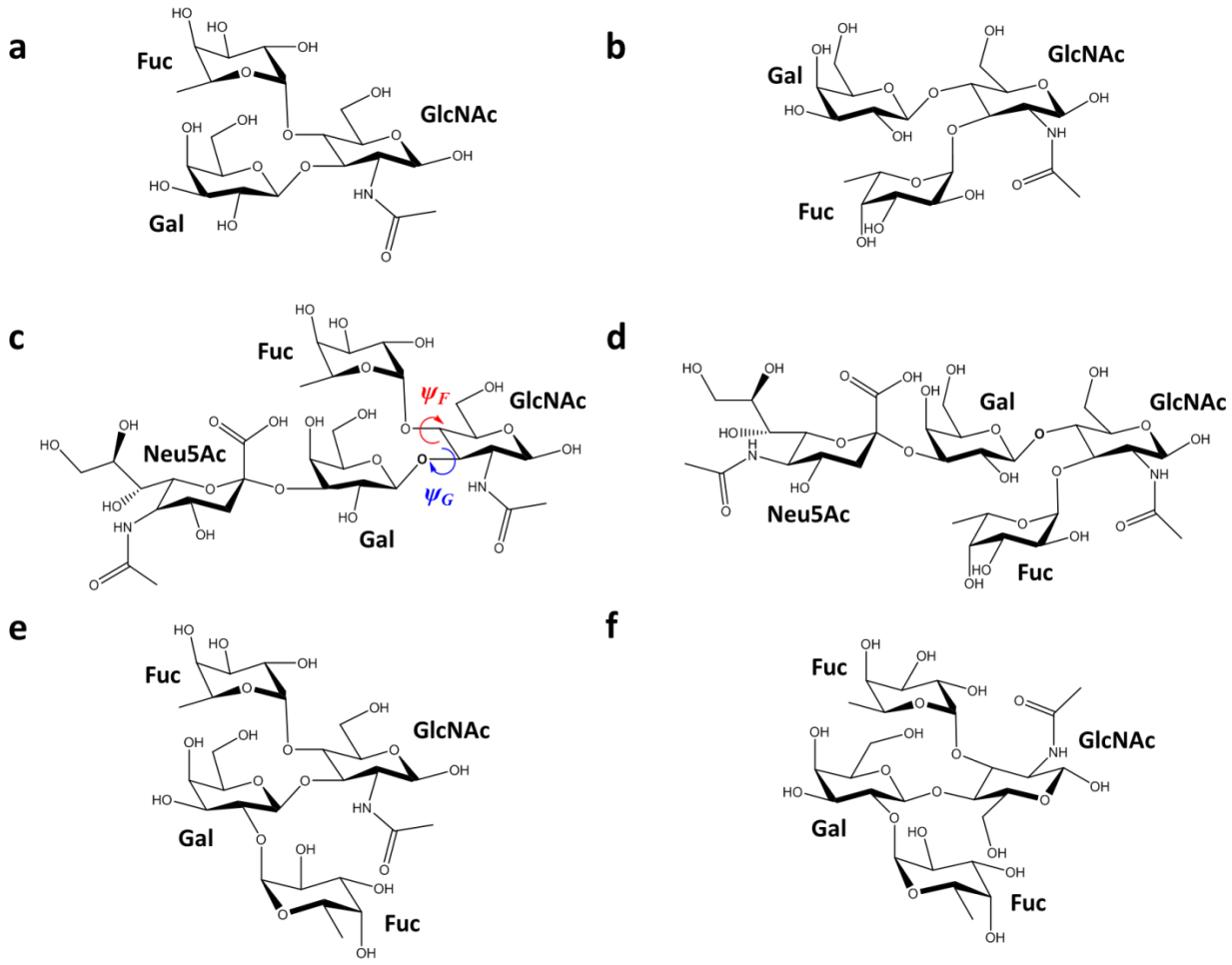
Figure 5. (a) PMF of ψ_F rotation as calculated by umbrella sampling, msesMD and unbiased MD. **b)** PMF of ψ_G rotation as calculated by umbrella sampling, msesMD and unbiased MD.

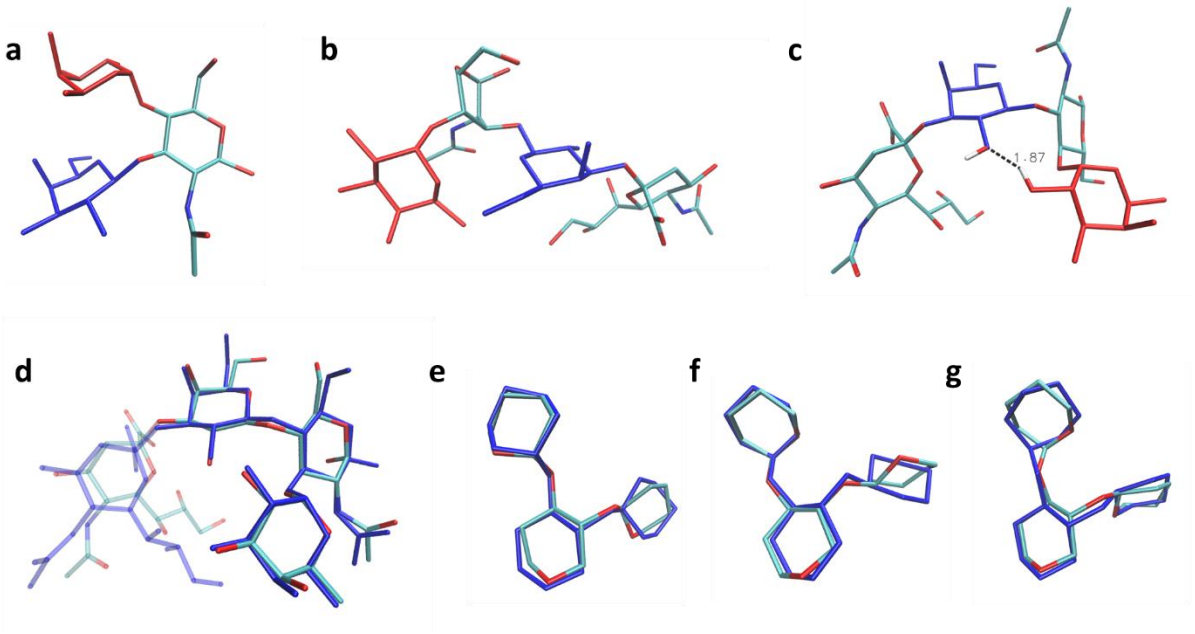
Figure 6. Free energy surfaces (in kcal mol^{-1}) of the $\phi\psi$ glycosidic torsions (degrees) of sLe^x computed via MD, msesMD and aMD. X-ray crystal positions are indicated on the triplicate 1 μs profile.

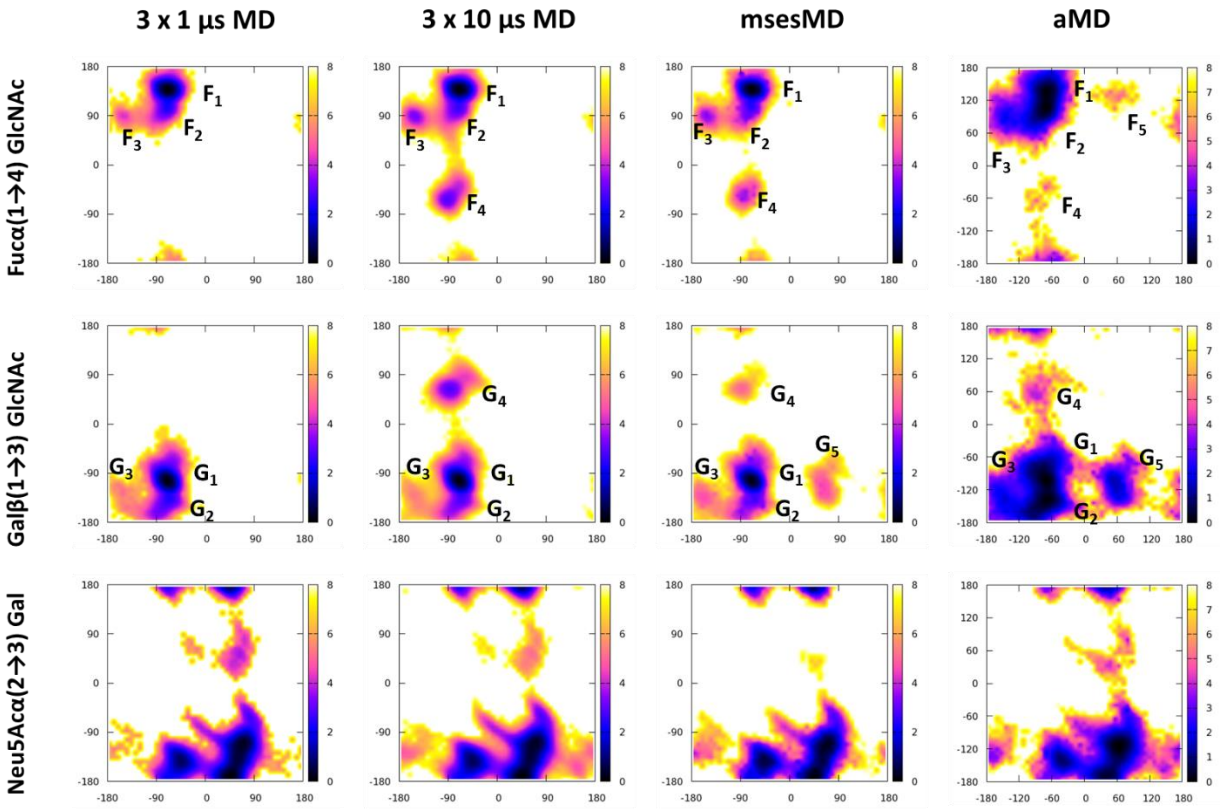
Figure 7. Cremer-Pople θ puckering profiles of sLe^x as calculated by triplicate 1 and 10 μs MD and msesMD for the four saccharide rings, (**a**) GlcNAc, (**b**) Neu5Ac, (**c**) Gal, and (**d**) Fuc.

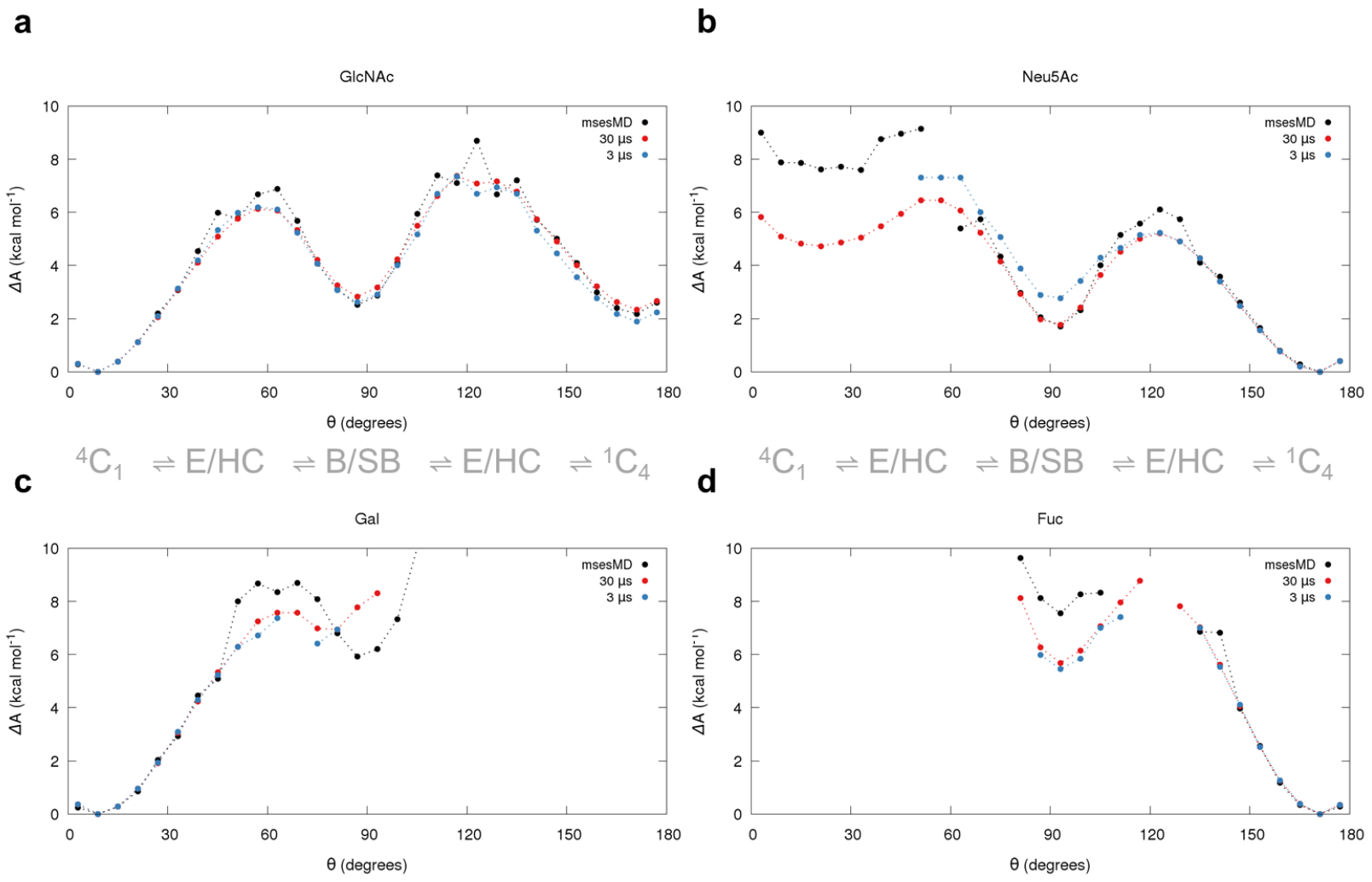
Figure 8. (**a**) Free energy surfaces (in kcal mol^{-1}) of the ϕψ glycosidic torsions (degrees) of (a) Le^a and (**b**) Le^x generated via each method with different regions labelled accordingly, z-axis in kcal mol^{-1}. X-ray crystal positions are indicated on the triplicate 1 μs profile.

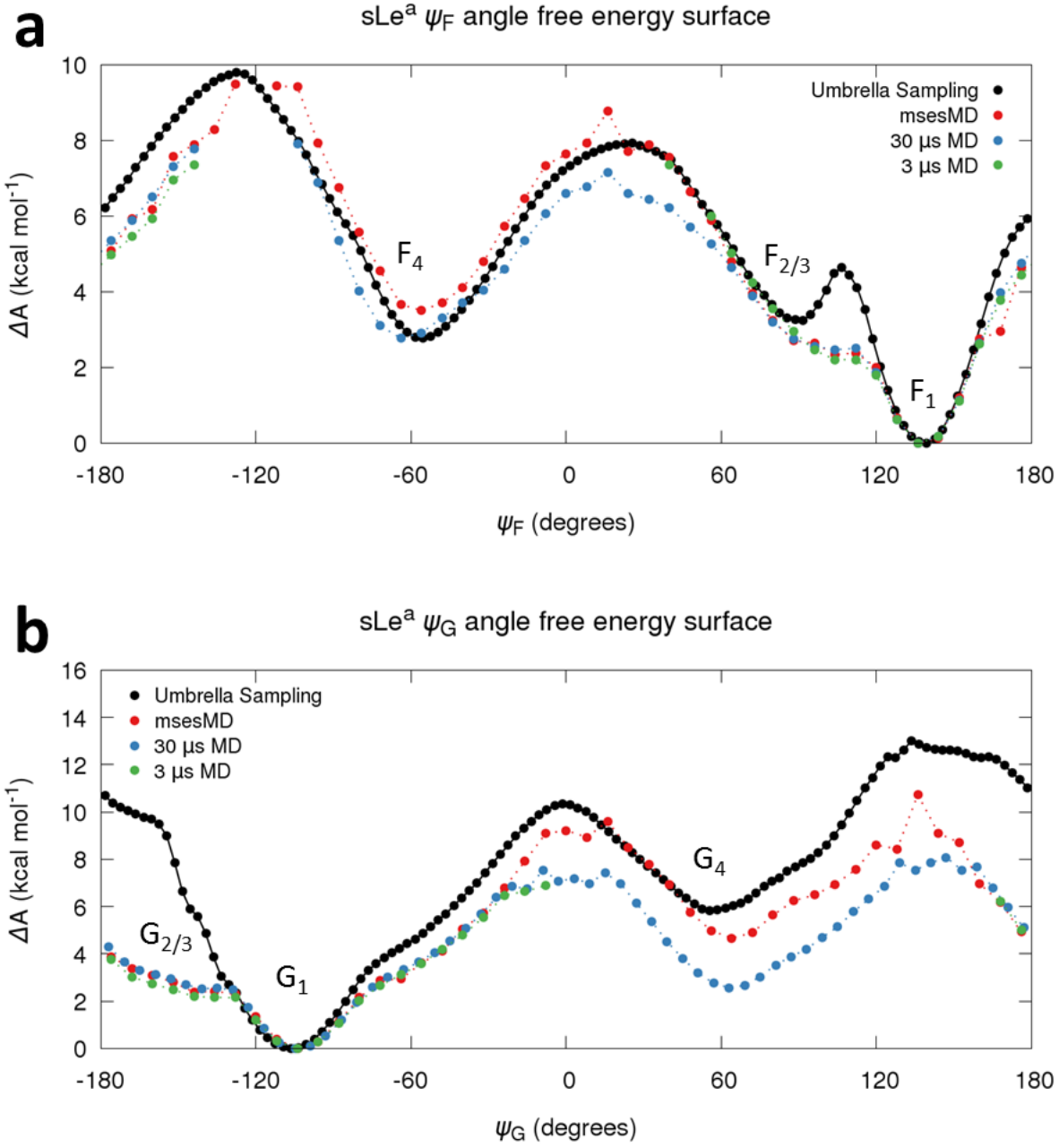

a
Fuc
GlcNAc
Gal
b
Fuc
GlcNAc
Neu5Ac
Gal
ψ_F
ψ_G
c
Fuc
GlcNAc
Gal
Fuc
d
Gal
GlcNAc
Fuc
e
Gal
GlcNAc
Neu5Ac
Fuc
f
Fuc
GlcNAc
Gal
Fuc

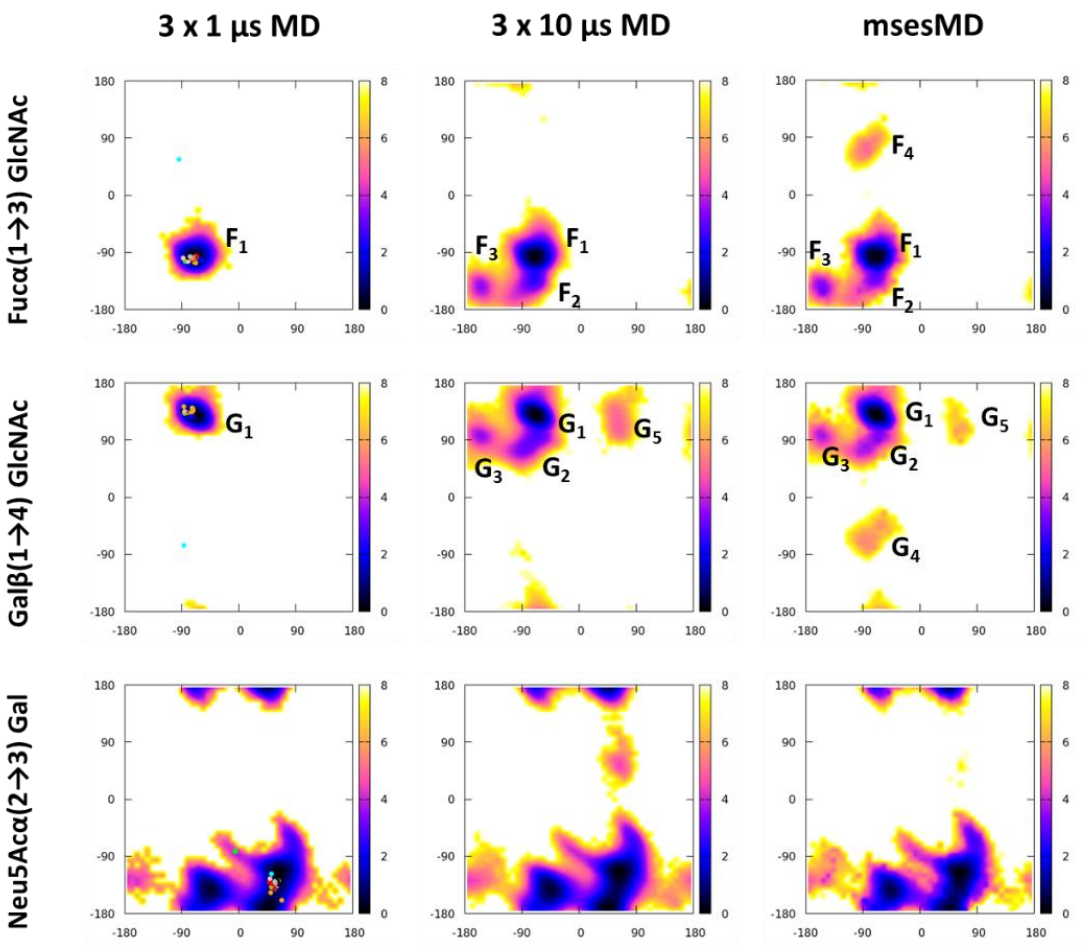

a
Fuc
GlcNAc
Gal
b
Gal
GlcNAc
Fuc
c
Fuc
GlcNAc
Neu5Ac
Gal
ψ_F
ψ_G
d
Neu5Ac
Gal
GlcNAc
Fuc
e
Fuc
GlcNAc
Gal
Fuc
f
Fuc
GlcNAc
Gal
Fuc

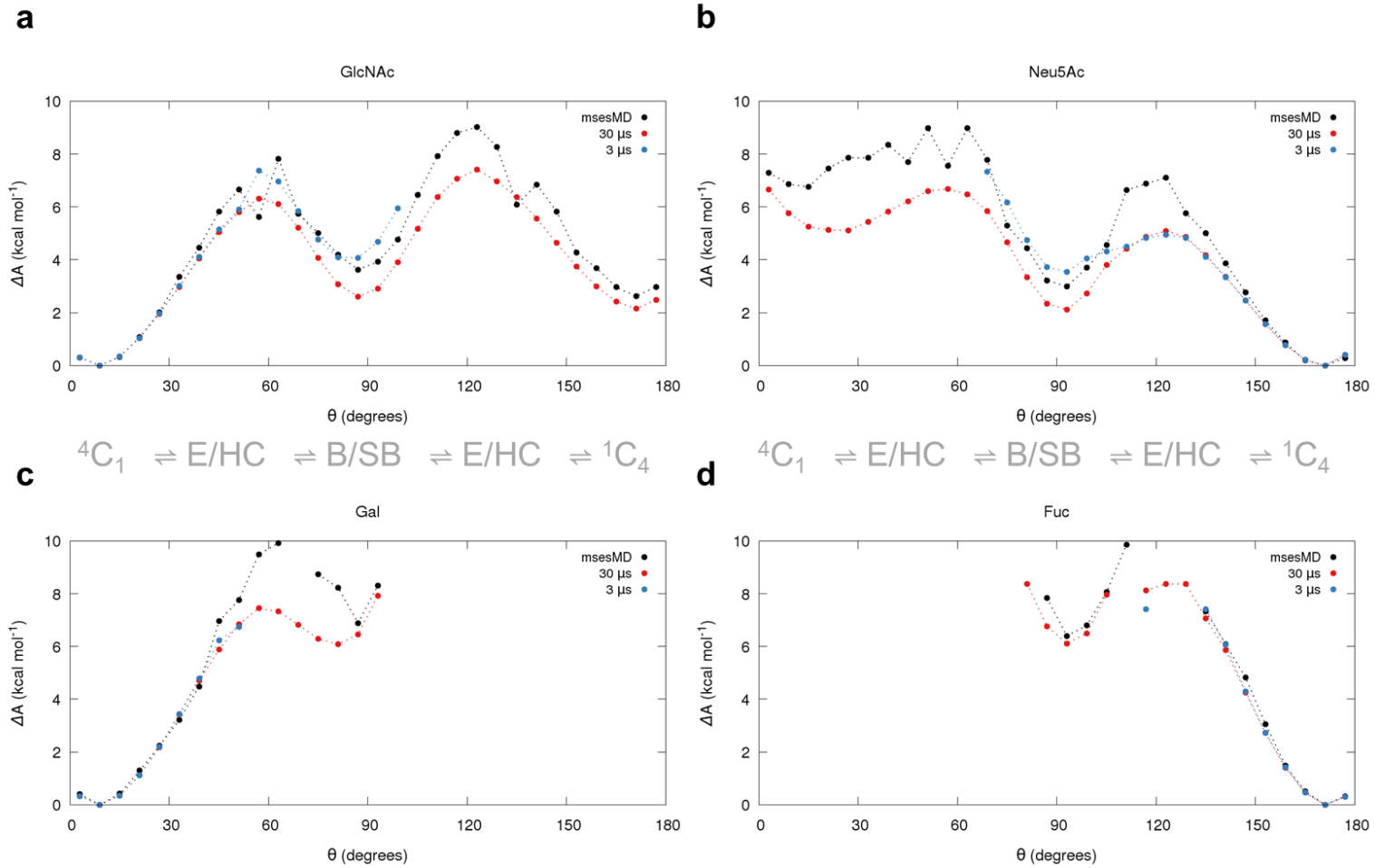

a
b
c
1.87
d
e
f
g

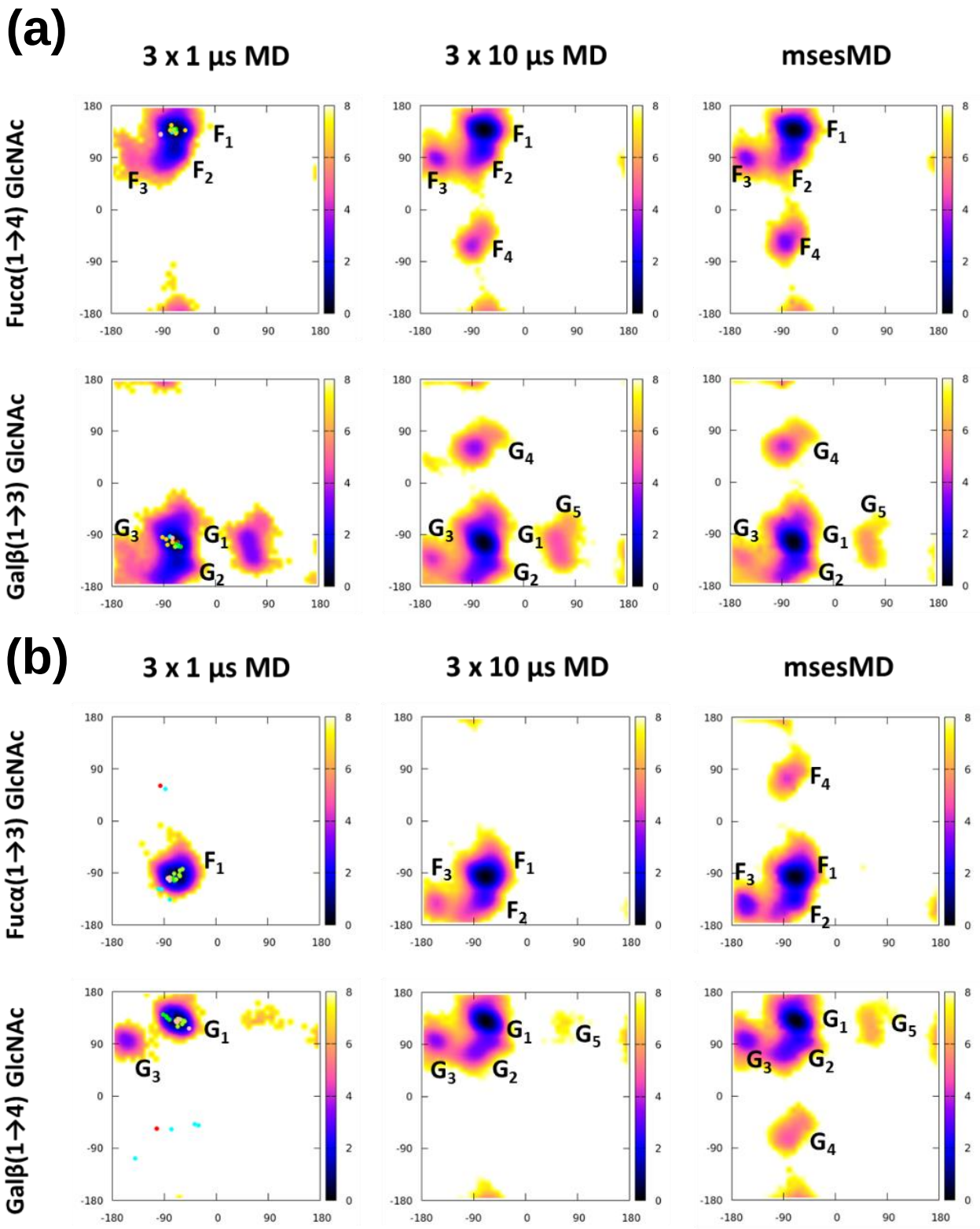

3 x 1 µs MD
3 x 10 µs MD
msesMD
aMD
Fucα(1→4) GlcNAc
Galβ(1→3) GlcNAc
Neu5Acα(2→3) Gal
F1
F2
F3
F4
F5
G1
G2
G3
G4
G5


a
GlcNAc
msesMD
30 µs
3 µs
ΔA (kcal mol-1)
θ (degrees)
b
Neu5Ac
msesMD
30 µs
3 µs
ΔA (kcal mol-1)
θ (degrees)
4C1 ⇌ E/HC ⇌ B/SB ⇌ E/HC ⇌ 1C4
4C1 ⇌ E/HC ⇌ B/SB ⇌ E/HC ⇌ 1C4
c
Gal
msesMD
30 µs
3 µs
ΔA (kcal mol-1)
θ (degrees)
d
Fuc
msesMD
30 µs
3 µs
ΔA (kcal mol-1)
θ (degrees)


a
sLe^a ψF angle free energy surface
Umbrella Sampling
msesMD
30 µs MD
3 µs MD
F4
F2/3
F1
ΔA (kcal mol-1)
ψF (degrees)
b
sLe^a ψG angle free energy surface
Umbrella Sampling
msesMD
30 µs MD
3 µs MD
G2/3
G1
G4
ΔA (kcal mol-1)
ψG (degrees)


3 x 1 µs MD
3 x 10 µs MD
msesMD
Fucα(1→3) GlcNAc
Galβ(1→4) GlcNAc
Neu5Acα(2→3) Gal
F1
F2
F3
F4
G1
G2
G3
G4
G5


a
GlcNAc
msesMD
30 µs
3 µs
ΔA (kcal mol-1)
θ (degrees)
4C1 ⇌ E/HC ⇌ B/SB ⇌ E/HC ⇌ 1C4
b
Neu5Ac
msesMD
30 µs
3 µs
ΔA (kcal mol-1)
θ (degrees)
4C1 ⇌ E/HC ⇌ B/SB ⇌ E/HC ⇌ 1C4
c
Gal
msesMD
30 µs
3 µs
ΔA (kcal mol-1)
θ (degrees)
d
Fuc
msesMD
30 µs
3 µs
ΔA (kcal mol-1)
θ (degrees)


(a)
3 x 1 µs MD
3 x 10 µs MD
msesMD
Fucα(1→4) GlcNAc
Galβ(1→3) GlcNAc
F1
F2
F3
F4
G1
G2
G3
G4
G5
(b)
3 x 1 µs MD
3 x 10 µs MD
msesMD
Fucα(1→3) GlcNAc
Galβ(1→4) GlcNAc
F1
F2
F3
F4
G1
G2
G3
G5