PCCP

www.rsc.org/pccp

PAPER

Assigning kinetic 3D-signatures to glycocodes†

Benedict M. Sattelle and Andrew Almond*

Reconciling glycocodes and their associated bioactivities, *via* 3D-structure, will rationalise burgeoning high-throughput functional glycomics data and underpin a new era of opportunity in chemical biology. A major impasse to achieving this goal is a detailed understanding of pyranose sugar ring 3D-conformation (or pucker) and the affiliated microsecond-timescale exchange kinetics. Here, we perform hardware-accelerated kinetically-rigorous equilibrium simulations of fundamental monosaccharides to produce the hypothesis that pyranoses have microsecond-timescale kinetic puckering signatures in water, classified as unstable (rare in the glycome), metastable (infrequently observed) and stable (prevalent). The predicted μs-metastability of β-D-glucose explained hitherto irreconcilable experimental measurements. Twisted puckers seen in carbohydrate enzymes were present in the aqueous 3D-ensemble (suggesting preorganization) and pyranose-water interactions accounted for the relative stability of β-D-galactose. Characteristic 3D-shapes for biologically- and commercially-important carbohydrates and new rules linking chemical modifications with pyranose μs-puckering kinetics are proposed. The observations advance structural-glycomics towards dynamic 3D-templates suitable for structure-based design.

Introduction

New 3D-insights are needed to reconcile carbohydrate sequences, or glycocodes (Fig. 1), with burgeoning functional data resulting from high-throughput glycomics.[1–4] Deciphering glycocodes by assigning them 3D-shapes, promises to impact contemporary challenges in chemical biology, ranging from disease amelioration to food sustainability. Unfortunately, glycomic analogues of widely recognisable 3D-motifs, such as the DNA double-helix and protein α-helices, have not yet been identified. It is clear that static 3D-snapshots will be insufficient due to the glycome's intrinsic molecular flexibility, which is fundamental to its functional diversity, as in the proteome.[5] Only when new breakthroughs in the understanding of carbohydrate 3D-shape kinetics are realized will this fundamental and information-rich class of molecule be finally harnessed. This would, for instance, revolutionize new chemical entity design for influenza targets (*e.g.*, neuraminidase), by permitting the desirable attribute of carbohydrate-protein specificity to be rationally married with slow off rates and strong binding (atypical for endogenous glycans).[6]

The main 3D-structural degrees of freedom in carbohydrates, which ultimately determine their biological function, are glycosidic linkage geometries and monosaccharide ring conformational exchange (or puckering). Linkage dynamics and five-membered furanose puckering (which differentiates the 3D-structures of DNA and RNA)[7] occur on nanosecond timescales; these phenomena are tractable to NMR spectroscopy and rigorous aqueous computer simulations that can probe kinetics. However, in the ubiquitous six-membered pyranoses (*e.g.*, glucose) puckering remains enigmatic, like so many important biological phenomena that possess relatively high ($\approx$ 10 kcal mol^{-1}) free energy barriers and consequently undergo microsecond (μs) exchange. This timescale is often out of reach by experiment alone and theoretical techniques that can readily traverse the associated barriers, such as enhanced sampling molecular dynamics, distort the very kinetics they set out to elucidate.[8]

Kinetically-rigorous simulations, wherein the sugar and solvent molecules are explicitly modelled and conformational sampling is not artificially enhanced (*e.g.*, by elevated temperature), has been a missing link to characterization of 3D-molecular exchange on μs-timescales and beyond. Such simulations can probe the dynamic 3D-signatures of individual pyranose puckers, facilitating derivation of relationships between puckering and bioactivity. Not only do they provide transition pathways (itineraries) and free energies, they crucially also reveal equilibrium puckering kinetics, which can be used to interpret experimental measurements. The main technical challenge that has limited its use is one of processing power. Iterative and computationally-intensive calculations (integrations) take place

Faculty of Life Sciences, Manchester Interdisciplinary Biocentre, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK. E-mail: Andrew.Almond@manchester.ac.uk; Tel: +44(0)161 306 4199
† Electronic supplementary information (ESI) available: Simulation data including computed pucker populations, relative free energies, average paramaters, convergence to equilibrium, phase-space projections, time series and canonical classifications. See DOI: 10.1039/c2cp40071e
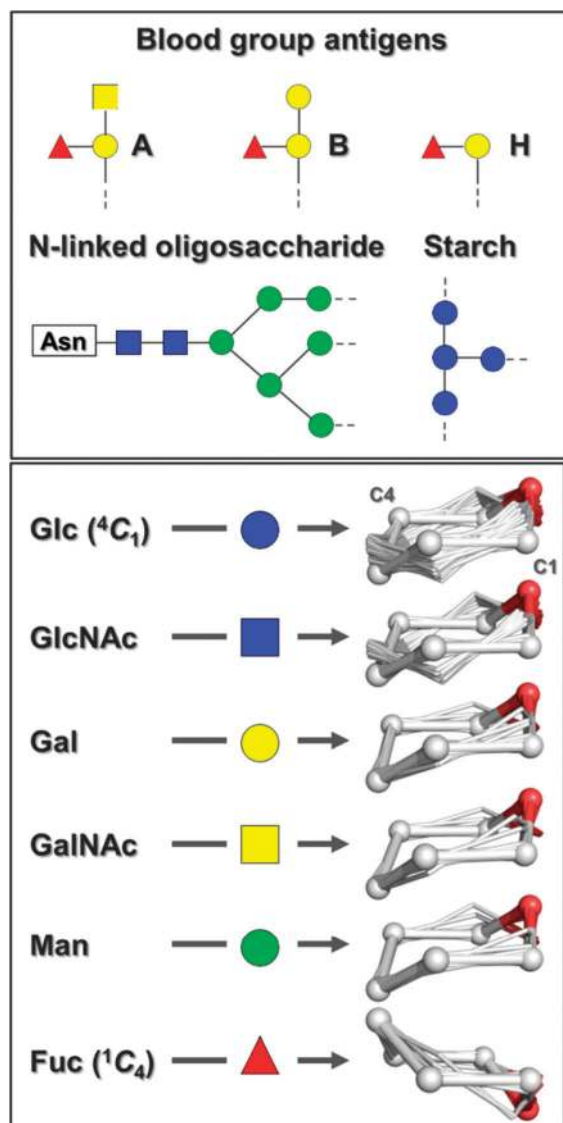

Fig. 1 Top: Carbohydrate sequences (glycocodes) that include monosaccharides considered in this study. The blood group antigens A, B and H are present on all mammalian cells, oligomannose is a key feature of *N*-linked glycosylation and starch is used by all green plants to store energy. Bottom: Assigning 3D-kinetic signatures to glycocodes is central to rationalising bioactivity. For the five pyranoses studied (and GlcNAc),[9] non-chair intermediate ring puckers (shown as lines) computed to be less than 5 kcal mol^{-1} above the lowest energy chair (ball-and-stick) are illustrated adjacent to their respective standard symbol notation, from which it is essential to progress in order to reconcile sequence and emergent functional glycomics data.

over very short time periods (femtoseconds) in order to capture the fastest molecular motions, while pyranose puckering occurs on a timescale that is ten orders of magnitude larger.

In order to finally reveal pyranose puckering µs-kinetics, we recently adopted massively-parallel graphics processing units (GPUs) to perform the first kinetically-rigorous aqueous carbohydrate simulations on the µs-timescale. In the uronic acid monosaccharides α-L-iduronic acid (IdoA) and 2-*O*-sulfo-IdoA, the 3D-ensembles were equilibrated, µs-exchange between $^4C_1$ and $^1C_4$ chair puckers (crucial for heparin bioactivity)[10] was observed

and the predicted populations of both chairs and non-chair (intermediate) puckers (Fig. 1) agreed with measurements from ultra-high-field NMR spectra (ring three-bond proton-proton vicinal couplings or $^3J_{H,H}$).[11] Subsequent work led to the prediction that *N*-acetyl-α-D-glucosamine (GlcNAc) undergoes substituent-dependent $^4C_1 \leftrightarrow {}^1C_4$ µs-transitions at equilibrium.[9] While unexpected (pyranoses are conventionally considered to be stable and chair-like in 3D-shape), this observation was not inconsistent with experiment.

To extend these initial observations, we have performed similar analyses on five monosaccharides with extremely fundamental and wide-ranging biological roles: β-D-glucose (Glc); β-D-galactose (Gal), β-D-*N*-acetyl-galactosamine (GalNAc), α-D-mannose (Man), and α-L-fucose (Fuc). In particular, we sought to explore the stability of their respective dynamic 3D-ensembles on µs-timescales in water, to advance our hypothesis in a range of monosaccharides and also to associate computational predictions with bioactivity.

Aqueous simulations of each 1-*O*-methylglycoside (as a model for polymerization) were extended to 5 µs and puckering was quantified as described previously.[9] Convergence of puckering was achieved after ≈2 µs (see ESI† Fig. S1) and calculated equilibrium $^3J_{H,H}$-values agreed with previous measurements. Our findings retrospectively explain biological and biophysical data and prospectively enable structure-based design of carbohydrate-based materials, chemical probes and drugs (*e.g.*, novel enzyme inhibitors).

Results and discussion

When compared with a previous kinetically-rigorous simulation of GlcNAc,[9] the simulation of Glc (Fig. 2) depicted a pyranose ring that was a $^4C_1$ chair for less of the trajectory (97% *versus* over 99%). A larger percentage of $^1C_4$ puckers were observed and more of the 36 canonical intermediates were sampled (29 contrasted with 20 in GlcNAc, see ESI† Table S1). Computed forward ($^4C_1 \rightarrow {}^1C_4$) and backward ($^1C_4 \rightarrow {}^4C_1$) rates of Glc pyranose exchange were 2 and 200 µs^{-1}, respectively (*cf.* 1 and 300 µs^{-1} in GlcNAc). Average lifetimes of the $^1C_4$ chair (5 ns) were also longer in Glc (3 ns in GlcNAc), which resulted from a lower computed free energy barrier separating the two chair puckers ($\Delta G = 3$ kcal mol^{-1} in Glc), a difference of $\Delta\Delta G = -0.7$ kcal mol^{-1} between the monosaccharides. The seven $^1C_4$ hemisphere puckers not sampled in the Glc simulation were estimated to have relative free energies (*cf.* $^4C_1$) of $\Delta G = 9$–10 kcal mol^{-1} (Fig. 3a). Strikingly, this observation points to potential benefits of performing even more extended molecular dynamics on simple monosaccharides. It also highlights the inadequacy of previous aqueous carbohydrate simulations, which have typically been hundreds of nanoseconds in duration.

In simulations of glucose's 4- and 2-epimers, Gal and Man, neither underwent $^4C_1 \leftrightarrow {}^1C_4$ exchange (Fig. 3); both were predominantly $^4C_1$ chairs (>99%) and had comparatively restricted pseudorotational itineraries, populating only a handful of adjacent equatorial ($\theta = 90°$, boat and skew-boat) puckers (these were $^1S_3/B_{3,O}$ in Gal and $B_{3,O}/{}^2S_O$ in Man). For Gal and Man, respectively, the populated half-chair and envelope puckers adjacent to the $^4C_1$ chair were centred on $^4H_3/{}^4E$ and $^2E/{}^2H_3$. Relative to the $^4C_1$ pucker, free energies of
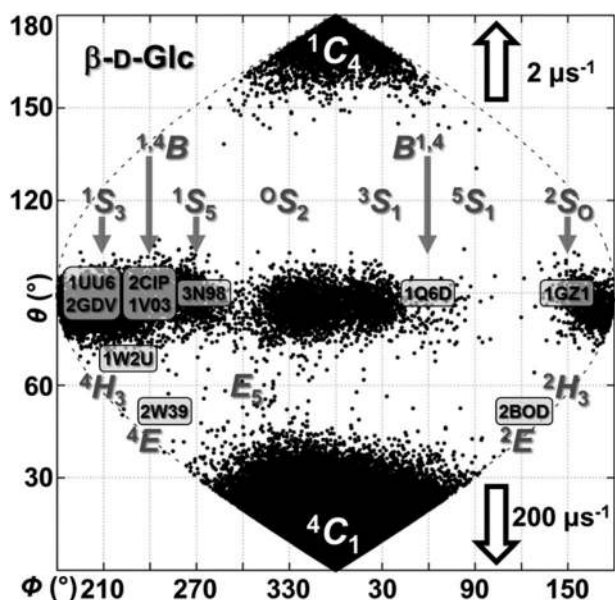
Fig. 2 A sinusoidal projection of 1-O-methyl-β-D-glucose (Glc) equilibrium puckering during a 5 µs kinetically-rigorous explicit solvent simulation. Each data point represents a single pucker using the Cremer-Pople angle nomenclature.[15] Boats (B) and skew-boats (S) are labelled above their respective positions on the equator. Half-chairs (H) and envelopes (E) are labelled at their respective locations. Protein Data Bank identifiers corresponding to high-resolution (≤ 2.0 Å) co-crystals including Glc in a non-chair intermediate pucker are marked at their respective sites. Analogous plots for all monosaccharides studied (and GlcNAc,[9] for comparison) are shown in ESI† Fig. S2.
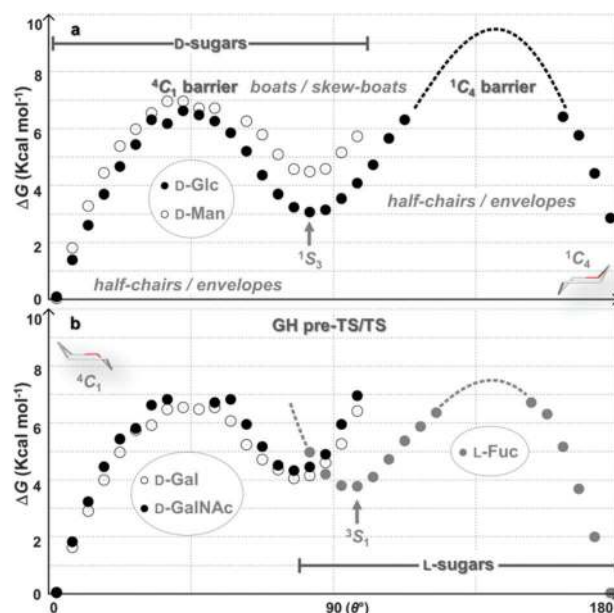

Fig. 3 One-dimensional equilibrium free energy (G) landscapes derived from populations of puckers during 5 µs kinetically-rigorous explicit solvent simulations. a: 1-O-methyl-β-D-glucose (●) and -α-D-mannose (○), b: -β-D-galactose (○), -β-D-N-acetyl-galactosamine (●) and -α-L-fucose (●). The lowest energy equatorial puckers, bioactive glycoside hydrolase (GH) transition-states (TS) and $^4C_1$ and $^1C_4$ hemisphere free energy barriers are noted. The free energy barriers of Glc and Fuc (--/--) were extrapolated using a suitable curve. Rarely populated puckers ($\Delta G > 7$ kcal mol^{-1}) are not shown.

their equatorial puckers were also greater than the corresponding values in Glc, by $\Delta G \approx 1$ kcal mol^{-1} (Fig. 3). This accords with our previous observation[11] that even small changes in monosaccharide stereochemistry impact both the puckering equilibrium and kinetics.

The fucose simulation revealed a stable $^1C_4$ chair that did not undergo $^1C_4 \leftrightarrow {}^4C_1$ transitions (Fig. 3) and occupied just 8/36 canonical intermediate puckers. Relative to $^1C_4$, only the $^3S_1$ pucker had a low energy ($\Delta G = 3$ kcal mol^{-1}); all other Fuc intermediates were greater than $\Delta G = 4$ kcal mol^{-1} above $^1C_4$. Also, the $^1C_4$ hemisphere free energy barrier (intermediate half-chairs and envelopes separating the lowest energy chair from the equator, see Fig. 3), was $\Delta G \approx 7$–8 kcal mol^{-1} at its highest point and greater in magnitude than the analogous $^4C_1$ barrier in the D-sugars Glc, Gal, GalNAc and Man ($\Delta G \approx 6$–7 kcal mol^{-1}). The free energy surfaces shown in Fig. 3 describe, in our opinion, the fundamental puckering signatures for all biologically-important pyranoses. These observations suggest that the $^4C_1 \leftrightarrow {}^1C_4$ µs-transitions predicted previously in GlcNAc apply also to Glc and, while Gal, GalNAc, Man and Fuc explored only one chair and a small number of boats and skew-boats, other pyranoses cannot be assumed to be rigid chairs on the µs-timescale.

Considering the thirteen pyranose monosaccharides on which we have performed kinetically-rigorous µs-simulations to date (in the range 5–20 µs), their 3D-kinetic signatures can be classified as either: (1) unstable, in which many (forward) transitions from the lowest energy chair to the highest energy chair occurred each µs, only observed in IdoA,[11] (2) metastable,

where fewer forward transitions (≈ 1–3 µs^{-1}) are predicted, as in IdoA2S,[11] GlcNAc[9] and Glc, and (3) stable, i.e., pyranoses that did not to undergo any chair-chair transitions. The latter group is the largest, comprising GlcA,[11] four sulfonated GlcNAc derivatives (GlcNS, GlcNS6S, GlcNS6S3S and GlcNS3S)[9] and Gal, GalNAc, Man and Fuc investigated in this study.

Even allowing for possible errors in the absolute populations of simulated puckers, all computed axial-axial $^3J_{H,H}$-values in Glc were smaller (indicative of greater flexibility) than in GlcNAc,[9] Gal, GalNAc, Man and Fuc, a trend that is also evident in the corresponding experimental data (Table 1). The predictions that Glc has, comparatively, a greater pseudorotational itinerary and also that it undergoes transient chair-chair interconversions, therefore reconcile the previous experimental NMR measurements. While the employed force field potentials, GLYCAM06[16] (sugar) and TIP3P[17] (water), have not been extensively tested for pyranose puckering on µs-timescales, they performed very well in our previous µs-simulations[9,11] and it is encouraging that here, on a wide range of monosaccharides, all calculated $^3J_{H,H}$-values deviate from experimental measurements by ≤ 2 Hz (Table 1). Better agreement may be achievable via small modifications to force field parameters.

Since no perceivable intramolecular steric hindrance results from the stereoisomerization that occurs between Glc, Gal and Man, their similar $^4C_1$ hemisphere free-energy barriers (Fig. 3) were in agreement with our expectations. We propose that the relatively higher-energy equatorial and $^1C_4$ hemisphere puckers of Gal and Man (cf. Glc) is attributable to different through-space interactions between monosaccharide and water, involving

Table 1 Experimental (Expt) and calculated (Calc) three-bond proton-proton vicinal couplings ($^3J_{H,H}$). Axial-axial $^3J_{H,H}$-values smaller in Glc are in bold

$^3J_{H,H}$	Glc		Gal		GalNAc		Man		Fuc	
	Expt[a]	Calc	Expt[a]	Calc	Expt[b]	Calc	Expt[a]	Calc	Expt[c]	Calc
$J_{1,2}$	**7.8**	**6.5**	**8.0**	**8.1**	**8.5**	**7.5**	1.8	2.3	4.0	3.5
$J_{2,3}$	**9.5**	**7.3**	**10.0**	**8.3**	**10.8**	**8.9**	3.8	2.6	**10.0**	**8.4**
$J_{3,4}$	**9.5**	**7.3**	3.8	2.5	0.6	2.3	**10.0**	**8.3**	4.2	2.4
$J_{4,5}$	**9.5**	**8.2**	1.0	0.8	2.4	0.8	**9.8**	**8.2**	0.9	0.9

Experimental values from. [a] Dowd *et al.*[12] (Glc, Gal and Man), and. [b] Sattelle *et al.*[13] (GlcNAc in a tetrasaccharide) and. [c] Otter *et al.*[14] (1-linked Fuc in a trisaccharide). Experimental errors were estimated to be 0.2 Hz, standard deviations of calculated values were all <0.05 Hz.

the epimerized (4- and 2-position) hydroxyls. This view is further supported by the finding that the calculated Glc hydroxymethyl rotamer distribution (2:49:49 for *tg*:*gt*:*gg* states, which agrees with prior experimental observations)[18,19] is dissimilar to the predictions for GlcNAc,[9] Gal, GalNAc and Man (see ESI† Table S2). Fig. 4 summarizes the predicted effect of 4-epimerization on the aqueous free energies of biologically-important 3D-intermediates in Glc and Gal.

The ubiquitous 2-position *N*-acetylamino group of Glc and Gal was predicted to have a negligible effect on pyranose μs-dynamics. In GlcNAc, compared to Glc, the $^4C_1$ chair was marginally stabilized and the $^4C_1 \rightarrow ^1C_4$ exchange rate was only a little slower.[9] The free energy puckering landscapes of GalNAc and Gal were also extremely similar (Fig. 3b). Slightly reduced populations of equatorial and $^4C_1$ hemisphere puckers were seen in GalNAc, however, $^1S_3$ and $B_{3,O}$ were the lowest energy intermediates in both sugars, having comparable relative free energies (*cf.* $^4C_1$, $\Delta G \approx 4$ kcal mol^{-1}), while the hydroxymethyl rotamer state distributions and average $^4C_1$ puckers were almost identical (see ESI† Table S2 and S3). In light of the significant effect of epimerization on simulated μs-puckering (kinetics and thermodynamics), it is notable that

Fig. 4 Puckers of 1-*O*-methyl-β-D-glucose (D-Glc) and -D-galactose (D-Gal) representing the lowest energy chair ($^4C_1$) and the glycoside hydrolase glycopyranosidic bond substitution pre-transition-state ($^1S_3$) and transition-state ($^4H_3$). This comparison highlights differences in computed equilibrium free energies between the 4-position epimers and also that there is no intramolecular steric hindrance in either itinerary. Non-polar hydrogen atoms are hidden for clarity.

addition of a bulky *N*-acetylamino substituent did not modify pyranose ring dynamics. This suggests that protein recognition of these sugars is likely due to interactions with the *N*-acetylamino group and not specific pyranose ring 3D-intermediates.

Conclusions

Previous studies have hinted at Glc exploring non-$^4C_1$ puckers,[20–26] however, equilibrium puckering μs-kinetics and free energies have not previously been quantified from kinetically-rigorous aqueous molecular dynamics. Two of the most important biomaterials, cellulose and starch, are polymers of Glc (Fig. 1). In glycoside hydrolase (GH) enzymes, which modify or degrade these Glc-based molecules (and a plethora of other carbohydrates), protein-mediated ring distortion optimizes pyranose atom positions with respect to catalytic residues.[22–24] In the Glc simulation, the pre-transition-state energy minimum in GH mediated glycopyranosidic bond substitutions, $^1S_3$,[27] was the predominant intermediate pucker ($\Delta G = 3$ kcal mol^{-1} *cf.* $^4C_1$). Furthermore, Glc puckers centred on the GH transition-state ($E_3/^4H_3/^4E$), in which the pyranose C5-O5-C1-C2 atoms are co-planar,[28] were all predicted to be accessible in water on the μs-timescale (see Fig. 2 and ESI† Fig. S3). These two Glc puckers and other explored intermediates are present in numerous high-resolution GH co-crystals; some examples are illustrated in Figs. 2 and 5. We therefore propose that in water Glc is a metastable $^4C_1$ chair (like IdoA2S[11] and GlcNAc),[9] it undergoes facile pseudorotation and also that it transiently occupies puckers recognized by GH enzymes.

Monosaccharides predicted to have comparatively stable pyranose rings, GalNAc, Gal and Fuc, are components of numerous *N*- and *O*-linked oligosaccharides, in particular, mammalian blood group antigens that are presented on all cells. The $^3S_1$ pucker was the most populated intermediate in the Fuc simulation, which is consistent with the presence of $^3S_1$ in a 2.3 Å resolution co-crystal of the GH α-L-fucosidase and a derivatized Fuc inhibitor (Fig. 5).[32] Absence of chair-chair transitions in the computed GalNAc, Gal and Fuc equilibrium 3D-ensembles leads us to propose that the blood group antigens A, B and H (Fig. 1) have relatively stable chair puckers, and that recognition by proteins is less likely to involve non-chair 3D-intermediates.

Kinetic subtleties of pyranose μs-puckering are only now beginning to emerge, enabled by the extended duration of our recent kinetically-rigorous GPU-accelerated simulations. The calculated free energy landscapes comprise definitive (equilibrium) predictions (for the validated GLYCAM06/TIP3P[16,17] model)
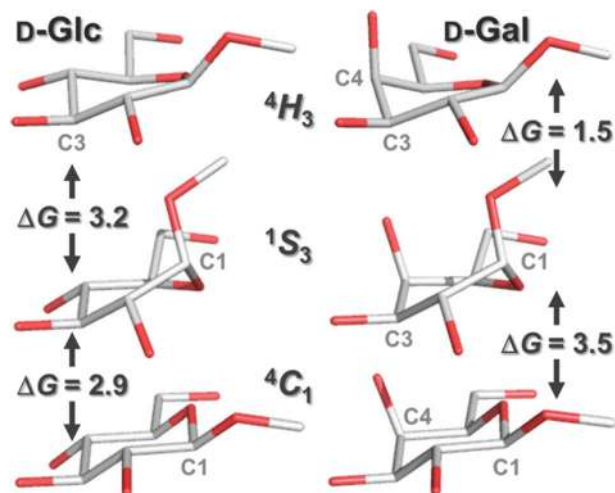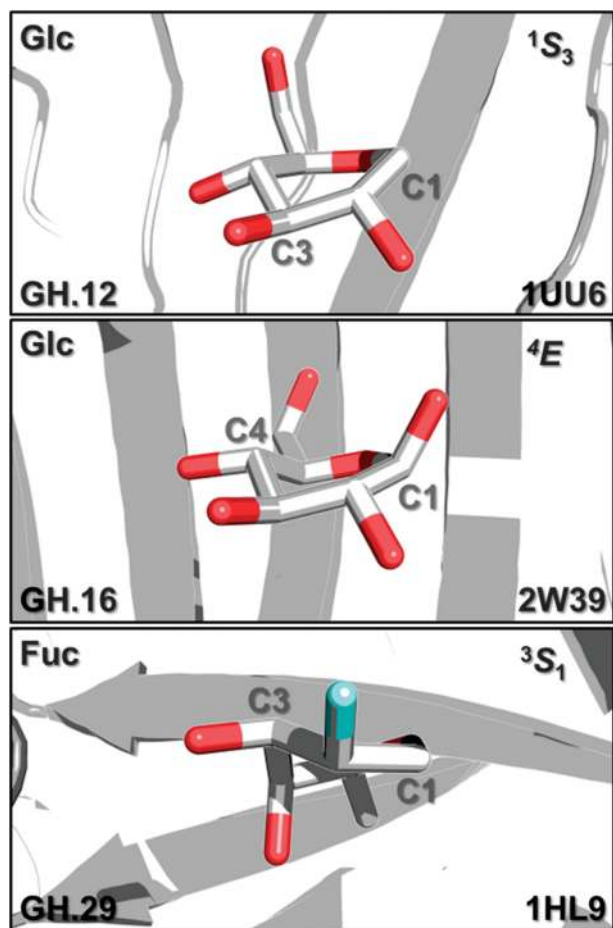
Fig. 5 Pre-transition-state and transition-state puckers of β-D-glucose and an inhibitory fluorinated fucose derivative in high-resolution co-crystals of glycoside hydrolase (GH) enzymes (families 12, 16 and 29).[29–31] Puckers and Protein Data Bank identifiers are labelled, background active site residues are shown as ribbons and sugar hydrogen atoms are hidden for clarity.

of aqueous µs-puckering in monosaccharides of biological- and commercial-importance. Alongside our hypothesis, that the glycome contains three pyranose µs-timescale 3D-kinetic signatures (unstable, metastable and stable), interpretation of the molecular dynamics reported here and previously has led to new rules for chemical modifications governing pyranose µs-puckering: (a) Glc 4- and 2-position epimerization stabilizes the $^4C_1$ chair, (b) N-acetylation has little effect on µs-puckering, (c) IdoA and GlcNAc sulfonation and (d) polymerization slow puckering µs-kinetics and modulate pucker free energies.[9,11] Topologies of equilibrium pyranose µs-puckering free energy landscapes in fundamental D- and L-sugars have been generalized to be $^4C_1$- and $^1C_4$-like, respectively, while other emergent principles have also been identified: (a) pyranoses cannot be assumed de facto to be rigid chairs, (b) water structure plays a key role in pyranose puckering, (c) carbohydrate-modifying enzymes select for pyranose puckers that are transiently populated in water, and (d) that Gal, GalNAc, Man and Fuc, which are commonly found in N- and O-linked glycosylation (and blood group antigens in particular), have 3D-shapes that are less flexible than IdoA, IdoA2S, Glc and GlcNAc.

Additional µs-simulations are required to investigate puckering in pyranoses not yet studied (e.g, sialic acid and xylose), to probe the role of water in molecular detail and to further study the effect of polymerization on µs-puckering. An experimental application of our predictions will be their use in a priori rational derivatization of pyranoses to achieve modified µs-puckering kinetics, populations of key 3D-intermediates and bioactivity (shown previously experimentally for conformationally-locked hexosamines,[33] IdoA,[34,35] and GH substrates).[36,37] In concert with hypothesis-driven experimentation and understanding the effects of pyranose chemical modifications on glycosidic linkage flexibility, the reported equilibrium free energy landscapes and our classification of pyranose µs-puckering 3D-kinetic signatures provide dynamic 3D-templates for structure-based design of new carbohydrates and lay the foundation for 3D-kinetic assignment of arbitrary glycocodes.

Methods

Simulations were performed as described previously.[9,11] Briefly, each monosaccharide was initiated as a $^4C_1$ chair (except Fuc, $^1C_4$) in a cubic water box ($\approx 25 \times 25 \times 25$ Å) using the AMBER[38] tool Leap. Dynamics were performed with ACEMD[39] on Nvidia GPUs using the GLYCAM06[16] and TIP3P[17] force-fields. Following initial conjugate-gradient energy minimisation (1000 steps), each system was equilibrated in the NPT ensemble (20 ns) prior to 5.25 µs NVT production dynamics, the first 250 ns were discarded and data were recorded at 10 ps intervals. The velocity-Verlet integration algorithm and a hydrogen mass repartitioning scheme allowed a 4 fs time-step without affecting the equilibrium distribution.[40] Hydrogen atoms were constrained using the M-SHAKE algorithm[41] and electrostatic interactions were calculated via the PME method with a grid spacing of 0.8 Å (in the X, Y and Z dimensions). Electrostatic and van der Waals interactions were truncated at 9 Å and the recommended[16] scaling factor for carbohydrate 1–4 interactions (1.0) was employed. All properties were computed using the complete 5 µs trajectories (500 000 data points). Analyses of the simulations (including free energy and rate calculations) and the Protein Data Bank co-crystals were performed as described previously.[9,11] Relative free energies (ΔG) of binned puckers were calculated using the standard equation $\Delta G = kT\ln(p_1/p_2)$, where k is the Boltzmann constant, T is temperature (298 K) and p is probability. Pyranose ring three-bond proton-proton vicinal couplings were calculated using substituent adjusted Karplus equations of Altona and Haasnoot[42] and puckers were quantified using calculated Cremer-Pople[15] parameters (θ, φ and Q). In Fig. 3, puckers were placed in bins of unequal spacing in the azimuthal angle θ, but equal spacing in the meridian angle φ (ensuring that each bin occupies an equal area on the sphere). In all other cases puckers were defined as being one of the 38 canonical puckers, viz.: $^4C_1$, ^OE, $^OH_1$, E_1, $^2H_1$, ^2E, $^2H_3$, E_3, $^4H_3$, ^4E, $^4H_5$, E_5, $^OH_5$, $^{3,O}B$, $^3S_1$, $B_{1,4}$, $^5S_1$, $^{2,5}B$, ^2S_O, $B_{3,O}$, $^1S_3$, $^{1,4}B$, $^1S_5$, $B_{2,5}$, $^OS_2$, E_O, ^1H_O, ^1E, $^1H_2$, E_2, $^3H_2$, ^3E, $^3H_4$, E_4, $^5H_4$, ^5E, ^5H_O, $^1C_4$ (definitions in terms of θ and φ angles can be found in ESI† Table S4).

References

1 X. Song, Y. Lasanajak, B. Xia, J. Heimburg-Molinaro, J. M. Rhea, H. Ju, C. Zhao, R. J. Molinaro, R. D. Cummings and D. F. Smith, Nat. Methods, 2010, **8**, 85–90.
2 J. E. Turnbull and R. A. Field, Nat. Chem. Biol., 2007, **3**, 74–77.

3 G. W. Hart and R. J. Copeland, *Cell*, 2010, **143**, 672–676.
4 M. Ly, F. E. Leach, 3rd, T. N. Laremore, T. Toida, I. J. Amster and R. J. Linhardt, *Nat. Chem. Biol.*, 2011, **7**, 827–833.
5 H. J. Dyson and P. E. Wright, *Nat. Rev. Mol. Cell Biol.*, 2005, **6**, 197–208.
6 J. Holgersson, A. Gustafsson and M. E. Breimer, *Immunol. Cell Biol.*, 2005, **83**, 694–708.
7 A. Rich, *Nat. Struct. Biol.*, 2003, **10**, 247–249.
8 T. Schlick, *F1000. Biol. Rep.*, 2009, **1**, 51.
9 B. M. Sattelle and A. Almond, *Glycobiology*, 2011, **21**, 1651–1662.
10 M. Hricovini, M. Guerrini, A. Bisio, G. Torri, M. Petitou and B. Casu, *Biochem. J.*, 2001, **359**, 265–272.
11 B. M. Sattelle, S. U. Hansen, J. Gardiner and A. Almond, *J. Am. Chem. Soc.*, 2010, **132**, 13132–13134.
12 M. K. Dowd, A. D. French and P. J. Reilly, *Carbohydr. Res.*, 1994, **264**, 1–19.
13 B. M. Sattelle, J. Shakeri, I. S. Roberts and A. Almond, *Carbohydr. Res.*, 2010, **345**, 291–302.
14 A. Otter, R. U. Lemieux, R. G. Ball, A. P. Venot, O. Hindsgaul and D. R. Bundle, *Eur. J. Biochem.*, 1999, **259**, 295–303.
15 D. Cremer and J. A. Pople, *J. Am. Chem. Soc.*, 1975, **97**, 1354–1358.
16 K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. Gonzalez-Outeirino, C. R. Daniels, B. L. Foley and R. J. Woods, *J. Comput. Chem.*, 2008, **29**, 622–655.
17 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
18 Y. Nishida, H. Ohrui and H. Meguro, *Tetrahedron Lett.*, 1984, **25**, 1575–1578.
19 R. Stenutz, I. Carmichael, G. Widmalm and A. S. Serianni, *J. Org. Chem.*, 2002, **67**, 949–958.
20 P. E. Marszalek, A. F. Oberhauser, Y. P. Pang and J. M. Fernandez, *Nature*, 1998, **396**, 661–664.
21 G. J. Davies, V. M. Ducros, A. Varrot and D. L. Zechel, *Biochem. Soc. Trans.*, 2008, **31**, 523–527.
22 E. J. Taylor, A. Goyal, C. I. Guerreiro, J. A. Prates, V. A. Money, N. Ferry, C. Morland, A. Planas, J. A. Macdonald, R. V. Stick, H. J. Gilbert, C. M. Fontes and G. J. Davies, *J. Biol. Chem.*, 2005, **280**, 32761–32767.
23 V. A. Money, N. L. Smith, A. Scaffidi, R. V. Stick, H. J. Gilbert and G. J. Davies, *Angew. Chem., Int. Ed.*, 2006, **45**, 5136–5140.
24 X. Biarnes, A. Ardevol, A. Planas, C. Rovira, A. Laio and M. Parrinello, *J. Am. Chem. Soc.*, 2007, **129**, 10686–10693.
25 V. Spiwok, B. Kralova and I. Tvaroska, *Carbohydr. Res.*, 2010, **345**, 530–537.
26 S. Fushinobu, B. Mertz, A. D. Hill, M. Hidaka, M. Kitaoka and P. J. Reilly, *Carbohydr. Res.*, 2008, **343**, 1023–1033.
27 W. Nerinckx, T. Desmet and M. Claeyssens, *Arkivoc.*, 2006, 90–116.
28 C. B. Barnett, K. A. Wilkinson and K. J. Naidoo, *J. Am. Chem. Soc.*, 2010, **132**, 12800–12803.
29 M. Sandgren, G. I. Berglund, A. Shaw, J. Stahlberg, L. Kenne, T. Desmet and C. Mitchinson, *J. Mol. Biol.*, 2004, **342**, 1505–1517.
30 J. Vasur, R. Kawai, E. Andersson, K. Igarashi, M. Sandgren, M. Samejima and J. Stahlberg, *FEBS J.*, 2009, **276**, 3858–3869.
31 G. Sulzenbacher, C. Bignon, T. Nishimura, C. A. Tarling, S. G. Withers, B. Henrissat and Y. Bourne, *J. Biol. Chem.*, 2003, **279**, 13119–13128.
32 B. Henrissat, G. Sulzenbacher, C. Bignon, T. Nishimura, C. A. Tarling, S. G. Withers and Y. Bourne, *J. Biol. Chem.*, 2004, **279**, 13119–13128.
33 J. L. Asensio, A. Bastida, A. Hidalgo, J. L. Chiara, M. Torrado, F. Corzana, J. M. Perez-Canadillas, P. Groves, E. Garcia-Junceda, C. Gonzalez and J. Jimenez-Barbero, *J. Am. Chem. Soc.*, 2006, **128**, 100–116.
34 S. K. Das, J. M. Mallet, J. Esnault, P. A. Driguez, P. Duchaussoy, P. Sizun, J. P. Herault, J. M. Herbert, M. Petitou and P. Sinay, *Angew. Chem., Int. Ed.*, 2001, **40**, 1670–1673.
35 T. J. Boltje, T. Buskas and G. J. Boons, *Nat. Chem.*, 2009, **1**, 611–622.
36 K. Sivapriya, S. Hariharaputran, V. L. Suhas, N. Chandra and S. Chandrasekaran, *Bioorg. Med. Chem.*, 2007, **15**, 5659–5665.
37 D. J. Vocadlo and G. J. Davies, *Curr. Opin. Chem. Biol.*, 2008, **12**, 539–555.
38 D. A. Case, T. E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.
39 M. J. Harvey, G. Giupponi and G. De Fabritiis, *J. Chem. Theory Comput.*, 2009, **5**, 1632–1639.
40 I. Buch, M. J. Harvey, T. Giorgino, D. P. Anderson and G. De Fabritiis, *J. Chem. Inf. Model.*, 2010, **50**, 397–403.
41 V. Krautler, W. F. Van Gunsteren and P. H. Hunenberger, *J. Comput. Chem.*, 2001, **22**, 501–508.
42 C. Altona and C. A. G. Haasnoot, *Org. Magn. Reson.*, 1980, **13**, 417–429.