Free Energy Landscapes of Iduronic Acid and Related Monosaccharides

Benedict M. Sattelle, Steen U. Hansen, John Gardiner, and Andrew Almond*

Manchester Interdisciplinary Biocentre, University of Manchester, 131 Princess Street, Manchester M1 7DN, U.K.

Received June 21, 2010; E-mail: Andrew.Almond@Manchester.ac.uk

Abstract: The pyranose ring of L-iduronic acid (IdoA), a major constituent of the anticoagulant heparin, is an equilibrium of multiple ring puckers that have evaded quantification by experiment or computation. In order to resolve this enigma, we have calculated the free energy landscape of IdoA and two related monosaccharides from extensive microsecond simulations. After establishing that the simulated puckers had reached equilibrium, hypotheses were confirmed that (a) IdoA $^1C_4$- and $^4C_1$-chair conformations exchange on the microsecond time scale, (b) C5 epimerization leads to a $^4C_1$-chair, and (c) IdoA 2-O-sulfation (IdoA2S) stabilizes the $^1C_4$ conformer. The IdoA and IdoA2S $^1C_4$ conformers were isoenergetic and computed to be 0.9 and 2.6 kcal mol^{-1} lower in free energy than their respective $^4C_1$-chair conformations. The simulations also predicted that the IdoA ^2S_O-skew-boat was less populated than previously thought. Novel chemical synthesis and ultra-high-field NMR supported these observations, but slight discrepancies in observed and predicted NMR vicinal couplings implied that the simulation overestimated the population of the IdoA $^4C_1$-chair with respect to $^1C_4$-chair due to small force field inaccuracies that only manifest in long simulations. These free-energy calculations drive improvements in computational methods and provide a novel route to carbohydrate mimetic biomaterials and pharmaceuticals.
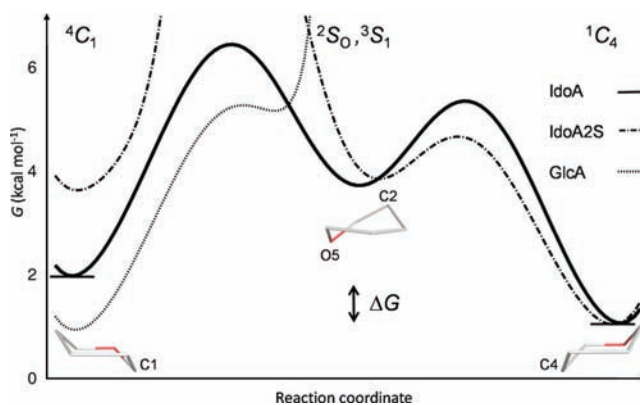
Figure 1. One-dimensional Bezier-smoothed free energy (G) landscapes derived from equilibrium populations in each monosaccharide simulation.

The pyranose ring conformation (or pucker) of L-iduronic acid (IdoA) is the subject of much controversy. An essential constituent of heparin, used to prevent blood clots during surgery, IdoA is postulated to be a dynamic equilibrium of three puckers ($^1C_4$, $^4C_1$, and ^2S_O, Figure 1) that exchange on the microsecond time scale. This time scale has been difficult to study experimentally and theoretically, but simulations are beginning to unravel microsecond enigmas and associated free-energy landscapes.[1,2] In this work, we explore the free energies of IdoA and two related monosaccharides for the first time using extensive equilibrium simulations, which resolve some of IdoA's controversies and allow new progress toward safer[3] synthetic analogues of animal-derived heparin.

The source of IdoA's conformational instability is C5 carboxyl epimerization, the only chemical difference between IdoA and its biological precursor D-glucuronic acid (GlcA, a stable $^4C_1$-chair). Decades of NMR studies have failed to reach a consensus on the exact nature of the IdoA conformational equilibrium. Ring three-bond $^1H-^1H$ vicinal coupling constant ($^3J_{H,H}$) measurements have suggested that in water the monomer (and IdoA at nonreducing oligosaccharide termini) is a dynamic equilibrium of $^1C_4$- and $^4C_1$-chairs, with a small ^2S_O-skew-boat component.[4,5] Internal heparin IdoA residues are considered conformationally biased such that only the $^1C_4$ and ^2S_O geometries contribute.[4,5] In sulfated glycosaminoglycans (heparin, heparan sulfate, dermatan), IdoA 2-O-sulfation (IdoA2S) lowers the vicinal couplings, indicating a $^1C_4$-chair.[4,5] These conformational changes, in concert with sulfate charge,

mediate heparin–protein interactions and anticoagulant properties, but the structural mechanism is unclear.

Here, the first equilibrated aqueous all-atom simulations of pyranoses (for energy states relevant at room temperature) were used to estimate the free energy (G) conformational landscape of IdoA and the effects of epimerization and sulfation without recourse to enhanced sampling. Six independent simulations of IdoA, IdoA2S, and GlcA, initiated as both $^1C_4$- and $^4C_1$-chairs, were each extended to 5 μs. In principle, complete phase-space sampling allows prediction of both ΔG and other experimental observables, provided conformations are accurately reproduced by the simulations. In the absence of experimental ΔG values, ring vicinal coupling constants were used to gauge performance.

Our simulations agreed with experimental data and previous hypotheses, namely that GlcA maintains a $^4C_1$-chair, IdoA is in exchange between $^1C_4$- and $^4C_1$-chair conformations, and IdoA2S adopts the $^1C_4$ pucker. The ^2S_O-skew-boat was rarely populated in IdoA but was an essential intermediate bridging the two chairs. Furthermore, the IdoA $^1C_4$ conformer was computed to be 0.9 kcal mol^{-1} more stable than the $^4C_1$-chair.

Aqueous molecular dynamics of IdoA, IdoA2S, and GlcA were performed at 298 K with ACEMD[6] on Nvidia graphics processors. Each simulation comprised 5.25 μs of NVT dynamics in a 32 Å periodic cube utilizing the GLYCAM06[7] force field, TIP3P water, counterions (Na$^+$), and hydrogen mass repartitioning to achieve a 4 fs time step without affecting the equilibrium distribution.[8] Data were recorded every 10 ps, and the first 250 ns of each trajectory was discarded. The $^1C_4$- and $^4C_1$-initiated simulations were concatenated (10 μs for each molecule) to calculate free energy surfaces, conformational exchange rates, and vicinal couplings (two Karplus relations were used).[9,10] 1-OMe-α-L-IdoA and 1-OMe-α-L-IdoA2S were synthesized, and 1-OMe-β-D-GlcA was acquired (Goldbio, St. Louis, MO). Samples were prepared for NMR in D$_2$O (99.99%, Sigma-Aldrich) at pD 7.0. High-resolution [^{1}H]-1D spectra (900 MHz, 32768 complex points) were recorded at 298 K for

accurate $^3J_{H,H}$ measurement. A sine bell window function was applied, and the data were zero-filled four times. Experimental $^3J_{H,H}$ values were measured from [^{1}H]-1D spectra or by fitting to strongly coupled resonances.

The simulations equilibrated after ∼3 μs, determined by the convergence of the average value of cos θ (θ is the Cremer−Pople polar angle,[11] which quantifies ring pucker) and comparison of its limiting value in two independent $^1C_4$- and $^4C_1$-initiated trajectories. Our simulations therefore represent the first thoroughly equilibrated trajectories for a pyranose in water; the previous longest was a 1 μs trisaccharide simulation.[12] Further analysis of θ quantified the simulated occupancy of $^1C_4$, $^4C_1$, and ^2S_O conformations. IdoA was a dynamic equilibrium of $^1C_4$, ^2S_O, and $^4C_1$ (70:10:20), and IdoA2S was biased toward a stable $^1C_4$-chair with a small contribution from ^2S_O (90:10), while GlcA was a stable $^4C_1$-chair. Analysis of the skew-boat region revealed two quasi-stable conformations in both IdoA and IdoA2S simulations: ^2S_O and $^3S_1$. The path between these was more populated in IdoA.

Free energy landscapes (Figure 1) were derived from relative populations, $\Delta G = RT \ln(p_1/p_2)$, as a function of θ. In IdoA, the $^4C_1$ and skew-boat conformers were 0.9 and 2.6 kcal mol^{-1} higher in energy than the $^1C_4$-chair, respectively. Exchange rates for $^1C_4$ ⇆ $^4C_1$ interconversion were 3.5 μs^{-1} (forward) and 19.4 μs^{-1} (back), giving an equilibrium constant of 5.6 (consistent with the 0.9 kcal mol^{-1} value). Small barriers (up to ∼5 kcal mol^{-1}) above the ground state separated the IdoA conformations, resulting in the fast exchange. Both IdoA and IdoA2S had identical $^1C_4$-chair and skew-boat energies, but in IdoA2S the $^4C_1$ energy and the barrier to this conformer were higher (1.7 and >10 kcal mol^{-1}, respectively). This manifested as slower $^4C_1$ ⇆ $^1C_4$ exchange in IdoA2S (*cf.* IdoA). The GlcA $^4C_1$ conformer had the lowest energy (1 kcal mol^{-1} less than IdoA), and the skew-boat and $^1C_4$ puckers were negligibly populated at 298 K ($RT \approx 0.6$ kcal mol^{-1}).

Computed IdoA2S and GlcA vicinal couplings were in good agreement with experimental data (∼1−2 Hz difference, Figure 2). The small and large vicinal couplings for IdoA2S and GlcA, respectively, were consistent with stable $^1C_4$- and $^4C_1$-chairs. Additionally, the NMR spectra confirmed four-bond *w*-couplings between H1 and H3 (∼1.0 Hz) and H2 and H4 (∼0.9 Hz) only in IdoA2S, confirming the $^1C_4$-chair. While calculated IdoA and IdoA2S vicinal couplings trended below their respective observed values, a different pattern was observed for GlcA. Here, the Altona−Haasnoot-derived vicinal couplings were smaller and the Hricovini−Bizik-derived couplings were bigger than the experimental values. For IdoA, the two Karplus relationships gave similar predictions but also resulted in the largest discrepancy with experiment. However, no single IdoA canonical conformer could give a better agreement to the data than the simulated equilibrium of chairs and skew-boat.

Compared to all other vicinal couplings, the difference between experimental data and predictions was relatively large for IdoA $^3J_{2,3}$ and $^3J_{3,4}$. Given the good accuracy of vicinal couplings computed from the IdoA2S and GlcA simulations and the fact that the Karplus equation coefficients are unlikely to be >10% inaccurate, the discrepancy in these two cases was attributed to the force field. In particular, the experimental values implied an increased population of IdoA $^4C_1$-chair (*cf.* the simulation). This would result from only a small error in computed ΔG with respect to $^1C_4$ (∼0.5 kcal mol^{-1}), reiterating that inaccuracies in biomolecular force fields, such as in GLYCAM06 used here, are evident only in long microsecond simulations.[1,2] Furthermore, while recent work showed microsecond time scale simulations are necessary to explain some carbohydrate NMR observables,[12] the small energetic values associated with the
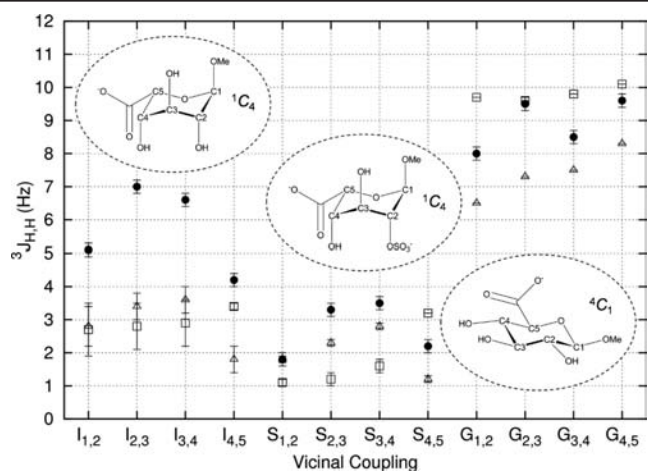

Figure 2. Experimental and calculated monosaccharide vicinal couplings. I, IdoA; S, IdoA2S; G, GlcA. $I_{1,2}$ denotes IdoA $^3J_{1,2}$, etc. ●, experimental; □, calculated (Hricovini and Bizik, ref 9); △, calculated (Altona and Haasnoot, ref 10).

errors found here highlight that exactly reproducing the IdoA conformational landscape will be a demanding theoretical exercise. However, identifying and rectifying such problems is essential because simulation is a vital complement of experimentation while material heterogeneity and NMR resonance overlap problems persist for carbohydrates.

The importance of dynamics in structural biology was highlighted by the prediction that ∼25% of mammalian proteins are fully disordered.[13] The skew-boat occupancy in the IdoA and IdoA2S conformational equilibriums is almost certainly biologically significant. Inhibition of the coagulation cascade is thought to be initiated by antithrombin binding the heparin IdoA2S ^2S_O conformer, and synthetic heparins including ^2S_O-biased IdoA analogues are highly potent.[14] Our simulations imply a smaller contribution of ^2S_O than previously thought, which reignites discussions around the role of conformational disorder in IdoA function, the structural mechanism of heparin−protein interactions, and hence how to develop safer anticoagulants.

In this Communication, we report the first simulated complete exploration of low-energy conformations for three fundamental biological building blocks. Aqueous all-atom molecular dynamics trajectories of ∼3 μs were required to achieve this, which is a crucial observation as almost all historical carbohydrate simulations are sub-microsecond in duration. Further, the monosaccharides studied were small in the context of functional biomolecules (oligosaccharides, peptides, proteins), which contain vastly more conformational degrees of freedom and other complexities (e.g., in heparin the IdoA conformational equilibrium is thought to be affected by 3-*O*-sulfation of the preceding hexosamine residue).[4]

These unique simulations of IdoA, IdoA2S, and GlcA reveal how enzymatic chemical modifications (epimerization and sulfation) fine-tune the free energy landscape of each and thereby mediate protein selectivity. We also showed how biological time scale simulations can be used to identify limitations of molecular modeling and drive theoretical improvements. Graphics processors provide an excellent platform on which to iteratively test and augment force fields so that complete and highly accurate free energy landscapes can be explored and their features exposed. These theoretical landscapes, not amenable to experiment, provide a new route to development of much needed carbohydrate mimetic biomaterials and pharmaceuticals.

Supporting Information Available: Simulation data, NMR spectra, and evidence of synthesis. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120–127.

(2) Orozco, M.; Noy, A.; Perez, A. *Curr. Opin. Struct. Biol.* **2008**, *18*, 185–193.

(3) Liu, H.; Zhang, Z.; Linhardt, R. J. *Nat. Prod. Rep.* **2009**, *26*, 313–321.

(4) Casu, B.; Petitou, M.; Provasoli, M.; Sinay, P. *Trends Biochem. Sci.* **1988**, *13*, 221–225.

(5) Ferro, D. R.; Provasoli, A.; Ragazzi, M.; Casu, B.; Torri, G.; Bossennec, V.; Perly, B.; Sinay, P.; Petitou, M.; Choay, J. *Carbohydr. Res.* **1990**, *195*, 157–167.

(6) Harvey, M. J.; Giupponi, G.; De Fabritiis, G. *J. Chem. Theory Comput.* **2009**, *5*, 1632–1639.

(7) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622–655.

(8) Buch, I.; Harvey, M. J.; Giorgino, T.; Anderson, D. P.; De Fabritiis, G. *J. Chem. Inf. Model.* **2010**, *50*, 397–403.

(9) Hricovini, M.; Bizik, F. *Carbohydr. Res.* **2007**, *342*, 779–783.

(10) Altona, C.; Haasnoot, C. A. G. *Org. Magn. Reson.* **1980**, *13*, 417–429.

(11) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.

(12) Landstrom, J.; Widmalm, G. *Carbohydr. Res.* **2010**, *345*, 330–333.

(13) Dunker, A. K.; Silman, I.; Uversky, V. N.; Sussman, J. L. *Curr. Opin. Struct. Biol.* **2008**, *18*, 756–764.

(14) Boltje, T. J.; Buskas, T.; Boons, G. J. *Nat. Chem.* **2009**, *1*, 611–622.

JA1054143