An ATLAS for understanding carbohydrate chemistry

Summary
Carbohydrates and sugars are biological molecules, which adopt various geometrices. In cells, carbohydrates change conformation when they bind to proteins. This project aims to develop methods for simulating the binding of sugars to proteins on a computer and to understand how these proteins function. Machine learning will be used to compare the conformers the sugar adopts when it is bound with the conformations that it adopts when it is unbound. Traditionally function in molecular biology is inferred from structure. However, dynamics are increasingly seen to play an important role. Developing tools for systematically studying these effects is thus critical.

Project
Carbohydrates, also known as saccharides, glycans, or sugars, are the most abundant and diverse set of biological molecules on Earth. The fundamental unit of these molecules are monosaccharides, which bind to form polysaccharides [1]. Carbohydrates are implicated in biological functions ranging from structure and energy storage to signalling, with many of these functions being linked to carbohydrates' interactions with proteins [1]. A better understanding of protein–carbohydrate interactions is thus essential for many fields. For example, glycoside hydrolases are thought to function by distorting the carbohydrate ring geometries in ways that make the cleavage of glycosidic bonds more favourable [2]. To understand protein–carbohydrate interactions we thus need more than the static crystal structures of these complexes that are found in the protein data bank. We also need to understand the effect proteins have upon the dynamical behaviour of the carbohydrate molecules that bind to them so that we can better understand how these proteins facilitate carbohydrate reactivity.

Carbohydrates can adopt a number of distinct metastable configurations and will interconvert between these configurations over a microsecond timescale [3]. Consequently, when we say that binding to a protein affects the dynamics of the carbohydrate one or several of four things must be occurring:

1. Binding stabilizes metastable configurations that are unstable when the carbohydrate is unbound.
2. Binding changes the relative free energies of the various metastable configurations that the carbohydrate can adopt both when it is bound and when it is unbound.
3. Binding changes the rates at which the carbohydrate interconverts between the various conformers.
4. When the carbohydrate binds the way that it fluctuates while it is trapped in the metastable configurations changes.

It is difficult to determine which of these mechanisms proteins are using to change the behaviours of carbohydrates because it is computationally expensive to run the microsecond-long molecular dynamics simulations that would be required to observe the interconversions between the metastable states that the carbohydrates can adopt. In addition, because the

simulator defines what constitutes a metastable state, it can be difficult to locate the boundaries between them.

In this project, we will apply our recently-developed ATLAS technique [4] to study the effect of protein binding on the dynamics of carbohydrates using molecular dynamics (MD) simulations. ATLAS is ideally suited to this problem because it uses machine learning to analyze trajectories and find the basins of attraction that correspond to the various metastable states. In other words, the trajectory is used to fit a Gaussian mixture model (GMM). This probability distribution is then used to construct a bias potential that is similar to that used in metadynamics [5] and which enhances the sampling of phase space. By using this technique, we are thus able to study long-timescale processes using relatively short MD simulations.

The great strength of ATLAS is that the Gaussian mixture is a probability distribution. We can thus use this probability distribution to calculate the likelihood that a given trajectory frame is a representative example of a particular conformer. Consequently, we can use a Gaussian mixture model that is fit based on simulations of an unbound carbohydrate to analyse a simulation of a carbohydrate that is bound to a protein. The likelihood information that we obtain from this analysis tells us whether item (1) in the list above occurs for the system of interest, i.e. if the protein stabilises configurations of the carbohydrate that would not be stable for the unbound carbohydrate. Furthermore, this likelihood information also makes partitioning the trajectory between the various metastable configurations straightforward and thus allows us to also determine the effect the protein has on the relative free energies of the conformers and whether binding has affected the local dynamics.

We can even potentially use ATLAS to determine when the system is transitioning between states. The underlying kinetics can be obtained by using the infrequent metadynamics technique. Furthermore, the GMM provides us with an obvious set of states that we can use when constructing a Markov state model of the interstate dynamics.

In this initial project, where we are adapting the ATLAS method for this application we will first study proteins such as the asialoglycoprotein receptor [6], FimH Adhesin carbohydrate-binding domain [7], MhLap carbohydrate-binding domain [8] and L-fucokinase/GDP-fucose pyrophosphorylase [9] that bind to mono and disaccharides. These small saccharides can adopt fewer metastable configurations and the proteins they bind to only have a few hundred residues. These systems are small enough to study with our local computational cluster and are well suited to a proposal whose principal aim is to develop new methodologies. We will run long MD simulations of the mono and disaccharides that bind to these proteins in water and use machine learning to extract representations of the trajectories that can be used to construct an ATLAS bias potential. We will then run ATLAS simulations in the presence and absence of the protein and do the analysis described above.

We would note that there are many problems in biochemistry that, like the saccharide binding we are studying here, involve small, flexible molecules binding to larger, less-flexible macromolecules. Consequently, all the methodologies we develop in this project will be

implemented in the PLUMED plugin for molecular dynamics [10]. Example input files for all the calculations that appear in our published papers will be deposited in the PLUMED nest [11], while tutorials on how to use the method will be provided through PLUMED tutorials [12].

Our overall objective is to provide a quantitative framework for determining the effect the environment has on the dynamics of molecules. As discussed in the section on why the Leverhulme Trust, the function of biomolecules is not solely determined by their structures. Dynamics are increasingly seen as providing a critical role. The impact of this project will be to provide a quantitative lens that can be used to probe biomolecular dynamics. This lens leverages the power of machine learning. However, the insight obtained from the algorithms is comprehensible by human users as the model used by it is based on a physical model that assumes the molecule can be in a small number of distinct metastable states. A coarse grained model for the behaviour of the biomolecule can thus be obtained by extracting relative free energies for these states and the transition rates between them.

References

1. *Essentials of Glycobiology.* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2022).
2. Mayes, H. B., Broadbelt, L. J. & Beckham, G. T. How Sugars Pucker: Electronic Structure Calculations Map the Kinetic Landscape of Five Biologically Paramount Monosaccharides and Their Implications for Enzymatic Catalysis. *J. Am. Chem. Soc.* **136**, 1008–1022 (2014).
3. Guvench, O., Martin, D. & Greene, M. Pyranose Ring Puckering Thermodynamics for Glycan Monosaccharides Associated with Vertebrate Proteins. *Int. J. Mol. Sci.* **23**, 473 (2022).
4. F. Giberti, G. A. Tribello and M. Ceriotti, Global free-energy landscapes as a smoothly joined collection of local maps. *J. Chem. Theory Comput.* **17** 3292-3308 (2021)
5. G. Bussi and A. Laio. Using metadynamics to explore complex free-energy landscapes. *Nature Reviews Physics* **2,** 200-212 (2020)
6. Meier, M., Bider, M. D., Malashkevich, V. N., Spiess, M. & Burkhard, P. Crystal Structure of the Carbohydrate Recognition Domain of the H1 Subunit of the Asialoglycoprotein Receptor. *J. Mol. Biol.* **300**, 857–865 (2000).
7. Vanwetswinkel, S. *et al.* Study of the Structural and Dynamic Effects in the FimH Adhesin upon α-D-Heptyl Mannose Binding. *J. Med. Chem.* **57**, 1416–1427 (2014).
8. Vance, T. D. R., Guo, S., Assaie-Ardakany, S., Conroy, B. & Davies, P. L. Structure and functional analysis of a bacterial adhesin sugar-binding domain. *PLOS ONE* **14**, e0220045 (2019).

9. Lin, S.-W. *et al.* Structural insight into the catalytic mechanism of the bifunctional enzyme l-fucokinase/GDP-fucose pyrophosphorylase. *J. Biol. Chem.* **301**, 108344 (2025).
10. G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi. PLUMED2: New feathers for an old bird, *Comp. Phys. Comm.* **185**, 604 (2014)
11. The PLUMED consortium. Promoting transparency and reproducibility in enhanced molecular simulations; *Nat. Methods.* **16**, 670 (2019)
12. The PLUMED consortium. PLUMED Tutorials: A collaborative, community-driven learning ecosystem *J. Chem. Phys.* **162**, 092501 (2025)