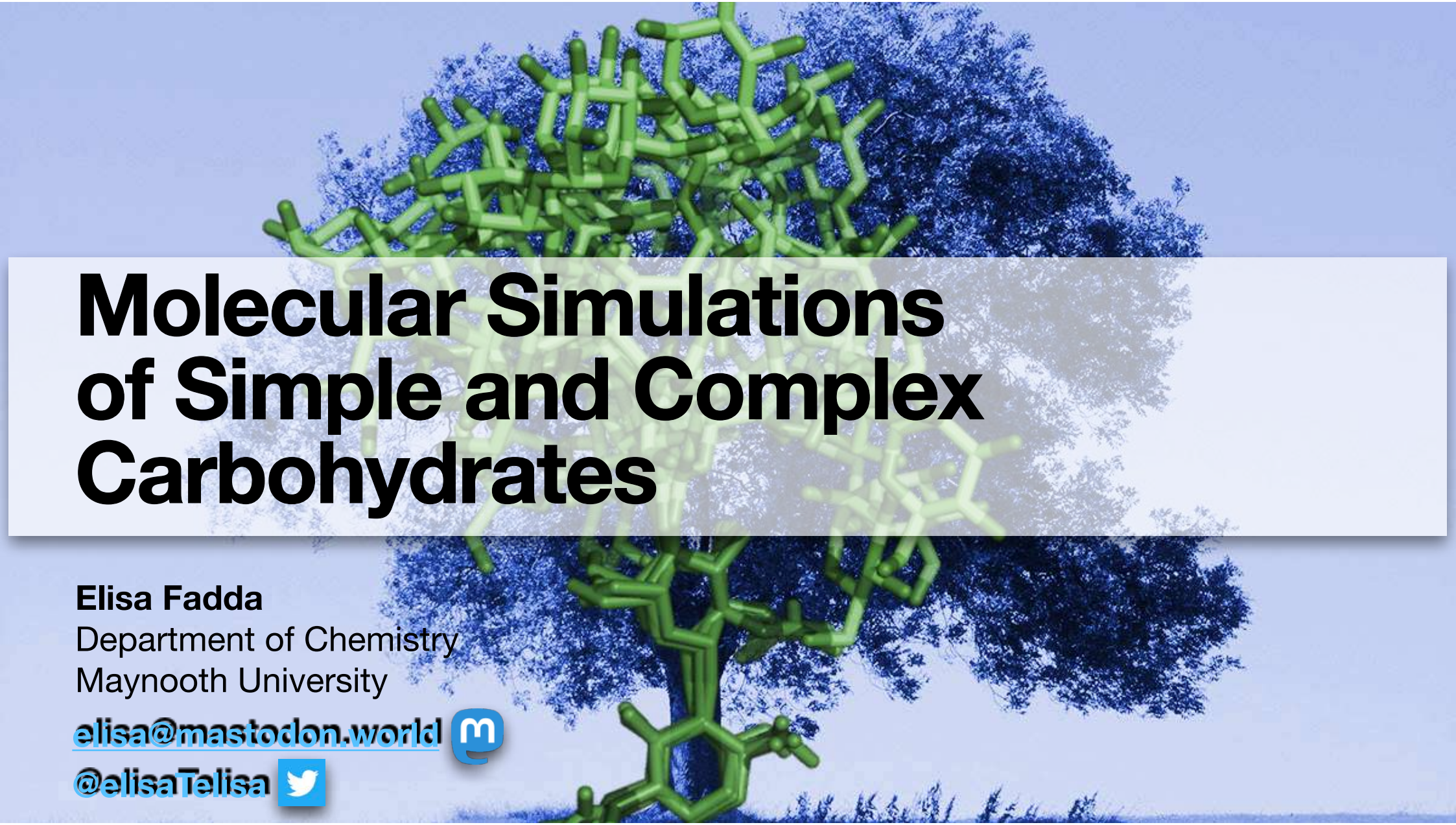


Molecular Simulations of Simple and Complex Carbohydrates

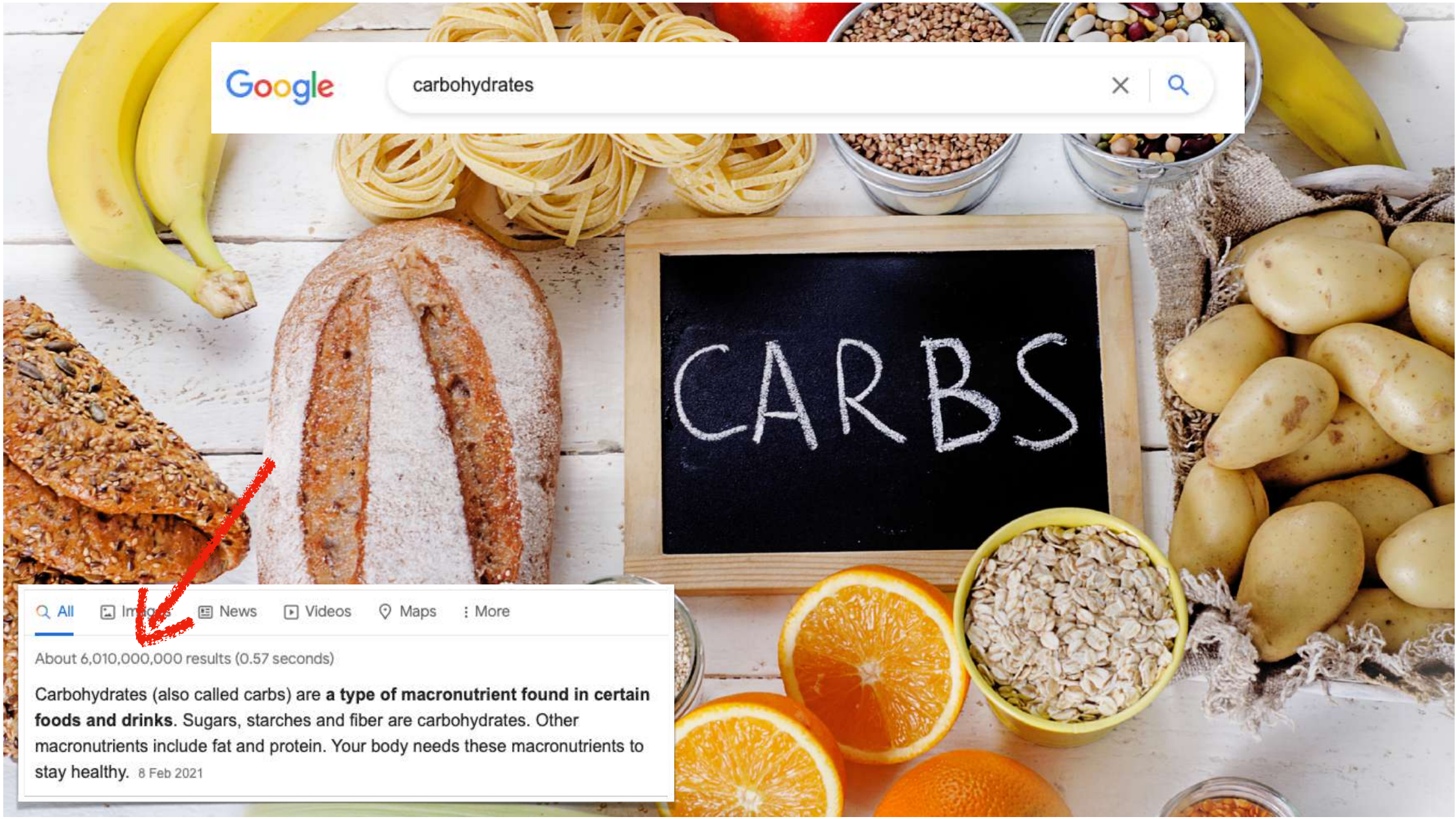
Elisa Fadda
Department of Chemistry
Maynooth University

elisa@mastodon.world
@elisaTelisa

Outline

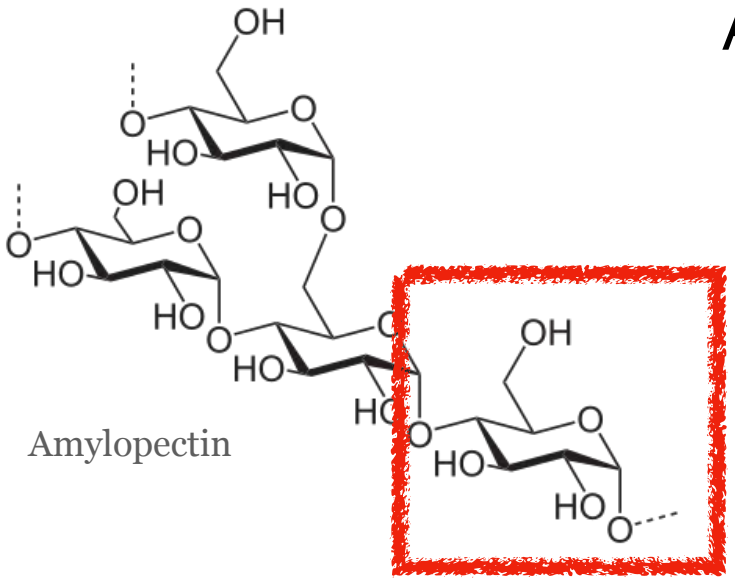
slido.com

- What are carbohydrates?
- Carbohydrates **3D** structure
- Intrinsic disorder and conformational ensembles
- Classical force fields representations
- Enhanced sampling and difficult cases
- Sequence-to-structure relationships through MD and molecular recognition
- The SARS-CoV-2 S as a case study

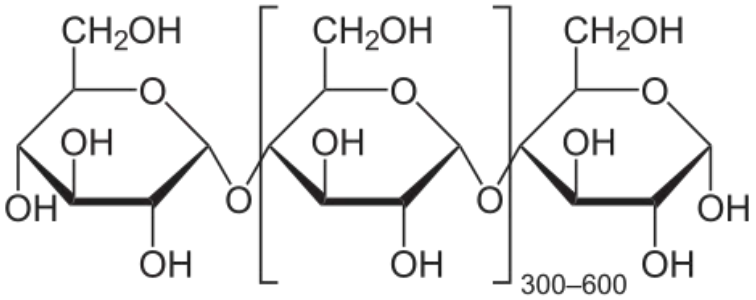

Google
carbohydrates
CARBS
All
News
Videos
Maps
More
About 6,010,000,000 results (0.57 seconds)
Carbohydrates (also called carbs) are a type of macronutrient found in certain foods and drinks. Sugars, starches and fiber are carbohydrates. Other macronutrients include fat and protein. Your body needs these macronutrients to stay healthy. 8 Feb 2021

Starch is <u>one type</u> of carbohydrate

Actually, two types…

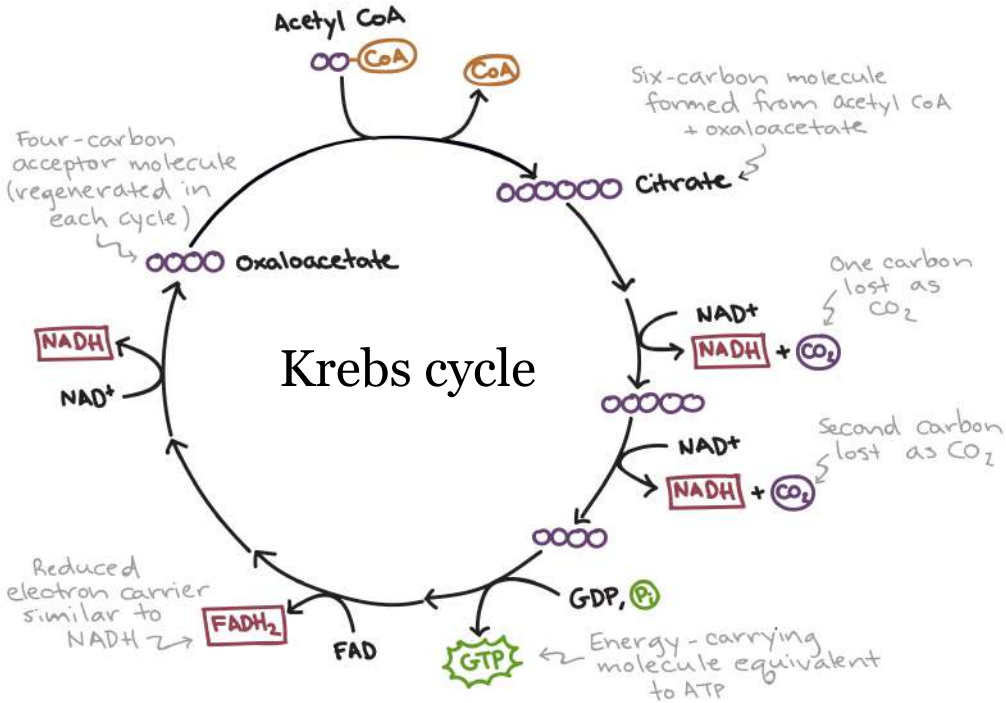


Amylopectin



Amylose

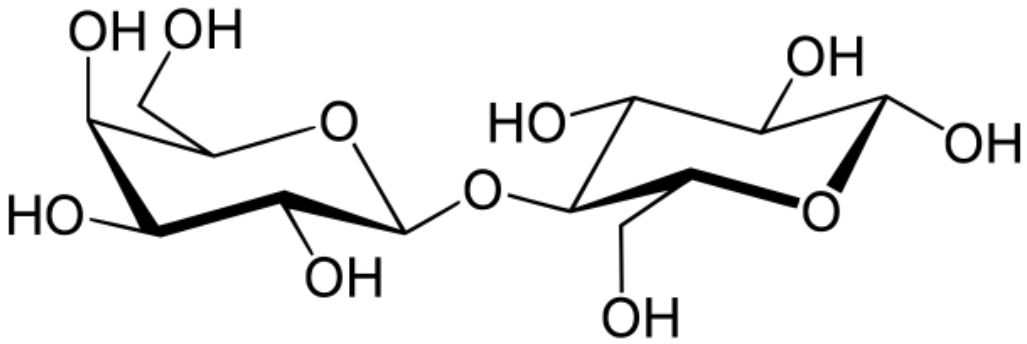
Glucose!!!



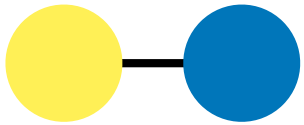
Glucose is the energy currency of the cell

There's plenty of other sugars!

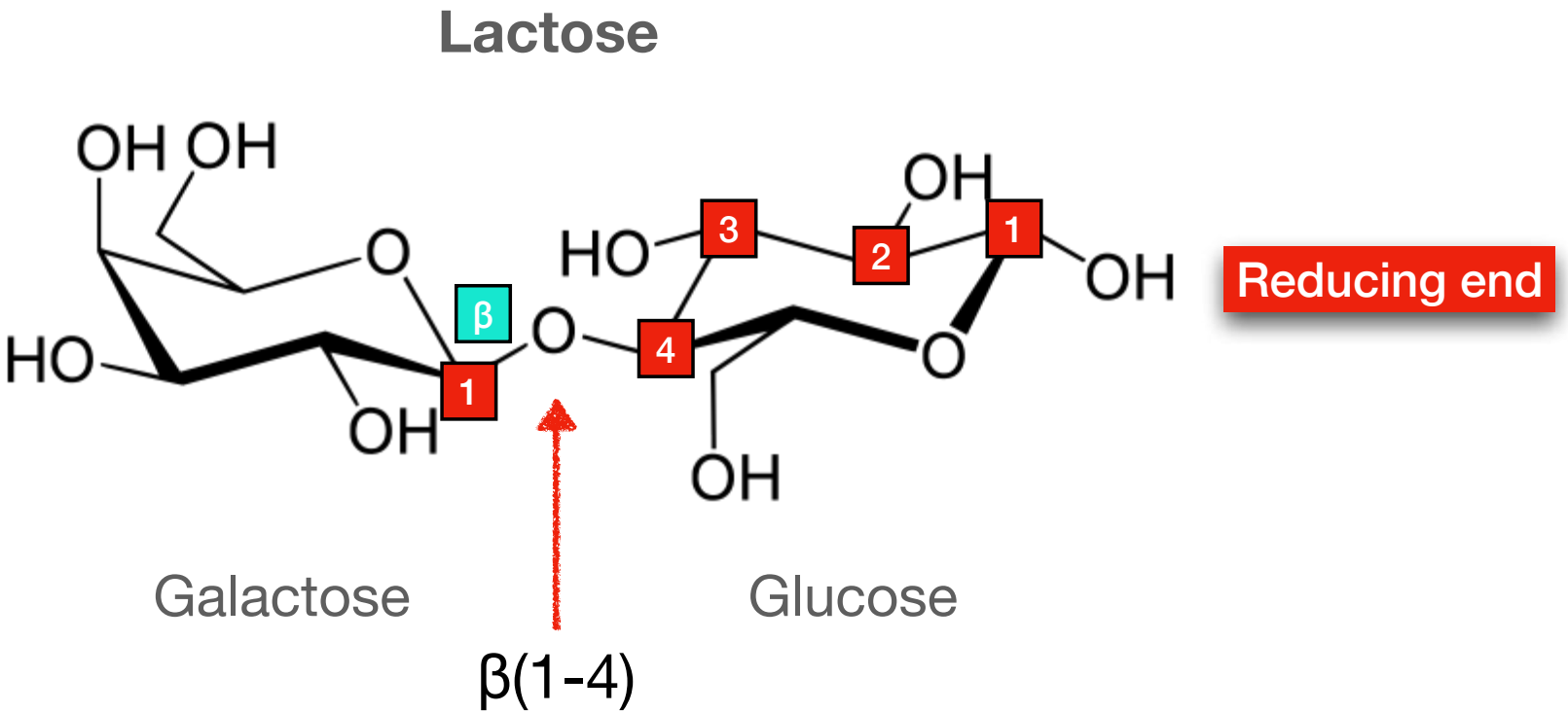
Galactose Glucose



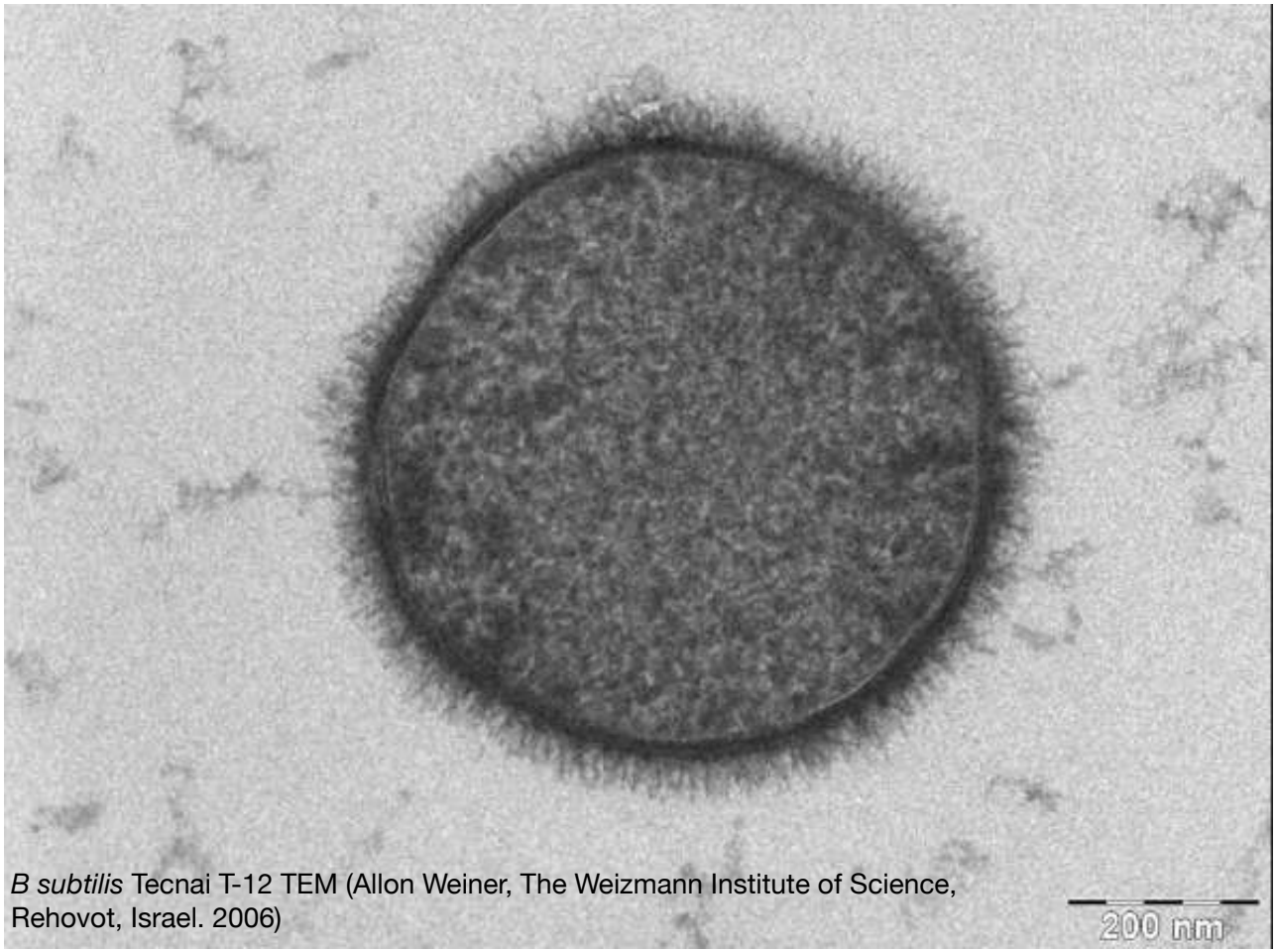
Lactose



Glycan 3D Structure: Glycosidic linkages

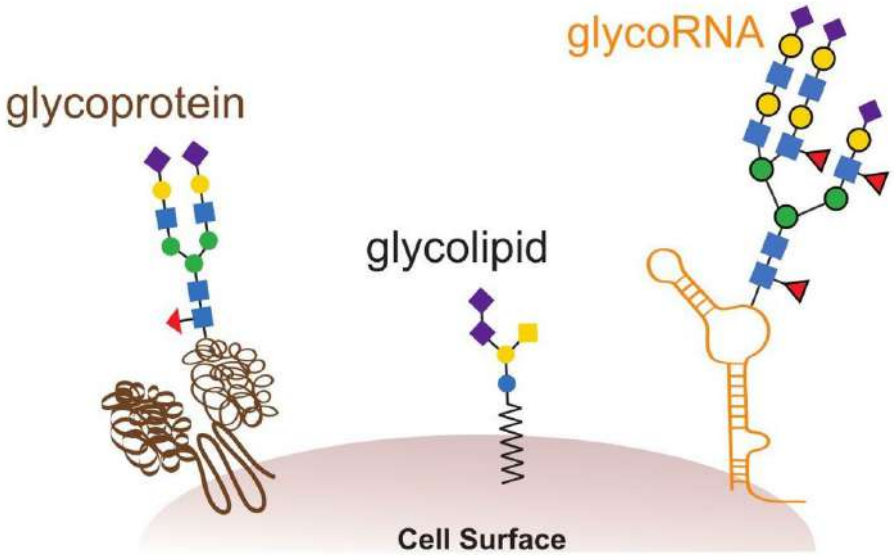


Glycans, the furry side of cells

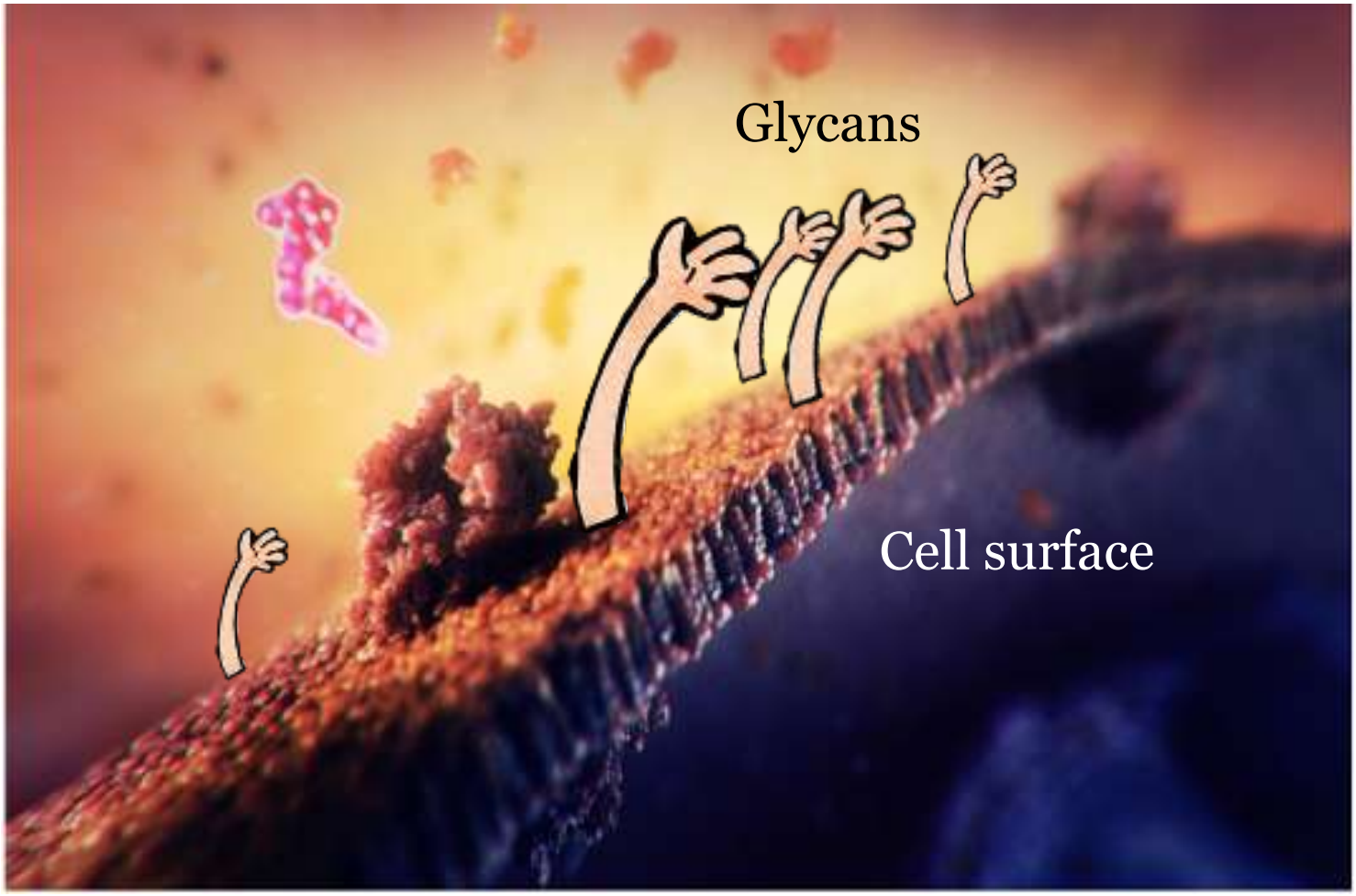


B subtilis Tecnai T-12 TEM (Allon Weiner, The Weizmann Institute of Science, Rehovot, Israel. 2006)

Mammalian cells have evolved a highly complex machinery **counting over 600 proteins** to build glycans, which 'decorate' the cell surface and most secreted proteins



Glycans, the furry side of cells



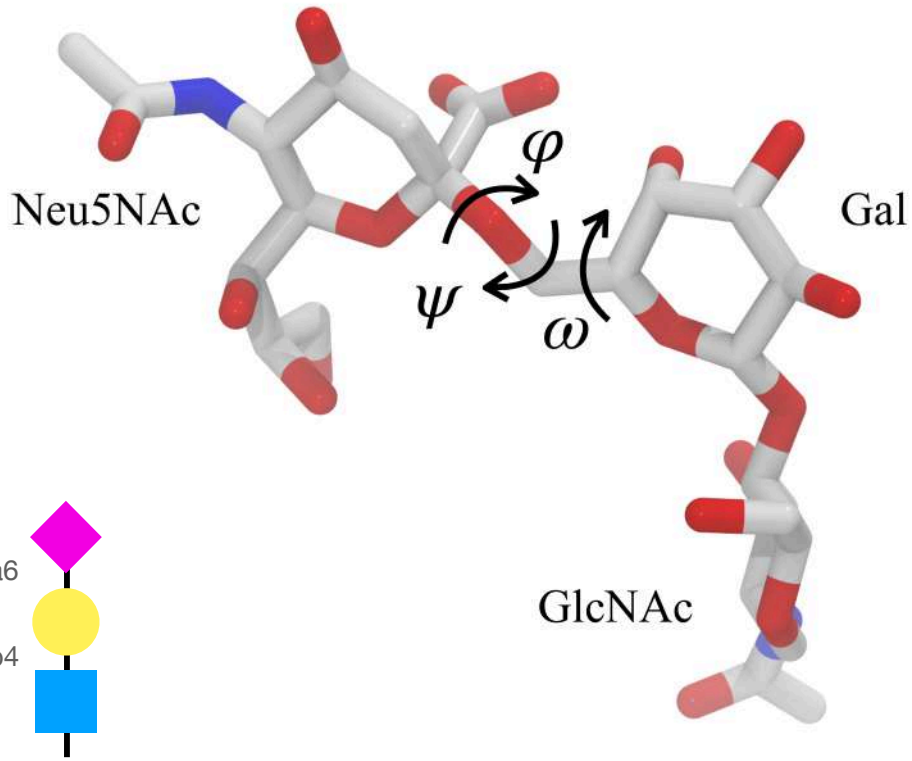
A friendly furry coat, that helps the cell move, and interact with its environment and with other cells and proteins

The immune system recognises its own glycans as "friends"

Exploited as the first port of entry for infection:

- Bacteria
- Toxins
- Viruses (e.g. flu viruses)

Glycan 3D Structure: Glycosidic linkages



The ϕ torsion is found preferentially in a ± gauche conformation relative to the O5 ring oxygen (ϕ = ± 70°)

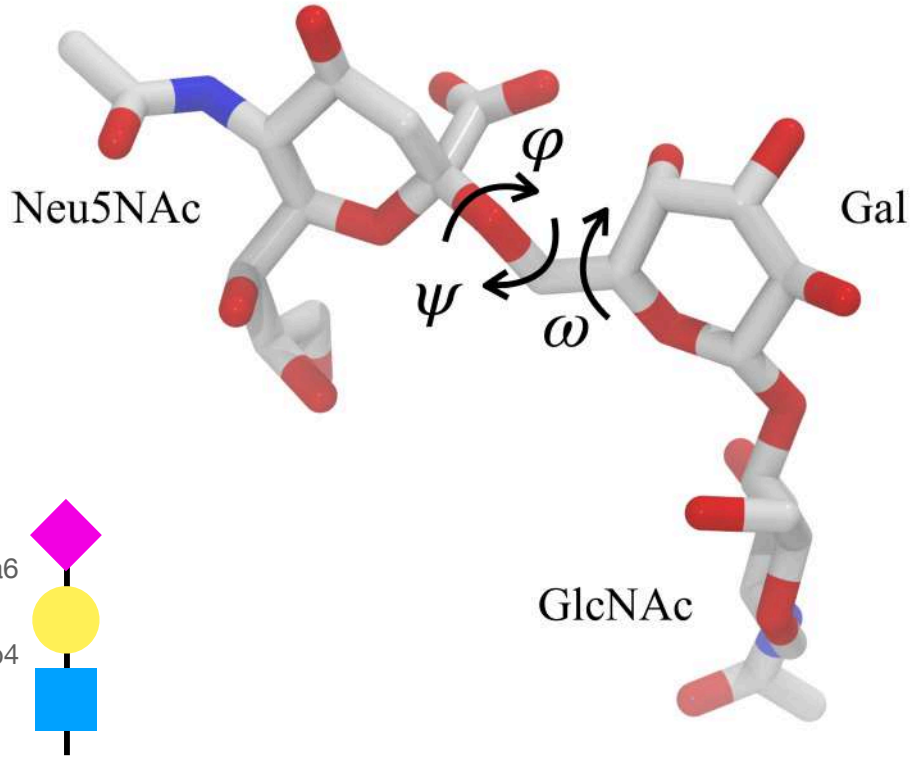
The ψ torsion will adopt value(s) corresponding to a minimum steric compression, usually around ψ = 180° in disaccharides

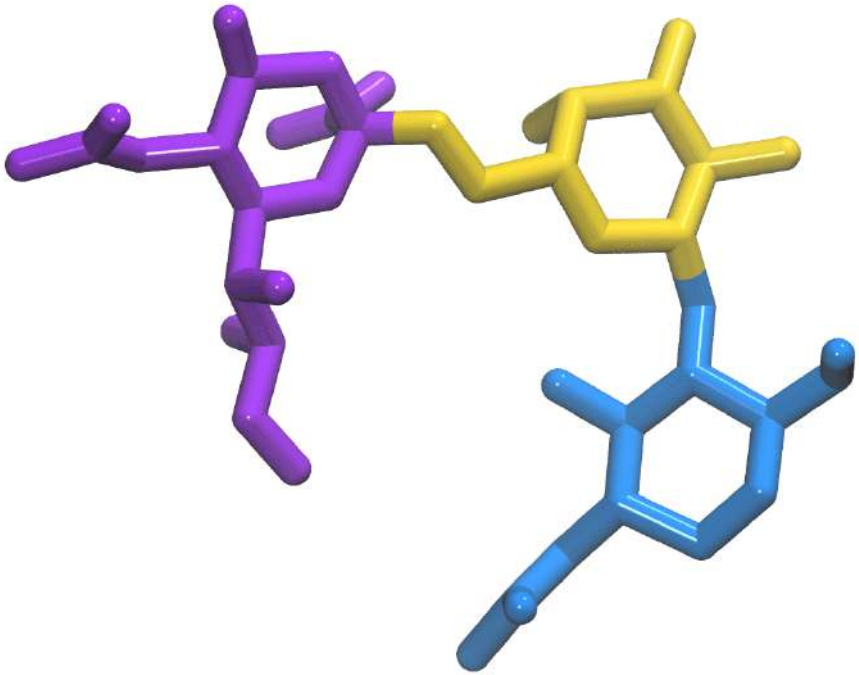
The ω torsion can populate two or three different energy basins at ω = ± 60° and 180°

Each energy basin is associated with a flexibility in the range of ± 10°-15°

Glycan 3D Structure: Glycosidic linkages

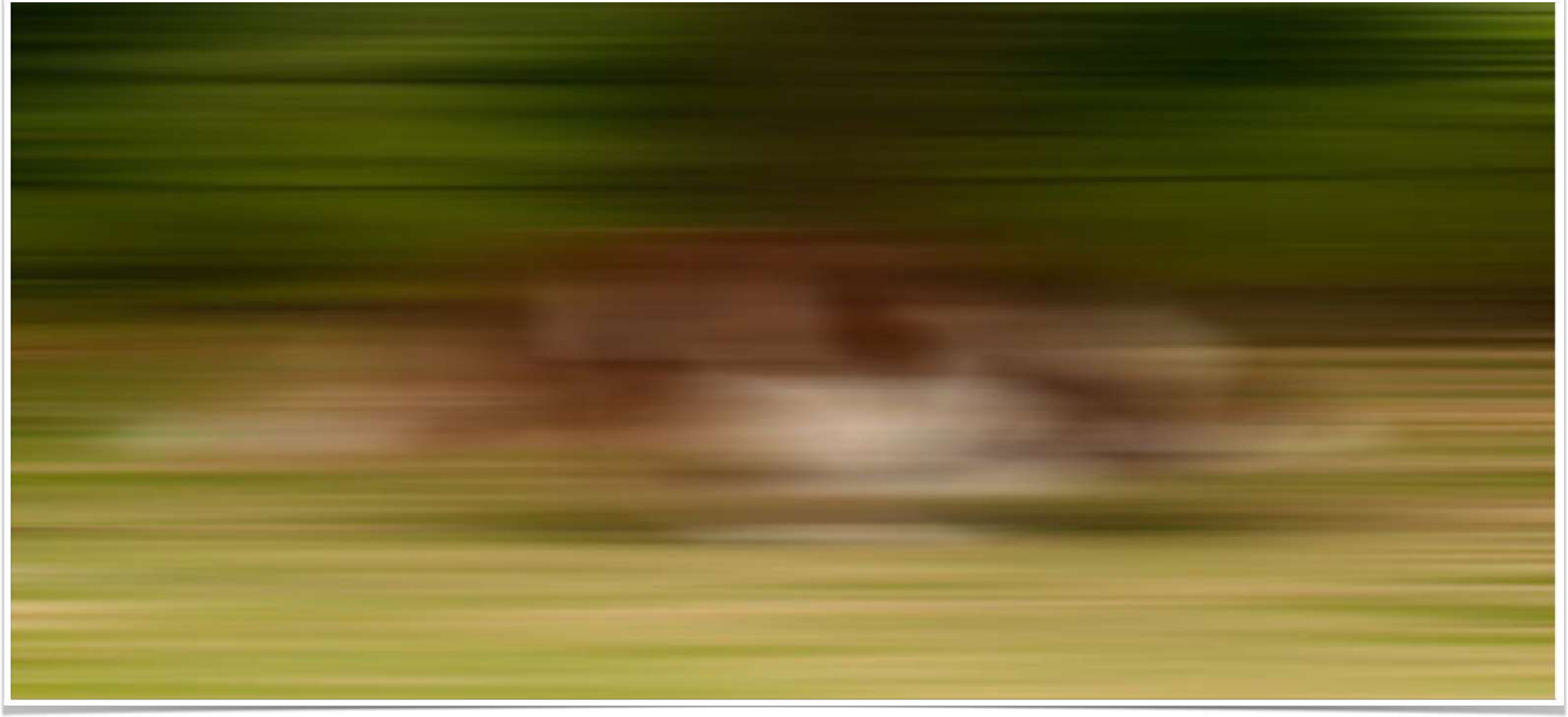
The chemical nature determines the glycans' intrinsic disorder



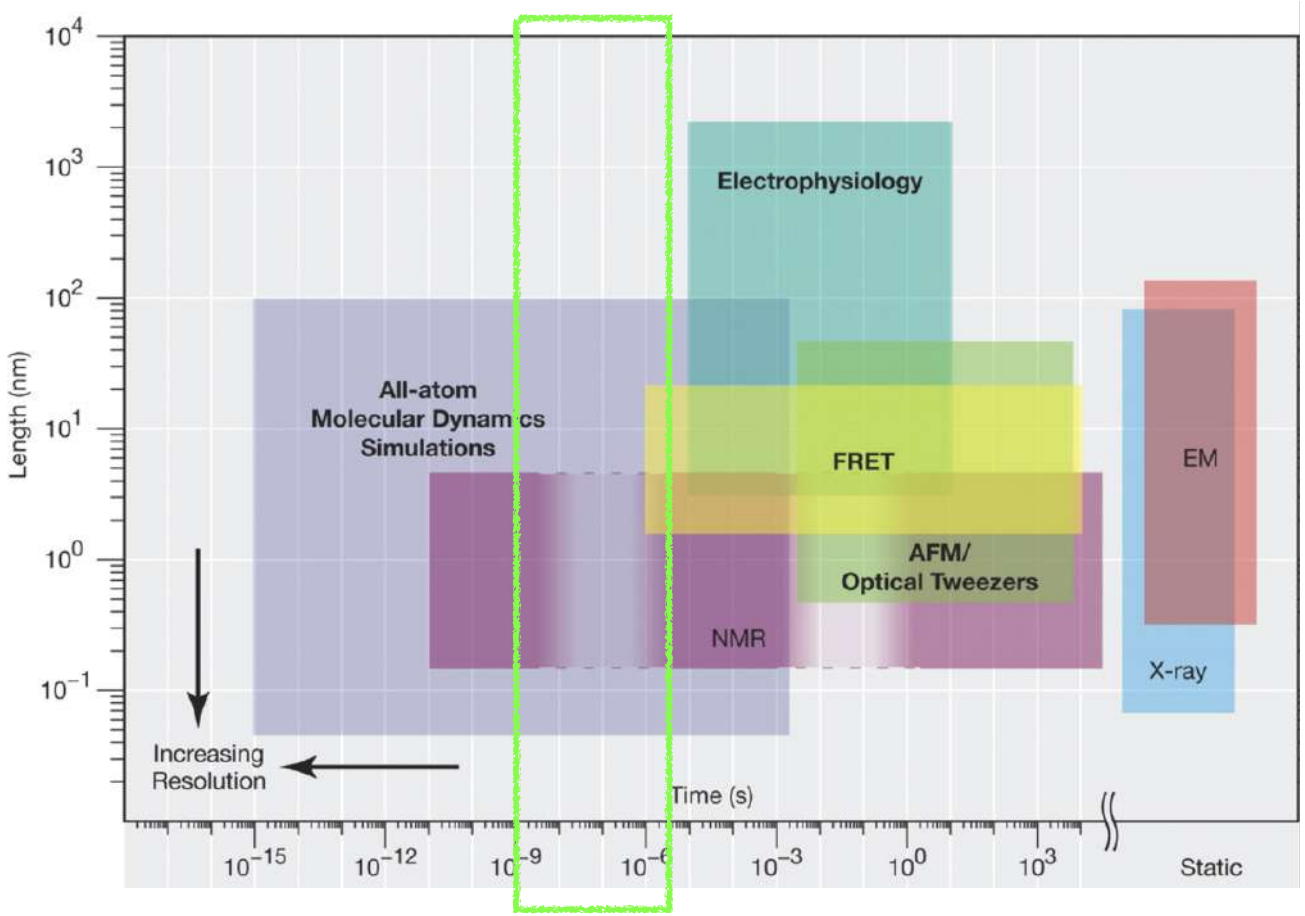


Fadda, *Curr Opin Chem Biol* (2022) DOI: /10.1016/j.cbpa.2022.102175
Harbison *et al, Glycobiology* (2019), DOI:10.1093/glycob/cwy097
Woods, *Chem Rev* (2019) DOI: 10.1021/acs.chemrev.8b00032

Structural (Intrinsic) Disorder

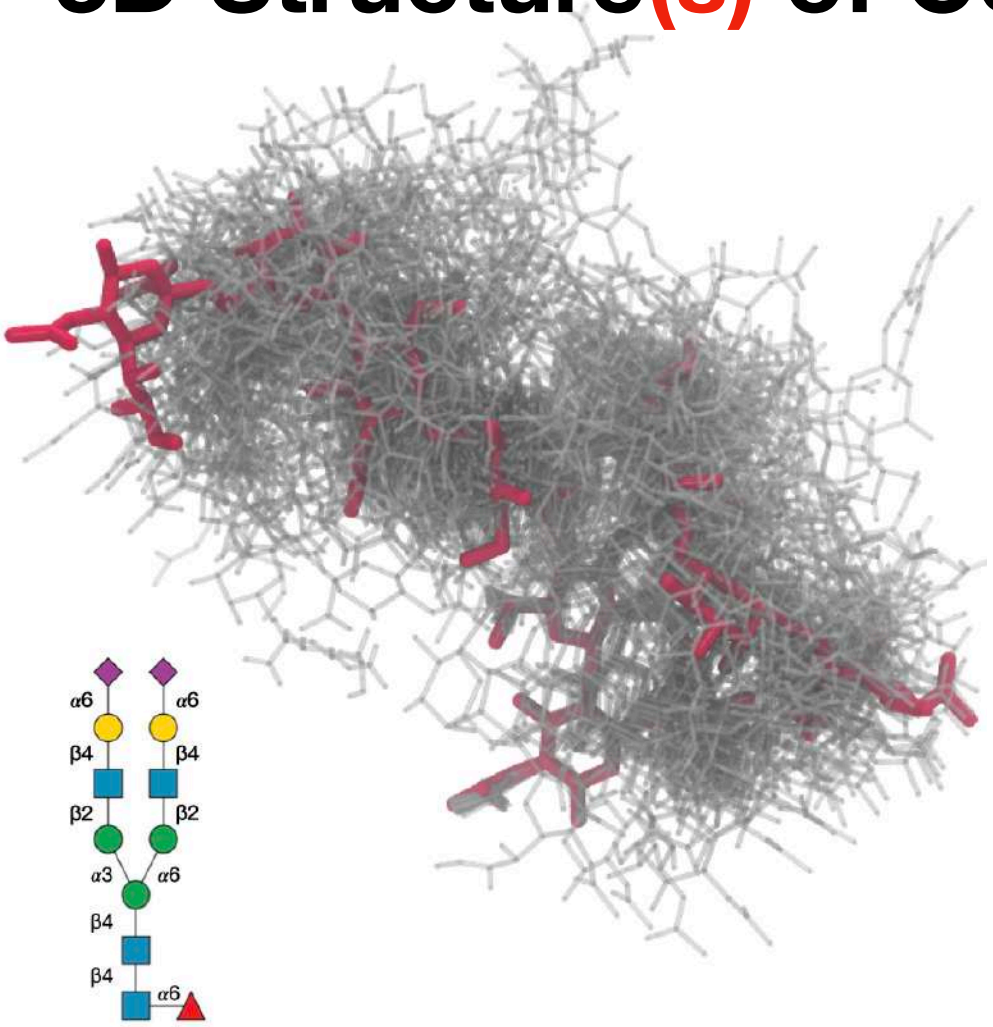


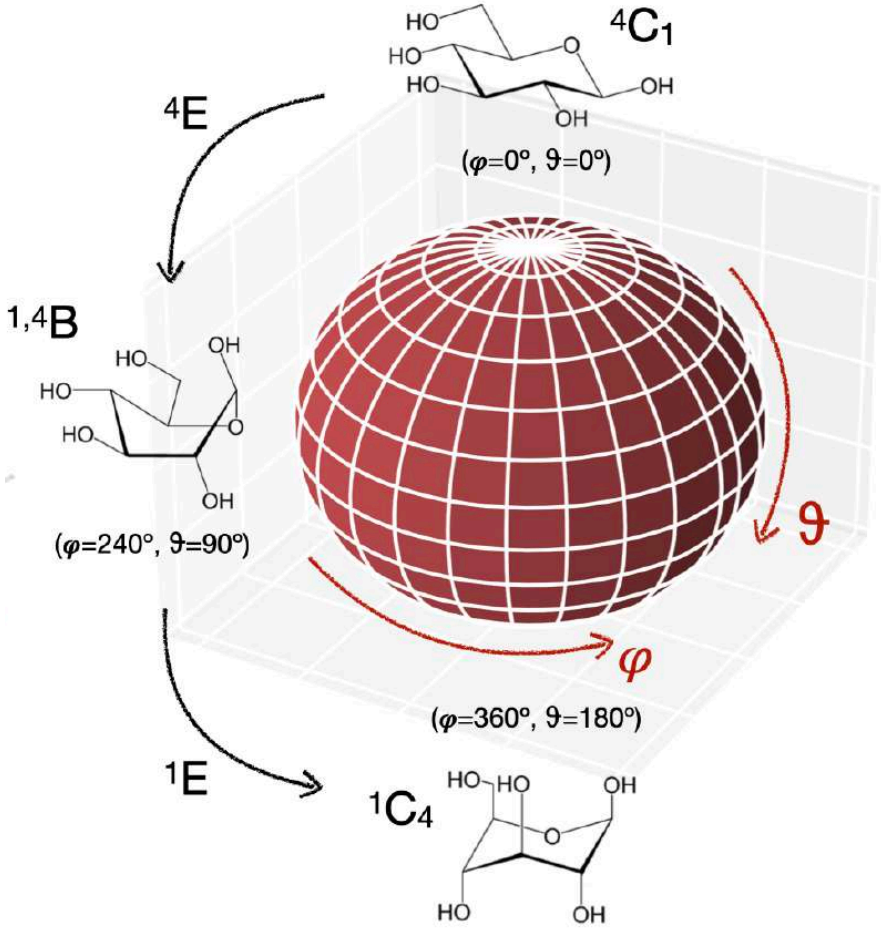
Spatiotemporal Resolution in Biophysics



Dror *et al, JGP* (2010): DOI 10.1085/jgp.200910373

3D Structure(s) of Complex Carbohydrates





Fadda, *Curr Opin Chem Biol* (2022) DOI: /10.1016/j.cbpa.2022.102175

Maynooth University
National University of Ireland Maynooth
Hamilton Institute

CINECA

PRACE
PARTNERSHIP FOR ADVANCED COMPUTING IN EUROPE
18.5 million core hrs

ORACLE

ICHEC
Irish Centre for High-End Computing
10 million core hrs

TACC
TEXAS ADVANCED COMPUTING CENTER
Powering Discoveries That Change The World

Collaboration with Rommie Amaro (UCSD)

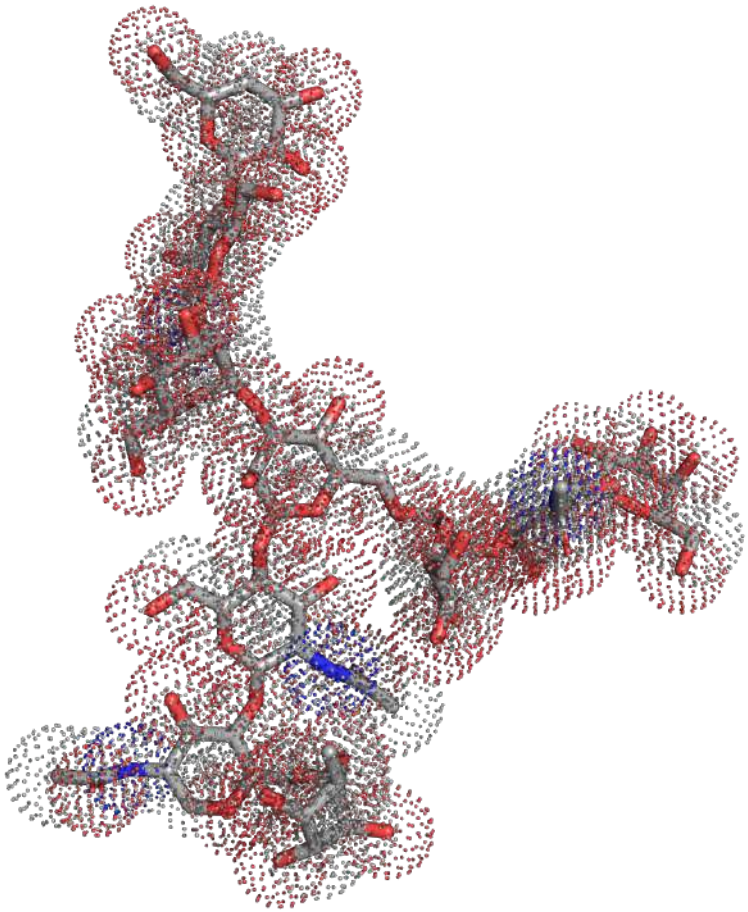
Citizen Scientists Create an Exascale Computer to Combat COVID-19

Maxwell I. Zimmerman[1,2], Justin R. Porter[1,2], Michael D. Ward[1,2], Sukrit Singh[1,2], Neha Vithani[1,2], Artur Meller[1,2], Upasana L. Mallimadugula[1,2], Catherine E. Kuhn[1,2], Jonathan H. Borowsky[1,2], Rafal P. Wiewiora[3,4], Matthew F. D. Hurley[5], Aoife M Harbison[6], Carl A Fogarty[6], Joseph E. Coffland[7], Elisa Fadda[6], Vincent A. Voelz[5], John D. Chodera[4], Gregory R. Bowman[1,2,*]

FOLDING @HOME

Nat Chem (2021) DOI: 10.1038/s41557-021-00707-0

A Classical Representation is Good Enough



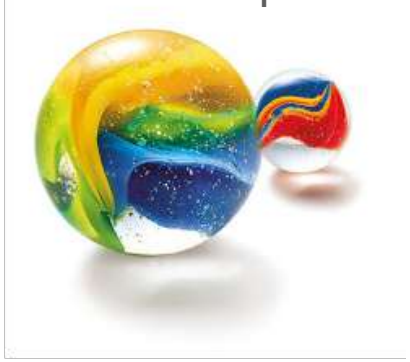
According to the Born Oppenheimer (BO) or adiabatic approximation, we can separate the motion of the nuclei from the motion of electrons and safely assume that **the electron cloud adjusts instantly to changes in the nuclear configuration**

This ‘translates’ into nuclei carrying a fixed net charge q (point charges) obtained by a projection of the QM electrostatic potential and long range attractive and short range repulsive, van Der Waals type potentials

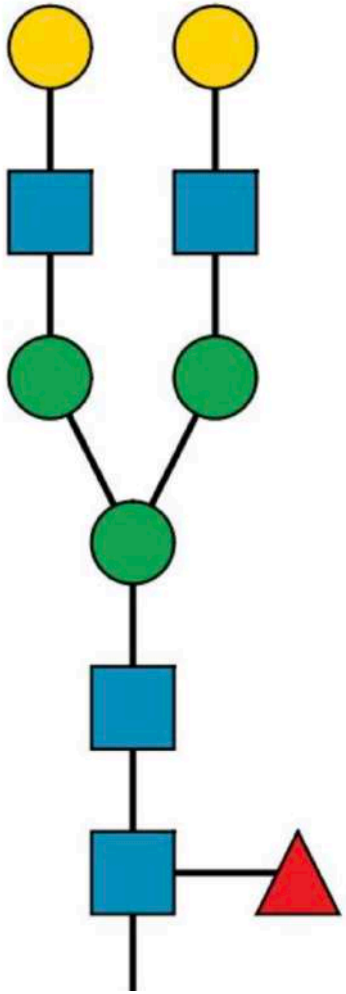
Fadda, *Curr Opin Chem Biol* (2022) DOI: /10.1016/j.cbpa.2022.102175

Classical Mechanics Representation

Atoms are represented by hard impenetrable spheres



Atoms are connected by springs representing bond stretching, angle and torsion bending



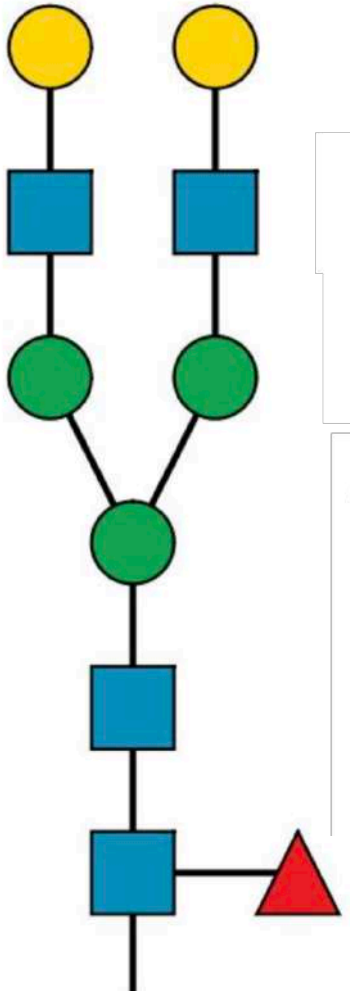
Electrostatics are represented by a **Coulomb potential**

Dispersion and short range steric repulsion are represented by a **Lennard-Jones potential**

Classical Mechanics Representation

Atoms are represented by hard impenetrable spheres

Atoms are connected by springs representing bond stretching, angle and torsion bending



THE FORCE FIELD

$$v^{\text{Coulomb}}(r) = \frac{Q_1 Q_2}{4\pi\epsilon_0 r},$$

$$v^{\text{LJ}}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right].$$

$$\begin{aligned}\mathcal{U}_{\text{intramolecular}} = &\frac{1}{2}\sum_{\text{bonds}} k_{ij}^{r}\left(r_{ij} - r_{\text{eq}}\right)^2 \\ &+ \frac{1}{2}\sum_{\substack{\text{bend}\\ \text{angles}}} k_{ijk}^{\theta}\left(\theta_{ijk} - \theta_{\text{eq}}\right)^2 \\ &+ \frac{1}{2}\sum_{\substack{\text{torsion}\\ \text{angles}}}\sum_{m} k_{ijkl}^{\phi,m}\left(1 + \cos(m\phi_{ijkl} - \gamma_m)\right)\end{aligned}$$

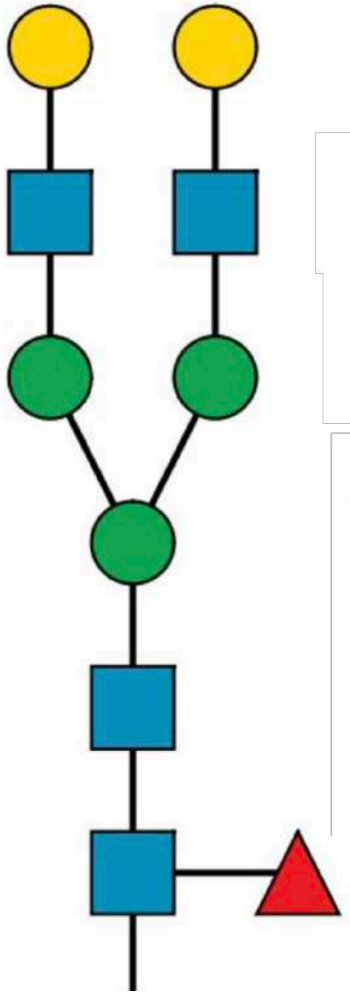
Classical Mechanics Representation

1. Positions of all atoms at *t(0)* given by the PDB

2. Corresponding potential energy *U(r)* given by the force field

$$\vec{F} = m\vec{a} = -\frac{dU}{dr}$$

These steps are iterated through integration over **2 fs** time steps to generate what we call **an MD trajectory**



THE FORCE FIELD

$$v^{\text{Coulomb}}(r) = \frac{Q_1 Q_2}{4\pi\epsilon_0 r},$$

$$v^{\text{LJ}}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right].$$

$$\begin{aligned}\mathcal{U}_{\text{intramolecular}} = &\frac{1}{2}\sum_{\text{bonds}} k_{ij}^{r}\left(r_{ij} - r_{\text{eq}}\right)^2 \\ &+ \frac{1}{2}\sum_{\substack{\text{bend}\\ \text{angles}}} k_{ijk}^{\theta}\left(\theta_{ijk} - \theta_{\text{eq}}\right)^2 \\ &+ \frac{1}{2}\sum_{\substack{\text{torsion}\\ \text{angles}}}\sum_{m} k_{ijkl}^{\phi,m}\left(1 + \cos\left(m\phi_{ijkl} - \gamma_m\right)\right)\end{aligned}$$

Some questions you should ask

- What information can I obtain from classical, all-atoms MD simulations?
- **Poll at slido.com select #3173816 to vote**
- Does the outcome of a simulation depend on the choice of force field? YES
- Is any force field good enough? NO
- What force field should I choose?
- Does the starting structure matter?
- How long should I run an MD simulation for?

GLYCAM06 Carbohydrate Force Field

A complete, self-contained and transferable set of parameters for the simulation of carbohydrates and glycoconjugates

- Carbohydrates of all ring sizes and conformations for both monosaccharides and oligosaccharides (not open ring conformations)
- Parameters for N- and O-glycosidic linkages, developed in combination with the *param99sb* version of the AMBER protein force field
- Parameters for sulfated (and not) GAGs

GLYCAM06 is a 'stand alone' force field that can in principle be combined to any other protein ff (non-covalent linkages)

It is the only ff that uses the same atom type (CG) for a and b anomers allowing for ring conformational changes (puckering)

Workshop: Building blood group epitopes with GLYCAM-Web

Kirschner *et al J Comp Chem* (2008) 29(4): 622–655
Fadda, *Curr Opin Chem Biol* (2022) DOI: /10.1016/j.cbpa.2022.102175
Fadda, *Comprehensive Glycoscience 2nd Ed* (2021) DOI:1/0.1016/B978-0-12-819475-1.00056-0

Maynooth University | Hamilton Institute
National University of Ireland Maynooth

Structural Glycoscience Summer School, June 2023

CHARMM36 Carbohydrate Force Field

CHARMM36 Carb parameter set (PAR_ALL36_CARB.PRM) to be used in combination with CHARMM36 parameters for proteins

42. Guvench O, Mallajosyula SS, Raman EP, Hatcher E, Vanommeslaeghe K, Foster TJ, Jamison FW, MacKerell AD., Jr CHARMM Additive All-Atom Force Field for Carbohydrate Derivatives and Its Utility in Polysaccharide and Carbohydrate-Protein Modeling. J Chem Theory Comput. 2011;7 (10):3162–3180.

43. Mallajosyula SS, MacKerell AD., Jr Influence of Solvent and Intramolecular Hydrogen Bonding on the Conformational Properties of O-Linked Glycopeptides. J Phys Chem B. 2011;115 (38):11215–11229.

44. Mallajosyula SS, Guvench O, Hatcher E, MacKerell AD., Jr CHARMM Additive All-Atom Force Field for Phosphate and Sulfate Linked to Carbohydrates. J Chem Theory Comput. 2012 doi: 10.1021/ct200792v.

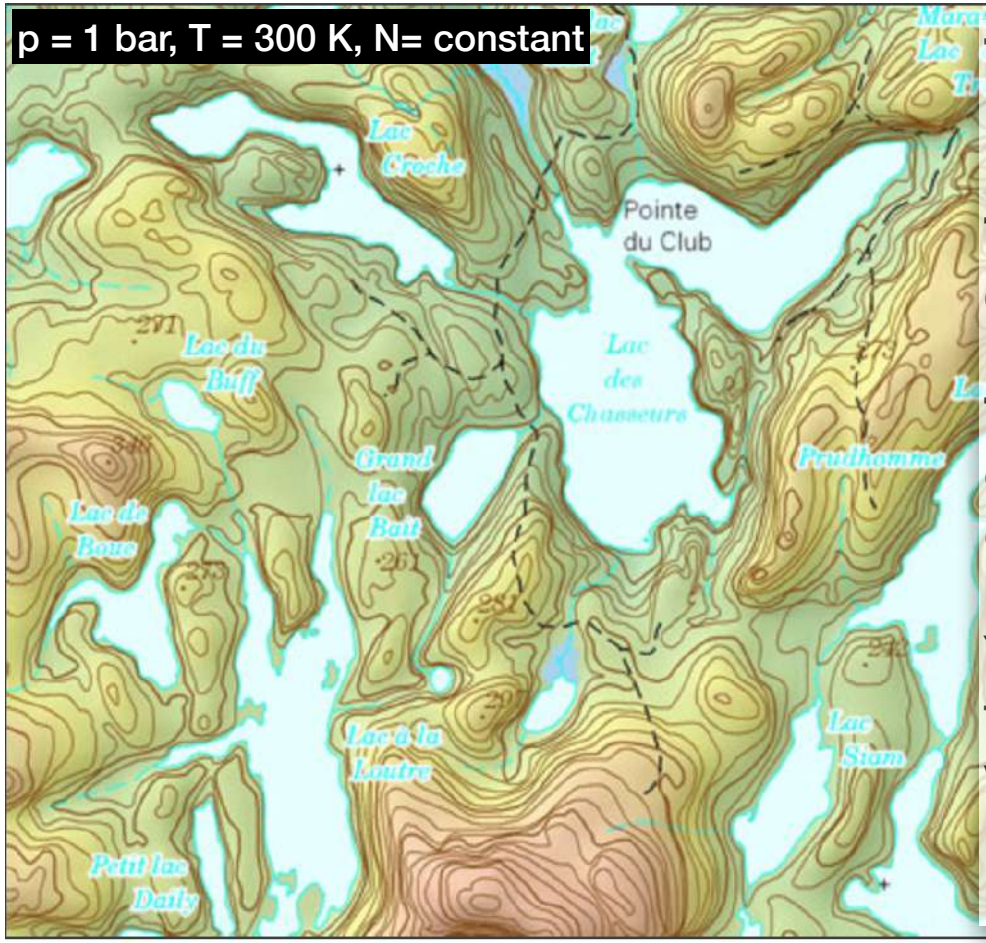
J Chem Theory Comput. 2018 Jun 12;14(6):3132-3143. doi: 10.1021/acs.jctc.8b00175. Epub 2018 May 4.

CHARMM Drude Polarizable Force Field for Glycosidic Linkages Involving Pyranoses and Furanoses.

Aytenfisu AH[1], Yang M[1,2], MacKerell AD Jr[1].

Fadda, *Curr Opin Chem Biol* (2022) DOI: /10.1016/j.cbpa.2022.102175
Fadda, *Comprehensive Glycoscience 2nd Ed* (2021) DOI:1/0.1016/B978-0-12-819475-1.00056-0

Does the starting structure matter?



This map is a "3D version" of a potential energy hypersurface

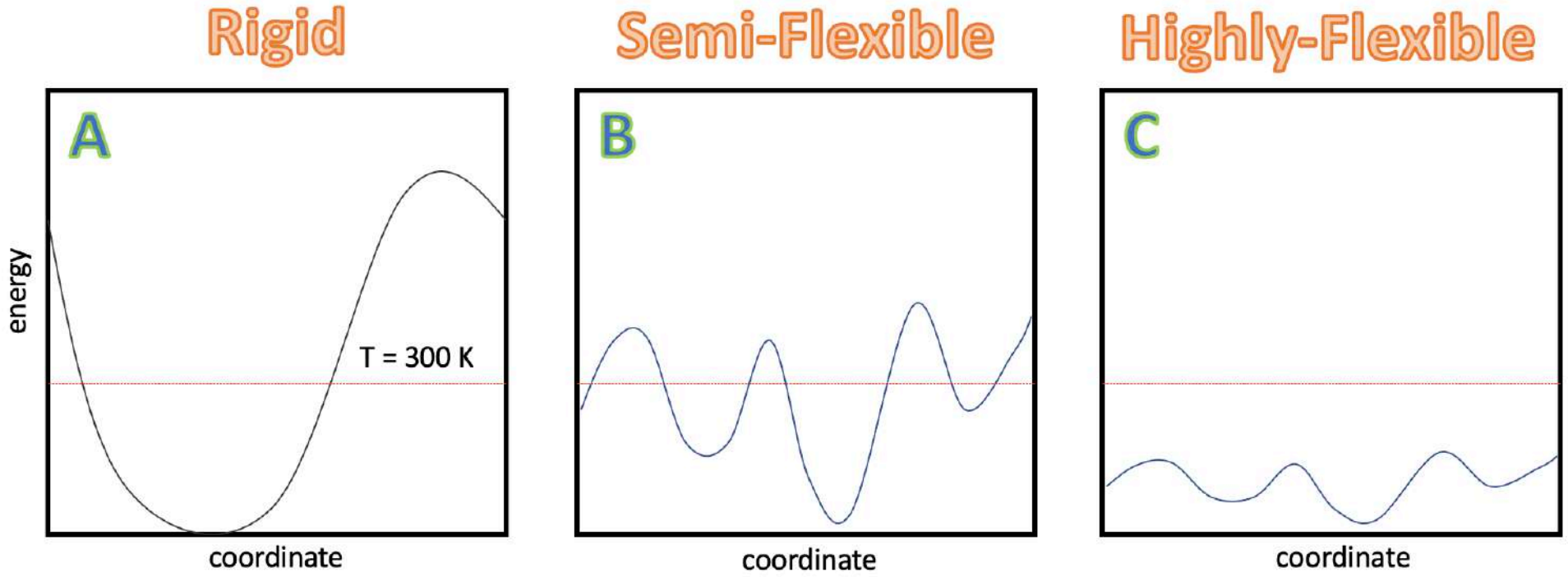
The height of each point indicates energy determined by the force field

The position in the map corresponds to the (collective) coordinates in the structure

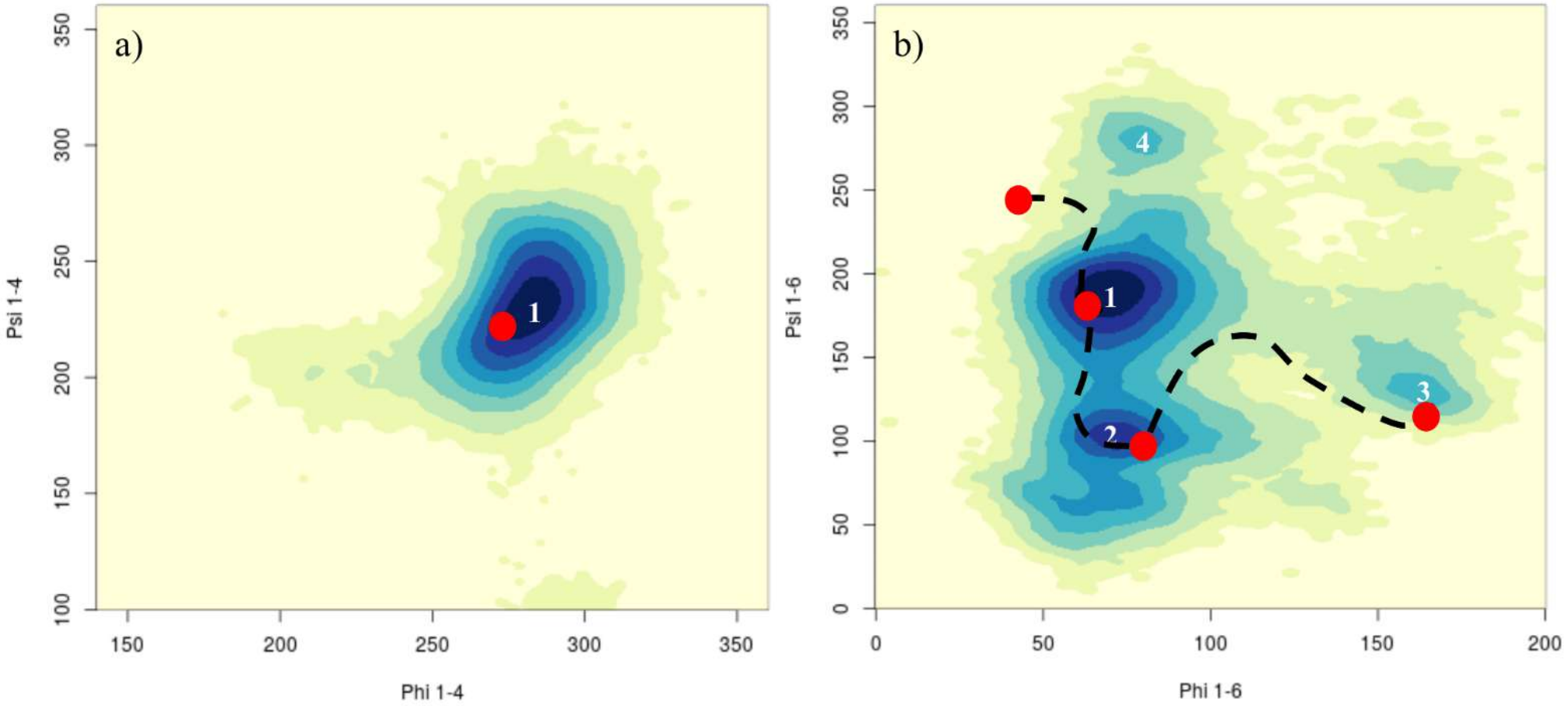
You can go from place to place through conformational sampling, via MD for example, where each new position is determined by the previous in the chosen path

Sampling: How long is long enough?

It depends on the system, on its intrinsic flexibility, on the chosen ff and on the question you would like to answer. **BUT** it all boils down to statistics, sampling is enough when you reach an answer supported by a **statistically relevant set of uncorrelated data**



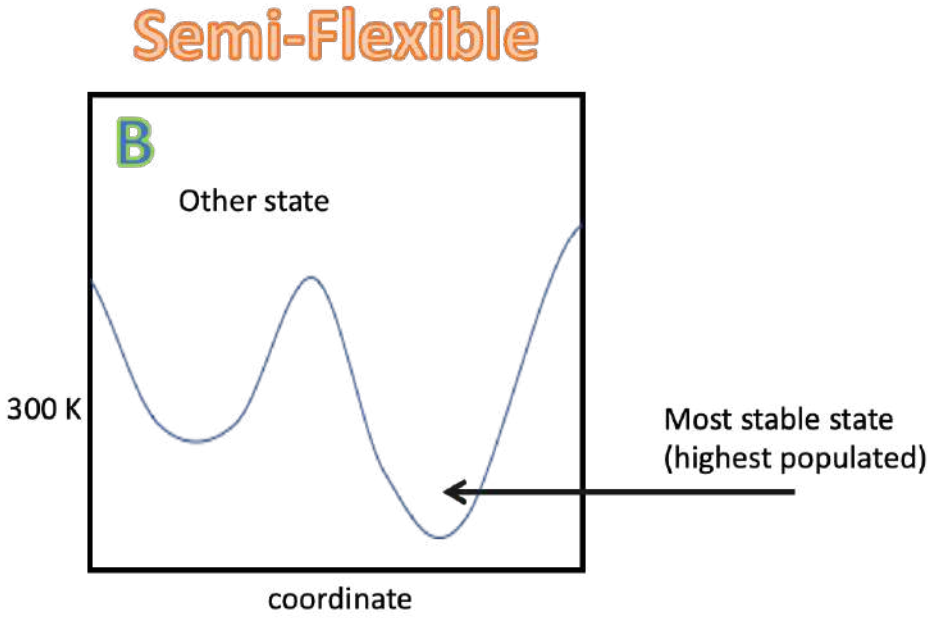
Sampling: How long is long enough?

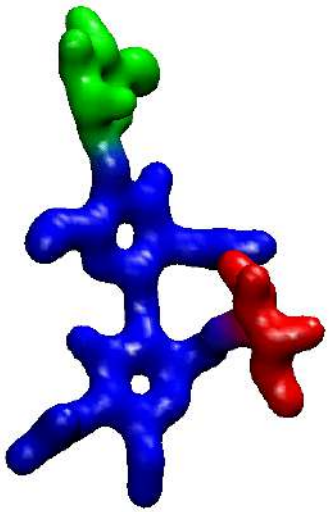


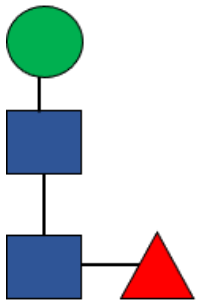
Fadda, *Curr Opin Chem Biol* (2022) DOI: /10.1016/j.cbpa.2022.102175
Fadda, *Comprehensive Glycoscience 2nd Ed* (2021) DOI:1/0.1016/B978-0-12-819475-1.00056-0

Enhanced Sampling Techniques

- **Deterministic**: Multiple independent (uncorrelated) parallel MD runs
- **Deterministic with stochastic exchanges**: Replica-Exchange, Fold-It
- **Biased-potentials Simulations**: Metadynamics, PMF, Umbrella Sampling
- **Stochastic:** Monte Carlo (MC) simulations, Brownian dynamics



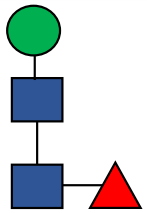




Harbison *et al, Glycobiology* (2019), 29:94
Mayes *et al, JACS* (2014), 136:1008

Confomational sampling from MD

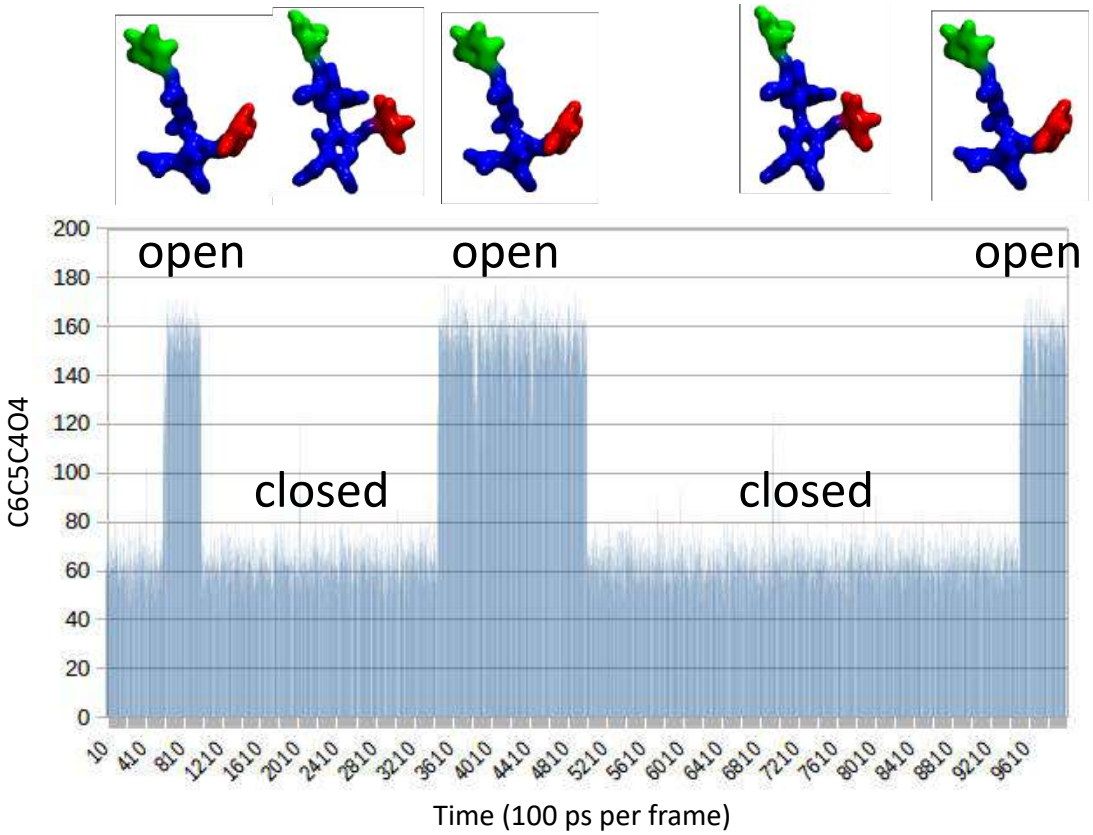
Man-β(1-4)-GlcNAc-β(1-4)-[α(1-6)-Fuc]-GlcNAc-β-OH



The MD trajectory was extended to 1 μs to analyze the conformational behavior of the sugar. A visual analysis has shown that there is **a significant conformational change occurring** at first around 60 ns.

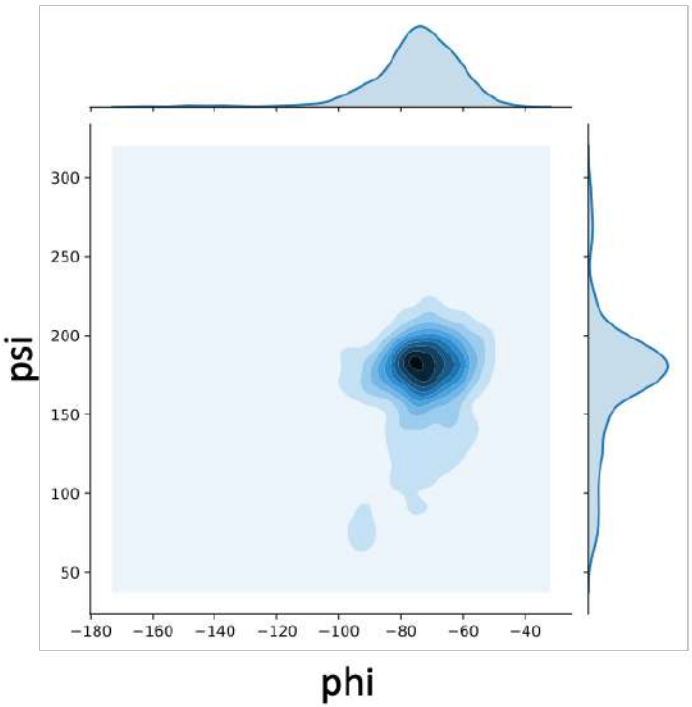
The conformational change lasts for about 40 ns. If the MD trajectory was of 100 ns, **the relative stability of the 2 conformations would have been 60:40**, largely overestimated.

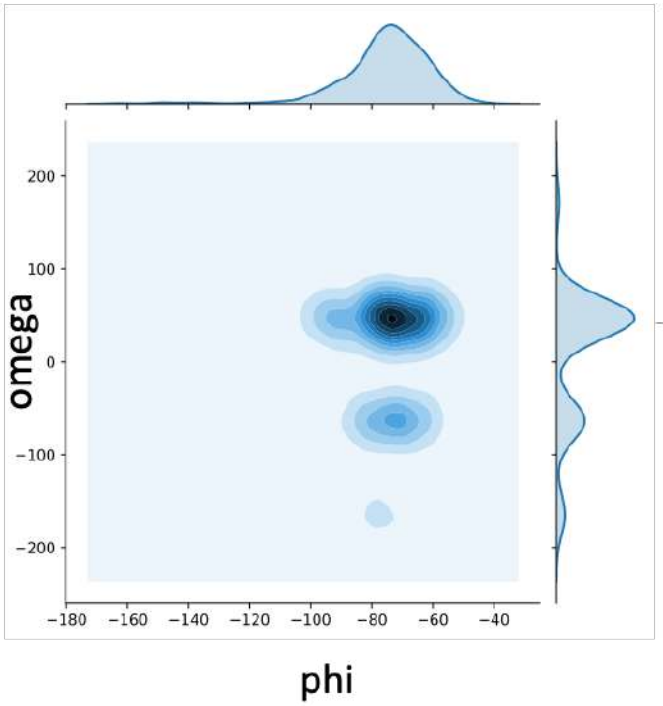
Over 1 μs the ratio is 76:24 => $|\Delta G°| \approx 3$ kJ/mol

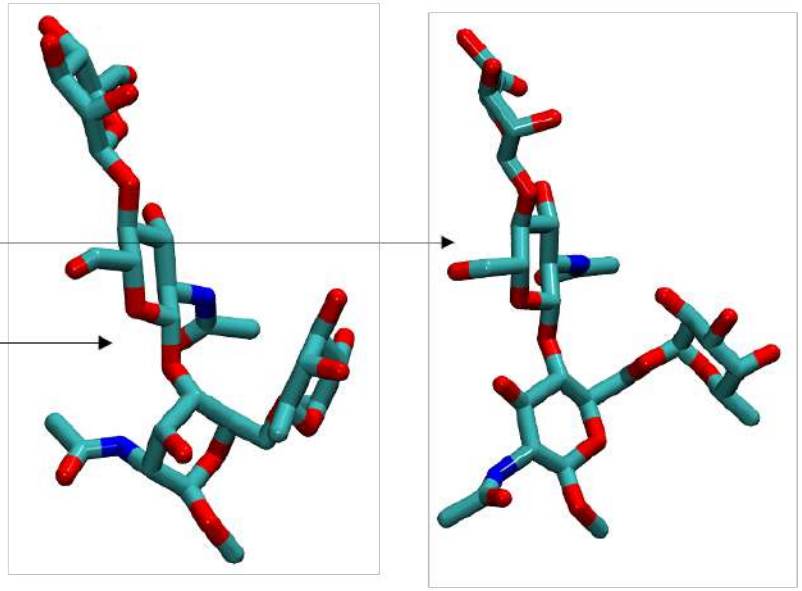


Confomational sampling from MD

Man-β(1-4)-GlcNAc-β(1-4)-[α(1-6)-Fuc]-GlcNAc-β-OH



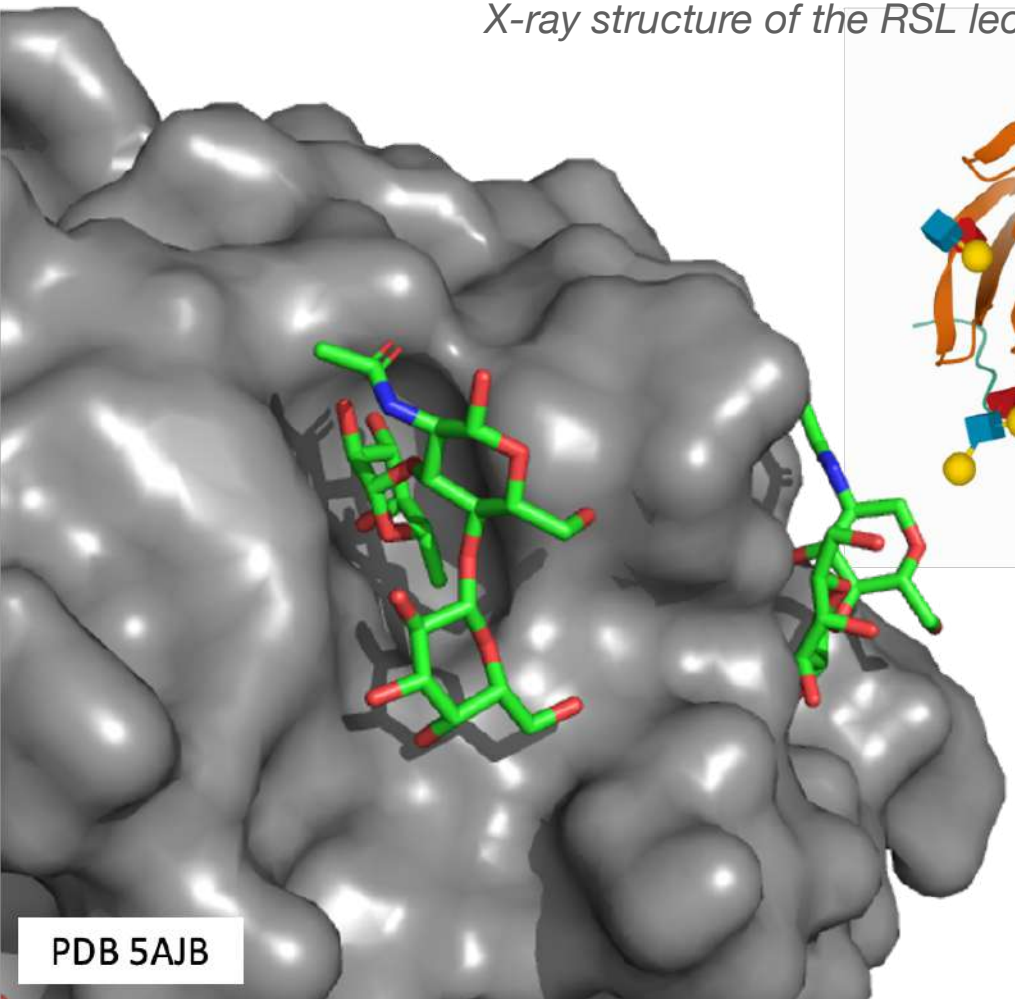


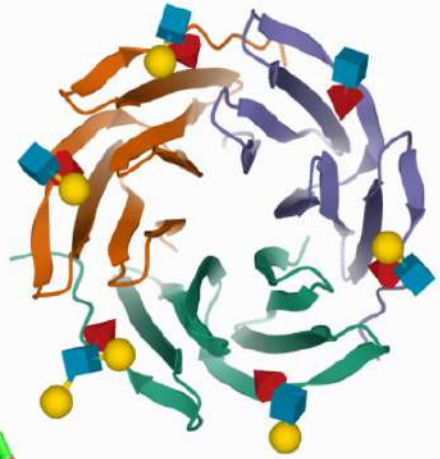


Harbison *et al, Glycobiology* (2019), 29:94
Mayes *et al, JACS* (2014), 136:1008

Recognition of rare conformations

X-ray structure of the RSL lectin in complex with Lewis X tetrasaccharide





Understand molecular recognition of carbohydrates:

- Structure
- Energetics

$$\Delta G^{\circ} = -RTln\frac{[conf\ 1]}{[conf\ 2]}$$

Topin *et al, ACS Chem Biol* (2016) DOI: 10.1021/acschembio.6b00333

What we have learned so far…

MD is a powerful structural biology tool that allows us to understand molecular structure, dynamics and function/behaviour at the atomistic level of detail

The above is true **ONLY** when the system is properly equilibrated and the simulation is converged, i.e. data obtained are statistically significant and reproducible within the given limitations of the model

Within these constrains, the characterisation of the glycans conformational propensity allows us to understand molecular recognition and related functions of glycans in biology

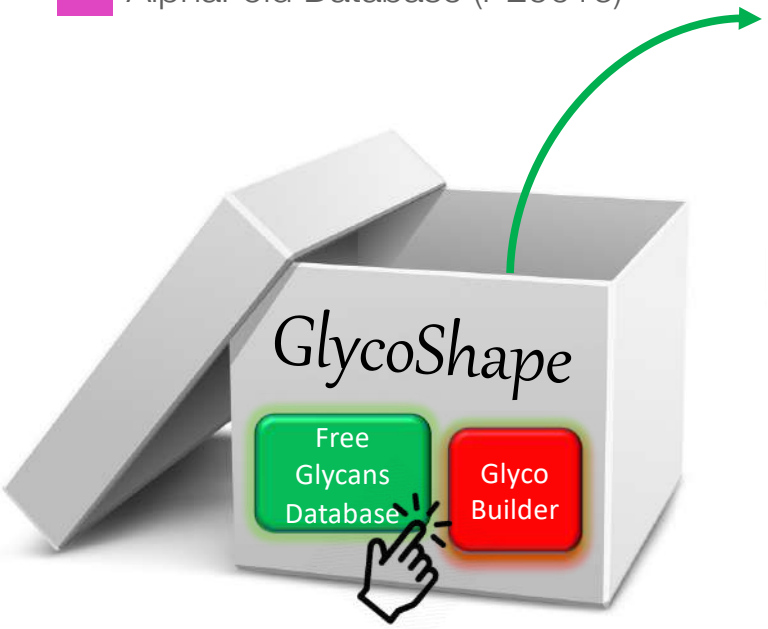
BUT, as any other research method, MD is far from an easy plug-in…

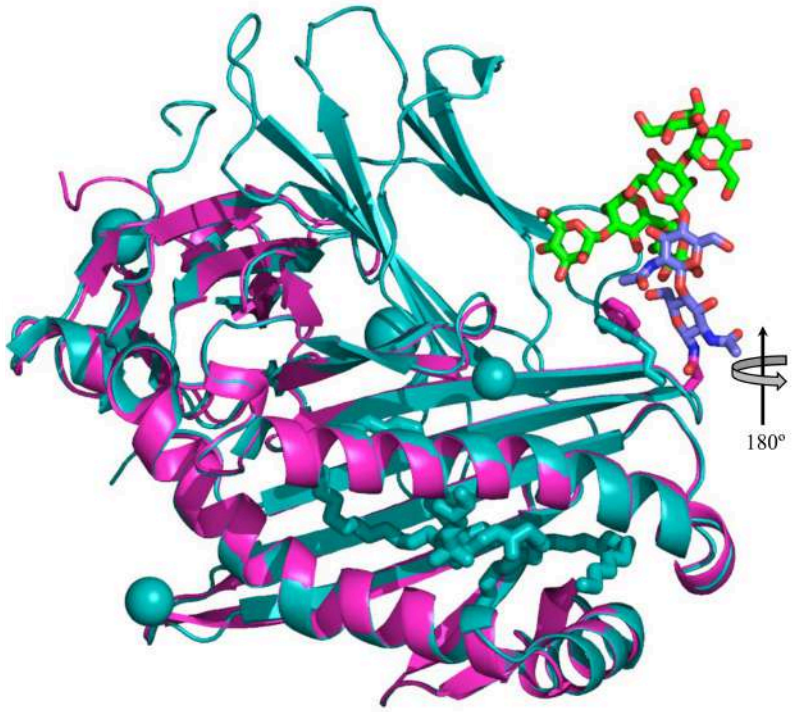
Maynooth University | Hamilton Institute
National University of Ireland Maynooth

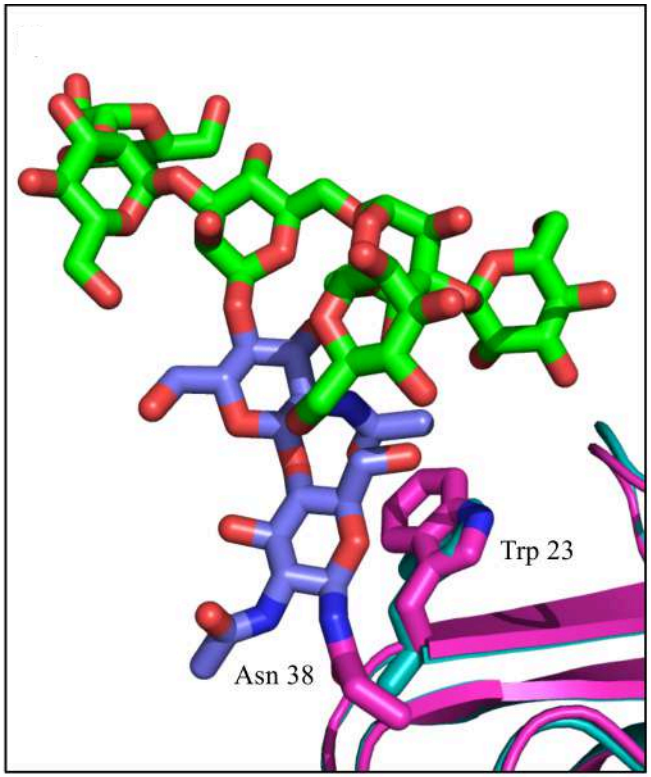
Structural Glycoscience
Summer School, June 2023

Rebuilding glycoproteins with MD libraries

PDB 5WL1 (1.38 Å)
AlphaFold Database (P29016)





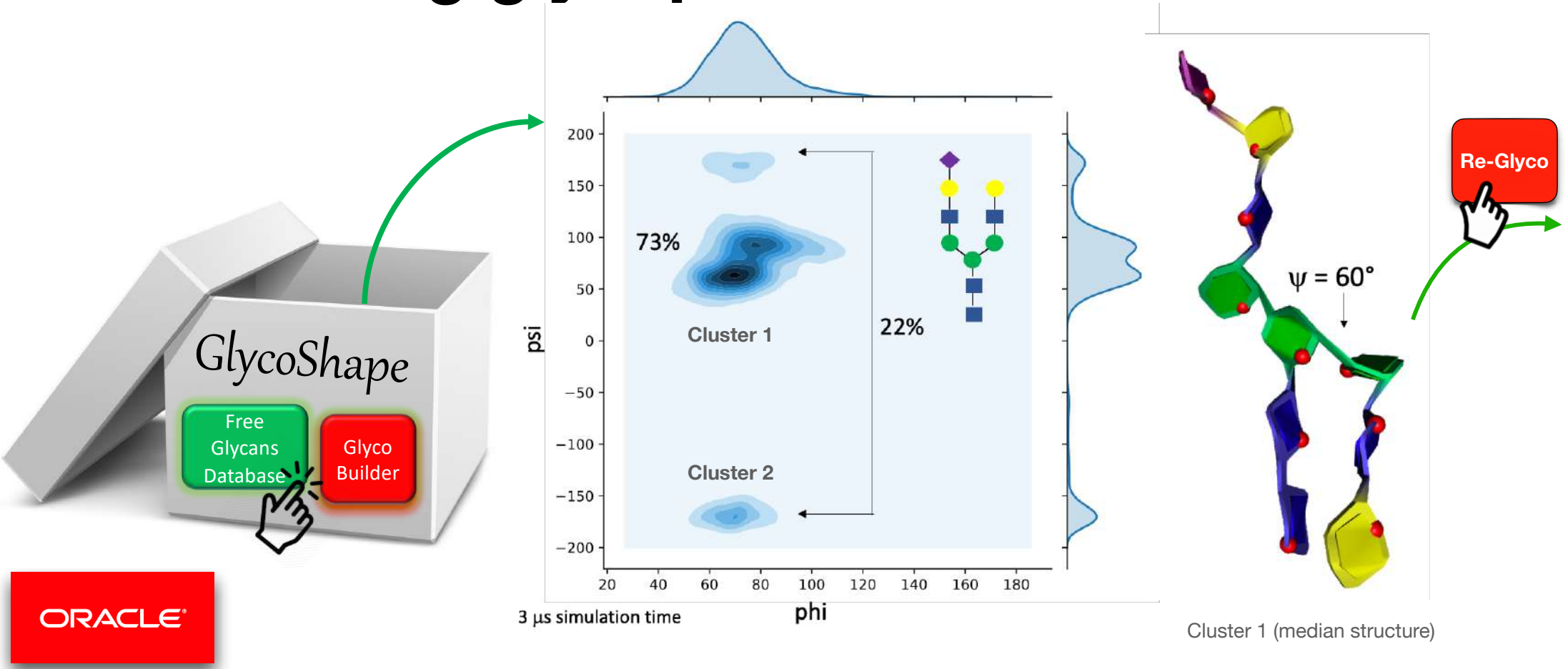


Human CD1b in Complex with PG

ORACLE

Bagdonas, Fogarty *et al*, *Nat Struct Mol Biol* (2021) DOI:10.1038/s41594-021-00680-9
GlycoShape.ie (under construction 🚧👷⛑️), **Ives C., Singh O.**, Fogarty, C, Harbison, A, Satheesan, A, D'Andrea S, Tropea B, Cuxart-Sanchez I, Khaleque M

Rebuilding glycoproteins with MD libraries



Cluster 1 (median structure)

ORACLE

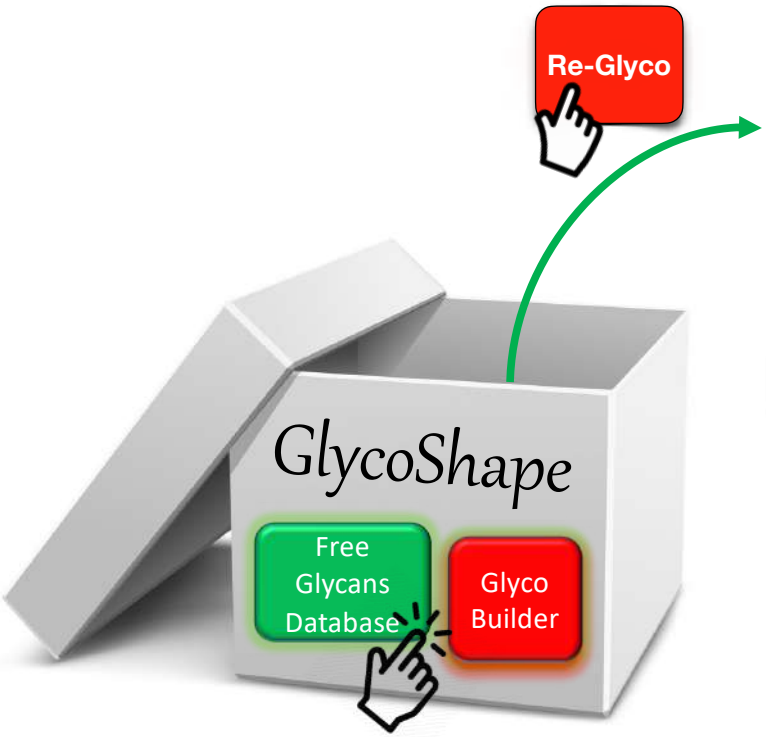
Bagdonas, Fogarty *et al*, *Nat Struct Mol Biol* (2021) DOI:10.1038/s41594-021-00680-9

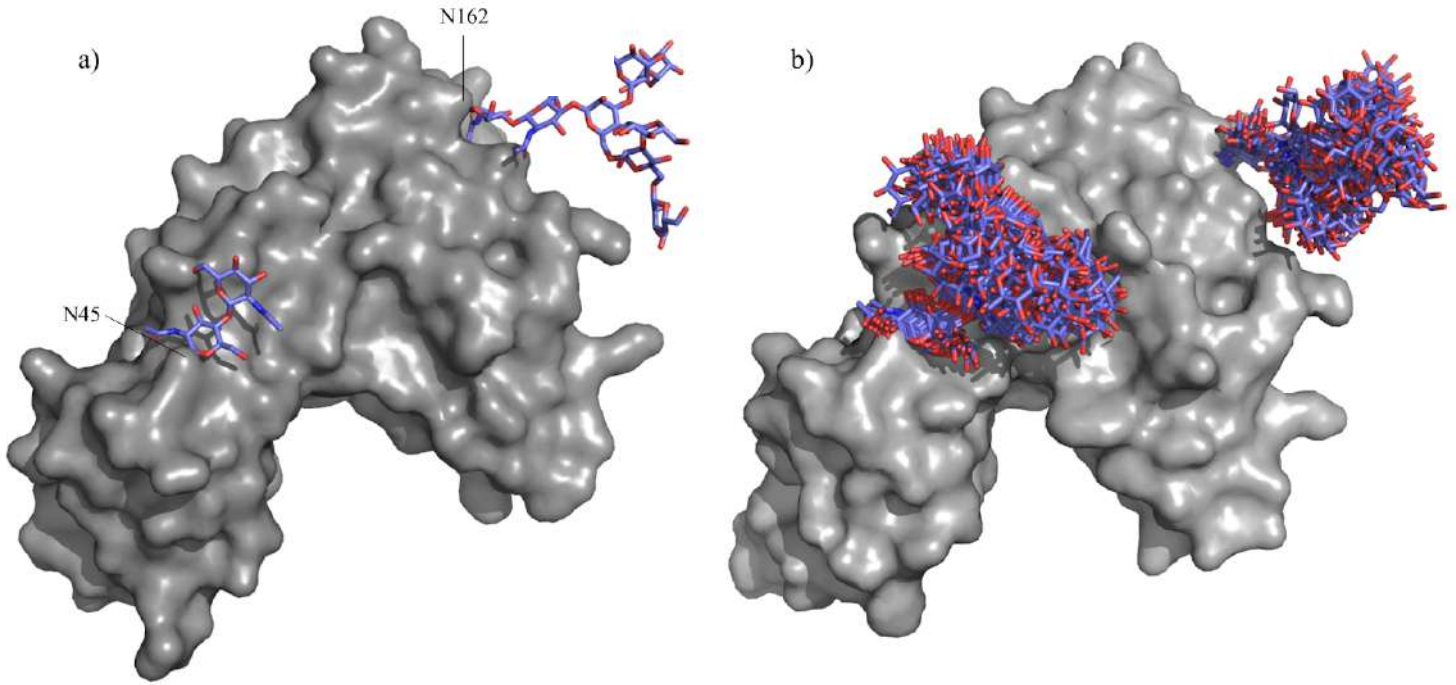
GlycoShape.ie (under construction 🚧👷⛑️), **Ives C., Singh O.**, Fogarty, C, Harbison, A, Satheesan, A, D'Andrea S, Tropea B, Cuxart-Sanchez I, Khaleque M

Maynooth University | Hamilton Institute
National University of Ireland Maynooth

Structural Glycoscience Summer School, June 2023

Rebuilding glycoproteins with MD libraries





a) Structure of the human CD16a (FcγRIIIa, **PDB 3SGK**) reglycosylated with Man5 from MD structural ensembles in **b)**. Only one ensemble (cluster) complements the protein landscape at N45, explaining the prevalence of hybrid N-glycosylation at that site (Fogarty and Fadda, *JPC-B* (2021) 125, 2607–2616; Fogarty *et al*, *RSC Glycoprotein Analysis* (2023 *in press*))

ORACLE

Bagdonas, Fogarty *et al*, *Nat Struct Mol Biol* (2021) DOI:10.1038/s41594-021-00680-9

GlycoShape.ie (under construction 🚧👷⛑️), **Ives C., Singh O.**, Fogarty, C, Harbison, A, Satheesan, A, D'Andrea S, Tropea B, Cuxart-Sanchez I, Khaleque M

Take home messages

Glycans are integral part of biomolecules regulating their folding, structural stability and biological functions. For this reason, glycans need to be re-integrated within the structure where present

Computer simulation methods, such as MD, are powerful tools to reconstruct and study the effects of glycosylation in biomolecular structure and function, **but** they need to be used correctly and the results interpreted within the limitations of the model systems

An appropriate choice of force field, simulation time and/or sampling technique and starting structure are determinant in obtaining meaningful insight, which should always be backed by statistics

Reconstructing protein glycosylation is not for the faint-of-heart, but *GlycoShape* will help, so stay tuned!


protein
glycosylation
@bnschlz