# In silico modelling of the function of disease-related CAZymes

Alba Nin-Hill, Beatriz Piniello, Carme Rovira

PORTLAND PRESS

## Review Article

# *In silico* modelling of the function of disease-related CAZymes

Alba Nin-Hill[1],*, Beatriz Piniello[1] and Carme Rovira[1,2]

[1]Departament de Química Inorgànica i Orgànica (Secció de Química Orgànica) & Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, 08028 Barcelona, Spain; [2]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08020 Barcelona, Spain

**Correspondence:** Carme Rovira (c.rovira@ub.edu)

OPEN ACCESS

*In silico* modelling of proteins comprises a diversity of computational tools aimed to obtain structural, electronic, and/or dynamic information about these biomolecules, capturing mechanistic details that are challenging to experimental approaches, such as elusive enzyme-substrate complexes, short-lived intermediates, and reaction transition states (TS). The present article gives the reader insight on the use of *in silico* modelling techniques to understand complex catalytic reaction mechanisms of carbohydrate-active enzymes (CAZymes), along with the underlying theory and concepts that are important in this field. We start by introducing the significance of carbohydrates in nature and the enzymes that process them, CAZymes, highlighting the conformational flexibility of their carbohydrate substrates. Three commonly used *in silico* methods (classical molecular dynamics (MD), hybrid quantum mechanics/molecular mechanics (QM/MM), and enhanced sampling techniques) are described for nonexpert readers. Finally, we provide three examples of the application of these methods to unravel the catalytic mechanisms of three disease-related CAZymes: β-galactocerebrosidase (GALC), responsible for Krabbe disease; α-mannoside β-1,6-N-acetylglucosaminyltransferase V (MGAT5), involved in cancer; and *O*-fucosyltransferase 1 (POFUT1), involved in several human diseases such as leukemia and the Dowling–Degos disease.

*Present address: Toulouse Biotechnology Institute, TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France. 135, avenue de Rangueil, F-31077 Toulouse Cedex 04, France.

## Introduction

Until the end of the 20th century, carbohydrates were mainly related to energy storage and structural support in living organisms. It was not until a few decades ago that scientists discovered that carbohydrates are also involved in much more complex biological processes, such as modulation of protein structures, signalling in multicellular systems, and cell–cell recognition. These processes are relevant for diseases including bacterial and virus infections, cancer [1–3], Alzheimer's [4,5], lysosomal storage diseases (LSD) [6], or disruption in the gut microbiota [7]. Nowadays, carbohydrates (commonly named 'sugars') are used in drug delivery strategies, vaccine development, and disease therapeutics [8–12].

Carbohydrates can exhibit many stereochemistries, configurations, and conformations (Figure 1), which makes them very complex molecules to study. The vast amount of carbohydrate-based structures in nature needs a larger number of enzymes responsible for their degradation, synthesis, and modification. This is the role of carbohydrate-active enzymes (CAZymes), which include glycoside hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs), and carbohydrate esterases (CEs) [1]. CAZymes are involved in many and various biological processes and thus they are keystones in human health. They are not only essential to life but also have important applications in food, detergent, oil, gas, and biotechnology industries [2,13]. Understanding the function of CAZymes at atomic detail can provide information on enzyme–substrate interactions that are critical for substrate recognition and catalysis, as well as the 'shape' (conformation) of the substrate along the catalytic reaction. These data can inform
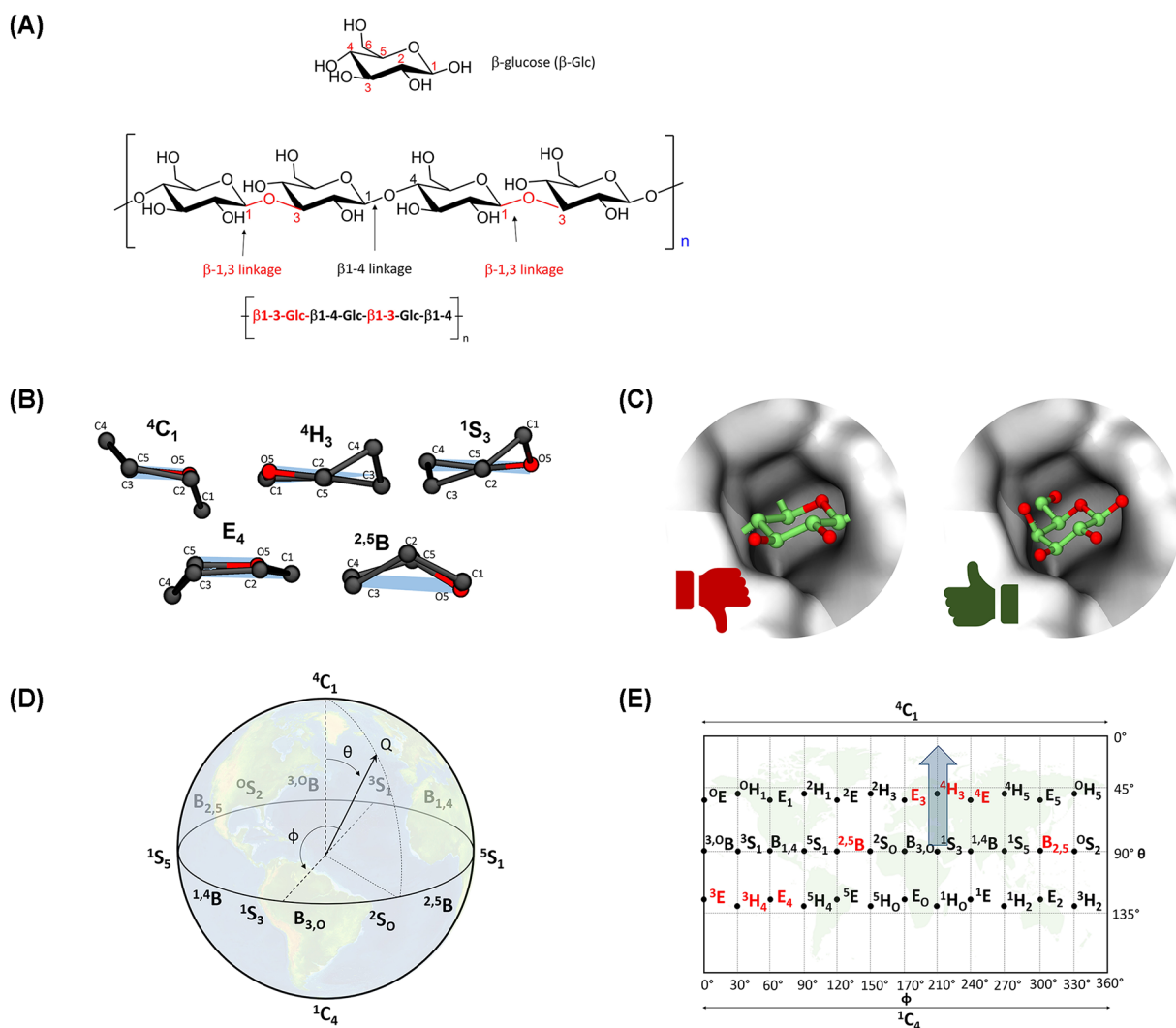
**Figure 1. Diversity of sugar linkages and conformations.**
(**A**) One of the most common β-glucans, a linear carbohydrate formed by glucose units with two types of glycosidic bond linkages, found in the cell wall of cereals, bacteria, and fungi. β-glucans can also display glycosidic linkage branching, such as β-1,6 branching in yeast and fungus. (**B**) The five possible types of conformations (C, H, S, E, and B) of a six-membered sugar ring. One example of each type is shown. The sugar reference plane is represented in light blue. The exocyclic groups of the sugars are not shown for clarity (**C**). The active sites of CAZymes have evolved to only fit a particular type of sugar conformation. In the example, β-glucose (left) is not able to fit in the active site of a β-galactosidase, while β-galactose (right) adapts to it. (**D**) The Cremer–Pople puckering sphere of pyranoses. The Earth is depicted as background to emphasise the similarity with an Earth map. (**E**) Mercator representation of the Cremer–Pople sphere. Conformations in which an oxocarbenium ion is stable are shown in red. The blue arrow shows one of the favoured conformational catalytic itineraries of β-glucosidases [ 14].

---

the design of substrate analogues or inhibitors that can efficiently bind to the active site of the enzyme, being potential drug candidates.

In the present review, we focus on the *in silico* modelling of CAZyme reaction mechanisms, highlighting the importance of sugar conformational changes during catalysis. We will start by describing how sugar conformations are determined and classified, and then introduce the two most abundant classes of CAZymes, GHs and GTs, along with their most common reaction mechanisms. Afterwards, we will briefly describe the basis of efficient computational methods used to uncover CAZyme reaction mechanisms, alongside a few examples of the application of these methods to unveil mechanisms of disease-related CAZymes. To make this text more understandable for the nonexperts, we provide standalone boxes to give more inside details of the concepts used in the present review.
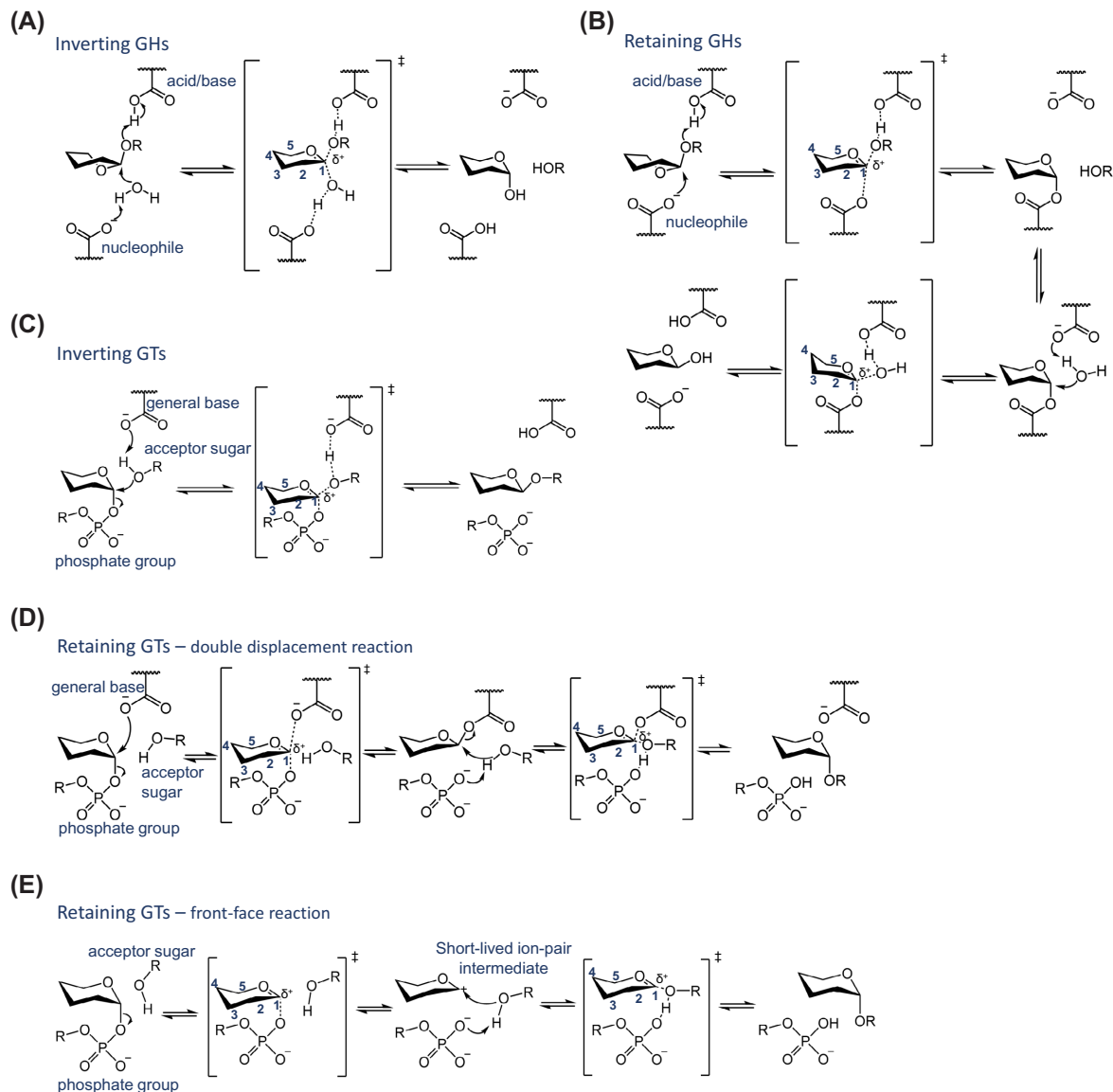
**Figure 2. Main catalytic mechanisms of the two most abundant CAZymes, GHs and GTs.**
Catalytic mechanisms of inverting GHs (**A**), retaining GHs (**B**), inverting GTs (**C**), retaining GTs operating by a double displacement reaction (**D**), and retaining GTs operating via a front-face type of reaction (**E**).

# Diversity of carbohydrate conformations

Carbohydrates can exhibit various types of linkages between their sugar units, as well as several stereochemistries (Figure 1A), leading to an enormous number of possible structures [1]. Ring conformation adds another level of complexity (Figure 1B). A six-membered sugar ring, or pyranose, can adopt 38 different ring conformations, which can be classified as chair (C), half-chair (H), skew-boat (S), envelope (E), or boat (B). These conformations are defined according to the position of their atoms with respect to a reference ring plane. When one or two atoms are placed below or above this plane, these atoms are denoted as subscripts or superscripts, respectively (e.g. $^4C_1$, $^4H_3$, etc.). From a mathematical point of view, all these conformations can be associated to points in the surface of a sphere of radius Q, the *puckering sphere*, defined by the sum of the perpendicular distance of all atoms to a reference ring plane. The 38 canonical conformations can be distinguished by their polar θ and φ coordinates, as introduced by

Cremer and Pople in 1975 [15]. Usually, the puckering sphere (Figure 1D) is projected in two-dimensions, leading to the so-called Mercator representation, in a similar way as Earth is represented in maps (Figure 1D,E), although other representations (*e.g.* Stoddart reresentation) have also been used. The position and orientations of the sugar exocyclic groups in each of the 38 Cremer–Pople conformations are essential to understand the binding of sugars in the active site of CAZymes, as will be discussed below.

## CAZymes

CAZymes are classified in families according to sequence similarities, and they are exquisitely curated in the CAZy database [16], which is steadily growing as genome sequencing projects increase. Complementary information about structure and catalytic mechanisms of CAZymes can also be found in CAZypedia [17]. The two most abundant types of CAZymes are GHs and GTs, which are responsible for the hydrolysis (GHs) and synthesis (GTs) of the glycosidic bond linkages between sugar units in carbohydrates.

### GHs

GHs catalyse the cleavage of glycosidic bonds in carbohydrates and glycoconjugates (carbohydrates bound to other biomolecules such as lipids or protein residues). They can be classified as retaining or inverting, depending on whether or not the stereochemistry of the anomeric carbon changes during the reaction, by the similarity in their amino acid sequences (with more than 173 families to date) [16] or by the location of the scissile glycosidic bond within the carbohydrate chain. Endo-GHs cleave glycosidic bonds at the middle of the chain, whereas exo-GHs cleave the terminal ends of carbohydrates.

### GTs

GTs catalyse the formation of glycosidic bonds in carbohydrates and glycoconjugates. These enzymes use activated sugars (i.e. sugars with good leaving groups such as phosphate) as glycosyl donors and transfer their glycosyl unit to an acceptor molecule (carbohydrate, lipid, or protein). Similar to GHs, they can either invert or retain the configuration of the donor anomeric carbon and are classified in families according to their amino acid sequence similarity, with 116 families to date [16]. GTs can also be classified in two groups depending on whether the donor sugar is activated by a nucleotide phosphate (Leloir GTs) or by another phosphate-substituted molecule (non-Leloir GTs) [18].

## The chemistry behind CAZymes: GHs

Most GHs share a common mechanism in which two residues with a carboxylic acid side chain (Asp or Glu) assist the reaction. One of these residues acts initially as an acid (the so-called *acid/base residue* in retaining GHs, or *catalytic acid* in inverting GHs) and another one acts as a general base (inverting GHs) or as a nucleophile (retaining GHs) [19,20] (see Figure 2).

Most retaining GHs operate via two chemical steps (double displacement mechanism), each one being a $S_N2$-type of reaction (Figure 2B). In the first step, the general base attacks the anomeric carbon to form a covalent glycosyl-enzyme intermediate (GEI), while the acid/base residue assists the reaction by protonation of the glycosidic oxygen. The order of events and the degree of nucleophilic attack and glycosidic oxygen protonation at the TS varies from one GH to another. In the second step, a water molecule activated by the acid/base residue, which now acts as a base, attacks the anomeric carbon, leading to the products of the reaction. Inverting GHs (Figure 2A) operate by a single-displacement mechanism (one chemical step) with the participation of a nucleophilic water molecule that is deprotonated by the general base (see e.g. [21] and [22]).

### Changes on sugar ring conformation along the chemical reaction

A number of structural analyses of GHs in complex with their carbohydrate substrates (Michaelis complexes, MCs) have shown that the substrate distorts upon binding to the enzyme. In particular, the reactive sugar (i.e. the one bearing the C–O bond to be hydrolysed by the enzyme) is distorted away from the ground-state $^4C_1$ conformation [21,23–25]. This change from a 'ground state' to a distorted – and typically higher energy – conformation promotes catalysis. In particular, the distortion orients the glycosidic bond axially and elongates it, making the sugar's overall shape resemble that of the transition state (TS) of the reaction, thus preactivating the substrate for catalysis (Figure 2B). The specific distorted conformation that a sugar adopts depends on the particular GH family. However, computer simulations have shown that the ring conformations adopted by the substrate in the enzyme active site are the ones that exhibit certain 'suitable' structural, energetic, and electronic properties (long and axial C1–O bond, large anomeric

charge, low energy, etc.) of the free sugars (i.e. in absence of the enzyme) [25]. In other words, GHs have evolved to use intrinsic properties of their sugar substrates for a most efficient catalysis [26].

The TS of the GHs reaction is characterised by a partial positive charge at the anomeric carbon and a partial double bond between the anomeric carbon and the ring oxygen [27–30]. These two properties are maximised when the C1, C2, O, and C5 atoms are on the same plane, *i.e.* the ring becomes planar around the anomeric carbon (e.g. conformation $^4H_3$ in Figure 1B). Only eight sugar conformations fulfil this condition in pyranoses, namely, $^3E$, $^3H_4$, $E_4$, $^{2,5}B$, $E_3$, $^4H_3$, $^4E$, $B_{2,5}$ (Figure 1E). Thus, the sugar ring at the TS of glycosylation reactions needs to adopt a conformation that is close to one of these eight conformations.

The specific conformations that the reactive sugar adopts along the chemical reaction, especially those of the MC, TS, and product (P) complexes delineate the so-called *catalytic conformational itinerary*. This itinerary is specific for each GH family. For instance, members of family GH16, which are mostly β-glucosidases, display a $^1S_3 \rightarrow [^4H_3]^\ddagger \rightarrow {}^4C_1$ itinerary [14], whereas GH38 enzymes, which are α-mannosidases, display a $^OS_2 \rightarrow [B_{2,5}]^\ddagger \rightarrow {}^1S_5$ itinerary [31]. Identifying conformational catalytic itineraries of GHs is of great importance when designing selective inhibitors. Molecules that mimic the properties of the MC or the TS of GHs are often powerful inhibitors [32].

# The chemistry behind CAZymes: GTs

GTs catalyse the synthesis of glycosidic bonds with retention or inversion of the anomeric configuration. Inverting GTs operate through a single-nucleophilic substitution reaction in which a general base assists the reaction by deprotonating the nucleophile acceptor (see Figure 2C) [17]. This mechanism is similar to that of inverting GHs (Figure 2A). In Leloir GTs, the nucleophile is a polar group of an acceptor (e.g. a hydroxyl group in the case of a sugar acceptor) and the leaving group is a phosphate group (Figure 2C). Many GTs have a divalent cation (typically $Mg^{2+}$ or $Mn^{2+}$) that coordinates to the nucleotide phosphates, although metal-independent enzymes have also been described [33].

The detailed mechanism of retaining GTs remains challenging, as two possible mechanisms have been proposed. A double displacement, Koshland-type mechanism was initially proposed by analogy with retaining GHs (Figure 2D). In this mechanism, an active site aspartate or glutamate plays the role of a nucleophile, reacting with the anomeric carbon of the donor sugar and forming a glycosyl—enzyme covalent intermediate. In a second step, an acceptor molecule attacks the anomeric carbon, breaking the glycosyl-enzyme covalent bond and forming a new glycosidic bond, with overall retention of stereochemistry. A similar mechanism has been recently proved by structural and mass spectrometry experiments on a GT99 [34]. There is theoretical evidence of a double-displacement mechanism for GT6 enzymes [35,36], although experimental detection of the covalent intermediate remains challenging [37,38].

The second proposed mechanism for retaining GTs involves a $S_Ni$-type of reaction, also termed *front-face* reaction [39]. In this mechanism, the nucleophile attacks from the same face of the donor sugar in which the phosphate group departs, without the direct intervention of an enzyme residue [16]. Whether the reaction is concerted or a stepwise remains controversial [40].

Sugar ring distortion is less relevant for GTs reactivity than for GHs since the phosphate leaving group is typically axially oriented, facilitating the nucleophilic displacement. However, ring distortions of the donor sugar to satisfy tight hydrogen-bond interactions have been reported [41]. For more information on GT structures and mechanisms, we recommend references [42,43].

# Dissecting catalytic reaction mechanisms in CAZymes

The modelling of reaction mechanisms on CAZymes relies on the initial determination of the structure of the enzyme in complex with its substrate/s (MC complex), obtained by e.g. X-ray crystallography. The MC structure can inform *in silico* modelling, allowing to 'visualise' changes on the structure as the system evolves towards the products of the reaction at atomic and electronic detail.

## Obtaining the structure of the enzyme in complex with its substrate/s

There are several methods to obtain the three-dimensional structure of a protein: NMR [44], cryo-electron microscopy [45], modelling via artificial intelligence (AlphaFold2 or RoseTTAFold) [46–48], and X-ray crystallography [49–51]. The latter has been traditionally used for the study of CAZyme mechanisms, as one can obtain a structure of the enzyme in complex with its substrate that is reliable enough to inform *in silico* modelling of the mechanism of action. For more information on obtaining protein structures, we refer to reference [52]. Nearly 200 000 protein structures determined using the above-mentioned methods (and more) are stored in the Protein Data Bank [53–55],

but only 2363 CAZymes had at least one 3D structure in the PDB (1688 GHs and 332 GTs) when this manuscript is being written (January 2023) [16].

Ideally, an *in silico* study of a catalytic mechanism requires the structure of the complex between the wild-type enzyme and its natural substrate as input. Such structure is rarely available, as the reaction timescale ($\approx$ms) is smaller than the time resolution of the experiment. However, structural biologists can use various strategies to slow down or knock-out the enzymatic reaction, being able to solve the structure of the MC. Among these strategies, the use of nonreactive analogue substrates, mutation of essential residues, and/or working at nonoptimal pH conditions have been used. These modifications perturb the system and may result in enzyme configurations that depart from the true MC (i.e. the complex of the wild-type enzyme with the natural substrate). However, the structure obtained can give essential clues to uncover the enzyme mechanism [17]. In addition, complementary techniques such as site-directed mutagenesis, spectroscopic, and *in silico* modelling can enormously contribute to decipher complex catalytic mechanisms [56].

## *In silico* modelling of CAZymes reaction mechanisms

*In silico* approaches such as molecular dynamics (MD), based on molecular mechanics (MM) force fields, are often used to ensure that the enzyme and/or substrate modifications used in the laboratory to trap the MC have led to a catalytically competent enzyme–substrate configuration, or to transform the noncompetent crystal structures into competent ones. Once these modifications (e.g. residue mutations or substrate alterations) are reverted to the wild-type form, the MC complex is subjected to energy minimisation and thermal equilibration by MD. Afterwards, the reaction mechanism is modelled using quantum mechanics-molecular mechanics (QM/MM) approaches, which combine a QM treatment of a certain region (e.g. the enzyme-active site) with a MM treatment of the rest of the system (rest of the enzyme and solvent molecules). The basis of these approaches and the specific procedure is provided below.

Experimental and computational approaches can work in synergy to reach the final objective of deciphering how CAZymes work and, consequently, understanding how they influence health and disease in living organisms.

# Specific procedure

To model CAZyme reaction mechanisms, we first focus on the MC, i.e. the enzyme in complex with the substrate, which is our 'reactants complex.' After building the initial model (see Box 1), we use methods based on MM (so-called 'classical' methods) to bring the system to a configuration that corresponds to the ensemble of structures accessible at the temperature of interest (e.g. 300 K). This procedure, called thermal equilibration, is part of the MD protocol (see Box 2). By using MD, we can also simulate how the system moves in the ns–$\mu$s timescale, capturing important protein structural rearrangements. Popular programs for performing MD of proteins are Amber [57,58], NAMD [59], Gromacs [60–62], and OpenMM [63] (Figure 3).

---

**Box 1 Model building**

To build a catalytically competent protein–substrate structure, we do the following:

1. Starting from the available experimental structure, add the missing residues and/or loops. Several programs can be used, we name here just a few ones such as Modeller [64], standalone or as implemented in Chimera [65], or the web service MolProbity [66].

2. Revert possible mutations of the structure that were needed to determine the MC structure, using *e.g.* VMD [67].

3. Add the missing hydrogen atoms of the structure. MolProbity [66], PyMol [68], Chimera [65], VMD [67], pdb4amber, or tleap [57,58] are a few software examples.

4. Inspect the protonation state of the titratable residues [69]. This can be done by visual inspection of the amino acid environment and/or using pKa predictor programs such as PropKa [70,71] and H++ [72].

5. Solvate the protein, placing it into a water box periodically repeated in space. This can also be done with several tools included in MD packages, such as tleap [57].

---
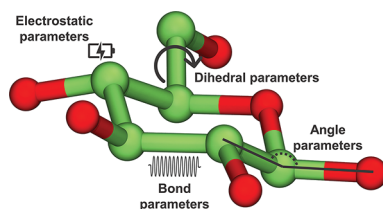
**Figure 3. Types of force field parameters used to model a sugar molecule in classical MD (see [73] for the mathematical expression).**

Each parameter describes interactions among pairs of atoms that are separated by one, two, or three covalent bonds (bond, angle, and dihedral parameters), or atoms that are further away in space (electrostatic and van der Waals interactions).

---

## Box 2 MD

The time propagation of the atoms in MD is obtained from the solution Newton's equations of motion.

$$\vec{F_I}\,(t) = m_i\,\vec{a_I}\,(t) \Rightarrow -\frac{\delta V\left(\vec{r}\right)}{\delta r_i} = m_i\frac{\delta^2\,\vec{r_1}\,(t)}{\delta t^2}$$

These equations can be solved by numerical methods, obtaining the coordinates of the system at small and consecutive time intervals, typically femtoseconds, but the simulation can be extended up to microseconds, even milliseconds. The procedure relies on a mathematical expression for the potential energy, $V(r)$. In the so-called *classical* MD, the potential energy is approximated with a set of functions that describe the interactions among atoms and include predefined parameters. The resulting potential energy expression is called a force field. For an extensive review, see [73]. In the so-called *ab initio MD* (AIMD) [74], the potential energy is computed by QM, typically using density functional theory (DFT).

One limitation of classical MD is that it is unable to account for electronic rearrangements in atoms, thus preventing its use for modelling reaction mechanisms. To do that, it is necessary to include explicitly the electrons in the MD formalism, specifically in the calculation of the potential energy. This approach is named as AIMD. There are various MD methods that can do it, including Ehrenfest MD (EMD), Born-Oppenheimer MD (BOMD), or Car-Parrinello MD (CPMD) [75], based on the seminal work by Car and Parrinello [74] that enormously influenced the AIMD field. The potential energy, which is needed to propagate the equation of motion (Box 2), can be computed by wave function-based methods, such as coupled cluster of Moller–Plesset methods. However, it is still virtually impossible to do it for large systems, due to the exponentially increasing computational cost [76]. Usually, DFT [77,78], based on the calculation of the electronic density, is used (see [75] for detailed information). Two very popular computer programs to run AIMD simulations are CPMD [79] and CP2K [80]. The main drawback of AIMD is that it is restricted to around a few hundred atoms (see Figure 6), which precludes simulating most proteins. On the contrary, classical MD allows simulating large systems, but it describes them less accurately and cannot account for bond breaking/formation processes. Combining both methods, capitalising on the fact that the atoms involved in bond breaking/bond forming are in a small region of the whole biomolecule, overcomes the size limitation of AIMD. This is the basis of the QM/MM methodology, for which Karplus, Levitt, and Warshel were awarded the Nobel Prize in 2013, Box 3 (Figure 4).

---

## Box 3 QM/MM

The QM/MM approach is based on dividing the system in two regions. The so-called QM region comprises the group of atoms for which large electronic rearrangements are expected to take place (e.g. active site atoms). Atoms in this region are described by a QM method, e.g. DFT. The so-called MM region comprises the rest of the system, in which bonds are not expected to break or form. Atoms in the MM region are described by MM. Within a MD formalism, the method is named QM/MM MD (i.e. AIMD in the QM region and classical MD in the MM region). As a result, the whole system evolves under the Newton's equations of motion and bonds are able to break and form in the QM region. In
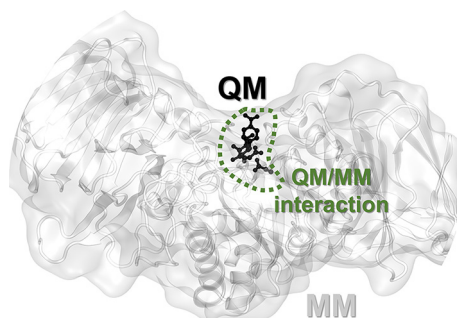
**Figure 4. The two regions (QM and MM) involved in QM/MM calculations**
The MM region is shown in grey, the QM region is shown in black, and the interface between the two regions is shown in green.

state-of-art QM/MM codes, the interaction between both systems is computed using the additive scheme:

$$E = E^{QM} + E^{MM} + E^{QM/MM}$$

in which $E^{QM}$ is the energy of the QM region, $E^{MM}$ is the energy of the MM region, and $E^{QM/MM}$ is the interaction energy between the atoms in the QM region and the atoms of the MM region. The former are represented by an electron density, whereas the later are represented as particles with point charges and van der Waals radii. Particular care needs to be taken when there are covalent bonds at the frontier between both regions. In this case, the QM atom does not have the right number of neighbors (e.g. four for an sp3 carbon atom) to saturate the electron density, as the neighbouring MM atom is modelled by a point charge, thus it does not have electron density. Several methodologies, such that the use of dummy link atoms to saturate the border QM atoms, have been developed to describe the QM-MM boundary. For more information, we refer to the following reviews [81,82].

Even though QM/MM methods can be used to describe larger enzymes with QM accuracy at the active site, chemical reactions cannot be observed in the time scale of QM/MM MD (tenths of picoseconds). This is because a sizable energy barrier, much higher than the energy available at room temperature, needs to be overcome when moving from reactants to products. In this scenario, the computational time needed to observe a transition from R to P is prohibitive [83]. Advanced methods have been developed to enhance the sampling of system configurations until the desired transition is observed. Some of these methods are based on the definition of a few degrees of freedom of the system (e.g. crucial distances or angles), named as collective variables (CVs), that describe the motion of interest with enough flexibility that reaction free energies and mechanisms can be obtained. Among these methods, steered MD, umbrella sampling, or metadynamics are commonly used [83–85]. Most of them are included in the Plumed software [86,87], which can be patched with standard MD packages. For more information, we recommend the following review [88] (Box 4).

**Box 4 Enhanced sampling: metadynamics**

Metadynamics is a MD-based method based on decreasing the probability that the system visits configurations/conformations (or compatible microstructures) that have already been visited during the MD simulation. In this way, it is easy for the system to escape from one energy minimum, such as the reactant state, to another one, such as the product state. This is achieved by adding a bias potential to the standard potential coming from the atomic/electronic system [89,90]. It can be demonstrated that in the time limit the total added biasing potential converges to the free energy change corresponding to the process of interest [91] (Figure 5).

$$V_{bias}[t \rightarrow inf](s, t) \approx -\Delta G(s)$$
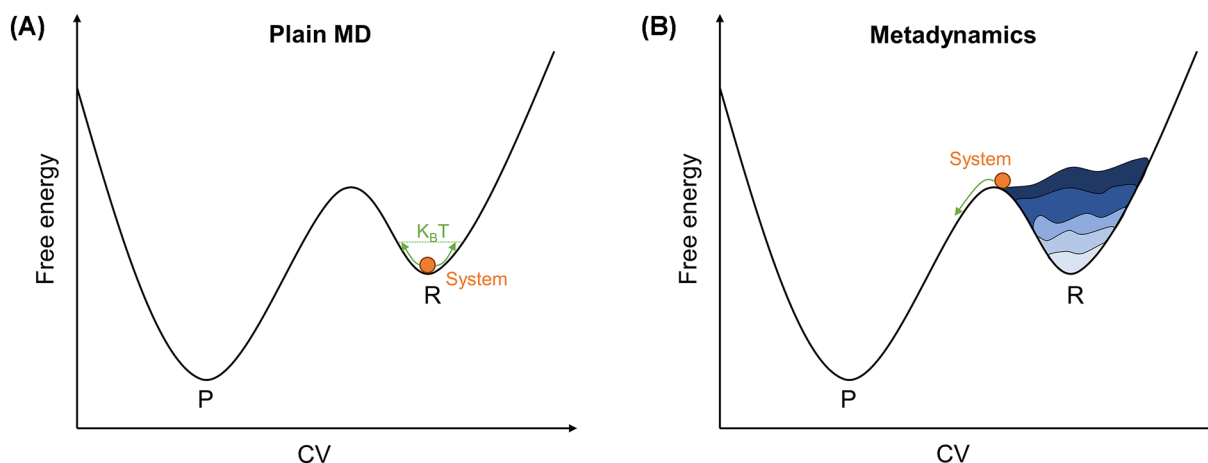
**Figure 5. Comparison of the system evolution between plain (i.e. standard) MD and metadynamics.**
Schematic representation of a fictitious particle in standard MD (**A**) and metadynamics (**B**). The system (orange circle) cannot escape from state R when using standard MD. In metadynamics, it can move through the 'hills' (in blue) that build up the bias potential, $V_{bias}$, crossing the energy barrier and sampling state *P*.

> A proper choice of CVs ensures a satisfactory physical description of the process of interest, avoiding a prohibitively large computational cost to achieve full convergence [92]. In the particular case of chemical reactions, several studies have shown that appropriate CVs can be taken as combinations of distances corresponding to the bonds being broken and formed during the reaction.

In summary, following equilibration of the system (enzyme MC complex) using classical MD, QM/MM MD along with an enhanced sampling technique is used to compute the reaction mechanisms, including the conformational catalytic itineraries of the substrate in GHs and GTs. We should note that these simulations require a significant computational cost; they cannot be performed in a personal desktop computer and often require access to high-performance computing (HPC) infrastructures. These centres are equipped with supercomputers (computers with more than one-hundred thousand processors, memories in the order of terabytes and disk storages in the order of petabytes) that can be accessed by academic and nonacademic groups according to specific policies from research centres, universities, governments, and other public bodies.

## Examples

*In silico* modelling has been used to investigate a number of catalytic mechanisms of CAZymes in the last few years. We refer to past and recent reviews for a list of interesting examples [22,93–96]. Here, we focus on three disease-related CAZymes that have been recently studied.

β-galactocerebrosidase (GALC) is a retaining GH that is responsible for the cleavage of certain glycosphingolipids, *i.e.* sugars attached to lipid molecules. Deficiencies of GALC lead to Krabbe disease, an incurable neurodegenerative disorder caused by the accumulation of the unhydrolysed substrates (mainly β-galactocerebroside, GalCer, and psychosine) in the nervous system [97].

The structure of GALC in complex with a hydrolysable substrate, reported in [98], enabled the modelling of the enzyme mechanism of action for the first time. Strikingly, the reactive Gal was found in a relaxed $^4C_1$ conformation in the crystal structure. This is not a catalytically preactivated sugar conformation, as it displays the leaving group in an equatorial orientation, thus it was not clear whether the complex with the substrate in $^4C_1$ was representative of a true 'snap-shot' of the enzyme along the catalytic reaction. However, QM/MM metadynamics simulations by Nin-Hill and Rovira [99] showed that it is the case. Due to the solvent-exposed nature of the leaving group in exo-acting GHs, such as GALC, the glycosidic bond can be efficiently cleaved when the galactoside substrate adopts either a $^4C_1$ conformation (Figure 7A) or a distorted $^1S_3$ conformation. Therefore, GALC can operate via two alternative conformational itineraries, either starting from a relaxed conformation ($^4C_1 \rightarrow [^4H_3]^{\ddagger} \rightarrow {}^4C_1$) or a distorted onen ($^1S_3 \rightarrow [^4H_3]^{\ddagger} \rightarrow {}^4C_1$) [99], with the latter being slightly favoured. These specific conformations, discovered by *in*
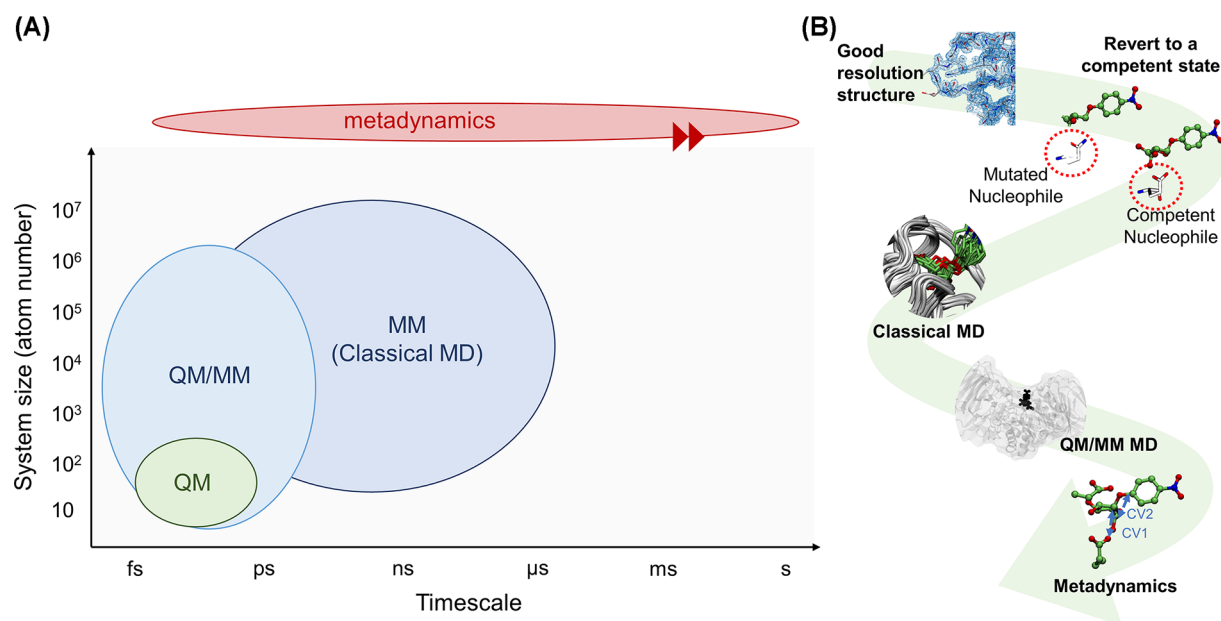
**Figure 6. Schematic representations of the methodologies described in this review.**
(**A**) Timescale vs system size representation of the methods described in this review. The accuracy and computational cost of the simulation decreases from bottom-left to top-right. Using enhanced sampling techniques such as metadynamics allows describing processes that would only be observed in long time scales, such as chemical reactions (ms-s time scale). (**B**) Computational protocol described in this review, from the static X-ray structure to the results of QM/MM MD and metadynamics simulations.

*silico* modelling, can be useful for the rational design of substrate analogues that can be used as conformational chaperones for Krabbe disease therapy.

Our second example focuses on α-mannoside β-1,6-*N*-acetylglucosaminyltransferase V (MGAT5, also GnT-V), a mammalian inverting GT involved in the formation of complex-type tetra-antennary *N*-glycans (carbohydrates linked to the nitrogen atom of asparagine residues in proteins). MGAT5 transfers *N*-acetylglucosamine (GlcNAc) from a UDP-GlcNAc glycosyl donor on to the core α-1,6 mannose (Man) of an *N*-glycan acceptor. The resulting branched GlcNAc-β-1,6-Man linkage is a precursor for the formation of complex tri- and tetra-antennary *N*-glycans, which are elaborated in the *trans*-Golgi by the addition of Gal-β-1,4-GlcNAc (LacNAc) disaccharides and sialic acids. Overexpression of MGAT5 strongly drives cancer [100–103], thus reducing the enzyme activity can inhibit tumour growth. To date, very few effective small-molecule inhibitors of MGAT5 have been developed [104–106] and understanding of its enzyme−substrate interactions and catalytic mechanisms is limited. The reaction mechanism of MGAT5 was recently uncovered by a combination of X-ray crystallography and QM/MM metadynamics simulations [107]. The results highlighted the key assisting role of Glu297 (the putative catalytic base) and revealed a distinct conformational itinerary for the GlcNAc ring during its transfer from donor to the acceptor (Figure 7B). This work provided a comprehensive molecular overview of MGAT5 catalysis that will guide inhibitor development efforts. In particular, it was suggested that pharmacological targeting of the MGAT5 donor subsite, using inhibitors inspired by conformational analysis, may also be effective for the development of compounds to control the activity of this GT.

Our last example concerns an inverting GT with a mechanism that departs from the classical $S_N2$-type of reaction. *O*-fucosyltransferase 1 (POFUT1) is an inverting GT involved in *O*-glycosylation, a protein modification essential to life. In particular, POFUT1 attaches L-fucose sugars to threonine or serine residues in certain protein sequences. POFUT1 is involved in the Notch signalling pathway (NSP), an essential cell–cell communication pathway conserved in all multicellular animals [108,109]. Malfunction of the NSP can cause several diseases in humans, from Dowling–Degos disease, to leukemia or colorectal cancer [110–112], thus POFUT1 is a relevant therapeutic target. Unlike most inverting GTs (e.g. the previous example, MGAT5), the active site of POFUT1 lacks a basic residue that can act as the catalytic base deprotonating the incoming nucleophile acceptor, thus the mechanism of action of POFUT1 remained puzzling. Using QM/MM metadynamics simulations, Piniello et al. [41] revealed that the reaction involves proton shuttling through an active site asparagine, conserved among species, which undergoes tautomerisation, i.e. it changes the side chain from amide to imidic acid form during the reaction. The enzyme retains the $S_N2$
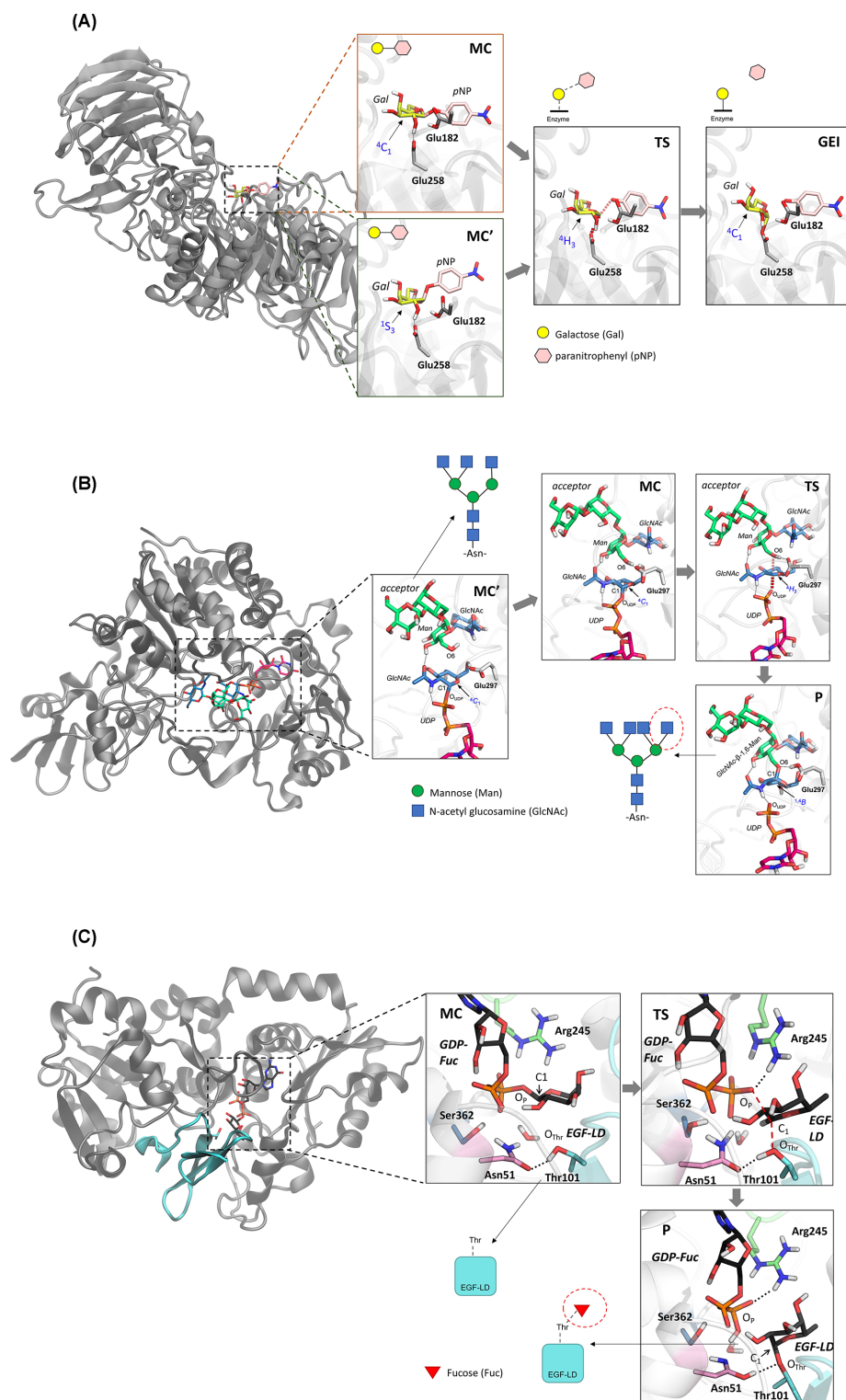
**Figure 7. Three disease-related CAZymes whose catalytic mechanisms have been recently investigated.**
(**A**) Representative structures of the GALC catalytic mechanism, starting with the substrate in either $^4C_1$ or $^1S_3$ conformation.
(**B**) Representative structures along the MGAT5 catalytic mechanism. The image has been adapted with permission from [107], Copyright (2011), The American Chemical Society. (**C**) Representative structures of the POFUT1 catalytic mechanism. The epidermal growth factor-like domain (EGF-LD) containing the Thr where fucose will be attached to is depicted in cyan. The image has been adapted with permission from [41]. Copyright (2011), The American Chemical Society. The results shown were obtained by QM/MM metadynamics simulations in all cases. Hydrogen atoms attached to carbon atoms have been omitted for clarity.

mechanism of inverting GTs (Figure 1C), but an asparagine residue rather than a basic residue (Glu or Asp) deprotonates the hydroxyl group of the acceptor nucleophile. This novel mechanism, recently invoked in a related GT [113], could be found in other GTs yet to be discovered.

This short review shows that the use of state-of-art *in silico* approaches, such as QM/MM metadynamics, is providing unprecedented insight into enzyme catalytic mechanisms of CAZymes. We think that, in the future, this will facilitate the development of inhibitors and activity-based probes for CAZymes involved in diseases.

## Summary

- CAZymes are the enzymes responsible for the processing of carbohydrates in nature. They are involved in many biological processes and thus they are keystones in human health.

- *In silico* modelling enables uncovering conformational catalytic itineraries and reaction mechanisms of CAZymes at atomic detail.

- We described three different examples in which *in silico* modelling was essential for the unravelling of complex enzymatic mechanisms.

## Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

## Author Contribution

C.R., A.N-H., and B.P. contributed to writing the manuscript. A.N-H. also produced the figures.

## Abbreviations

B, boat; BOMD, Born-Oppenheimer molecular dynamics; C, chair; CAZymes, carbohydrate-active enzymes; CE, carbohydrate esterase; CPMD, Car-Parrinello molecular dynamics; CV, collective variable; DFT, density functional theory; E, envelope; EMD, Ehrenfest molecular dynamics; GALC, $\beta$-galactocerebrosidase; GEI, glycosyl-enzyme intermediate; GH, glycoside hydrolase; GT, glycosyltransferase; H, half-chair; HPC, high-performance computing; LSD, lysosomal storage disease; MC, Michaelis complex; MD, molecular dynamics; MGAT5 or GnT-V, $\alpha$-mannoside $\beta$-1,6-N-acetylglucosaminyltransferase V; NSP, Notch signalling pathway; PL, polysaccharide lyase; POFUT1, *O*-fucosyltransferase 1; QM/MM, quantum mechanics/molecular mechanics; S, skew-boat; TS, transition state.

## References

1. Stick, V.R. and Williams, J.S. (2009) *Carbohydrates: The Essential Molecules of Life*, 2nd edn, Elsevier, Amsterdam

2. Hart, G. and Cell, R.C. (2010) Glycomics hits the big time. *Cell* **143**, 672–676, https://doi.org/10.1016/j.cell.2010.11.008

3. Zhou, J.Y., Oswald, D.M., Oliva, K.D., Kreisman, L.S.C. and Cobb, B.A. (2018) The glycoscience of immunity. *Trends Immunol.* **39**, 523–535, https://doi.org/10.1016/j.it.2018.04.004

4. Gloster, T.M. and Vocadlo, D.J. (2012) Developing inhibitors of glycan processing enzymes as tools for enabling glycobiology. *Nat. Chem. Biol.* **8**, 683–694, https://doi.org/10.1038/nchembio.1029

5. McCaddon, A., Hudson, P., Davies, G., Hughes, A., Williams, J.H.H. and Wilkinson, C. (2001) Homocysteine and cognitive decline in healthy elderly. *Dement. Geriatr. Cogn. Disord.* **12**, 309–313, https://doi.org/10.1159/000051275

6. Platt, F.M., D'Azzo, A., Davidson, B.L., Neufeld, E.F. and Tifft, C.J. (2018) Lysosomal storage diseases. *Nat. Rev. Dis. Primers* **4**, 1–25, https://doi.org/10.1038/s41572-018-0025-4

7  Wardman, J.F., Bains, R.K., Rahfeld, P. and Withers, S.G. (2022) Carbohydrate-active enzymes (CAZymes) in the gut microbiome. *Nat. Rev. Microbiol.* **20**, 542–556, https://doi.org/10.1038/s41579-022-00712-1

8  Koeller, K.M. and Wong, C.H. (2000) Emerging themes in medicinal glycoscience. *Nat. Biotechnol.* **18**, 835–841, https://doi.org/10.1038/78435

9  Spratley, S.J. and Deane, J.E. (2016) New therapeutic approaches for Krabbe disease: the potential of pharmacological chaperones. *J. Neurosci. Res.* **94**, 1203–1219, https://doi.org/10.1002/jnr.23762

10  Ghazarian, H., Idoni, B. and Oppenheimer, S.B. (2011) A glycobiology review: carbohydrates, lectins and implications in cancer therapeutics. *Acta Histochem.* **113**, 236–247, https://doi.org/10.1016/j.acthis.2010.02.004

11  Ranzinger, R., Herget, S., Adibekian, A. and von der Lieth, C.-W. (2007) Synthesis and medical applications of oligosaccharides. *Nature* **446**, 1046–1051, https://doi.org/10.1038/nature05819

12  Graziano, A.C.E., Pannuzzo, G., Avola, R. and Cardile, V. (2016) Chaperones as potential therapeutics for Krabbe disease. *J. Neurosci. Res.* **94**, 1220–1230, https://doi.org/10.1002/jnr.23755

13  André, I., Potocki-Vé Ronè Se, G., Barbe, S., Moulis, C. and Remaud-Simé On, M. (2014) CAZyme discovery and design for sweet dreams. *Curr. Opin. Chem. Biol.* **19**, 17–24, https://doi.org/10.1016/j.cbpa.2013.11.014

14  Biarnés, X., Ardèvol, A., Iglesias-Fernández, J., Planas, A. and Rovira, C. (2011) Catalytic itinerary in 1,3-1,4-$\beta$-glucanase unraveled by QM/MM metadynamics. Charge is not yet fully developed at the oxocarbenium ion-like transition state. *J. Am. Chem. Soc.* **133**, 20301–20309, https://doi.org/10.1021/ja207113e

15  Cremer, D. and Pople, J.A. (1975) A general definition of ring puckering coordinates. *J. Am. Chem. Soc.* **97**, 1354–1358, https://doi.org/10.1021/ja00839a011

16  Drula, E., Garron, M.L., Dogan, S., Lombard, V., Henrissat, B. and Terrapon, N. (2022) The carbohydrate-active enzyme database: functions and literature. *Nucleic. Acids. Res.* **50**, D571–D577, https://doi.org/10.1093/nar/gkab1045

17  Contributors Caz (2019) Main Page. CAZypedia, © 2007-2019 the authors and curators of CAZypedia. 13510, http://www.cazypedia.org/index.php?title=Main_Page&oldid=13510 [Accessed 16 November 2022]

18  Ünligil, U.M. and Rini, J.M. (2000) Glycosyltransferase structure and mechanism. *Curr. Opin. Struct. Biol.* **10**, 510–517, https://doi.org/10.1016/S0959-440X(00)00124-X

19  Koshland, D.E. (1953) Stereochemistry and the mechanism of enzymatic reactions. *Biological Rev.* **28**, 416–436, https://doi.org/10.1111/j.1469-185X.1953.tb01386.x

20  Sinnott, M.L. (1990) Catalytic mechanisms of enzymic glycosyl transfer. *Chem. Rev.* **90**, 1171–1202, https://doi.org/10.1021/cr00105a006

21  Vasella, A., Davies, G. and Böhm, M. (2002) Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* **4**, 573–580, https://doi.org/10.1016/S1367-5931(02)00380-0

22  Ardèvol, A. and Rovira, C. (2015) Reaction mechanisms in carbohydrate-active enzymes: glycoside hydrolases and glycosyltransferases. Insights from ab initio quantum mechanics/molecular mechanics dynamic simulations. *J. Am. Chem. Soc.* **137**, 7528–7547, https://doi.org/10.1021/jacs.5b01156

23  Speciale, G., Thompson, A.J., Davies, G.J. and Williams, S.J. (2014) Dissecting conformational contributions to glycosidase catalysis and inhibition. *Curr. Opin. Struct. Biol.* **28**, 1–13, https://doi.org/10.1016/j.sbi.2014.06.003

24  Biarnés, X., Nieto, J., Planas, A. and Rovira, C. (2006) Substrate distortion in the Michaelis complex of Bacillus 1,3-1,4-beta-glucanase. Insight from first principles molecular dynamics simulations. *J. Biol. Chem.* **281**, 1432–1441, https://doi.org/10.1074/jbc.M507643200

25  Davies, G.J., Planas, A. and Rovira, C. (2012) Conformational analyses of the reaction coordinate of glycosidases. *Acc. Chem. Res.* **45**, 308–316, https://doi.org/10.1021/ar2001765

26  Biarnés, X., Ardèvol, A., Planas, A., Rovira, C., Laio, A. and Parrinello, M. (2007) The conformational free energy landscape of $\beta$-D-glucopyranose. Implications for substrate preactivation in $\beta$-glucoside hydrolases. *J. Am. Chem. Soc.* **129**, 10686–10693, https://doi.org/10.1021/ja068411o

27  Zechel, D.L. and Withers, S.G. (2000) Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc. Chem. Res.* **33**, 11–18, https://doi.org/10.1021/ar970172

28  Denekamp, C. and Sandlers, Y. (2005) Anomeric distinction and oxonium ion formation in acetylated glycosides. *J. Mass Spectrom.* **40**, 765–771, https://doi.org/10.1002/jms.848

29  Huang, M., Retailleau, P., Bohé, L.B. and Crich, D. (2012) Cation clock permits distinction between the mechanisms of $\alpha$-and $\beta$-O-and $\beta$-C-glycosylation in the mannopyranose series: evidence for the existence of a mannopyranosyl oxocarbenium ion. *J. Am. Chem. Soc.* **134**, 14746–14749, https://doi.org/10.1021/ja307266n

30  Martin, A., Arda, A., Désiré, J., Martin-Mingot, A., Probst, N., Sinaÿ, P. et al. (2015) Catching elusive glycosyl cations in a condensed phase with HF/SbF5 superacid. *Nat. Chem.* **8**, 1–6

31  Petersen, L., Ardévol, A., Rovira, C. and Reilly, P.J. (2010) Molecular mechanism of the glycosylation step catalyzed by Golgi $\alpha$-mannosidase II: a QM/MM metadynamics investigation. *J. Am. Chem. Soc.* **132**, 8291–8300, https://doi.org/10.1021/ja909249u

32  Gloster, T.M. and Vocadlo, D.J. (2012) Developing inhibitors of glycan processing enzymes as tools for enabling glycobiology. *Nat. Chem. Biol.* **8**, 683–694, https://doi.org/10.1038/nchembio.1029

33  Lairson, L.L., Henrissat, B., Davies, G.J. and Withers, S.G. (2008) Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77**, 521–555, https://doi.org/10.1146/annurev.biochem.76.061005.092322

34  Forrester, T.J.B., Ovchinnikova, O.G., Li, Z., Kitova, E.N., Nothof, J.T., Koizumi, A. et al. (2022) The retaining $\beta$-Kdo glycosyltransferase WbbB uses a double-displacement mechanism with an intermediate adduct rearrangement step. *Nat. Commun.* **13**, 1–13

35  Rojas-Cervellera, V., Ardèvol, A., Boero, M., Planas, A. and Rovira, C. (2013) Formation of a covalent glycosyl-enzyme species in a retaining glycosyltransferase. *Chem. Eur. J.* **19**, 14018–14023, https://doi.org/10.1002/chem.201302898

36  Gómez, H., Lluch, J.M. and Masgrau, L. (2013) Substrate-assisted and nucleophilically assisted catalysis in bovine $\alpha$1,3-galactosyltransferase. Mechanistic implications for retaining glycosyltransferases. *J. Am. Chem. Soc.* **135**, 7053–7063, https://doi.org/10.1021/ja4024447

37　Soya, N., Fang, Y., Palcic, M.M. and Klassen, J.S. (2011) Trapping and characterization of covalent intermediates of mutant retaining glycosyltransferases. *Glycobiology* **21**, 547–552, https://doi.org/10.1093/glycob/cwq190

38　Monegal, A. and Planas, A. (2006) Chemical rescue of α3-galactosyltransferase. Implications in the mechanism of retaining glycosyltransferases. *J. Am. Chem. Soc.* **128**, 16030–16031, https://doi.org/10.1021/ja0659931

39　Sinnott, M.L. and Jencks, W.P. (1980) Solvolysis of D-glucopyranosyl derivatives in mixtures of ethanol and 2,2,2-trifluoroethanol. *J. Am. Chem. Soc.* **102**, 2026–2032, https://doi.org/10.1021/ja00526a043

40　Ardèvol, A. and Rovira, C. (2011) The molecular mechanism of enzymatic glycosyl transfer with retention of configuration: evidence for a short-lived oxocarbenium-like species. *Angew. Chem., Int. Ed.* **50**, 10897–10901, https://doi.org/10.1002/anie.201104623

41　Piniello, B., Lira-Navarrete, E., Takeuchi, H., Takeuchi, M., Haltiwanger, R.S., Hurtado-Guerrero, R. et al. (2021) Asparagine tautomerization in glycosyltransferase catalysis. The molecular mechanism of protein O-fucosyltransferase 1. *ACS Catal.* **11**, 9926–9932, https://doi.org/10.1021/acscatal.1c01785

42　Ardèvol, A., Iglesias-Fernández, J., Rojas-Cervellera, V. and Rovira, C. (2016) The reaction mechanism of retaining glycosyltransferases. *Biochem. Soc. Trans.* **44**, 51–60, https://doi.org/10.1042/BST20150177

43　Kornberg, S., Zimmerman, S. and Kornberg, A. (1961) Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage-infected *Escherichia coli*. *J. Biol. Chem.* **236**, 1487–1493, https://doi.org/10.1016/S0021-9258(18)64202-4

44　Hu, Y., Cheng, K., He, L., Zhang, X., Jiang, B., Jiang, L. et al. (2021) NMR-based methods for protein analysis. *Anal. Chem.* **93**, 1866–1879, https://doi.org/10.1021/acs.analchem.0c03830

45　Danev, R., Yanagisawa, H. and Kikkawa, M. (2019) Cryo-electron microscopy methodology: current aspects and future directions. *Trends Biochem. Sci* **44**, 837–848, https://doi.org/10.1016/j.tibs.2019.04.008

46　Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R. et al. (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science (1979)* **373**, 871–876, https://doi.org/10.1126/science.abj8754

47　Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O. et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589, https://doi.org/10.1038/s41586-021-03819-2

48　Akdel, M., Pires, D.E.V., Pardo, E.P., Jänes, J., Zalevsky, A.O., Mészáros, B. et al. (2022) A structural biology community assessment of AlphaFold 2 applications. *Nat. Struct. Mol. Biol.* **29**, 1056–1067, https://doi.org/10.1038/s41594-022-00849-w

49　Wlodawer, A., Minor, W., Dauter, Z. and Jaskolski, M. (2008) Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures . *FEBS J.* **275**, 1–21, https://doi.org/10.1111/j.1742-4658.2007.06178.x

50　Afonine, P.v. and Adams, P.D. (2013) Crystallographic structure refinement in a nutshell. In *Advancing Methods for Biomolecular Crystallography. NATO Science for Peace and Security Series A: Chemistry and Biology* (Read, R., Urzhumtsev, A. and Lunin, V., eds), Springer, Dordrecht, https://doi.org/10.1007/978-94-007-6232-9˙20

51　Wlodawer, A., Minor, W., Dauter, Z. and Jaskolski, M. (2013) Protien crystallography for aspiring crystallographers. *FEBS J.* **280**, 5705–5736, https://doi.org/10.1111/febs.12495

52　Stollar, E.J. and Smith, D.P. (2020) Uncovering protein structure. *Essays Biochem.* **64**, 649–680, https://doi.org/10.1042/EBC20190042

53　Burley, S.K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G.v. et al. (2021) RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic. Acids. Res.* **49**, D437–D451, https://doi.org/10.1093/nar/gkaa1038

54　Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H. et al. (2000) The protein data bank. *Nucleic. Acids. Res.* **28**, 235–242, https://doi.org/10.1093/nar/28.1.235

55　Bhat, T.N., Bourne, P., Feng, Z., Gilliland, G., Jain, S., Ravichandran, V. et al. (2001) The PDB data uniformity project. *Nucleic. Acids. Res.* **29**, 214–218, https://doi.org/10.1093/nar/29.1.214

56　Sobala, L.F., Speciale, G., Zhu, S., Raich, L., Sannikova, N., Thompson, A.J. et al. (2020) An epoxide intermediate in glycosidase catalysis. *ACS Cent. Sci.* **6**, 760–770, https://doi.org/10.1021/acscentsci.0c00111

57　Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M. et al. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688, https://doi.org/10.1002/jcc.20290

58　Salomon-Ferrer, R., Case, D.A. and Walker, R.C. (2013) An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **3**, 198–210, https://doi.org/10.1002/wcms.1121

59　Phillips, J.C., Hardy, D.J., Maia, J.D.C., Stone, J.E., Ribeiro, J.v., Bernardi, R.C. et al. (2020) Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **153**, 044130, https://doi.org/10.1063/5.0014475

60　Páll, S., Zhmurov, A., Bauer, P., Abraham, M., Lundborg, M., Gray, A. et al. (2020) Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *J. Chem. Phys.* **153**, 134110, https://doi.org/10.1063/5.0018516

61　Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. et al. (2015) Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19–25, https://doi.org/10.1016/j.softx.2015.06.001

62　Berendsen, H.J.C., van der Spoel, D. and van Drunen, R. (1995) GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56, https://doi.org/10.1016/0010-4655(95)00042-E

63　Eastman, P., Swails, J., Chodera, J.D., McGibbon, R.T., Zhao, Y., Beauchamp, K.A. et al. (2017) OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, e1005659, https://doi.org/10.1371/journal.pcbi.1005659

64　Webb, B. and Sali, A. (2016) Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* **54**, 5.6.1–5.6.37, https://doi.org/10.1002/cpbi.3

65　Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I. et al. (2021) UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82, https://doi.org/10.1002/pro.3943

66  Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N. et al. (2018) MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315, https://doi.org/10.1002/pro.3330

67  Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, https://doi.org/10.1016/0263-7855(96)00018-5

68  Schrödinger, L.L.C. (2015) The PyMOL molecular graphics system.

69  Montgomery, A.P., Xiao, K., Wang, X., Skropeta, D. and Yu, H. (2017) *Computational Glycobiology: Mechanistic Studies of Carbohydrate-Active Enzymes and Implication for Inhibitor Design* , 1st ed.,, vol. **109**  Elsevier Inc

70  Olsson, M.H.M., Søndergaard, C.R., Rostkowski, M. and Jensen, J.H. (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pK a predictions. *J. Chem. Theory Computation 2* **7**, 525–537, https://doi.org/10.1021/ct100578z

71  Søndergaard, C.R., Olsson, M.H.M., Rostkowski, M. and Jensen, J.H. (2011) Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pK a values. *J. Chem. Theory Comput.* **7**, 2284–2295, https://doi.org/10.1021/ct200133y

72  Anandakrishnan, R., Aguilar, B. and Onufriev, A.V. (2012) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic. Acids. Res.* **40**, W537–W541, https://doi.org/10.1093/nar/gks375

73  Braun, E., Gilmer, J., Mayes, H.B., Mobley, D.L., Monroe, J.I., Prasad, S. et al. (2019) Best practices for foundations in molecular simulations. *Living J. Comput. Mol. Sci.* **1**, 1–28, https://doi.org/10.33011/livecoms.1.1.5957

74  Car, R. and Parrinello, M. (1985) Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **55**, 2471, https://doi.org/10.1103/PhysRevLett.55.2471

75  Marx, D. and Hutter, J. (2009) Ab initio molecular dynamics: basic theory and advanced methods. Cambridge University Press, Cambridge

76  Jensen, F. (2016) *Introduction to Computational Chemistry*, p. 664, John Wiley & Sons, New Jersey

77  Kohn, W. and Sham, L.J. (1964) Quantum density oscillations in an inhomogeneous electron gas. *Phys. Rev. Lett.* **137**, 1697–1705

78  Hohenberg, P. and Kohn, W. (1964) Inhomogeneous electron gas. *Phys. Rev. B.* **136**, 864–871, https://doi.org/10.1103/PhysRev.136.B864

79  CPMD program (2001) Copyright IBM Corp 1990-2008. *Copyright MPI für Festkörperforschung*, Stuttgart, https://github.com/CPMD-code [accessed 8 February 2023]

80  Kühne, T.D., Iannuzzi, M., Del Ben, M., Rybkin, V.V., Seewald, P., Stein, F. et al. (2020) CP2K: an electronic structure and molecular dynamics software package - Quickstep: efficient and accurate electronic structure calculations. *J. Chem. Phys.* **152**, 194103, https://doi.org/10.1063/5.0007045

81  Ahmadi, S., Barrios Herrera, L., Chehelamirani, M., Hostaš, J., Jalife, S. and Salahub, D.R. (2018) Multiscale modeling of enzymes: QM-cluster, QM/MM, and QM/MM/MD: a tutorial review. *Int. J. Quantum Chem.* **118**, 1–34, https://doi.org/10.1002/qua.25558

82  Raich, L., Nin-Hill, A., Ardèvol, A. and Rovira, C. (2016) Enzymatic cleavage of glycosidic bonds: strategies on how to set up and control a QM/MM metadynamics simulation. *Methods Enzymol.* **577**, 159–183, https://doi.org/10.1016/bs.mie.2016.05.015

83  Valsson, O., Tiwary, P. and Parrinello, M. (2016) Enhancing important fluctuations: rare events and metadynamics from a conceptual viewpoint. *Annu. Rev. Phys. Chem.* **67**, 159–184, https://doi.org/10.1146/annurev-physchem-040215-112229

84  Gullingsrud, J.R., Braun, R. and Schulten, K. (1999) Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations. *J. Comput. Phys.* **151**, 190–211, https://doi.org/10.1006/jcph.1999.6218

85  Kästner, J. (2011) Umbrella sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 932–942, https://doi.org/10.1002/wcms.66

86  Bonomi, M., Bussi, G., Camilloni, C., Tribello, G.A., Banáš, P., Barducci, A. et al. (2019) Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **16**, 670–673, https://doi.org/10.1038/s41592-019-0506-8

87  Tribello, G.A., Bonomi, M., Branduardi, D., Camilloni, C. and Bussi, G. (2014) PLUMED 2: new feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613, https://doi.org/10.1016/j.cpc.2013.09.018

88  Hénin, J., Lelièvre, T., Shirts, M.R., Valsson, O. and Delemotte, L. (2022) A LiveCoMS perpetual review enhanced sampling methods for molecular dynamics simulations. *Living J. Comput. Mol. Sci.* **4**, 1583, https://doi.org/10.33011/livecoms.4.1.1583

89  Laio, A. and Parrinello, M. (2002) Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562–12566, https://doi.org/10.1073/pnas.202427399

90  Bussi, G. and Laio, A. (2020) Using metadynamics to explore complex free-energy landscapes. *Nat. Rev. Phys.* **2**, 1–13, https://doi.org/10.1038/s42254-020-0153-0

91  Dama, J.F., Parrinello, M. and Voth, G.A. (2014) Well-tempered metadynamics converges asymptotically. *Phys. Rev. Lett.* **112**, 240602, https://doi.org/10.1103/PhysRevLett.112.240602

92  Ensing, B., Laio, A., Parrinello, M. and Klein, M.L. (2005) A recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions. *J. Phys. Chem. B* **109**, 6676–6687, https://doi.org/10.1021/jp045571i

93  Coines, J., Raich, L. and Rovira, C. (2019) Modeling catalytic reaction mechanisms in glycoside hydrolases. *Curr. Opin. Chem. Biol.* **53**, 183–191, https://doi.org/10.1016/j.cbpa.2019.09.007

94  Mendoza, F. and Masgrau, L. (2021) Computational modeling of carbohydrate processing enzymes reactions. *Curr. Opin. Chem. Biol.* **61**, 203–213, https://doi.org/10.1016/j.cbpa.2021.02.012

95  Coines, J., Cuxart, I., Teze, D. and Rovira, C. (2022) Computer simulation to rationalize "rational" engineering of glycoside hydrolases and glycosyltransferases. *J. Phys. Chem. B* **126**, 802–812, https://doi.org/10.1021/acs.jpcb.1c09536

96  Morais, M.A.B., Nin-Hill, A. and Rovira, C. (2022) Glycosidase mechanisms: sugar conformations and reactivity in endo and exo enzymes. *Curr. Opin. Chem. Biol.*, (in press)

97  Graziano, A.C.E. and Cardile, V. (2015) History, genetic, and recent advances on Krabbe disease. *Gene* **555**, 2–13, https://doi.org/10.1016/j.gene.2014.09.046

98  Hill, C.H., Graham, S.C., Read, R.J. and Deane, J.E. (2013) Structural snapshots illustrate the catalytic cycle of β-galactocerebrosidase, the defective enzyme in Krabbe disease. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20479–20484, https://doi.org/10.1073/pnas.1311990110

99   Nin-Hill, A. and Rovira, C. (2020) The catalytic reaction mechanism of the β-galactocerebrosidase enzyme deficient in Krabbe disease. *ACS Catal* **10**, 12091–12097, https://doi.org/10.1021/acscatal.0c02609

100  Croci, D.O., Cerliani, J.P., Dalotto-Moreno, T., Méndez-Huergo, S.P., Mascanfroni, I.D., Dergan-Dylon, S. et al. (2014) Glycosylation-dependent lectin-receptor interactions preserve angiogenesis in anti-VEGF refractory tumors. *Cell* **156**, 744–758, https://doi.org/10.1016/j.cell.2014.01.043

101  Partridge, E.A., le Roy, C., di Guglielmo, G.M., Pawling, J., Cheung, P., Granovsky, M. et al. (2004) Regulation of cytokine receptors by golgi N-glycan processing and endocytosis. *Science (1979)* **306**, 120–124, https://doi.org/10.1126/science.1102109

102  Lau, K.S., Partridge, E.A., Grigorian, A., Silvescu, C.I., Reinhold, V.N., Demetriou, M. et al. (2007) Complex N-glycan number and degree of branching cooperate to regulate cell proliferation and differentiation. *Cell* **129**, 123–134, https://doi.org/10.1016/j.cell.2007.01.049

103  Nabi, I.R., Shankar, J. and Dennis, J.W. (2015) The galectin lattice at a glance. *J. Cell Sci.* **128**, 2213–2219, https://doi.org/10.1242/jcs.151159

104  Hassani, Z., Saleh, A., Turpault, S., Khiati, S., Morelle, W., Vignon, J. et al. (2017) Phostine PST3.1a targets MGAT5 and inhibits glioblastoma-initiating cell invasiveness and proliferation. *Mol. Cancer Res.* **15**, 1376–1387, https://doi.org/10.1158/1541-7786.MCR-17-0120

105  Hanashima, S., Inamori, K.I., Manabe, S., Taniguchi, N. and Ito, Y. (2006) Systematic synthesis of bisubstrate-type inhibitors of N-acetylglucosaminyltransferases. *Chem. Eur. J.* **12**, 3449–3462, https://doi.org/10.1002/chem.200501348

106  Brockhausen, I., Reck, F., Kuhns, W., Khan, S., Matta, K.L., Meinjohanns, E. et al. (1995) Substrate specificity and inhibition of UDP-GlcNAc:GlcNAcβ1-2Manα1-6R β1,6-N-acetylglucosaminyltransferase V using synthetic substrate analogues. *Glycoconj. J.* **12**, 371–379, https://doi.org/10.1007/BF00731340

107  Darby, J.F., Gilio, A.K., Piniello, B., Roth, C., Blagova, E., Hubbard, R.E. et al. (2020) Substrate engagement and catalytic mechanisms of N-acetylglucosaminyltransferase . *ACS Catal* **10**, 8590–8596, https://doi.org/10.1021/acscatal.0c02222

108  Okajima, T., Xu, A., Lei, L. and Irvine, K.D. (2005) Chaperone activity of protein O-fucosyltransferase 1 promotes notch receptor folding. *Science (1979)* **307**, 1599–1603, https://doi.org/10.1126/science.1108995

109  Bray, S.J. (2006) Notch signalling: a simple pathway becomes complex. *Nat. Rev. Mol. Cell Biol.* **7**, 678–689, https://doi.org/10.1038/nrm2009

110  McMillan, B.J., Zimmerman, B., Egan, E.D., Lofgren, M., Xu, X., Hesser, A. et al. (2017) Structure of human POFUT1, its requirement in ligand-independent oncogenic Notch signaling, and functional effects of Dowling-Degos mutations. *Glycobiology* **27**, 777–786, https://doi.org/10.1093/glycob/cwx020

111  Jundt, F., Schwarzer, R. and Dorken, B. (2008) Notch signaling in leukemias and lymphomas. *Curr. Mol. Med.* **8**, 51–59, https://doi.org/10.2174/156652408783565540

112  Du, Y., Li, D., Li, N., Su, C., Yang, C., Lin, C. et al. (2018) POFUT1 promotes colorectal cancer development through the activation of Notch1 signaling. *Cell Death Dis.* **9**, 1–12, https://doi.org/10.1038/s41419-018-1055-2

113  Bao, Y.O., Zhang, M., Qiao, X. and Ye, M. (2022) Functional characterization of a C-glycosyltransferase from Pueraria lobata with dual-substrate selectivity. *Chem. Commun.* **58**, 12337–12340, https://doi.org/10.1039/D2CC04279G