Tutor/s

Dra. Carme Rovira Virgili
Departament de Química Inorgànica i Orgànica

Grau de Química

Treball Final de Grau

Computational study of the active center of an endo-β-1,4-mannosidase, a carbohydrate-degrading enzyme.
Estudi computacional del centre actiu d'una endo-β-1,4-mannosidase, un enzim que degrada carbohidrats.

Ricard Silvestre Barceló

June 2020

UNIVERSITAT DE BARCELONA

B:KC Barcelona Knowledge Campus
Campus d'Excel·lència Internacional

CONTENTS

REPORT

1. SUMMARY

Glycoside hydrolases, also named glycosidases (GHs), are enzymes responsible for processing carbohydrates by hydrolyzing their glycosidic bond linkages. These enzymes operate either by retention or inversion of the configuration of the anomeric carbon. Generally speaking, two carboxylate-based protein residues, which perform the functions of acid/base and nucleophilic catalysts, are directly involved in the reaction. In this project, the pKa of the catalytic residues will be calculated by means of different computational methods and the active site of a bacterial *endo*-β-1,4-mannanase of GH family 134 will be studied also by computer simulation to understand how the protonation state of these two carboxylate residues can impact the structure and dynamics of the enzyme. In addition, the configuration of the cleaved sugar ring during the simulation will be mapped.

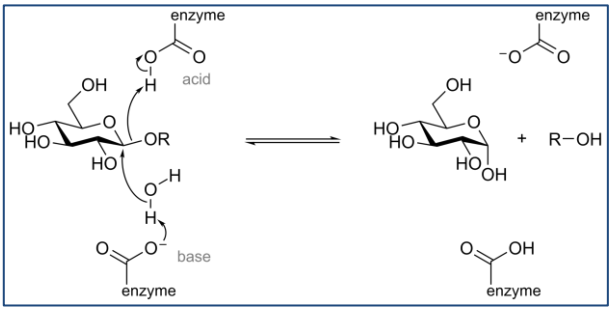


Figure 1 Mechanism for an inverting GH
Image courtesy of Wikimedia Commons

Keywords: Computational study, Glycosyl hydrolases, Sugar conformation, Carbohydrates

2. RESUM

Les glucosidases (GH's) són enzims responsables de la hidròlisis i el processament dels carbohidrats. Aquests enzims trenquen l'enllaç glicosídic dels sucres i poden funcionar amb retenció o inversió de la configuració del carboni anomèric. Els residus involucrats en aquesta reacció són, generalment, dos residus carboxílics, els quals realitzen la funció de catàlisi àcid/base i catàlisi nucleofílica cadascun. En aquest treball es realitzarà primer el càlcul del pKa dels residus catalítics a partir de diferents mètodes computacionals i després s'analitzarà un enzim *endo*-β-1,4-manosidasa bacterià, de la família GH134, a partir de simulacions computacionals. L'objectiu serà el de comprovar com l'estat de protonació dels residus catalítics pot afectar a l'estructura i el comportament del seu lloc actiu. Complementàriament es farà un seguiment de la conformació de l'anell de sucre que pateix la hidròlisis i es representarà en un diagrama de Stoddart.

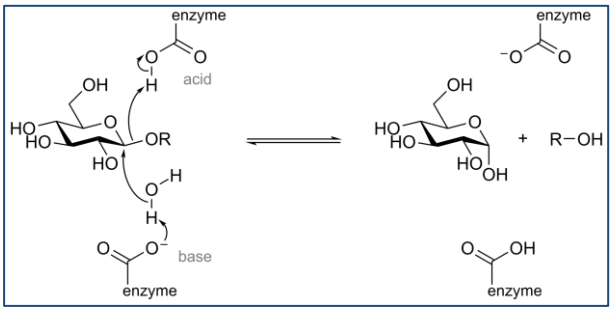


Figura 1 Mecanismes de reacció d'una GH d'inversió
Imatge prestada de Wikimedia Commons

Paraules clau: Estudi computacional, Glucosidases, Conformacions de monosacàrids, Carbohidrats

3. INTRODUCTION

3.1. GLYCOSIDE HYDROLASES

Glycoside hydrolases (GHs) are highly specific enzymes responsible for the hydrolysis of glycosidic bonds in carbohydrates. The knowledge of their enzymatic mechanism at the molecular level is crucial to understand how carbohydrates are processed in the organisms, as well as developing new drugs[1]. Despite the large number of GH's known, classified into more than 150 families[2][3][4], their catalytic mechanism is similar: they typically operate by means of a combination of nucleophilic and acid/base catalysis.

3.1.1. GH's mechanisms

The catalytic mechanisms of GHs can generally be classified in two types, depending on whether they retain or invert the anomeric carbon configuration upon the hydrolysis reaction. (figure 2).

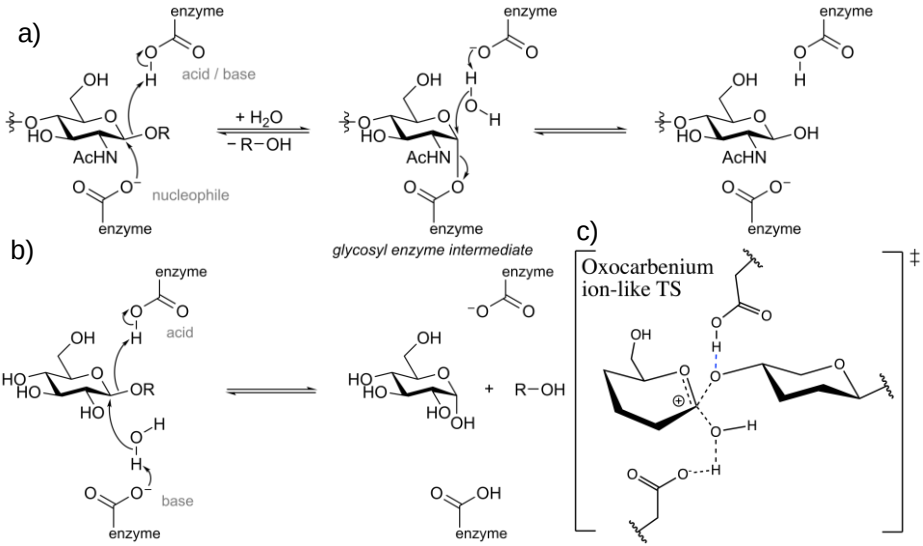


Figure 2 Different mechanisms for GH's: a)Retaining; b)Inverting; c) Transition state
Image courtesy of Wikimedia Commons

As seen in figure 2a, inverting GH's operate by a single nucleophilic substitution, where a carboxylate residue (the general base catalyst) activates the water molecule which attacks the anomeric carbon; in coordination, the acid/base catalyst residue donates a proton to the leaving sugar in order to ease the nucleophilic substitution. In the case of retaining mechanism, it consists of a double S_N2 with an enzyme-substrate covalent intermediate (figure 2b). In addition, some mechanisms of retaining GH's have been reported in which the nucleophile that attacks the anomeric carbon in the first place is a neighbor substituent of the same sugar, e.g. N-acetamide substituents or even one of its hydroxyls to form an epoxide[5][6]. Because of the function each catalytic residue performs, it is obvious that the one acting as the acid catalyst must be protonated when the reaction occurs; and the general base or nucleophilic residue must be deprotonated. Also, regardless of the mechanism, each nucleophilic substitution involves an oxocarbenium ion-like transition state (TS)(figure 2c) as evidenced by kinetic measurements of the isotope effect[7]. The oxocarbenium TS has a sp2 hybridization and development of positive charge at the anomeric carbon, which is partially stabilized by electron donation from the ring oxygen. In this structure, the sugar ring is distorted from the relaxed $^4C_1$ conformation of a pyranose in solution because the sp2 hybridization requires the C2, C1, O5 and C5 atoms to be as coplanar as possible.

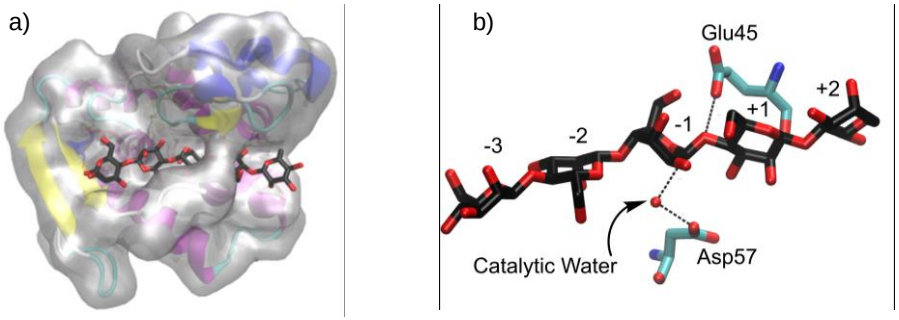


Figure 3 a) Michaelis Complex f SsGH134 and 5-Mannan; b) Catalytic residues in SsGH134

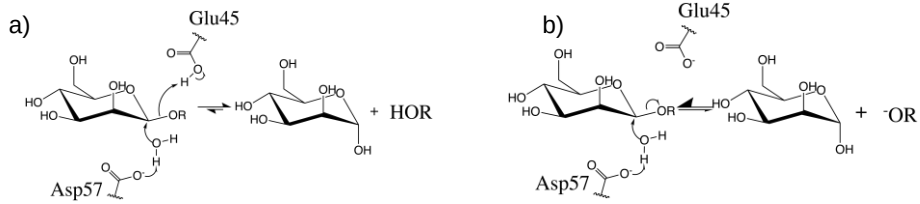


Figure 4 Mechanism for the SsGH134
a) correct protonation state; b) wrong protonation state

3.1.2. Effect of the protonation state

The glycosyl hydrolase that will be the object of this study is a bacterial inverting *endo*-β-1,4-mannanase: *Ss*GH134 from *Streptomyces* sp, GH family 134[8]. Structural analysis of *Ss*GH134 suggested roles for Glu45 as a general acid catalyst and for Asp57 as a general base catalyst. These roles were supported by mutagenesis with the corresponding Glu45Ala and Asp57Ala single mutants inactive on β-mannopentaose and hexose[9]. However, although the pKa values for the catalytic residues can be determined experimentally[10] it is rarely done. Usually, the protonation state of each catalytic residue is determined by a qualitative visual inspection of the surroundings of the active site and the orientation of the substrate in the crystallographic structure. Nonetheless, some computational methods for the calculation of the pKa of the titratable residues of an enzyme are available in the internet in the form of computational servers[11][12][13]. These servers, which are widely used in computational biology, will be used in this work to test their applicability and reliability to predict the function of the two catalytic residues in SsGH134. Additionally, we will test the consequences of a wrong assignation of the protonation state of the residue Glu45 (general acid catalyst) in a molecular dynamics simulation. Presumably, because the proton in the residue acting as the acid catalyst is needed, the deprotonation of Glu45 should signify an impossibility for the reaction to happen. This is so because the oxygen of the glycosidic bond that is cleaved needs to protonate during the reaction, otherwise it will represent a very poor leaving group and therefore the reaction will not happen (figure 4).

3.1.3. Sugar configurations

3.1.3.1. Ring distortions

It was found by X-ray[14] and confirmed by NMR[15] studies in the 1990s that the substrate binds to GHs in a distorted conformation, usually a different conformation from the relaxed $^4C_1$. This is, in fact, because the distortion aims for a configuration leading to that of the transition state. In the reactive conformation, aglycon (i.e., the leaving group) is pushed into a pseudo axial position. This causes the glycosidic oxygen to approach the acid/base residue, facilitating the proton transfer and thus increasing its ease to leave; the distortion also opens the space to permit a better

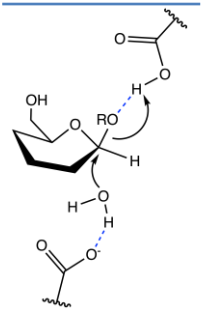


Figure 5 Distorted ring facilitating hydrolisis

nucleophilic attack on the anomeric carbon and modifies some bond distances in a way that favors the cleavage (figure 5)[16][17]. Not only does the distortion happen before the reaction but also during the reaction (TS). The reasons for the change in conformation in the TS have been discussed previously, arguing that is the sp2 hybridization of the anomeric carbon which forces the entire ring to a planarity among C2, C1, O5 and C5, increases the partial charge of the anomeric carbon and makes some bonds change in length. Also, it is usual to find the product of the hydrolysis in a different configuration from the most stable chair configurations, usually differing into a skew-boat configuration.

3.1.3.2. Puckering coordinates and Stoddart and Mercator representation

Pyranoses adopt 38 main conformations that can be classified into chair, half-chair, boat, skew-boat, and envelope conformations. These conformations feature four atoms at a shared plane and the other two atoms out of the plane; this is so for all conformations except for the envelope, which have only one out-of-plane atom. All the canonical conformations can be specified with a letter denoting which conformation the ring is in. Two indexes (one for envelope) denote the atoms out of plane and the direction in which they are out of the plane. The different letters are C for chair H for half-chair, B for boat, S for skew-boat and E for envelope and each configuration has a different number of non-equivalent configurations (for example a $^{4}C_{1}$ configuration is equivalent to $^{O}C_{3}$ or $^{2}C_{5}$ and thus the only different configurations are $^{4}C_{1}$ and $^{1}C_{4}$). All the possible configurations are: $^{4}C_{1}$, $^{1}C_{4}$, $^{3,0}B$, $B_{1,4}$, $^{2,5}B$, $B_{3,0}$, $^{1,4}B$, $B_{2,5}$, $^{O}H_{5}$, OH_{1}, $^{2}H_{1}$, $^{2}H_{3}$, $^{4}H_{3}$, $^{4}H_{5}$, $^{3}H_{2}$, $^{1}H_{O}$, $^{5}H_{4}$, $^{5}H_{O}$, $^{3}H_{4}$, $^{1}H_{2}$, ^{1}E, ^{2}E, ^{3}E, ^{4}E, ^{5}E, ^{0}E, $^{3}S_{1}$, $^{5}S_{1}$, $^{2}S_{O}$,$^{1}S_{3}$, $^{1}S_{5}$, and $^{O}S_{2}$.

Mapping all these conformations in a single diagram in order to be able to represent the conformational itineraries or the substrate distortions on GH's is therefore very important, for it permits an easy-following discussion of these changes. As so, two main ways of representing them have been widely used in GH and related research papers: The Stoddart representation and the Mercator representation of the Cremer-Pople puckering coordinates[18][19].

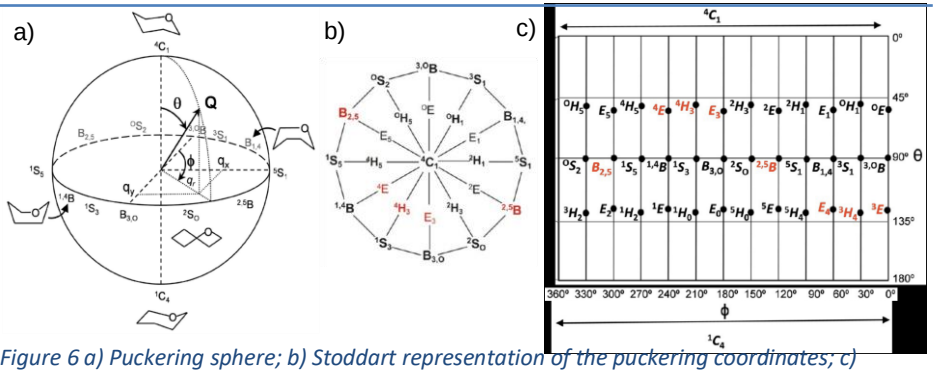


Figure 6 a) Puckering sphere; b) Stoddart representation of the puckering coordinates; c) Mercator representation

The Stoddart representation (figure 6b) is a diagram that James Fraser Stoddart (Nobel Prize in Chemistry 2016 for the design and synthesis of molecular machines) introduced in 1971. The Stoddart representation integrates all possible conformations of a pyranose ring together with their inner relations and has become very popular for representing conformational itineraries of carbohydrates during catalysis. Because the Stoddart representation is the circular projection of the sphere in which all conformation can be put, there are two Stoddart representations, one for the north pole and one for the south pole. Although the Stoddart diagram is just a diagram and cannot show the relative energy of each conformation, theoretical simulations can be used to obtain this information.

Along with the Stoddart diagram, the Mercator representation (figure 6c) is also used to map and relate all the possible conformations of a pyranose ring. In the case of the Mercator representation it is the rectangular projection of the whole sphere before mentioned and thus, there is only one possible diagram. Again, theoretical simulations can be used to obtain the relative energy of each conformation in the Mercator diagram.

The sphere from which the Stoddart and Mercator diagrams are projected can be described mathematically so that, assigning three coordinates to each ring conformation, which are functions of the distances and angles of the atoms and bonds of the sugar ring, all possible structures lay within this sphere. This general mathematical definition is what Dieter Cremer and John Anthony Pople introduced in 1975, naming them 'Puckering coordinates'[19].

The puckering coordinates constitute a polar coordinate system in which the three coordinates Q, φ, and θ are calculated solving the following system of equations:

$$\begin{cases} Q \sin\theta \cos\phi = \sqrt{\frac{1}{3}} \sum_{j=1}^{6} z_j \cos\left[\frac{2\pi}{6} 2(j-1)\right] \\ Q \sin\theta \sin\phi = \sqrt{\frac{1}{3}} \sum_{j=1}^{6} z_j \sin\left[\frac{2\pi}{6} 2(j-1)\right] \\ Q \cos\theta = \sqrt{\frac{1}{6}} \sum_{j=1}^{6} (-1)^{j-1} z_j \end{cases}$$

In this equations j designates each ring atom and z_j is the perpendicular distance of each atom to the ring average plane. Also, the coordinate Q is expressed as the sum of all the z_j: $Q = \sum_{j=1}^{6} z_j$. With these coordinates, any ring conformation would fall within the puckering sphere (figure 6a), where the two chair conformers are located on the poles.

3.2. MOLECULAR SIMULATIONS

The basic idea of any molecular simulation method is straightforward; a particle-based description of the system under investigation is constructed and then the system is propagated by either deterministic or probabilistic rules to generate a trajectory describing its evolution over the course of the simulation[22]. If the probabilistic method is chosen then it is called a Monte Carlo (MC) simulation, in which, by probabilistic algorithms, new configurations of a system are generated. Those configurations do not represent a dynamical and ordered set of configurations (this is, no configuration sampled has any relation with the previous or next ones), instead, they are a collection of configurations that could be obtained dynamically. In the case of the molecular dynamics method (MD), which is the one used in this work, the equations of motion are numerically integrated to generate a dynamical trajectory of the system's motion. In classical MD. Either of the methods (MC or MD) can also be used with a quantum mechanical approach. In this case, the energy is computed from first principles (the wavefunctions of each atom are calculated and the electronic interactions are computed) instead of using an empirical force-field.

The time and computational resources required to perform a QM simulation is much larger than that of a MM simulation and that is why a MM simulation (classical MD) is done in this work. It should be noted as well that, for large systems like GH's a QM simulation would be unfeasible. But, as the most important part of a GH is its active site, a hybrid simulation (quantum mechanics/molecular mechanics, QM/MM) can be performed. In this type of

simulation, the most important atoms of the active site are treated with quantum mechanics and the rest of the system is treated classically[20].

3.2.1. Force fields

The forces that determine the system dynamics are calculated employing what are called 'force fields': force fields are the functional forms and parameter sets needed to calculate the potential energy of a system of atoms. The terms included in the functional forms range from Coulombic interactions to non-bonded Lennard-Jones interactions and bonds, angles, and torsional (dihedral) terms[22].

3.2.2. Periodic Boundary conditions (PBC)

Real life systems are, obviously, finite. However, in the scale of a protein, even for the smallest real systems, it can be assumed that they are infinite. However simulating an infinite system or one of real dimensions is obviously utterly unfeasible and simulating only one would mean that the particles on the edges of the solvating box are exposed to the void. Because of this, a single enzyme is modelled and periodic boundary conditions are applied to them. PBC's periodically repeat the system in every direction allowing more accurate estimations of bulk properties and removing the unreal interactions that the atoms on the edges would experiment when exposed to the void. In PBC's each particle interacts with periodic images of particles in the same system and, as it is obvious, it is not computationally possible for a single particle to interact infinitely with other particles. For this reason, a cut-off length of many non-bonded interactions (excluding electrostatic) is chosen that is, at least, less than half the length of the simulation box in any dimension[22].

3.2.3. Steps in the MD simulation[22]

While each system studied may have its particularities, the protocol followed in the MD simulations performed in this work will be discussed.

3.2.3.1. System preparation

The system preparation focuses on preparing the starting state of the system in a correct form so that it serves as an input for an appropriate simulation. In our case, the preparation consisted in: first a manual adaptation of the PDB file of the SsGH134 enzyme in complex with a pentaasaccharide substrate, then the force field for proteins was chosen for both the enzyme,

the carbohydrate substrate and the solvent water molecules. The whole system was then solvated and counterions were added to neutralize the total change of the system.

3.2.3.2. Energy minimization

The purpose of the minimization step is to find a local energy minimum, with the structure close to the starting one, so that the system does not immediately "blow up", i.e., the forces on any one atom are not so large that the atoms move an unreasonable distance in a single timestep.

3.2.3.3. Assignment of velocities

Because during the minimization the output only gives the positions of the atoms and not their velocities, for the posterior simulation it is necessary to assign those starting velocities. It must be noted that, although the individual atoms are moving, the momentum of the center-of-mass of the system must remain 0 because a value different than zero would mean that the simulation box would move. Those velocities are given in the heating stage that will be discussed next. The assignation of the velocities is random, and it is usually non important because the Maxwell-Boltzmann distribution will quickly arise from the equations of motion.

3.2.3.4. Equilibration

The ultimate objective for the simulation is to collect data of the behavior of our system in a particular state. Thus, before that data is collected, the system must be driven to the desired conditions (i.e. temperature, pressure, density and volume) and some relaxation must be done to avoid metastable states. That is what the equilibration stage consists on.

In our case the equilibration featured a heating of the system (done with a thermostat, which will be discussed later), followed by a run at constant pressure to reach the desired pressure (done with a barostat, which will be discussed later) and a final run at constant volume to ensure the equilibration of the system.

For the equilibration to be complete, it must be ensured that none of the properties of the system (energy and structure) changes systematically. At equilibrium, a system may still undergo slow fluctuations but key properties should not show systematic trends away from their starting structure. Usually, a good way to know that our system is equilibrated is to check that the time evolution of the root mean square displacement (RMSD) reaches a plateau.

3.2.3.5. Production

Once the equilibration is completed, the system is prepared for the data collection. This step is the production. The production step is the longest one and for the production to be considered finished, the properties that are being measured must converge into a stable value, this is, it should be ensured that properties no longer depend on the length of the simulation.

3.2.4. Thermostats

In molecular dynamics, thermostats seek to bring a simulation to a desired temperature and maintain it during the rest of the simulation, just as a real system would do.

The temperature of a molecular dynamics simulation is typically measured using the kinetic energies of their molecules as formulated by the equipartition theorem. The temperature is defined as a time-averaged quantity; therefore, it may fluctuate between the target value. Thermostats work by altering the Newtonian equations of motion and therefore modifying the forces acting on the atoms of the system. Also, the algorithms of the thermostats may either be deterministic or stochastic. Although the thermostats alter the Newtonian dynamics, some have been found to have little effect on the calculation of particular dynamical properties[23].

3.2.5 .Barostats

Typically, thermodynamic properties of interest are measured under open-air conditions, meaning that they are measured at essentially constant temperature and pressure. The constant temperature can be achieved with the thermostat and, analogously, the constant pressure is achieved with a barostat. Much of the working of barostats is analogous to thermostats and the pressure of a molecular dynamics simulation is commonly measured using the virial theorem. Again, as the thermostats with the temperature, the pressure is defined as a time-averaged quantity and it in fact fluctuates over time.

3.3. pKa CALCULATION SERVERS

In this section the methods that each pKa computing server use will be briefly explained.

3.3.1. The H++ server

The approach that the H++ server utilizes is based on classical continuum electrostatics and basic statistical mechanics. As such, it contains several approximations.

The continuum electrostatics methodology has been widely used to calculate the proton transfer energetics and it is based in the following premises: Firstly, a protonation is represented as the addition of a unit charge to a point in the titrating site. It is assumed that the protonation will only have significant contributions in the electrostatic interactions and, therefore, the essential step in determining a pKa is the calculation of the electrostatic work of charging a site in the protein from zero to one unit (positive or negative) of electronic charge. It also relies in the assumption that polarization effects are represented by continuum dielectric constants, treating the protein as a continuous medium. This approximation is called the generalized Born model[24].

3.3.2. The PROPKA server

The latest version of PROPKA, which is PROPKA3, calculates the pKa of the titratable residues with an empirical method that relies in a parametrization of the pKa values for those same residues in known conditions and samples, and utilizes those parameters to calculate the pKa of the specific residue in the given conditions. PROPKA3 also takes into account how deeply a residue is buried in the protein for the calculation of its pKa[25].

3.3.3. The Rosetta server

In Rosetta, the pH is titrated from 1 to 14 with Monte Carlo sampling of protonated and deprotonated side chains. The samples given by the MC method are evaluated with the Rosetta score function and accepted or rejected following the Metropolis criterion[26].

4. OBJECTIVES

In the first place the pKa values of the catalytic residues will be calculated using the three computational methods previously mentioned. In doing so, the applicability and reliability of those methods will be tested, because, although the exact values of pKa of each catalytic residue are not known, their role in the mechanism is known and thus it is expected that the acid/base residue remains protonated and the general base deprotonated before the reaction. Also, a comparison between the pKa values given for the catalytic residues and those of the

residues of the same type but exposed to the solvent will be done. This comparison aims to quantify in some way how affected the acidity of the catalytic residues is. Additionally, a study of the chemical environment of the active site will be performed in order to logically support the chosen pKa value.

The dynamics of the GH134 Michaelis complex were already investigated in a previous study of the group of Prof. Rovira[9]. Here, as a second objective, we will investigate the system dynamics when a different protonation state of one of the two catalytic residues is chosen. Specifically, the residue E45, which acts as the acid/base catalyst, will be forced into a deprotonated state. It is expected that the system will evolve towards an unreactive conformation of the active site, but we do not know a priori.

The simulated system will include the catalytic water of the active site featured in the crystallization data; alternatively, another simulation will be done in which the catalytic water is removed from the catalytic site, checking whether a water molecule from the bulk is able to enter the catalytic site. A conformational analysis of the reactive sugar will be performed in all simulations.

5. METHODS

5.1. WORKING STRUCTURES

All the structures used in this work are listed below, together with their PDB codes, and it is discussed how they differ from one another and the different ligands they include.

Some of these working structures (5jts.pdb and 5ju9.pdb) have been directly extracted from the PDB. These structures do not contain hydrogen atoms and have crystallization waters (only the oxygen atom is present).

- **5jts.pdb** $\rightarrow$ SsGH134 + 3 glycerol + 1 ethylene glycol + 1 Cl^-. Named **WT** (wild type enzyme)

This structure does not contain the sugar to be cleaved, thus having an arguably different environment in the catalytic site. It contains some crystallization molecules such as glycerol or ethylene glycol.

- **5ju9.pdb** → SsGH134 + 3-Mannan + 2 Cl^-. Named **WTP** (wild type + product)

  In this case, the product of the cleavage of a 5-Mannan is featured, a 3-Mannan.

- **5jug_nomut.pdb** → SsGH134 + 5-Mannan + 1 glycerol + 1 Cl^-. Named **MC1** (Michaelis Complex)

  This structure corresponds to an original structure (PDB code 5jug) of the sugar in complex with a mutant form of the enzyme (E45Q). This mutation, necessary for the structural determination of the Michaelis complex, was reverted during the preparation of the system for MD. This structure corresponds to the Michaelis complex, this is, the SsGH134 and the substrate it hydrolyses (5-Mannan), as well as some crystallization molecules.

The following structures correspond to other processed files of the system, necessary for the simulations, in some of which the hydrogen atoms have been added or the water molecules removed. Also, some residues have been protonated so that the computational servers do not try to calculate their pKa. Specifically, the protonation states specified are those of the residues Cys 8 and 14 and it was done so because they form a disulfide bond. These Cys residues were named as CYX during system preparation.

- **man_a.pdb** → SsGH134 with specified protonation states (cysteine residues 8 and 14) + 5-Mannan + solvating waters + catalytic water. Named **MC2** (Michaelis Complex).

  As it can be seen, the *man_a.pdb* structure does not yet contain hydrogens and it features the specified protonation states of the disulfide bond and all the solvating waters including the one in the catalytic site, apart, obviously, from the substrate.

- **gh134.pdb** → SsGH134 with specified protonation states (Cys 8 and 14) + 5-Mannan + catalytic water (Residues shifted by -7). Named **MC3** (Michaelis complex)

  In the case of *gh134.pdb* the solvation waters are removed and only the catalytic water remains. It is noted that the residue numbers are shifted 7 numbers. This is so because the original PDB files from the protein data bank do not include the residues

from 1 to 7 and the files start in the residue number 8, however in the gh134.pdb the residue numbers start in 1 so all the residues are the same but are shifted by 7 units. This is the effect of some processing software.

- **man134.pdb** → Used as a replacement of gh134.pdb for compatibility purposes. (Residues not shifted)

  This final structure was used as a replacement of *gh134.pdb* for compatibility purposes, the modification applied were the elimination of the naming CYX and substitution for CYS and the substitution of the name WAT for HOH for the catalytic water.

5.2. pKA CALCULATION SERVERS

The H++ and PROPKA servers calculated the pKa of the ionizable residues for a chosen value of the pH of the medium. The pH used as an input was 5, which is the optimal pH reported for the activity of the SsGH134[9]. In the case of Rosetta, no inputs were needed.

Also, in the H++ server, an input for the salinity of the medium, the internal and external dielectric constant was needed. The inputs selected were the standard values given by the same method, which are a salinity of 0.15M, an internal dielectric constant of 10 and an external dielectric constant of 80 (the dielectric constant has no units).

5.3. MOLECULAR SIMULATIONS

5.3.1. System preparation

5.3.1.1. Structure processing

The system preparation of our GH was done from the structure 5jug_nomut.pdb described above.

The processes followed to prepare the structure for the simulation were the following. First, the header of the original PDB file where information about the crystallographic data is given was removed. Then, form each line in the PDB of the positions of the atoms, other lines starting with the word ANISOU were removed. Also, the PDB file features different possible conformations for some residues, naming them conformation A, B, etc. Each of those conformations has a probability ranging from 0 to 1 to be present in the actual molecule but for

the dynamics only one conformation has to be given. Thus, the conformations of less probability were eliminated as well. Finally, the residue names of the CYS 8 and 14 were changed to CYX. Additionally, we have simulated two different systems: one in which the catalytic water and all crystal water molecules are removed and one in which they are not.

5.3.1.2. Force fields, solvation box and counterions

Once the PDB file was prepared, it was given as the input of a software which loads the corresponding force fields and prepares the solvation box. In the AMBER Molecular Dynamics Package, the name of this software is LeaP.

The force field used for the enzyme atoms is called 'ff14SB' (the most recent Amber force field for proteins) and the one corresponding to the sugar atoms is 'GLYCAM_06j-1'. The water model chosen for the simulations is the so called 'TIP3P' model.

After the force fields are set, the PDB is sourced and the bond between Cys8 and Cys14 is defined. Finally, a solvating box of 15 Angstrom is added and it is checked whether some counterions are necessary for the system to be neutral in charge. In our case, two Na^+ ions were added. To end up, the files corresponding to the atom parameters and coordinates of the system and the PDB featuring the solvating waters are saved.

5.3.2. Molecular dynamics simulations

5.3.2.1. Energy minimization

The minimizations of the energy of the systems were done in two steps. The first minimization was done restraining the movement of the protein, this is, the energy of the system is decreased but the movement of the Michaelis complex is restrained; the second one was a global minimization. For both minimizations, the algorithm of the steepest descent (SD) was used, in which the atomic displacements at the beginning of the simulations are forced to be quicker (i.e., the molecule is forced to a lower energy in bigger intervals of energy) and after a number of cycles is reached the displacements become smaller.

For the restrained minimization the number of SD steps was 5000 and the total number of steps 10000. The initial step length was 0.01 and the convergence value of difference in energy to assume a completely optimized conformation was 0.0001. Constant volume PBC's were applied and a non-bonded cut-off of 12 Angstrom was set. Also, the value for the restraint constant of the enzyme was set to 500.

In the case of the global minimization, the steps of SD minimization were 10000 and the number of total steps was 20000. The initial step length and the convergence value was the same as for the previous minimization as well as the cut-off length, and the PBC.

5.3.2.2. Heating

The heating stage of the two different systems was again equal for each system as well as all the following stages. It was divided into 4 steps, the first of which consisted on a heating from 0K to 100K of exclusively the solvent. The third remaining steps consisted on a global heating from 0 to 300K in intervals of 100K each.

During the heating, velocities are applied to the atoms. Those velocities are applied according to the reference temperature of the thermostat. Because the reference temperature is the temperature that is wanted to achieve through the heating step, its value was 100K, the temperature corresponding to the first heating. Again, as in the first minimization, the restraining constant for the protein and the sugar was of 500. The non-bonded cut-off length was 12 Angstrom and constant volume PBC were applied in all of the heating steps. The number of MD steps for the first heating was 50000 of a timestep of 0.001ps. A weak-coupling thermostat was applied with a time constant for the bath coupling of 1.

The weak-coupling thermostat, also called the Berendsen[27] thermostat, was used. For the understanding of the Berendsen thermostat it must be recalled how the temperature is a time average magnitude. The temperature is measured as an average of the kinetic energy of all the particles with the equipartition theorem, therefore the temperature of a dynamical system fluctuates around a given value. This means that if the instantaneous temperature is to be measured, it may give a value different from the average temperature. With this in mind, what the Berendsen temperature does is periodically measure the instantaneous temperature of the system and rescales the momentum of its particles so that it matches the target temperature. This rescaling is not done abruptly but, instead, a relaxation time is given for the transition to be smooth. As a result, the temperature fluctuations are significantly reduced.

For the rest of the heating steps, the velocities are not applied randomly but they are read from the previous step. All the conditions for the second, third and fourth heating steps are the same as for the first step, but they differ in that no restrains were applied.

5.3.2.3. Equilibration

Until now, all the simulation has been done in what is called the NVT ensemble or canonical ensemble, which basically means that the probability distribution of microstates of our system depends on its temperature (T), its volume (V) and the number of particles on the system (N). For this reason, and because the production will also be done in the NVT ensemble, an equilibration step in the NPT ensemble, to adjust the volume of the box size, is recommended[22]. The NPT ensemble or isothermal-isobaric ensemble is the one in which the probability distribution of microstates is calculated as a function of its (P), its temperature (T) and the number of particles in the system (N). That equilibration in the NPT ensemble is the constant pressure run. In this equilibration phase, not only the pressure is monitored but also a constant density is achieved and the volume of the simulation box is rescaled.

The constant pressure run consisted of 500000 steps of 0.002ps length, giving a total of 1ns of simulation. As it is obvious, constant pressure PBC are applied and constraints for the vibration of bonds involving hydrogen (SHAKE algorithm) are introduced, which will be explained next. The constant pressure stage also has a reference temperature of 300K. The pressure rescaling of the barostat is isotropic and the barostat chosen is the Berendsen barostat, with a coupling constant of 1, of course the Berendsen barostat works the same as the Berendsen thermostat, with the pressure rescaling to the target pressure, which in this case is 1bar.

The size of a MD timestep is determined by the fastest motion in the system, which in our case is the vibration of the bonds involving hydrogen. The hydrogen atom is so light that for a given vibrational energy, its velocity at some points of the vibration will be very large compared to that of heavier atoms. That means that, if a short enough time step is not used, because of the numerical integration, the hydrogen atoms can travel unreal distances, which would give wrong results. Thus, SHAKE removes that bond stretching freedom, allowing for a larger time step to be used and modifying the results of the simulation very slightly.

After the equilibration in the NPT ensemble and because the production is going to be done in the NVT ensemble, it is desired to do an equilibration in the same ensemble as the production before the actual production. This step is the constant volume step.

The constant volume stage also consists of 500000 MD steps of 0.002 ps, giving 1ns of total simulation time. PBC's are again set to constant volume and the reference temperature remains at 300K. The SHAKE algorithm is also used. In both the constant pressure and the constant

volume, we used an algorithm that wraps the coordinates written from the simulation to its primary box, this is, if a water molecule from the initial simulation leaves the primary box, one of its periodic image will enter it and therefore, the coordinates written in the output will be those of the molecule that is in the primary box. This is recommended for long simulations and does not affect the forces or energies.

5.3.2.4. Production

The production is the longest of the stages of the MD protocol, in terms of simulation time. It consisted in a total of 10ns of simulation time, using a timestep of 2 fs. In the production steps, the reference temperature was maintained at 300K with the Berendsen thermostat and constant volume PBC's were applied as well as the wrapping of the coordinates.

It must be mentioned that we considered more than one replica of the system. For the multiple replicas simulation, the results of all the previous stages were reused and the production was initiated from random velocities .

6. PROTONATION STATES

6.1. PKA CALCULATIONS

6.1.1. Catalytic residues

The computed pKa values of the catalytic residues are given in Table 1, considering the three diferent servers and initial structures used. The pKa values resulted to be systematically higher for the E45 residue (the catalytic acid/base residue) than for the D57 residue (the catalytic base), consistent with their

Structure	Residue	PROPKA	H++	Rosetta
WT	E45	5.68	5.834	4.4
	D57	3.50	5.216	2.6
WTP	E45	7.80	6.308	5.2
	D57	3.53	5.524	3.2
MC1	E45	8.67	7.644	4.8
	D57	4.04	5.397	3.4
MC2	E45	8.62	7.682	5
	D57	3.98	5.408	3.4
MC3	E45	8.81	7.339	5
	D57	4.12	4.891	3.4

Table 1 Computed pKa values for E45 and D57

function in the reaction mechanism, in which E45 is protonated (Figure 4a. This is true even in the WT structure, which does not include the 5-mannan chain neither any of the products of the cleavage. This indicates higher basicity of E45 than that of D57 even before the Michaelis complex is formed. However, the difference in pKa between both catalytic residues in some cases is not sufficient to predict a difference in the protonation states of both residues.

6.1.2. Comparison with non-buried residues

In parallel, a comparison between the pKa values of the catalytic residues and those of residues of the same type, but exposed to the solvent, can be seen in table 1 of appendix 1. In this table it is shown how the pKa of the catalytic residues was raised in comparison to that of the superficial residues. The latter had values very close to the pKa of the corresponding isolated aminoacids, while the catalytic residues presented higher values. This way it was shown how affected the catalytic residues are by their environment.

Regarding not buried residues, PROPKA gives a percentage of how buried a residue is in the protein and therefore not exposed to the solvent. All the residues chosen (Table 1 of appendix 1) had a 0% of burial and three different aspartates and two different glutamates were compared.

It must be noted how the most affected catalytic residue is the E45, and how, in the wild type structure, its pKa is not highly affected but when the substrate is added into the structure, its basicity highly raises.

6.2. QUALITATIVE ANALYSIS

The chemical environment surrounding both E45 and D57 is analyzed here, aiming to contrast whether the former pKa results are logical in terms of nonbonding interactions.

Firstly, the chemical environment of the enzyme without the substrate is analyzed.

Regarding the D57 chemical environment (figure 7a), the residue is basically surrounded by two lysine chains and some crystallization waters, therefore the lysines, which are highly basic and will be protonated and positively charged (pKa[K]=10.54), provide the aspartic acid with a protonated environment, which will stabilize the unprotonated form of D57 (COO^-) versus the protonated one (-COOH), giving the D57 residue no chance to be protonated. This is consistent with its role of general base during catalysis.

On the other hand, the E45 (figure 7b) is surrounded by a histidine residue, a tyrosine and a crystallization ethylene glycol (it should be noted that the ethylene glycol will not be present in a real solution). Both the tyrosine and the ethylene glycol will be in their neutral state, not contributing to the electrostatic attraction or repulsion of protons. However, the His124 residue, despite appearing slightly basic in the WT structure, the pKa values for the other structures (table 2) show numbers around the optimal pH for SsGH134 activity. This means that the His124 residue, will contribute to attract protons to the proximities of the E45 and, additionally, will be willing to share those protons, all of this favoring a rise of the basicity of E45 and, in correlation, its pKa. Again, this is consistent with the role of this residue as general acid during catalysis (Figure 4).

His124	PROPKA	H++	Rosetta
WT	7.02	5.238	6.9
WTP	5.07	5.227	6.2
MC1	3.26	5.172	5.5
MC2	-(*)	4.983	-(*)
MC3	-(*)	4.912	-(*)

Table 2 pKa values of His124.
(*)Residue name HIE not recognised.

All the former analysis applies to the rest of the structures that do contain the sugar (figure8) taking into account the correlation that was discovered from the analysis of table 1 regarding an increase of the basicity of the E45 with the presence of the mannan polysaccharide.

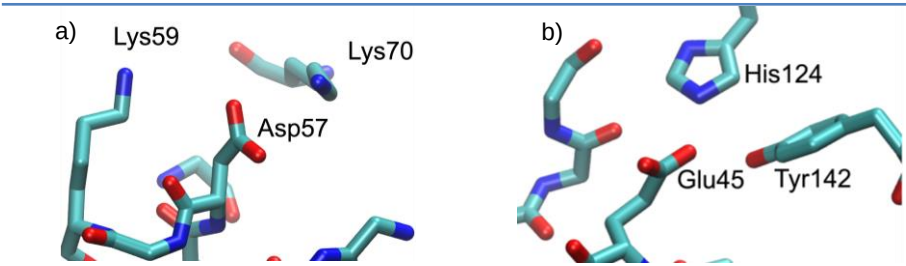


Figure 7 Chemical environment of the catalytic residues in WT a) Asp57; b) Glu45

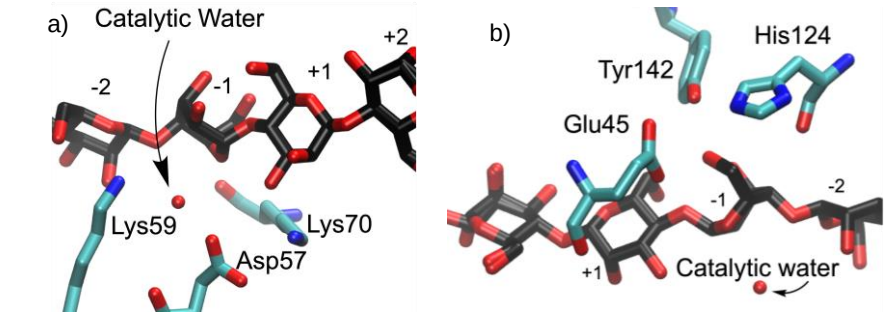


Figure 8 Chemical environment of the catalytic residues in MC1 a) Asp57; b) Glu45

In sumary, although the exact values of the pKa of each residue did not match form one server to another, the trend of D57 being more acid than the E45 did prevail for all of them, which can be useful when a visual inspection is not sufficient to ensure a protonation state over the other.

7. SIMULATIONS

In the first place some thought will be given regarding the expected behaviour of the system:

The function that the proton of the protonated E45 residue performs is to start bonding to the oxygen atom of the glycosidic bond between mannose -1 and mannose +1 (see Figure 3b) in order to assist the reaction by facilitating leaving group departure. Without that proton, the 3-mannan leaving group would be a deprotonated alcohol, highly basic and thus a very poor leaving group.

The E45 carboxylic residue, if protonated, would face the oxygen of the glycosidic bond because the proton is attracted to the oxygen atom that has a negative charge density, forming a hydrogen bond. However, now that the E45 is not protonated it will present a negative charge that will make it point somewhere else, in this case it will be seen that it ends up pointing to the hydroxyl group of the C6 of the -1 mannose.

Thus, as seen before, the leaving group that would appear if the reaction took place would be a deprotonated hydroxyl, which is itself very unstable, meaning that the expected behaviour is for the reaction not to happen.

7.1. RESULTS

For each step of the MD protocol, with the exception of the minimization, the energies, RMSD, and other properties will be discussed. Their graphic representations can be found either in this section or in the appendix 2.

7.1.1. Energy minimization

The most interesting aspect of the minimization is that the Glu45 residue points to the hydroxyl groups of both Tyr142 and the -1 mannose (figure 9), and not with the glycosidic

oxygen, forming hydrogen bonds with their protons. This is so for both the system with the catalytic water and the one without.

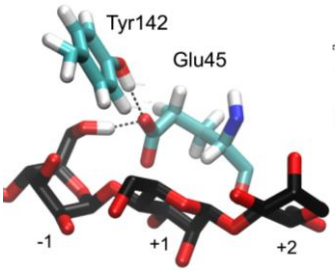


Figure 9 Glu45 orientation after the minimization

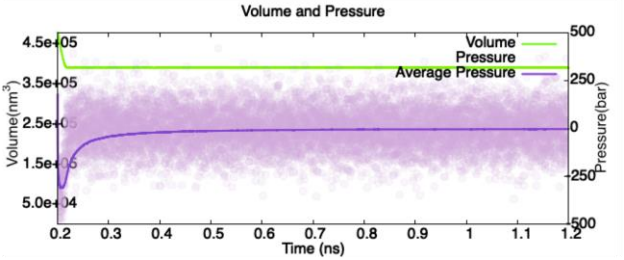


Figure 10 Volume and pressure of the preproduction

7.1.2. Preproduction

The preproduction consists of the heating steps and the constant pressure and constant volume equilibrations.

The temperature and energies of both the system with the catalytic water and the one without it show small oscillations around a constant value, indicating a good behavior. With respect to the volume and pressure of the constant pressure run, it it is shown in figure 10 how at the beginning of the constant pressure run the volume contracts and the pressure fluctuates until an equilibrium is reached. The density also reached a steady value at approximately 1g/cm^3.

In the case of the system with no catalytic water the behaviour was identical.

7.1.3. Production

For the production steps, values of all the properties analyzed were coherent and stable (appendix 2).

However, some data will be shown here for the sake of better comprehension.

First of all the global root mean square deviation (RMSD) of the global simulation of the system with the catalytic water is shown in figure 11. The RMSD of the protein backbone stabilizes at a value around 0.75 Å, indicating that the system is equilibrated and do not deviate much from the crystal structure. This value was also obtained for another MD replica and for the system without the catalytic water.

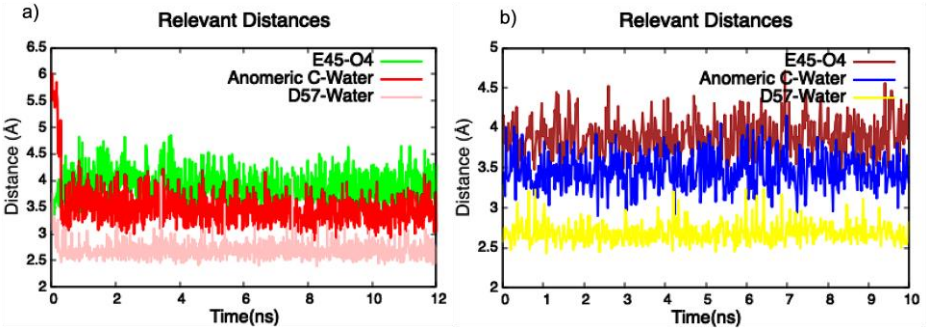


Figure 10 Relevant distances between catalytic residues on a) the system with the catalytic water and b) the system without the catalytic water

To continue, some data regarding the most important distances between the atoms in the catalytic domain are shown in figure 10 a and b.

As it can be seen, the largest of the analysed distances is always the distance between the residue E45 and the oxygen of the glycosidic bond, of course due to the fact that this residue is negatively charged, unlike what its pKa and analysis of its environment (section 6) indicate. This distance maintains values of around 4Å which is, overall, not an enormous distance. Also, the fact that the oxygen of the glycosidic bond is not interacting with E45 makes the anomeric carbon (C1) keep its normal charge density, which means that the catalytic water is not as attracted to it as it would be in a reactive setting, thus, maintaining a distance of around 3.5Å.

Finally, it can be observed that the lowest of the distances is the one between the residue D57 and the catalytic water, which form a hydrogen bond interaction and keep a distance of between 2.5 and 3Å. Of course, it must be noticed in figure 10b that a bulk water approaches the catalytic site. In fact, detailed analysis shows that the D57 residue is the one that brings the water to its corresponding place. At the beginning of the heating the distance between that bulk water and D57 is around 3.5Å, thus they are already interacting, and the distance between the water and the anomeric carbon is around 6Å, meaning that they are quite far apart. As soon as the distance D57-water gets closer (2.5Å), the distance C1-water drops to 3.5Å, indicating that is the D57 residue which drives the water molecule to the catalytic site and makes it interact with the anomeric carbon. Figure 12 shows all the hydrogen bonds between the catalytic residues and the sugar in the final state of the system in which the catalytic water was initially not included.

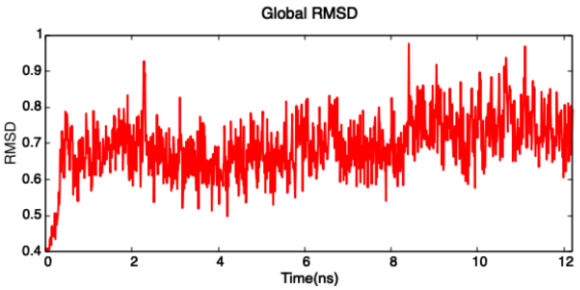


Figure 11 Global RMSD

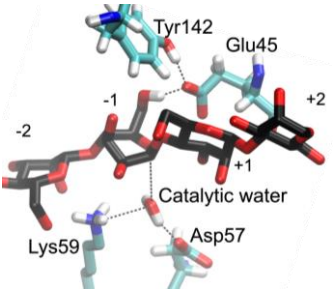


Figure 12 Chemical environment of the catalytic site

7.1.4. Puckering coordinates

Here, the puckering coordinates of the -1 sugar ring on each system during the whole simulation will be displayed, showing how there has not been a major conformational change (figure 13).

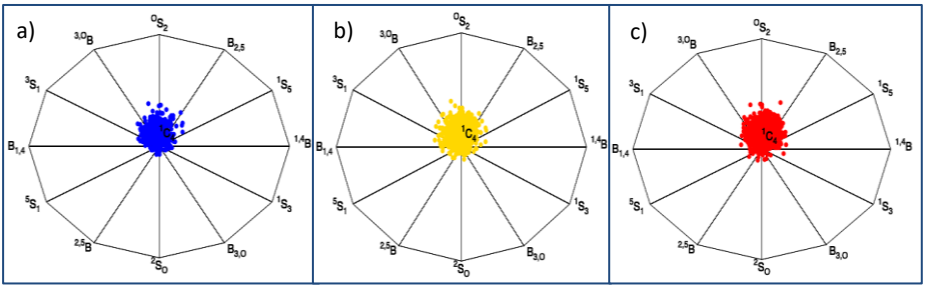


Figure 13 Puckering coordinates of the -1 Mannose on a) the system with catalytic water, b) the system without catalytic water and c) the replicate

8. CONCLUSIONS

First and regarding the computational servers of the pKa calculation, we can conclude that they have proven to be somewhat reliable at least in the fact that all of them were able to differentiate which of the two carboxylic residues had a higher acidity.

Secondly, in the case of the dynamics of the SsGH134 Michaelis complex, we can conclude that the reaction will not happen without the correct protonation state of the carboxylic residues, because the readiness of the leaving group to leave is one of the driving forces of the reaction, and a deprotonated hydroxyl is a very poor leaving group. The reactive sugar, however, does not experiment any conformational change, remaining in its reactive conformation. Extending these results to other GHs, our results indicate that the protonation state of the general acid residue will not influence the conformation of the sugar. This is consistent with the fact that sugar distortions have been observed when this residue is mutated to glutamine or asparagine in GHs.

9. REFERENCES AND NOTES

[1]. Gloster, T. M., & Vocadlo, D. J. (2012). Developing inhibitors of glycan processing enzymes as tools for enabling glycobiology. In *Nature Chemical Biology* (Vol. 8, Issue 8, pp. 683–694). Nature Publishing Group.

[2]. 'Carbohydrate Active Enzymes database' at URL: http://www.cazy.org/, accessed 9 June 9, 2020.

[3]. PM, Henrissat B (2014) The Carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res

[4]. *CAZypedia*, available at URL http://www.cazypedia.org/, accessed July 8th,2020.

[5]. Sobala, Lukasz F.; Speciale, Gaetano; Zhu, Sha; Raich, Lluís; Sannikova, Natalia; Thompson, Andrew J.; Hakki, Zalihe; Lu, Dan; Shamsi Kazem Abadi, Saeideh; Lewis, Andrew R.; Rojas-Cervellera, Víctor; Bernardo-Seisdedos, Ganeko; Zhang, Yongmin; Millet, Oscar; Jiménez-Barbero, Jesús; Bennet, Andrew J.; Sollogoub, Matthieu; Rovira, Carme; Davies, Gideon J.; Williams, Spencer J. (16 April 2020). "An Epoxide Intermediate in Glycosidase Catalysis". *ACS Central Science*, 6, 5, 760–770.

[6]. Terwisscha van Scheltinga A, Armand S, Kalk KH, Isogai A, Henrissat B, Dijkstra BW (1995) Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and x-ray structure of a complex with allosamidin. Evidence for substrate assisted catalysis. Biochemistry 34:15619-15623.

[7]. Zechel, D. L.; Withers, S. G. Acc. Chem. Res. 2000, 33,11−18.

[8]. Williams, S. "Glycoside Hydrolase Family 134" in *CAZypedia*, available at URL http://www.cazypedia.org/, accessed 8 July 2020.

[9]. Jin, Y., Petricevic, M., John, A., Raich, L., Jenkins, H., De Souza, L. P., Cuskin, F., Gilbert, H. J., Rovira, C., Goddard-Borger, E. D., Williams, S. J., & Davies, G. J. (2016). A β-mannanase with a lysozyme-like fold and a novel molecular catalytic mechanism. *ACS Central Science*, *2*(12), 896–903.

[10]. McIntosh, L. P., Hand, G., et al., (1996). The pK(a) of the general acid/base carboxyl group of a glycosidase cycles during catalysis: A 13C-NMR study of Bacillus circulans xylanase. *Biochemistry*, *35*(31), 9958–9966.

[11]. Rapid calculation of protein pKa values using Rosetta, at URL https://rosie.rosettacommons.org/pka/, accessed June 9, 2020.

[12]. *PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations, at URL* http://server.poissonboltzmann.org/pdb2pqr, accessed June 9, 2020.

[13]. *H++ a web-based computational prediction of protonation states and pK of ionizable groups in macromolecules*, at URL http://biophysics.cs.vt.edu/, accessed June 9, 2020.

[14]. (a) Tews, I.; Perrakis, A.; Oppenheim, A.; Dauter, Z.; Wilson, K. S.; Vorgias, C. E. Nat. Struct. Biol. 1996, 3, 638−648. (b) Sulzenbacher, G.; Driguez, H.; Henrissat, B.; Schulein, M.; Davies, G. J. Biochemistry 1996, 35, 15280−15287.

[15]. Espinosa, J. F.; Montero, E.; Vian, A.; Garcia, J. L.; Dietrich, H.; Schmidt, R. R.; Martin-Lomas, M.; Imberty, A.; Canada, F. J.; Jimenez- Barbero, J. J. Am. Chem. Soc. 1998, 120, 1309−1318.

[16]. Kirby, A. J. Acc. Chem. Res. 1984, 17, 305−311.

[17]. Biarnés, X.; Ardèvol, A.; Iglesias-Fernández, J.; Planas, A.; Rovira, C. J. Am. Chem. Soc. 2011, 133, 20301−20309.

[18]. Stoddart, J. F. Stereochemistry ofcarbohydrates; Wiley-Interscience: New York, 1971.

[19]. Cremer, D., & Pople, J. A. (1975). *Journal of the American Chemical Society*, *97*(6), 1354–1358

[20]. Ardèvol, A., & Rovira, C. (2015). Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. In *Journal of the American Chemical Society* (Vol. 137, Issue 24, pp. 7528–7547). American Chemical Society.

[21]. Ardèvol, A., Biarnés, X., Planas, A., & Rovira, C. (2010). The conformational free-energy landscape of β-d-mannopyranose: Evidence for a 1S5 → B2,5 → OS2 catalytic itinerary in β-mannosidases. *Journal of the American Chemical Society*, *132*(45), 16058–16065.

[22]. Braun, E., Gilmer, J., Mayes, H. B., Mobley, D. L., Monroe, J. I., Prasad, S., & Zuckerman, D. M. (2019). Best Practices for Foundations in Molecular Simulations [Article v1.0]. *Living Journal of Computational Molecular Science*, *1*(1), 5957.

[23]. Basconi JE, Shirts MR. Effects of Temperature Control Algorithms on Transport Properties and Kinetics in Molecular Dynamics Simulations. J Chem Theory Comput. 2013; 9(7):2887– 2899.

[24]. Bashford,D. and Karplus,M. (1990) pKa's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. Biochemistry, 29, 10219–10225.

[25]. *Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Research 32 W665-W667 (2004).*

[26]. Kilambi, K. P., & Gray, J. J. (2012). Rapid calculation of protein pKa values using rosetta. *Biophysical Journal*, *103*(3), 587–595.

[27]. Berendsen HJ, Postma Jv, van Gunsteren WF, DiNola A, Haak J. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984; 81(8):3684–3690

[28]. D.A. Case, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F.Pan, S. Pantano, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N.R.Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, L. Wilson, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York and P.A. Kollman (2020), AMBER 2020, University of California, San Francisco.

10. ACRONYMS

- GH: Glycoside Hydrolase
- *SsGH134: Streptomyces* sp, GH family 134 (the enzyme object of the study)
- TS: Transition state
- MD: Molecular Dynamics
- MC: Monte Carlo simulation
- QM: Quantum Mechanics
- MM: Molecular Mechanics
- PBC: Periodic Boundary Conditions
- RMSD: Root mean square deviation
- SD: Steepest descent

APPENDICES

APPENDIX 1: PKA CALCULATIONS

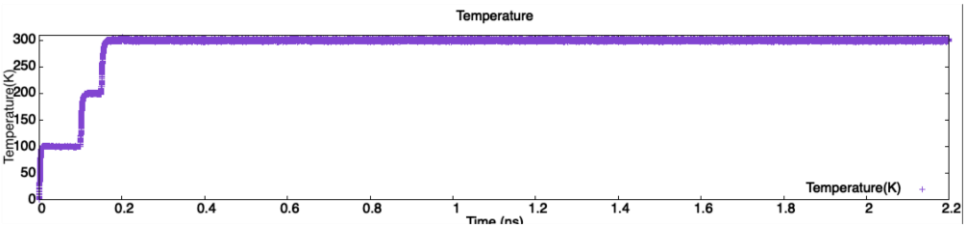
Structure	Residue	PROPKA	H++	Rosetta
WT	E45	5.68	5.834	4.4
	E47	3.60	4.187	3.6
	D57	3.50	5.216	2.6
	D92	3.41	3.867	3.3
	D156	3.27	3.825	3.4
WTP	E45	7.80	6.308	5.2
	E47	3.44	4.565	3.7
	D57	3.53	5.524	3.2
	D92	3.41	4.022	3.3
	D156	3.32	3.901	3.4
MC1	E45	8.67	7.644	4.8
	E47	3.39	4.141	3.6
	D57	3.98	5.408	3.4
	D92	3.41	3.918	3.3
	D156	3.28	3.886	3.4
MC2	E45	8.62	7.682	5
	E47	3.21	4.128	3.5
	D57	3.98	5.408	3.4
	D92	3.41	3.983	3.3
	D156	3.28	3.819	3.4
MC3	E45	8.81	7.339	5
	E47	4.10	4.091	3.5
	D57	4.12	4.891	3.4
	D92	3.22	3.997	3.3
	D156	3.27	3.814	3.4

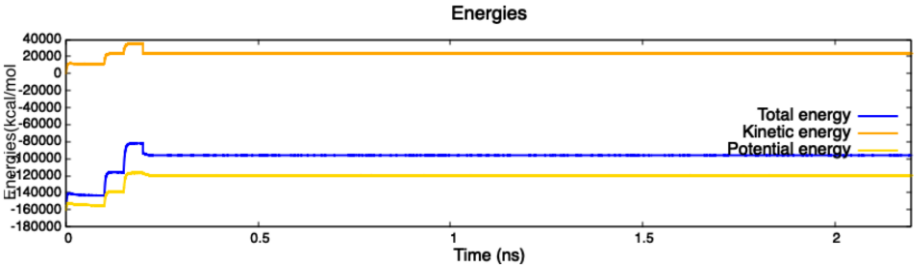
Table 1 pKa comparison of non-buried carboxylic residues

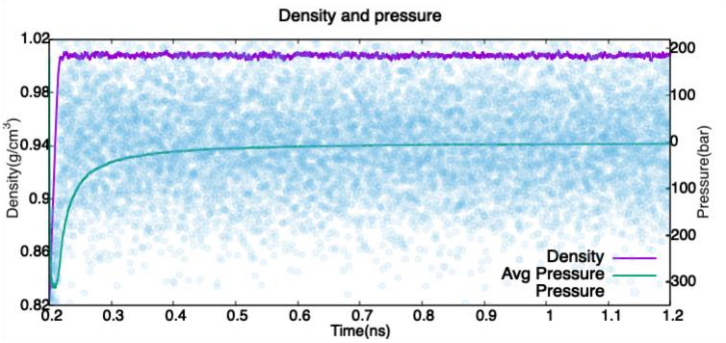
APPENDIX 2: GRAPHIC REPRESENTATIONS

Magnitudes of the Preproduction

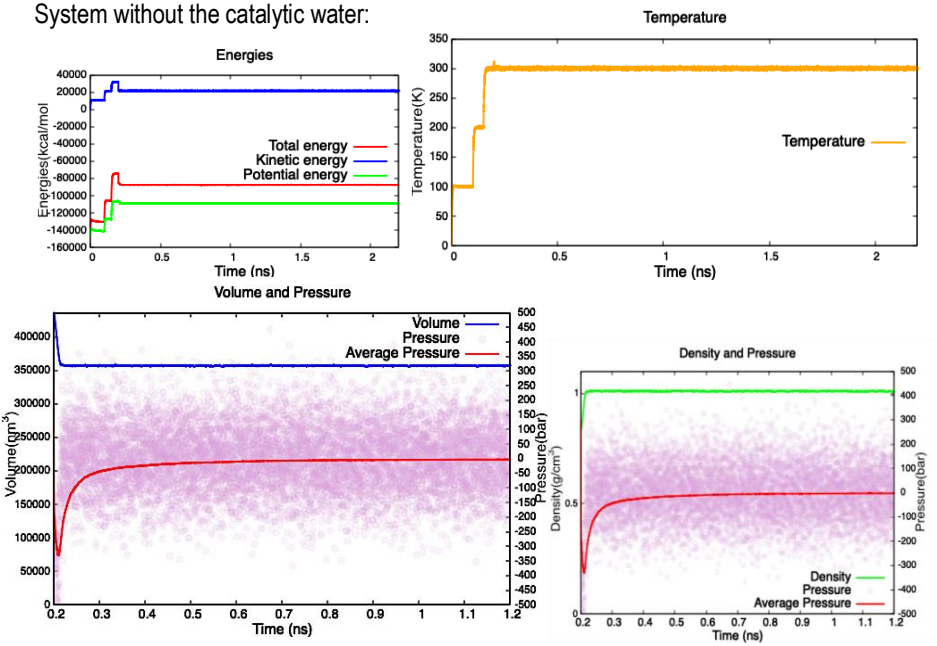
System with the catalytic water:





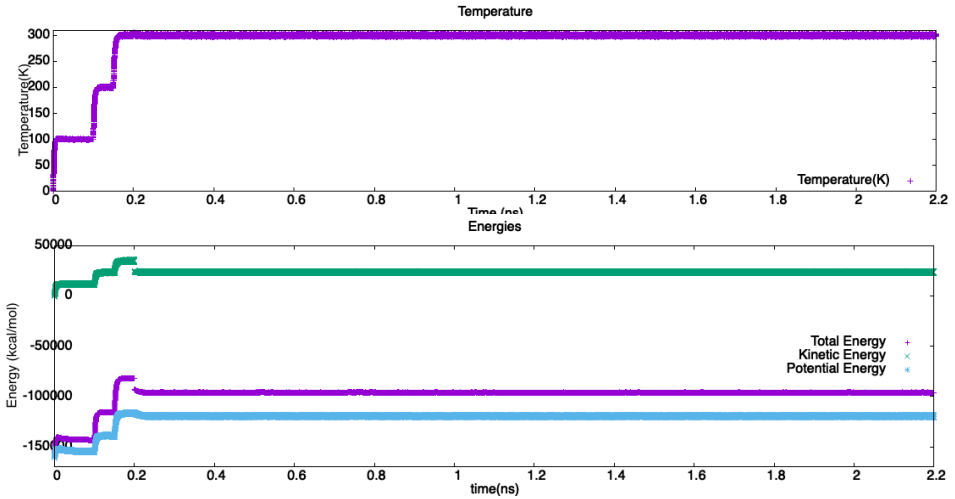


System without the catalytic water:

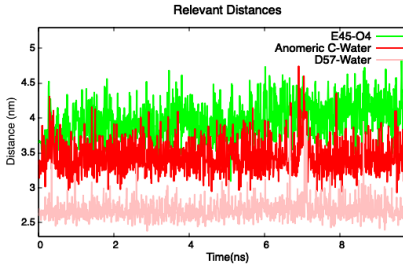


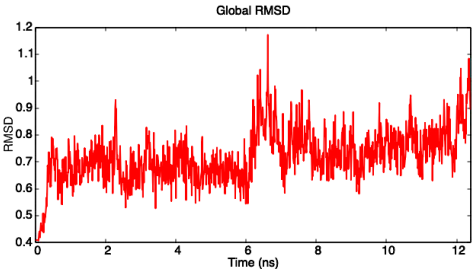
Magnitudes of the Production

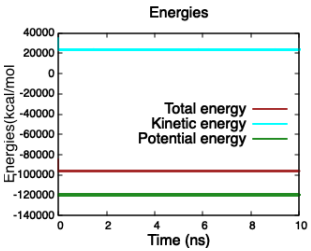
System with the catalytic water:

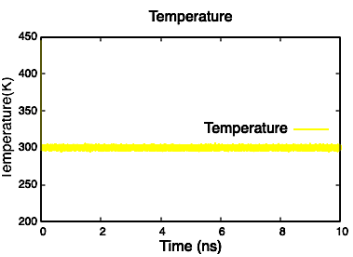


Replicate:









System without catalytic water: