PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Insights into the reaction mechanism of 3-*O*-Sulfotransferase through QM/MM calculations

*Rui P. Sousa, Pedro A. Fernandes, Maria J. Ramos and Natércia F. Brás**

REQUIMTE/Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal.

* corresponding author, NFB (nbras@fc.up.pt)

KEYWORDS: Sulfotransferases, Heparan sulfates, OST inhibitors, QM/MM, MD simulations

ABSTRACT

3-*O*-sulfotransferase (3-OST) is one of the enzymes involved in the Heparan Sulfates (HS) biosynthesis. HS are polysaccharides with variable patterns of sulfation and acetylation that serve as entry receptors for the Herpes Simplex Virus type 1 (HSV-1). 3-OST is responsible for the transfer of a sulfate group from 3'-phosphoadenosine-5'-phosphosulfate (PAPS) to glucosamine units of HS.

In this work, the catalytic mechanism of 3-OST was studied with atomic detail, using computational methods. We investigated the protonation state of key residues using the H++ web-based pK_a prediction tool and Molecular Dynamics (MD) simulations and estimated the most relevant protonation state of the catalytic residues during catalysis. The catalytic histidine (His186) is predominantly protonated, while the catalytic aspartate and glutamate (Asp189 and Glu184) are predominantly deprotonated. Subsequently, to study the catalytic mechanism, we applied a QM/MM method with the ONIOM(B3LYP/6-31G(d):ff94) level, starting from three geometries extracted from the 3, 6 and 8 ns point on the MD simulation. The results show that the reaction mechanism of 3-OST occurs by a single elementary step, consisting of an associative S_N2 transfer of the sulfate group from PAPS to the HS glucosamine units, with the transfer of a proton from glucosamine to the catalytic Glu184. The activation free energies for this reaction were determined at the ONIOM(M06-2X-D3/6-311++G(2d,2p):ff94//B3LYP/6-31G(d):ff94) level of theory. Despite the free energy differences among the three conformations (10.2, 20.9 and 16.1 kcal/mol), our results are consistent with the upper limit determined experimentally for the full cycle (20.4 kcal/mol).

The data obtained in this study will be useful for further studies on the inhibition of this enzyme, which is a useful target for drugs that block HSV-1 viral infections.

INTRODUCTION

Heparan sulfates (HS) are linear polysaccharides that consist of repeating tetrasaccharide units of D-glucuronic acid-(1-4)-D-glucosamine-(1-4)-D-iduronic acid-(1-4)-D-glucosamine. The different tetrasaccharide units can have variable patterns of *N*-sulfation, *O*-sulfation and *N*-acetylation, which result in a large number of different sequences. HS are members of the glucosaminoglycan (GAG) family of polysaccharides, along with chondroitin sulfate, dermatan sulfate and keratan sulfate [1]. Usually, the HS chains are attached to a core protein, forming heparan sulfate proteoglycans (HSPG). These molecules are located at the cell surface and in the extracellular matrix of most mammalian cells [2]. Figure 1 shows a representative segment of a HS molecule.

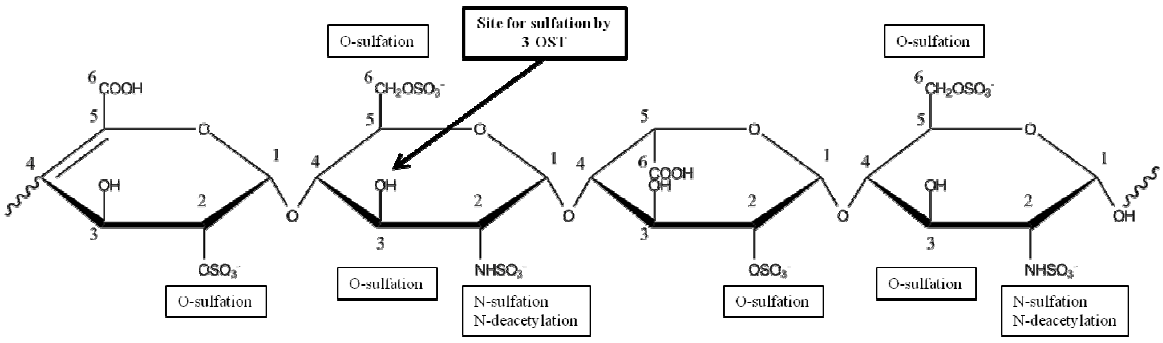

Figure 1. Representation of a segment of a Heparan Sulfate molecule. From left to right are represented one unit of glucuronic acid, followed by one glucosamine unit, one iduronic acid unit and finally another glucosamine unit. Relevant modification sites are highlighted.

HS biosynthesis occurs mainly in the Golgi apparatus and is initiated by the formation of a linkage tetrasaccharide bound to a specific serine residue [3]. During the elongation stage, which consists in the alternating addition of *N*-acetylglucosamine and glucuronic acid residues [4, 5], the chain is modified by several different substituents and in diverse positions, in an orderly fashion [6]. *N*-acetylglucosamines suffer *N*-deacetylation and *N*-sulfation, followed by C5-epimerization of glucuronic acid residues adjacent to the newly formed *N*-sulfated glucosamine units into iduronic

acid. Iduronic acid residues are then 2-*O*-sulfated, whilst the glucosamine residues are then 6-*O*-sulfated and/or 3-*O*-sulfated.

HS 3-*O*-sulfotransferase (3-OST) is a Golgi sulfotransferase, similarly to other enzymes involved in this biosynthetic pathway, responsible for the catalysis of the last step previously mentioned. Modification in the C3 position of specific glucosamine units is the rarest in this process and is responsible for 0.5% of sulfate in the HS chains. However, it is this specific sulfation that allows for the events that culminate in the entry of the herpes simplex virus type 1 (HSV-1) into the cells. So far, seven 3-OST isoforms have been identified, each with different substrate specificity, and conferring different physiological roles onto their catalytic targets. HS modified by isoforms 1 and 5 bind to antithrombin. Additionally, modification by isoforms 2, 3, 4, 5 and 6 generates HS, which serves as an entry receptor for HSV-1, through interaction with its glycoprotein D (gD). [7-9]. The physiological roles of HS modified by the isoforms other than 1 and 5 have not, hitherto, been determined.

The different 3-OST isoforms are responsible for the transfer of a sulfate group from 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to the 3-*O* position of glucosamine units. Their catalytic domain features a large open cleft, which runs across the surface of its spherical shape. There is an α/β motif at the core of this sphere, which consists on a five-stranded parallel β-sheet, laterally flanked by α-helices. The phosphosulfate-binding loop is located at the center of this motif, contained within a strand-loop-helix consisting of the first β-strand and the first α-helix. The residues of this loop establish several strong hydrogen bonding interactions with the 5'-phosphate of the PAPS molecule, which are thought to help position the sulfate group donor (PAPS) in the correct location for catalysis [10, 11].

In order for the sulfotransferase reaction to occur, the substrate binds in an extended conformation along the open cleft, providing access to the sulfate group of PAPS. The 3-OH group of the

glucosamine unit (which will be sulfated) is positioned in close proximity to the PAPS-binding site, in the open cleft. The uronic acid adjacent to the glucosamine unit establishes more interactions with the isoform 3 protein, which suggests its significant contribution to the substrate specificity [10, 11]. The 3-OST substrate is highly sulfated and therefore has a high negative charge, which is countered by the presence of a large number of positively charged residues around the active cleft.

Heparan sulfotransferases (in which 3-OST is included) are thought to catalyze their reaction through a S_N2-like in-line displacement mechanism, a mechanism that is common to other cytosolic sulfotransferases. A conserved glutamate should function as a catalytic base, deprotonating the 3-OH of the glucosamine unit, enabling nucleophilic attack on the sulfate group. Within this hypothesis, an associative bi-pyramidal trigonal transition state (TS) would be formed, with the PAPS and the $3\text{-}O^-$ in the axial positions. A conserved lysine (Lys215) on the phosphosulfate-binding loop would stabilize the negative charge on the PAP group, caused by the leaving of the sulfate group. This lysine residue forms a hydrogen bond with the bridging oxygen to the sulfur atom. Figure 2 depicts the proposed catalytic mechanism.

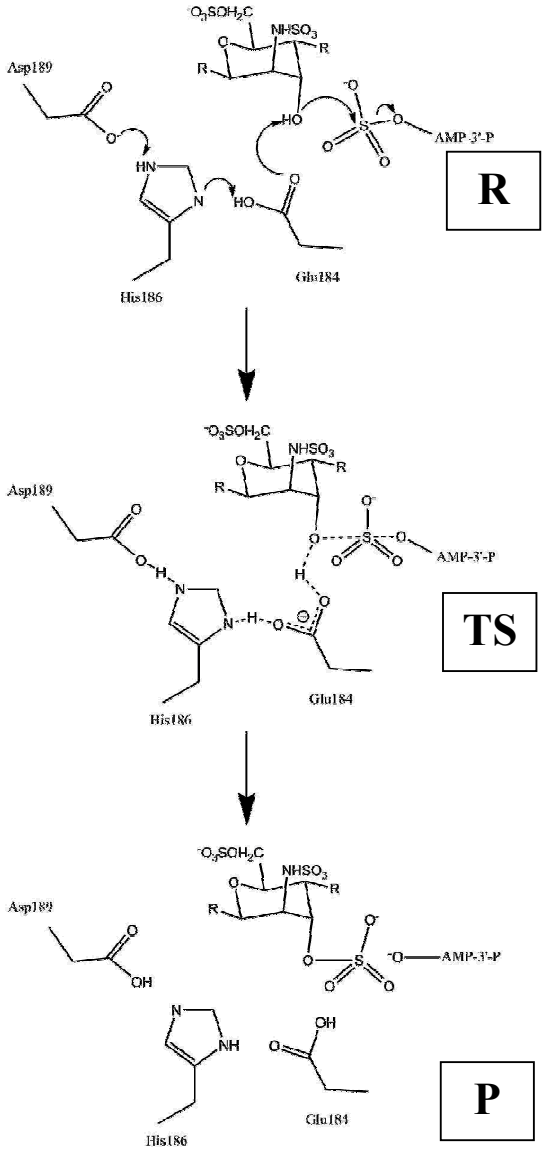

Figure 2. Simplified representation of the proposed catalytic mechanism for the 3-OST enzyme.

In the specific case of isoform 3, a hydrogen bonding network involving the glutamate residue that acts as a catalytic base (Glu184), a histidine (His186) and an aspartate (Asp189) is observed. Two possibilities for the function of this hydrogen bond network have been suggested: It might help to correctly position the glutamate for catalysis, or it might serve as a charge relay system to regulate

its pK_a. [11] Figure 3 illustrates the 3-O-sulfotransferase active site, evidencing the main catalytic residues, the substrate and co-substrate PAPS.

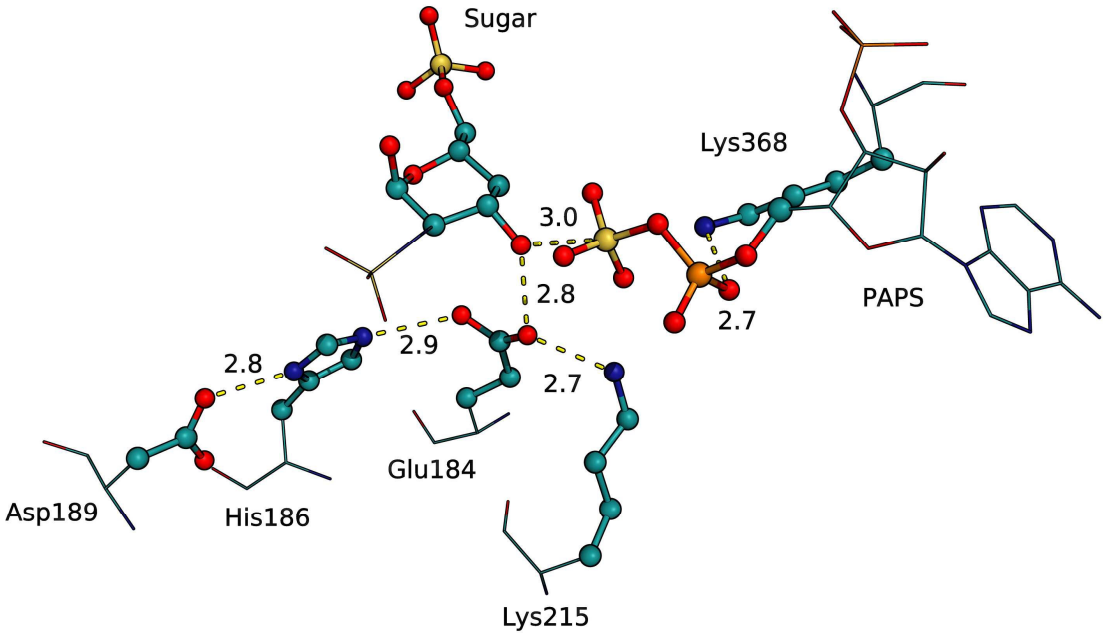

Figure 3. Three-dimensional representation of 3-*O*-sulfotransferase active site, taken from the 1T8U structure, displaying the three catalytic residues and the two lysine residues which establish relevant interactions with the reacting fragments, as well as the substrate and co-substrate PAPS. Hydrogen atoms are not represented.

This work aims to unravel the catalytic mechanism of the reaction catalyzed by 3-OST isoform 3, which is involved in the 3-*O*-sulfation of a glucosamine unit, using computational methods. Inhibition of this enzyme will prevent the formation of HS with gD-binding activity, thus preventing HSV-1 cell entry and infection by this virus. Determination of the reactants, transition state (TS) and products structures, as well as of the role of key residues in the catalytic reaction provide an important insight into the final step of the HS biosynthesis. This allows for the search of transition state analogues with a potential to inhibit this enzyme.

METHODS

The X-ray crystallographic structure corresponding to isoform 3 of human 3-OST, at 1.95 Å resolution, was obtained from the Protein Data Bank [12] (ID: 1T8U). The crystal corresponds to a ternary complex containing the enzyme, 3'-phosphoadenosine 5'-phosphate (PAP) and a tetrasaccharide HS substrate. This HS behaves as two separate disaccharide units, one on the reducing end consisting of iduronic acid-glucosamine and one on the non-reducing end consisting of glucosamine-uronic acid. Uronic acid establishes the most extensive interactions with the protein and the glucosamine unit is the one that will be sulfated. As the disaccharide unit on the reducing end presents fewer hydrogen bonding donors and acceptors, and has only a spectator role during the catalysis, it was not considered in the model used in the calculations. The glucosamine-uronic acid disaccharide was used as substrate in the study of the catalytic mechanism of 3-OST isoform 3 [11]. A sulfate group was also modeled onto PAP in order to form the sulfate donor PAPS. Since this group is small (4 atoms), the modelling tool contained within the GaussView program from the Gaussian 09 software [13] was used. The group was added to the oxygen atom closest to the position that would be sulfated in the catalytic reaction. Furthermore, the MD simulations also had the purpose of relaxing the PAPS position, since it was closely positioned to the glucosamine unit.

The chain B from this crystallographic structure was used. Although the crystal behaves as a dimer, the contact interface between the two chains present in the crystallographic structure is located far from the active site, which suggests that the activity of each active site center is independent to the presence of the other chain. Chain B was chosen due to the fact that it already has the substrate molecule modeled.

The pK_a for the residues which directly participate in the catalytic reaction were estimated using the web-based server H++ (Salinity = 0.15, Internal Dielectric = 10, External Dielectric = 80 and pH = 6.5). [14-16]

In order to obtain a clearer picture of the protonation states of the relevant residues (Glu184, His186 and Asp189), Molecular Dynamics (MD) simulations were performed on the system. Three models were built: one with neutral Glu184 (GLH) and His186 (HIE) and negatively charged Asp189 (A); one with neutral His186 (HID) and negatively charged Glu184 and Asp189 (B) and one with positively charged His186 (HIP) and negatively charged Glu184 and Asp189 (C). All the other histidine residues were kept at the HIE state. Other options for the protonation states were discarded as they would not be adequate for the catalytic reaction to take place. Ideally, constant pH calculations could be made, in order to assess the pKa of the mentioned residues. However, since the computational power required to so would be significantly high, when compared to the scope and goals of the work, the fact that two independent methods provided similar results indicates that they are sufficiently accurate for this work.

Two force fields were used: the AMBER 2003 force field (parm03) [17] for proteins and the general AMBER force field (gaff) [18] for the substrate and PAPS, which were parameterized using the Antechamber program [19]. Explicit solvent, consisting of TIP3P water molecules in a truncated rectangular box with a minimum 12 Å around the protein, was used and chloride counter-ions were added in order to neutralize the charge of the system, leaving the final model atom count at 51881 atoms. Atomic charges were calculated using a QM optimization calculation at the HF/6-31G(d) level, followed by a RESP (Restrained Electrostatic Potential) calculation. [20].

Two separate minimization stages were performed, first keeping the protein fixed and minimizing the water and counter-ions and then, in the second step, minimizing the whole system. The minimized structure was subsequently used to run a MD simulation for 200 ps using the NVT ensemble and periodic boundary conditions, where the temperature was increased from 0 K to 310.15 K. This equilibration stage preceded the second stage, where the NPT ensemble was used, with 10 ns duration. The temperature was kept constant throughout the MD simulations by use of a Langevin thermostat [21], with a collision frequency of 1 ps^{-1}. The Particle-Mesh Ewald (PME)

method [22] was used to treat long-range interactions, and the non-bonded interactions were accounted for in a 10 Å radius. The SHAKE algorithm was used in order to constrain the bond lengths involving hydrogen atoms [23]. The time step used was 2 fs. The simulations were run using the sander module and analyzed using the ptraj tool, both in the AMBER 9 package [17].

The structures used to perform hybrid quantum mechanics/molecular mechanics (QM/MM) calculations, in order to determine the enzymatic catalytic mechanism, were extracted from the equilibrated aforementioned MD simulation, at 3 ns, 6 ns and 8 ns. The system was divided into two layers, following the ONIOM subtractive scheme. The high-layer consists of the glucosamine residue that will be sulfated in the reaction catalyzed by 3-OST, the phosphosulfate termination of PAPS and the side chains of six residues involved in the reaction or in strong hydrogen bonds with the reacting system (Glu184, His186, Asp189, Lys162, Lys215 and Lys368). The positive lysine residues balance the active site negative charge [11]. The high-layer has 83 atoms (shown in Figure 4 and listed in detail in table 1 of S.I.) and was treated with density functional theory (DFT) at the B3LYP/6-31G(d) level for geometry optimizations [24, 25], whereas the rest of the system was treated at the MM level, using the AMBER force field. All the calculations were run using the Gaussian 09 software [13]. Residues more than 15 Å away from the active site were frozen during the calculations. The partial charges from the MM region were included in the QM Hamiltonian (Electronic Embedding), allowing for an accurate description of the electrostatic interaction between both layers and polarization of the QM layer by the MM layer. A Potential Energy Surface (PES) was obtained as a result of calculations, using the distance between the sulfate S atom and the glucosamine 3-O atom as reaction coordinate, that was diminished successively, while the free system was fully optimized for every different reaction coordinate value, therefore building a potential energy profile from the reactants to the products. The energy maximum obtained from the PES along the reaction path corresponded to the approximate Transition State (TS) geometry. The

increment was -0.05 Å in each step, decreasing to -0.01 Å once the TS geometry was nearly achieved.

Since the reaction path obtained through this one dimensional scan was not fully smooth, a bidimensional scan was performed, with an additional reaction coordinate consisting of the shortening of the distance between the proton from Lys162 and the phosphate group from PAPS. This bidimensional scan was performed in the region where the reaction coordinate could not be approximated correctly by the one dimensional approach.

Subsequently, the TS structure was freely optimized, and the vibrational frequencies were calculated, to confirm the existence of a single imaginary frequency, to determine the zero point energy (ZPE), and to calculate the entropic and thermal contributions to the free energy. Afterwards, we manually moved the reaction coordinate slightly towards the reactants and products and fully optimized these two structures. The stationary points were used to determine the activation and reaction free energies of the catalytic mechanism.

Single-point energy calculations were performed on the optimized geometries of reactants, TS and products, using different density functionals and a more complete basis set (6-311++G(2d,2p)). The density functionals tested were B1B95 [26] M06 [27] and M062X [27], BMK [28], CAM-B3LYP [29], B98 [30, 31], wB97XD [32], PBE1PBE [33, 34] and mPW1PBE [33, 35]. Dispersion corrections were applied to the tested functionals [36].

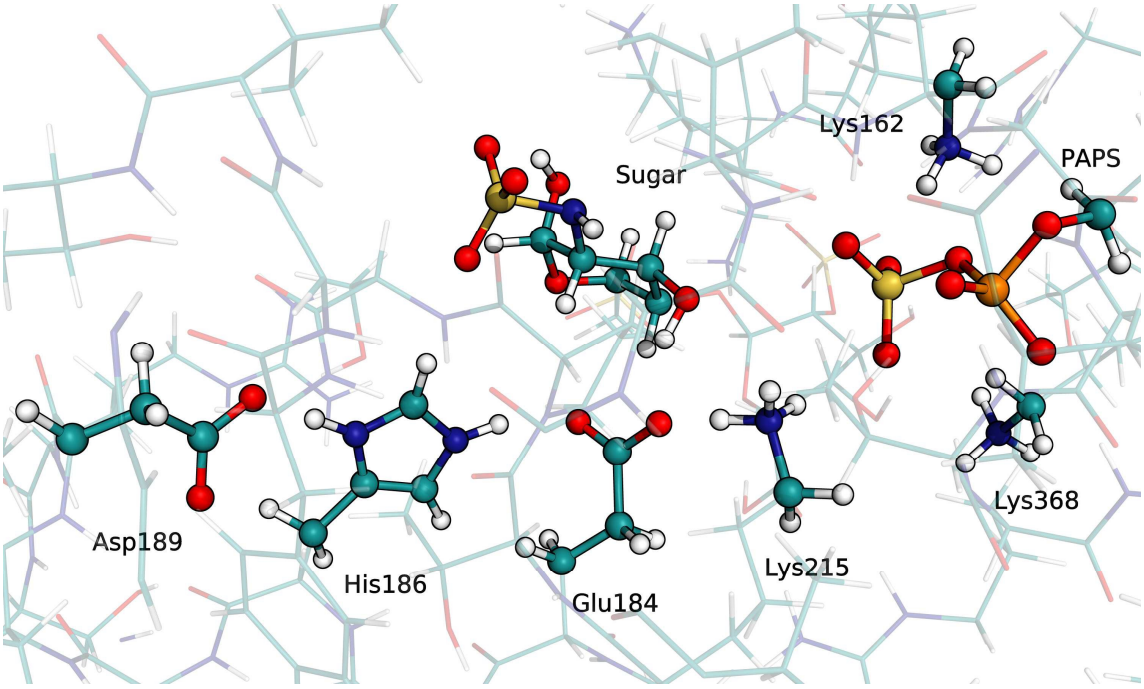

Figure 4. Representation of the QM region used throughout the QM/MM calculations performed on the studied model. Atoms in the QM region are represented in CPK.

RESULTS AND DISCUSSION

The crystallographic structure was used to compute the pK_a of the catalytic residues (Glu184, His186 and Asp189) in the web-based prediction tool, H++. According to this, the estimated pK_a for Glu184 is <0, for His186=6.8 and for Asp189<0. This would mean that the Glu184 and Asp189 would be deprotonated and the His186 could be doubly protonated as well (in particular if we consider that it is located between two Asp/Glu residues, meaning that its pKa is probabaly higher than estimated by the empiric algorithm of H++), at biologically relevant pH values. However, since this calculation does not take into account the ligand (due to limitations of the tool itself), further confirmation is necessary.

The initial minimization protocol performed on the 3-OST crystallographic structure was done in order to minimize the effects that could arise due to the addition of the sulfate group to PAPS, which was not present in the initial X-ray structure. Three different hypotheses for the protonation state of the catalytic residues, at the reactant state, were tested by Molecular Dynamics simulations, as previously described.

Root Mean Square Deviations (RMSD) were calculated for the protein backbone in each of the models, throughout the MD simulations. As shown in figure 5, RMSD values obtained for the three models are small (models A, B and C have average RMSD values of 2.3 Å, 2.4 Å and 2.4 Å, respectively). These RMSD values were obtained upon averaging in the time span 5-10 ns. The systems seemed equilibrated at this stage, despite a small raise in the RMSD value. The RMSD values of the active site residues and substrate molecules were also calculated and shown in Figure S1 in Supporting Information. As observed, this region also seemed equilibrated after 5 ns and, in general, the RMSD values are small. The model A has an average value slightly higher than the one obtained for the entire protein, due to a reorganization of the sugar and PAPS positions, but the opposite occurs in models B and C (models A, B and C have average RMSD values of 3.0 Å, 1.8 Å and 1.6 Å, respectively). A visual comparison between structures extracted from the three MD simulations and the crystallographic structure was made, by superimposition, as can be seen in Figure S2. In this figure, no differences in the actual folding and key residues can be discerned, whereas some differences in the sugar molecule are found.

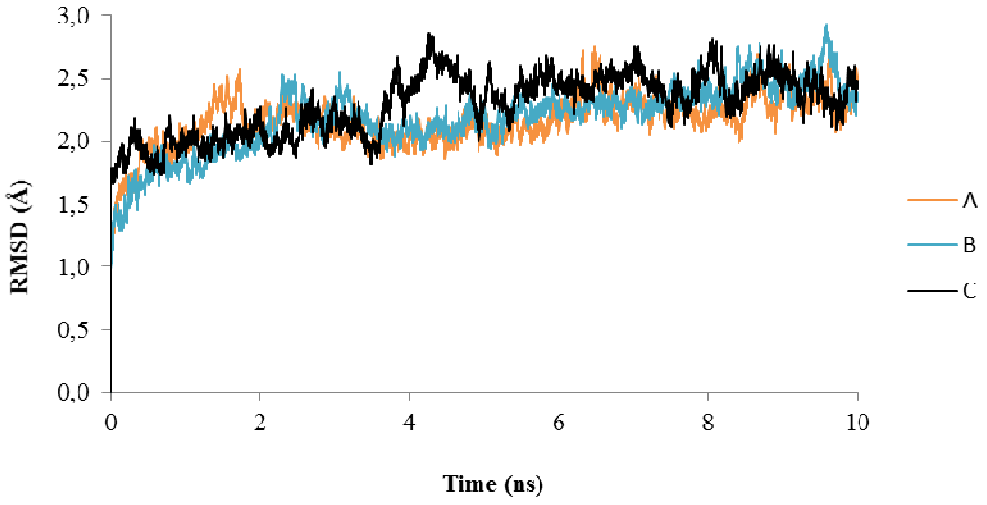

Figure 5. RMSD of the protein backbone of the three different models throughout the MD simulations.

Figure 6 shows the geometry of the active site residues at the beginning and the end of each MD simulation. As shown in figure 6, for model A (neutral Glu184 and His186, negative Asp189), the sugar molecule moved away from the PAPS molecule. The key active site residues changed quickly to a conformation in which the catalytic reaction was not possible to occur. For model B (neutral His186, negative Glu184 and Asp189), the His186 side chain rotated, moving away from Asp189. Furthermore, the sugar and PAPS molecules moved apart throughout the simulation. Model C (positive His186, negative Glu184 and Asp189) presented the most relevant results, since all the key residues positions were kept constant along the simulation. The behavior of these four key H-bonds during the MD simulation of model C is displayed in Figure S3 in Supporting Information.

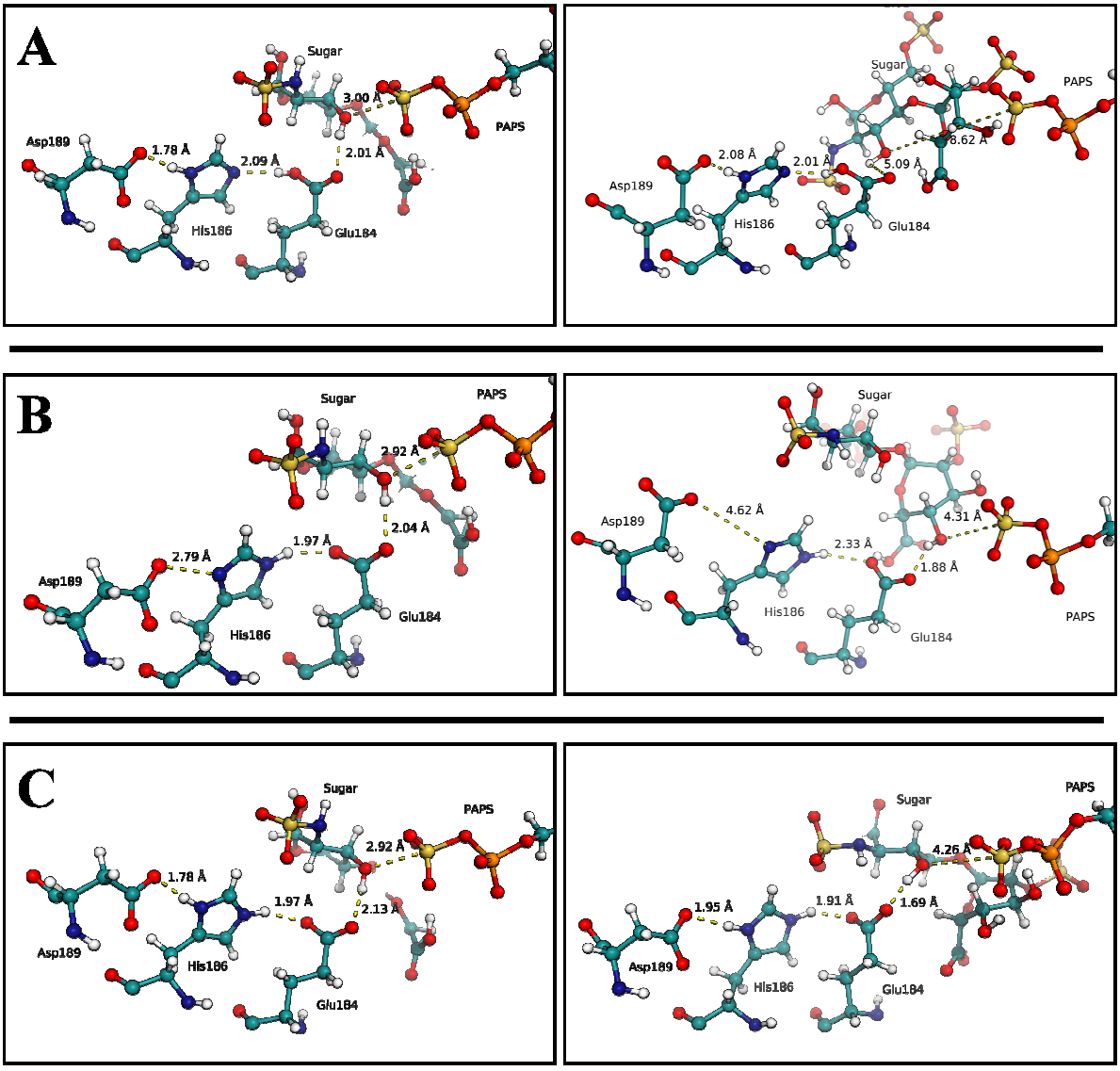

Figure 6. Representation of the structures at the beginning (left) and end (right) of all tested MD simulations.

Therefore, the only protonation state of the active site residues that is viable for catalysis has a negatively charged Glu184, a positively charged His186, and a negatively charged Asp189. The result obtained in the MD simulations is in agreement with the previously mentioned pK_a estimation results. These protonation states discard the previously mentioned charge relay system hypothesis,

since Glu184 has to be negatively charged and therefore cannot transfer a hydrogen atom to His186. This protonation state makes sense from a chemical point of view as it clearly maximizes the electrostatic interactions between the three residues. The two negative charges around His186 will raise its pK_a, promoting its protonation. An accurate calculation of the pK_a of each residue would be the best answer for the problem. However, there is presently no computational method that is accurate enough to give a definitive answer to this problem.

We have paid special attention to the protonation state of model A, as it is has been pointed out as the most abundant and relevant for catalysis. Even though the MD simulations provided support for the protonation state of model C, we moved the proton from His186 to Asp189 and optimized the structure at the ONIOM(B3LYP/6-31G(d):ff94)) level. The proton moved back to His186 immediately and spontaneously. The same happened when we moved the proton from His186 to Glu184. The enzyme definitely favours the "double salt bridge" configuration, with a positive His salt-bridged to two negative carboxylates, which makes sense from an electrostatic point of view.

Assuming this protonation state, the geometry of the enzyme was optimized and the resulting structure and the hydrogen bonding network were very similar to the one presented in the X-ray structure [11]. The catalytic residues, the PAPS and the sugar molecule only presented very small differences. The RMSD value for the protein backbone between the X-ray structure and the optimized QM/MM geometry (extracted from the MD simulation) is also small (1.3 Å), which indicates the preservation of the overall folding of the enzyme.

Reactants. In the structure of the reactants, the sulfur atom is positioned at 3.26 Å from the 3-O position on the glucosamine unit (O3). The O3 proton is hydrogen-bonded to the Glu184 carboxylate (1.89 Å), which in turn is hydrogen-bonded, albeit weakly, to His186 (2.59 Å) and to Lys215 (1.82). The latter makes a short H-bond with Asp189 (1.82 Å), as can be seen in figure 7.

The active site negative charge is stabilized by the presence of three lysine residues (Lys162, Lys215 and Lys368). These residues do not directly participate in the catalytic reaction, but they form hydrogen bonds with both PAPS and the catalytic Glu184. The protonation state of the catalytic residues has changed during the optimization, relatively to what was previously mentioned: His186 remains protonated, Asp189 remains negatively charged, whereas Glu184 is now neutral, due to the transfer of a proton from Lys215. Although this change in protonation in the reactants should have no influence in the overall reaction, we performed the same MD simulations protocol that had been previously mentioned, using these new protonation states (doubly protonated His186, negatively charged Asp189, neutral Glu184 and neutral Lys215) and the results can be seen in Figure S4. The RMSD values throughout the MD simulation for this protonation state are 2.1 Å for the protein and 1.8 Å for the active site. However, during the course of the MD simulation, the sugar and PAPS molecules drift far apart, and the resulting conformations are not catalytically viable. Even taking into account this protonation state, model C is still the most relevant for the catalysis. Figure 7 shows the main optimized catalytic core and the most important interactions established at the reactants.

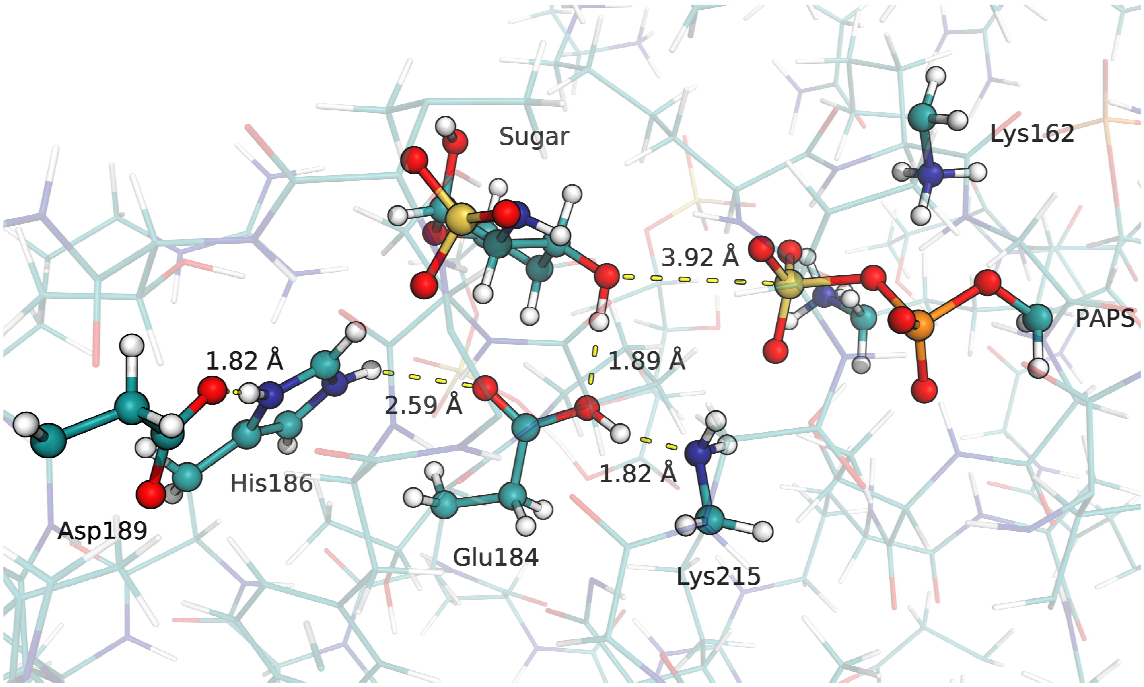

Figure 7. Representation of the catalytic core and the most important interactions established at the reactants.

Transition State. Figure 8 shows the main catalytic core and the most important interactions established at the transition state (TS) structure. The sulfate group is involved in an associative bipyramidal trigonal TS with a dissociative character, with the oxygen atom from the sugar and from PAPS in the axial positions. This group is positioned completely apart from PAPS, at 2.85 Å and closer to the O3 atom, at 2.41, thus evidencing the aforementioned dissociative character. The leaving sulfo group leads to a higher negative charge in the phosphate group, which acts as the driving force to form the TS structure. The Glu184 and His186 have moved slightly apart, when compared to the reactants structure, from 2.59Å in the reactants structure to 2.68 Å in the TS structure, whereas the distances between His186 and Asp189, and Glu184 and Lys215 residues remain similar. The protonation state of the key catalytic residues remains unchanged, with the O3 atom still protonated. The bond length of the proton in the O3 oxygen is only elongated from 0.98

Å to 0.99 Å at the TS. The hydrogen bond between O3 and Glu184 has shortened slightly, from 1.89 Å to 1.77 Å.

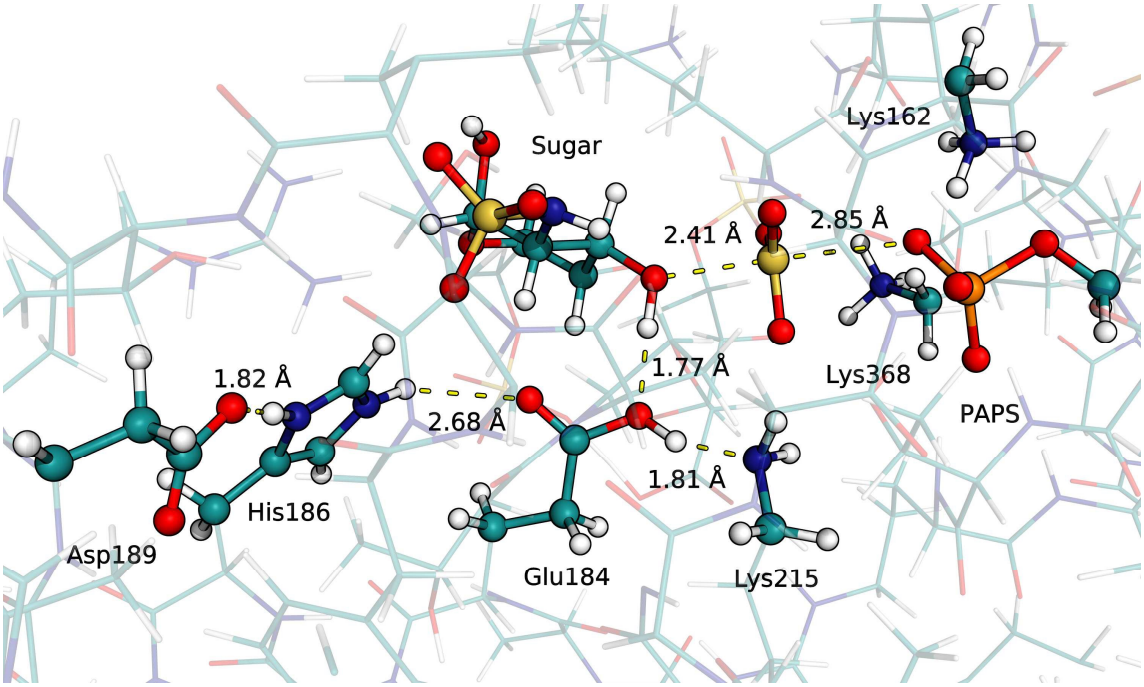

Figure 8. Transition state (TS) structure, emphasizing the main interactions established.

Products. The optimized products geometry is shown in Figure 9. The distance between PAPS and the sugar has increased, and the sulfur atom from the transferred sulfate group is positioned 5.21 Å away from the oxygen atom from PAPS to which it was previously bound to, and is now bound to O3 (1.70 Å), thus finishing the catalytic reaction. The proton from O3 has been fully transferred to Glu184 (0.99 Å) and is now at a distance of 1.80 Å from the O3 atom. The relative His186 and Asp189 positions and the protonation state of the active site residues remain unchanged, when compared to the TS structure. During the reaction, the proton from Glu184 was transferred back to Lys215, to accomodate for the transfer of the proton from O3. One surprise was that Lys162 also participated in the reaction. In fact, during the formation of the products, one proton from Lys162

amine group was fully transferred to an oxygen atom of the PAP phosphate group, thus stabilizing the leaving phosphate.

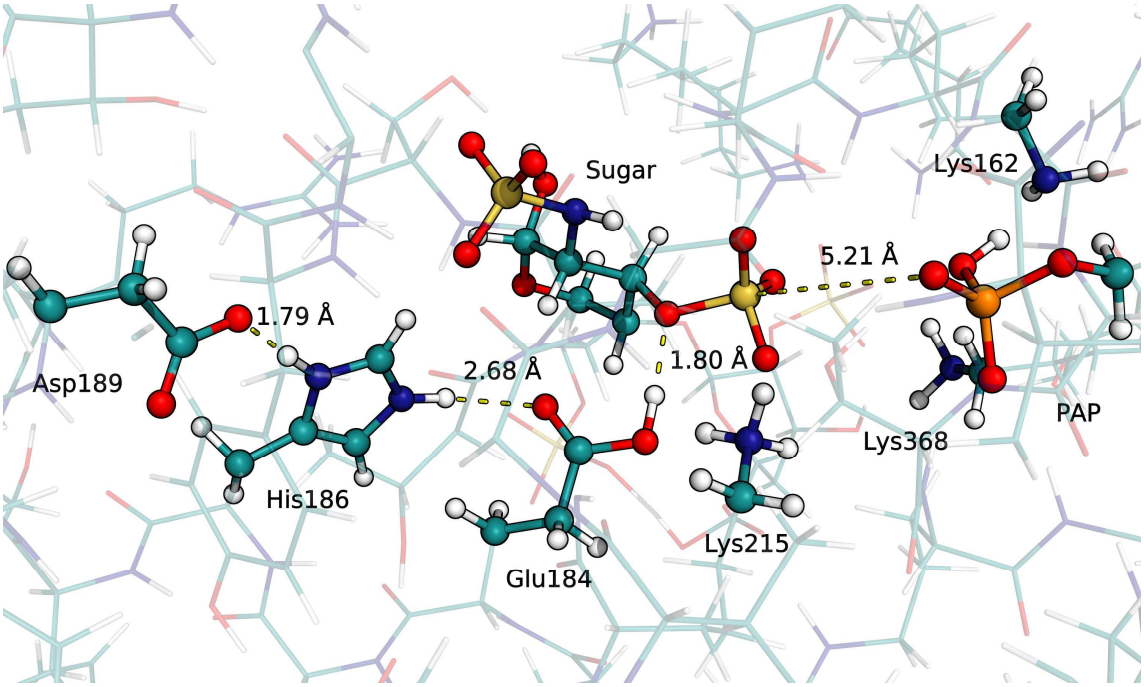

Figure 9. Representation of the products structure, emphasizing the main interactions established.

At the end of the reaction, two residues have a protonation state which is different from that of the reactants. Deprotonation of Glu184 and reprotonation of Lys162 has to take place before the next catalytic cycle begins. This probably occurs after product dissociation and solvation of the active site.

In order to further confirm the protonation state of the reactants and products, a more thorough study of the proton transfers which occur during the reaction was performed. We made a three-layer ONIOM calculation, in which we described the pairs of residues with a questionable protonation state (His186/Asp189, His186/Glu184, Lys215/Glu184, Sugar glycosidic oxygen/Glu184 and Lys162/phosphate), with the 6-311++G(2d,2p) basis set (the remaining atoms kept their original theoretical levels) and performed geometry optimizations in every of the two possible protonation

states. The obtained results suggest that the protonation states described in both the reactants and products are the most stable structures.

Energies associated with the mechanistic pathway

The optimization resulted in a 16.3 kcal/mol activation free energy and a -17.2 kcal/mol reaction free energy at the ONIOM(B3LYP/6-31G(d):ff94) level. Furthermore, the reaction coordinate resulted in a smooth path connecting the reactant and product valley through the TS, as can be seen in Figure S5 of the Supporting Information, which indicates that the chosen reaction coordinate is appropriate to describe this catalytic mechanism.

It is well known that the results of density functional theory have an undesirable dependence on the specific functional chosen for the calculations. To check the impact that the specific density functional has on the results we have recalculated the energy profile of the reaction using several density functionals of different types. The D3 dispersion correction was also included, and a more complete basis-set (6-311++G(2d,2p)) was used. All results from the single-point calculations are shown in Table 1. Activation free energies obtained for this reaction range from 11.2 kcal/mol (for the dispersion-corrected B3LYP functional) to 17.8 kcal/mol (for the BMK functional). Dispersion correction also appears to play a clear role in the observed barriers, lowering activation free energy when compared to their non-dispersion corrected counterparts. This decrease is less noticeable for functionals that already take into account dispersion implicitly, at least partially (such as M06, which shows a decrease of 0.6 kcal/mol, or M062X, which shows a decrease of 0.2 kcal/mol). Reaction free energies range from -24.8 kcal/mol (for the dispersion-corrected BMK functional) to -20.4 kcal/mol (for the B3LYP functional). As is the case for the activation free energy, dispersion correction also lowers the reaction free energy, although to a lesser extent.

Table 1 - Activation and reaction energies obtained for the catalytic mechanism of the 3-OST enzyme with the QM region treated with different density functionals (with dispersion corrections where applicable), using the 6-311+G(2d,2p) basis set and the electrostatic embedding scheme.

Functional	Functional Type	$\Delta G^{\ddagger}$	$\Delta G^{\ddagger}$ (D3)	ΔG_R	ΔG_R (D3)
B3LYP	h-GGA	13.7	11.2 (-2.5)	-20.4	-20.5 (-0.1)
BMK	hm-GGA	17.8	15.5 (-2.3)	-24.3	-24.8 (-0.5)
B1B95	hm-GGA	16.4	14.3 (-2.1)	-22.7	-23.1 (-0.4)
M06	hm-GGA	13.8	13.2 (-0.6)	-23.1	-23.0 (-0.1)
M062X	hm-GGA	16.3	16.1 (-0.2)	-23.8	-23.8 (=)
CAM-B3LYP	h-GGA	15.9	14.1 (-1.8)	-22.2	-22.4 (-0.2)
B98	h-GGA	14.8	-	-20.3	-
wB97XD	h-GGA	14.7	-	-22.3	-
PBE1PBE	h-GGA	16.4	14.7 (-1.7)	-21.1	-21.1 (=)
mPW1PBE	hm-GGA	16.7	-	-21.5	-

It is difficult to say, without further benchmarking tests, which should be the functional to adopt. For example, if we consider the Minnesota functionals M06 and M06-2X, which are well known to be accurate for thermodynamics and kinetics, we get a barrier of 13.2 and 16.1 kcal/mol, respectively. Previous studies describe an activation barrier, obtained through experimental methods, of 20.4 kcal/mol (k_{cat} = 0.77 min^{-1}) [11]. For this comparison, the thermal fluctuations of the MM environment are neglected, since, in order for them to be considered, the Hamiltonian would have to be lowered; this would, in turn, lead to the introduction of a factor of error which would be comparable to the one introduced by the approximation used by the current method. This comparison is, therefore, not entirely correct, since the experimental energy is obtained by an average of several different conformations, whereas the calculated one only takes into account one conformation. However, we are confident that the crystallographic structure conformation is an accurate depiction of the biological environment, and, thus, this comparison is a valid one. In order

to try and minimize the effect of the approximations mentioned, we have also repeated the QM/MM protocol previously described on other two conformations extracted from the 3 ns and 6 ns of the MD simulations, using the B3LYP functional with the 6-31G(d) basis set for the geometry optimizations. The results obtained for these conformations are slightly different than the other previously described, when concerning the activation free energy and reaction energy (17.90 and 11.70 kcal/mol for the barriers, and -25.70 and -32.90 kcal/mol for the reaction energies, respectively). However, all these energies are within the 16±6 kcal/mol interval and below the upper limit determined experimentally. Furthermore, as can be seen in Tables S2 and S3 in SI, we also performed the single-point energy calculations described previously, using the several different functionals, with the larger basis set (6-311++G(2d,2p) and with the inclusion of D3 dispersion, in order to assess the influence of the chosen functional in the activation free and reaction energies. We can observe that the energy variations obtained for each conformation were similar to the ones observed for the first described mechanistic pathway.

The combination of the three activation free energies into one "observed" rate depends on the probabilities of their occurrence, which in turn depends on both the reactant's energy and the probability density for the (classically continuous) energy levels, which is unknown in such a complex system. To calculate the observed rate constant we have to consider the different probability of the reaction to take place at each conformation, using transition state theory. As such, the weight of lower barriers is very large, and these barriers weigh the most for the combined rate constant. Therefore, the average activation free energy obtained through the ONIOM(M06-2X-D3/6-311++G(2d,2p):ff94//B3LYP/6-31G(d):ff94) level of theory was 10.86 kcal/mol. The QM, MM and ONIOM energies for the different stationary points are shown in table S4 in SI. One relevant point is that the final ONIOM energies (~16±6 kcal/mol) are within a much narrower range than its individual QM and MM components, than can change up to 40. The geometries of reactants, transition state and products for the conformations extracted from 3ns and 6ns of MD

simulation can be seen in Figures S6 and S7 in supporting information, respectively. As can be seen when comparing these geometries with the previously mentioned results, the three structures obtained from the MD simulations are very similar. At the reactants, the interactions established between the key residues are consistent throughout the different structures, with the slight difference that Glu184 is negatively charged and Lys215 is positively charged in the 6 ns structure. At the TS structure, the 6 ns conformation presents, once again, slight differences, with the sulfo group being slightly closer to PAPS than the sugar and with a proton from Lys162 edging closer to PAPS. At the products, the three structures are very similar, with Lys162 transferring its proton to PAPS in every case studied here.

The results obtained are slightly different from the previously observed experimental barrier. The small differences observed in the activation barrier may be due to the fact that the sulfate group had to be modeled into the PAP molecule, therefore decreasing the overall distance from the PAPS molecule to the sugar, and due to the fact that the tetrasaccharide present in the original crystallographic structure was cleaved and only the disaccharide portion relevant to the catalytic mechanism was kept. Furthermore, there are several factors implicit in the calculations which also might have influence in the results, such as the fact that the functionals and basis sets cannot fully accurately describe the interactions occurring in the analyzed model and the fact that there is a division between two layers, with part of the interactions measured at the MM level of theory, which might also negatively translate into the results obtained. The use of non-polarizable force fields in a reaction involving highly charged species may also be a factor that impacts on the accuracy of the calculated energies. Finally, it is possible that the limiting step is not the chemical reaction. Product release is frequently rate-limiting. In the absence of further information, the experimental kinetics only represents an "upper limit" for the chemical reaction. Hence, despite all limitations that any computational calculation possesses, the overall result can be considered as accurate and within the upper limit given by experimental kinetics.

CONCLUSIONS

The catalytic mechanism of the isoform 3 of the 3-OST enzyme was studied using hybrid QM/MM theoretical methods and MD simulations. Comparison between the optimized model and the crystallographic structure confirms that the models used are an accurate depiction of the structure of this enzyme. Several MD simulations using different ionization states at the active site were performed. Our results indicate that the model that generates a stationary point with a productive geometry at the active site, and that simultaneously leads to a structure that is similar to the one observed in the crystallographic structure, has a positively charged His186 and a negatively charged Glu184 and Asp189. These results are further confirmed by the QM/MM calculations.

The QM/MM calculations allowed us to confirm the catalytic mechanism that had been previously proposed using experimental methods. We observed that the reaction mechanism of 3-OST occurs through only one elementary, associative S_N2 reaction. The transfer of the sulfate group from the PAPS co-substrate to a specific glucosamine residue involves a pentacoordinated transition state structure, in which the sulfate group is in a bi-pyramidal trigonal geometry. The determined mechanism falls in line with the established conviction that this enzyme follows the same in-line displacement mechanism proposed and observed for other sulfotransferases, due to its use of PAPS as co-substrate. [11, 37] The obtained activation free energy is consistent with the previously observed experimental values. An interesting finding is that Lys162 participates in the reaction as a general acid, protonating the leaving phosphate group. The charge relay mechanism proposed earlier was shown to be unfavorable.

Clarification of the catalytic mechanism through which this enzyme acts, is a crucial first step in its successful inhibition, since the isolation of the transition state structure provides a valuable opportunity for searching and developing transition state analogues capable of inhibiting its activity.

Since this enzyme has no known inhibitors and has the potential to be an important target for drug design, all these studies provide a crucial framework which will allow further studies in the discovery and synthesis of inhibitors for it.

ACKNOWLEDGMENTS

The authors would like to acknowledge the program FEDER/COMPETE and the Fundação para a Ciência e Tecnologia (FCT) for financial support (Projects Pest-C/EQB/LA0006/2013, PTDC/QUI-QUI/122916/2010 and EXCL/QEQCOM/0394/2012). Natércia F. Brás would like to thank the FCT for her IF grant (IF/ 01355/2014).

SUPPORTING INFORMATION AVAILABLE

A list of all atoms included in the high-layer of the QM/MM calculations is shown in Table S1 of the SI. Tables S2 and S3 display the activation and reaction energies obtained for the catalytic mechanism obtained with the conformations extracted after 3 ns and 6 ns of MD simulation (with the QM region treated with different density functionals). Table S4 shows the energy values from the QM/MM calculations performed with the structures extracted from the MD simulations, at 3, 6 and 8 ns. Figure S1 shows the RMSd values for the active site residues and substrate molecules throughout the MD simulations, for the three models studied. Figure S2 illustrates the superimposition of the resulting structures extracted from the end of the MD simulations and the crystallographic structure. Figure S3 displays the key distances studied throughout the MD simulation of model C. Figure S4 presents the structures at the beginning and end of the MD simulation with doubly protonated His186, negatively charged Asp189, neutral Glu184 and Lys215. Figure S5 exhibits the Potential Energy Surface of the QM/MM calculations (for the structure extracted at 8 ns from the MD), along the chosen reaction coordinate. Geometries for the optimized

reactants, transition state and products structures for the structures extracted at 3 ns and 6 ns from the MD simulations are also included in the Figures S6 and S7 at SI.

REFERENCES

1. D. L. Rabenstein, *Natural product reports*, 2002, **19**, 312-331.
2. J. Turnbull, A. Powell and S. Guimond, *Trends in cell biology*, 2001, **11**, 75-82.
3. S. Gulberti, V. Lattard, M. Fondeur, J.-C. Jacquinet, G. Mulliert, P. Netter, J. Magdalou, M. Ouzzine and S. Fournel-Gigleux, *Journal of Biological Chemistry*, 2005, **280**, 1417-1425.
4. B.-T. Kim, H. Kitagawa, J.-i. Tamura, T. Saito, M. Kusche-Gullberg, U. Lindahl and K. Sugahara, *Proceedings of the National Academy of Sciences*, 2001, **98**, 7176-7181.
5. M. Busse, A. Feta, J. Presto, M. Wilén, M. Grønning, L. Kjellén and M. Kusche-Gullberg, *Journal of Biological Chemistry*, 2007, **282**, 32802-32810.
6. J. M. Whitelock and R. V. Iozzo, *Chemical reviews*, 2005, **105**, 2745-2764.
7. C. D. O'Donnell, V. Tiwari, M.-J. Oh and D. Shukla, *Virology*, 2006, **346**, 452-459.
8. V. Tiwari, C. D. O'Donnell, M.-J. Oh, T. Valyi-Nagy and D. Shukla, *Biochemical and biophysical research communications*, 2005, **338**, 930-937.
9. D. Xu, V. Tiwari, G. Xia, C. Clement, D. Shukla and J. Liu, *Biochem. J*, 2005, **385**, 451-459.
10. S. C. Edavettal, K. A. Lee, M. Negishi, R. J. Linhardt, J. Liu and L. C. Pedersen, *Journal of Biological Chemistry*, 2004, **279**, 25789-25797.
11. A. F. Moon, S. C. Edavettal, J. M. Krahn, E. M. Munoz, M. Negishi, R. J. Linhardt, J. Liu and L. C. Pedersen, *Journal of Biological Chemistry*, 2004, **279**, 45185-45193.
12. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Research*, 2000, **28**, 235-242.
13. M. e. Frisch, G. Trucks, H. B. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, B. Mennucci and G. e. Petersson, Gaussian 09, Gaussian, Inc. Wallingford, CT, 2009.
14. J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath and A. Onufriev, *Nucleic Acids Research*, 2005, **33**, W368-W371.
15. R. Anandakrishnan, B. Aguilar and A. V. Onufriev, *Nucleic acids research*, 2012, **40**, W537-W541.
16. J. Myers, G. Grothaus, S. Narayanan and A. Onufriev, *PROTEINS: Structure, Function, and Bioinformatics*, 2006, **63**, 928-938.
17. D. A. Case, T. Darden, T. E. Cheatham III, C. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman and M. Crowley, *University of California, San Francisco*, 2006.
18. Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo and T. Lee, *Journal of computational chemistry*, 2003, **24**, 1999-2012.
19. J. Wang, W. Wang, P. A. Kollman and D. A. Case, *Journal of molecular graphics and modelling*, 2006, **25**, 247-260.
20. C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *The Journal of Physical Chemistry*, 1993, **97**, 10269-10280.
21. J. A. Izaguirre, D. P. Catarello, J. M. Wozniak and R. D. Skeel, *The Journal of chemical physics*, 2001, **114**, 2090.
22. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577-8593.
23. K. D. Hammonds and J.-P. Ryckaert, *Computer Physics Communications*, 1991, **62**, 336-351.
24. A. D. Becke, *The Journal of Chemical Physics*, 1993, **98**, 5648.
25. C. Lee, W. Yang and R. G. Parr, *Physical Review B*, 1988, **37**, 785-789.
26. A. D. Becke, *The Journal of chemical physics*, 1996, **104**, 1040.
27. Y. Zhao and D. G. Truhlar, *Theoretical Chemistry Accounts*, 2008, **120**, 215-241.
28. A. D. Boese and J. M. Martin, *The Journal of chemical physics*, 2004, **121**, 3405-3416.

29. T. Yanai, D. P. Tew and N. C. Handy, *Chemical Physics Letters*, 2004, **393**, 51-57.
30. A. D. Becke, *The Journal of chemical physics*, 1997, **107**, 8554-8560.
31. H. L. Schmider and A. D. Becke, *The Journal of chemical physics*, 1998, **108**, 9624-9631.
32. J.-D. Chai and M. Head-Gordon, *Physical Chemistry Chemical Physics*, 2008, **10**, 6615-6620.
33. J. P. Perdew, K. Burke and M. Ernzerhof, *Physical review letters*, 1996, **77**, 3865.
34. C. Adamo and V. Barone, *The Journal of chemical physics*, 1999, **110**, 6158-6170.
35. C. Adamo and V. Barone, *The Journal of chemical physics*, 1998, **108**, 664.
36. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *The Journal of Chemical Physics*, 2010, **132**, -.
37. Y. Kakuta, E. V. Petrotchenko, L. C. Pedersen and M. Negishi, *Journal of Biological Chemistry*, 1998, **273**, 27325-27330.