

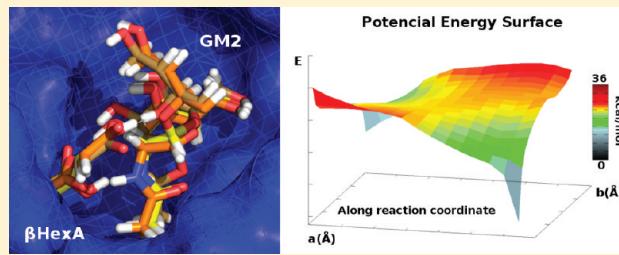
# QM/MM Study of the Catalytic Mechanism of GalNAc Removal from GM2 Ganglioside Catalyzed by Human $\beta$ -Hexosaminidase A

Óscar Passos, Pedro Alexandrino Fernandes, and Maria João Ramos\*

Requimte, Faculty of Sciences, Porto University, Rua do Campo Alegre S/N, 4169-007 Porto, Portugal

Supporting Information

**ABSTRACT:** This work is based on the glycosidase  $\beta$ -hexosaminidase, in particular, on  $\beta$ -HexA (HexA), involved in the emergence of the Sandhoff disease. Its function is to cleave the *N*-acetylgalactosamine (GalNAc) or *N*-acetylglucosamine moieties from its substrates. Here we reveal and consolidate many important aspects of the catalytic mechanism of GalNAc removal from the GM2 ganglioside. The reaction mechanism for this reaction is proposed to be substrate-assisted. It is expected to evolve through a mechanism similar to others glycosidase mechanisms, even though this fact has never been confirmed and the proposal still lacks atomic level detail in its description. To overcome these two limitations we have used the ONIOM formalism (B3LYP:Amber and M06-2X:Amber) with electrostatic embedding to calculate a bidimensional potential energy surface for the rate-limiting step. The potential energy surface reveals the mechanism with atomic detail. The formation of the covalent bond within the substrate has been confirmed, and its energy of stabilization has been calculated as 5.1 kcal/mol. The electronic energy barrier for this step is 22.5 kcal/mol at the hybrid ONIOM(M06-2X:Amber) level, well in line with the typical experimental values for similar reactions in glycosylases. The conformational transitions of the substrate GalNAc ring have been mapped in the PES for the first time. They support the hypothesis of a  $^4S_2$  skew boat– $^4E$  envelope– $^4C_1$  chair pathway and provide much finer detail to the earlier proposals. In general the results are in line with earlier predictions.



## INTRODUCTION

The interest of the scientific community in glycosidases first dates to centuries ago, not only because carbohydrates are a major source of energy but also because they are involved in many other biological processes in both prokaryotes and eukaryotes. For instance, oligosaccharides at the surface of mammalian cells are key elements for recognition during embryogenesis, metastasis, inflammation, and pathogen recognition.<sup>1</sup> The understanding of the catalytic mechanism promoted by glycosidases to break the extremely stable glycosidic bond is very important to clarify the nature and mechanisms of most of these processes. Our work will focus on a particular glycosidase,  $\beta$ -hexosaminidase (Hex, from glycosidases family 20 [EC 3.2.1.52]<sup>1</sup>), involved in neurologic devastating diseases with the function of removing a terminal *N*-acetylgalactosamine (GalNAc) or *N*-acetylglucosamine (GlcNAc) from its substrates.<sup>2</sup>

Diseases such as Tay-Sachs (TSD), Sandhoff (SD), and AB variant disease (ABv) are a consequence of the complete lack of activity of one of the isoforms of Hex, the enzyme  $\beta$ -hexosaminidase A (HexA). The most common form of these disorders has an infantile onset leading to an early death. They are a result of a lysosomal storage disorder, characterized by the accumulation of GM2 ganglioside ( $[\alpha\text{Neu5Ac}(2-3)-\beta\text{GalNAc}(1-4)]-\beta\text{Gal}(1-4)-\beta\text{Glc}(1-1)$ —ceramide) on the central and peripheral nervous system tissues.<sup>3</sup>

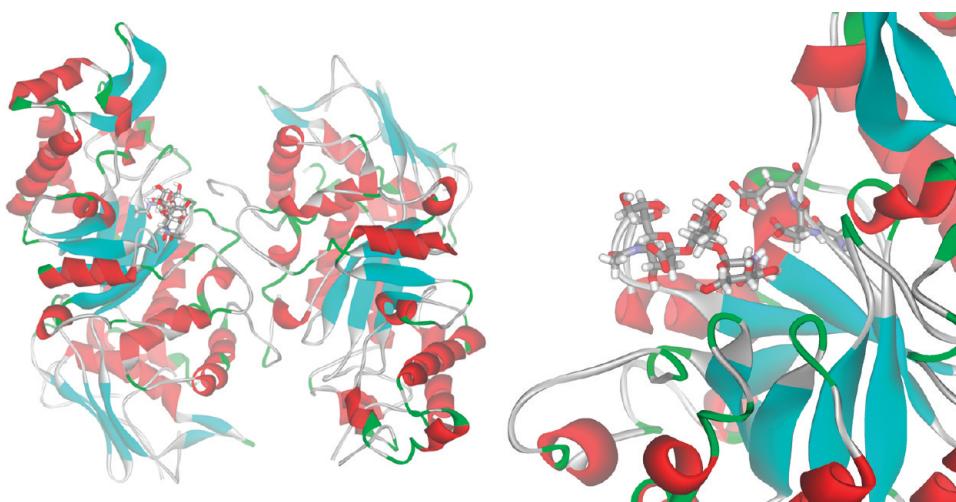
Hex is a dimer with three different isoenzymes: HexA (Figure 1) constituted by one  $\alpha$  and one  $\beta$  subunit,  $\beta$ -hexosaminidase B (HexB) constituted by two  $\beta$  subunits, and the nonfunctional form

$\beta$ -hexosaminidase S (Hex S), a homodimer with two  $\alpha$  subunits. HexA and HexB are detectable in comparable amounts in normal human tissue, and HexS ( $\alpha\alpha$ ) is present in small amounts and only in tissue from patients suffering from SD. Only the dimeric forms are active.<sup>4</sup> There is no scientific evidence of active monomers (exception made for some bacterial species in which  $\beta$ -hexosaminidase is only active as monomer).<sup>5</sup> The genes HEXA and HEXB are the precursors for the  $\alpha$  and  $\beta$  subunits, which share 60% sequence similarity. HexA and HexB have very similar subunits albeit they present important specific differences. There is also a third important gene, GM2AP, which codifies the GM2 activation protein (GM2Ap). GM2Ap is essential for the degradation of the substrate GM2 by HexA.<sup>6</sup> It presents the substrate GM2 ganglioside to the enzyme, assists the binding, and accelerates up to 180-fold the enzymatic activity by preventing hydrophobic interactions with the environment.<sup>7</sup> Any mutation on these genes might reduce or even completely abolish the enzymatic activity. The mutations can affect the transcription, translation, monomer folding and/or dimerization, and consequently the catalytic function.<sup>8</sup> The TSD disease is associated with mutations in the  $\alpha$  subunit, the SD disease with mutations in the  $\beta$  subunit, and the ABv disease with mutations in GM2Ap.<sup>6</sup> However, a small activity of HexA should be sufficient to avoid the more aggressive forms of these diseases.

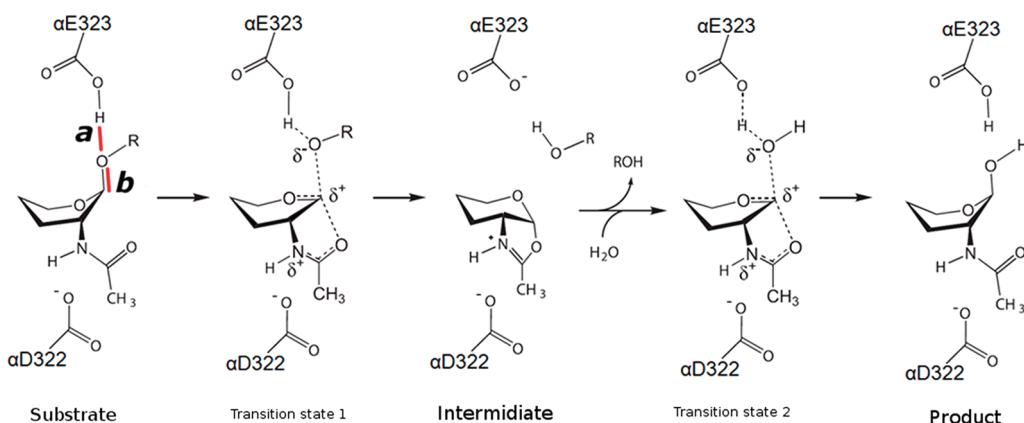
**Received:** June 21, 2011

**Revised:** October 17, 2011

**Published:** October 30, 2011



**Figure 1.** Left: HexA heterodimer. Right: close-up view of the catalytic center. The substrate model used in this work (NGN; Methods section) and the catalytic residues  $\alpha$ -Glu323 and  $\alpha$ -Asp322 are shown in stick presentation.



**Figure 2.** Full reaction mechanism proposed for HexA.<sup>11</sup> Residue numbers correspond to the  $\alpha$  subunit. Only the GalNAc portion of the substrate is explicitly shown. The mechanism is proposed for all forms of Hex. *a* and *b* correspond to the internal coordinates used to calculate the reaction potential energy surfaces (Results section).

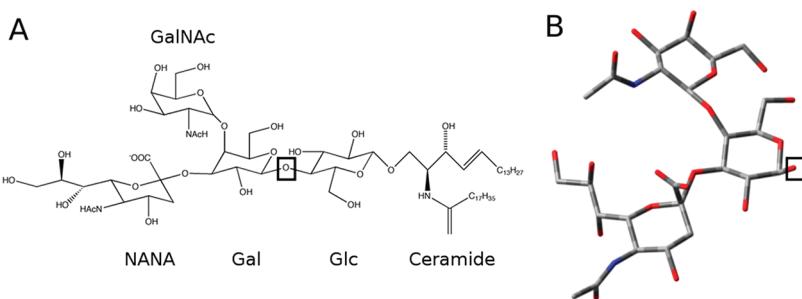
In TSD, the homodimer HexB is fully active, ensuring a nearly normal activity of total  $\beta$ -hexosaminidase. In this case only the  $\alpha$  subunit is affected. In the SD disease, where the  $\beta$  subunit is affected, only 3% of total  $\beta$ -hexosaminidase activity is retained, making it the most severe form of GM2 gangliosidoses.<sup>8,9</sup> There is also the possibility of mutations on the GM2AP gene. Mutations on this gene result in the rare variant AB of GM2 gangliosidoses. In this case both HexA and HexB are produced in normal levels but have small activity to the substrate GM2.<sup>10</sup>

HexA and HexB are both capable of removing GalNAc or GlcNAc from their substrates. The active sites in the  $\alpha$  and  $\beta$  subunits are located at the opening of a ( $\beta/\alpha$ ) 8 TIM barrel near the dimer interface and have the same catalytic residues.<sup>4</sup>

A generic proposal for the catalytic mechanism of this enzyme has been put forward, based on the general catalytic mechanism proposed for all retaining glycosidases and the structure of an HexA:inhibitor complex obtained by X-ray crystallography. Within this proposal, the general acid–base residue  $\alpha$ -Glu323/ $\beta$ -Glu455 protonates the oxygen of the scissile glycosidic bond oxygen, and  $\alpha$ -Asp322/ $\beta$ -Asp454 provides a negative charge that stabilizes the positive charge developed on the nitrogen of

the oxazolinium ion after the intramolecular attack of the *N*-acetamido oxygen atom on the anomeric carbon (Figure 2).<sup>11</sup> The active site protects the oxazolinium ion intermediate from secondary attacks and guides an incoming water molecule for the correct attack at the anomeric carbon in order to maintain the double displacement mechanism, resulting in a product with net retention of the  $\beta$  configuration.<sup>5</sup>

The aromatic residues  $\alpha$ -Trp373,  $\alpha$ -Trp392,  $\alpha$ -Trp460,  $\beta$ -Trp405,  $\beta$ -Trp424, and  $\beta$ -Trp489 are proposed to orient the nucleophilic carbonyl oxygen of the sugar molecule to be trimmed and to stabilize the oxazolinium ion. This last feature is also shared by residues  $\alpha$ -Tyr421 and  $\beta$ -Tyr450 but with the particular characteristic that they come from the adjacent subunit. The main structural differences between the  $\alpha$  and  $\beta$  subunits are present in the region near the active site.<sup>4</sup> The  $\alpha$  subunit has the capacity to catalyze reactions with negatively charged substrates, while the  $\beta$  subunit can only bind and convert neutral substrates.<sup>12</sup> The major reason for this lies on three residues ( $\beta$ -Asp426 vs  $\alpha$ -Glu394,  $\beta$ -Asp452 vs  $\alpha$ -Asn423, and  $\beta$ -Leu453 vs  $\alpha$ -Arg424), which are different in the two subunits. It is believed that the positive guanidine group of  $\alpha$ -Arg424



**Figure 3.** (A) GM2 ganglioside, which is the substrate used in this study. The portion that was deleted from the model corresponds to the Glc and ceramide residues. (B) Structure of our substrate model without hydrogen atoms (NGN), including the NANA, GalNAc, and Gal residues. The squares mark the oxygen atom where the truncation was carried out.

stabilizes the GM2 sialic acid residue (NANA, N-acetylneurameric acid) while  $\alpha$ Glu349 contributes to its correct binding position. Another difference between the isoforms is the  $\alpha$ 280GSEP283 loop (posttranslationally cleaved on the  $\beta$  subunit). This loop is believed to be important to the interaction with GM2Ap.<sup>11</sup>

The dimer interface does not participate directly in the reaction but is very important to the enzyme activity. The interface is located near the active site. The HexA interface probably forms the docking site for the essential GM2Ap. The main function of GM2Ap is to present the GM2 ganglioside to the enzyme. For that purpose it disrupts inhibitory H-bonds in the oligosaccharide portion of GM2 and solubilizes the lipid portion of the ceramide tail, removing GM2 from its membrane environment.<sup>13</sup> It is the combination of several factors that turns possible the degradation of GM2 only by HexA.

The complexity of this system in particular demands for further developments to enlighten understanding of the reaction mechanism with atomic level detail. Nevertheless, recently published works developed by He et al.<sup>14</sup> and by Bottoni et al.<sup>15</sup> shed light onto the subject. Using respectively theoretical and experimental methods, they managed to provide support for the reaction mechanism of a glycosidase from family 84, which, similarly to families 20, 25, 56, and 85, undergoes substrate-assisted catalysis.<sup>15</sup> These and other similar works suggest that a boat conformation is essential for the reaction, but how does the enzyme promote the conformational rearrangement starting from a much more stable chair conformation? Another important aspect, the transition state structure, is somewhat speculative, as the cocrystallized inhibitor that was modeled to be an analogue of the proposed transition state is significantly different from the substrate.

The computational study we have done provides the first atomic-level glimpse into the structure and energetics of the catalytic cycle of a human form of HexA. It clarifies the above-mentioned aspects, adds atomic insight, and validates earlier proposals on this subject.

## METHODS

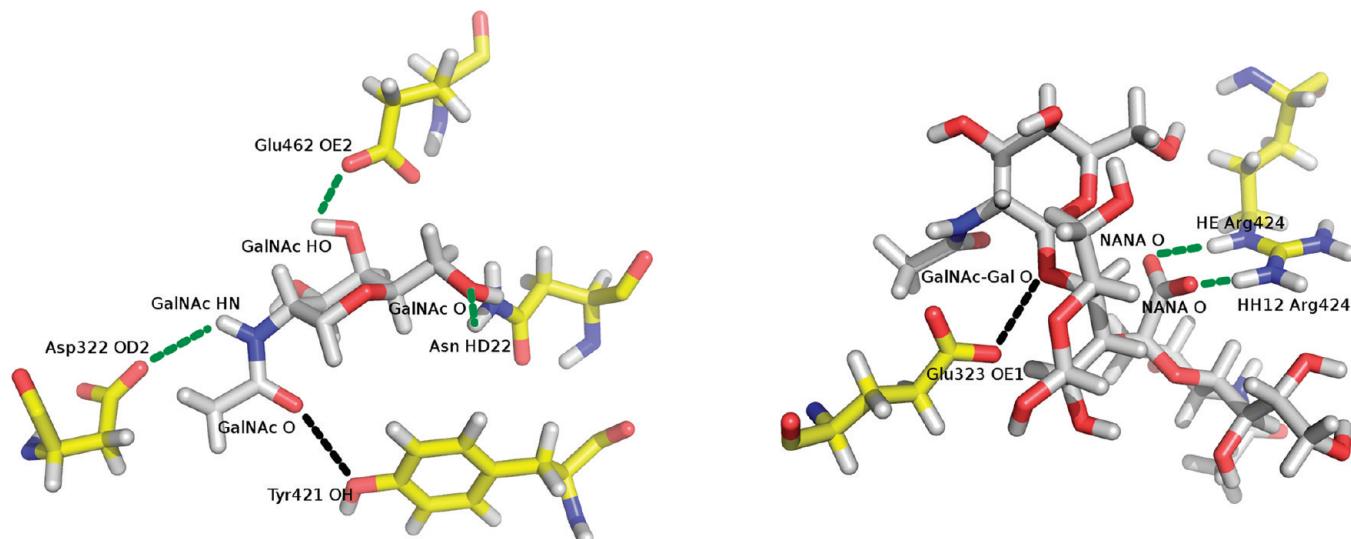
We have used a crystallographic structure of human HexA with the PDB ID:2GJX<sup>4</sup> (Figure 1) at 2.8 Å resolution. From this structure which has HexA bounded to a substrate-analogue inhibitor (*N*-acetyl-D-glucosamine thiazoline), we kept only a single HexA dimer, deleting all other molecules from the 2GJX file. The hydrogen atoms were added to the dimer using the software GaussView4.0.<sup>16</sup> Standard protonation states were assumed for all amino acids with the exception of the acid/base

catalytic glutamate ( $\alpha$ Glu323) that was modeled in the protonated form. GM2, constituted by four carbohydrate molecules and a ceramidic hydrophobic tail (Figure 3), was not completely included in our model. The lipid portion is important for its function but not for the catalytic mechanism being during the reaction completely surrounded by the activation protein. The ceramidic tail and the glucose were not included in the substrate model nor was the activation protein.<sup>13</sup> The modeled substrate was thus a  $\beta$ GalNAc(1–4) $-\beta$ Gal(3–2) $-\alpha$ Neu5Ac molecule (NGN). As shown in Figure 3, NGN is identical to the substrate portion that binds the active site pocket. NGN was modeled with the aid of the Glycam Biomolecule Builder.<sup>17</sup> This procedure improves the computational efficiency without compromising the results.

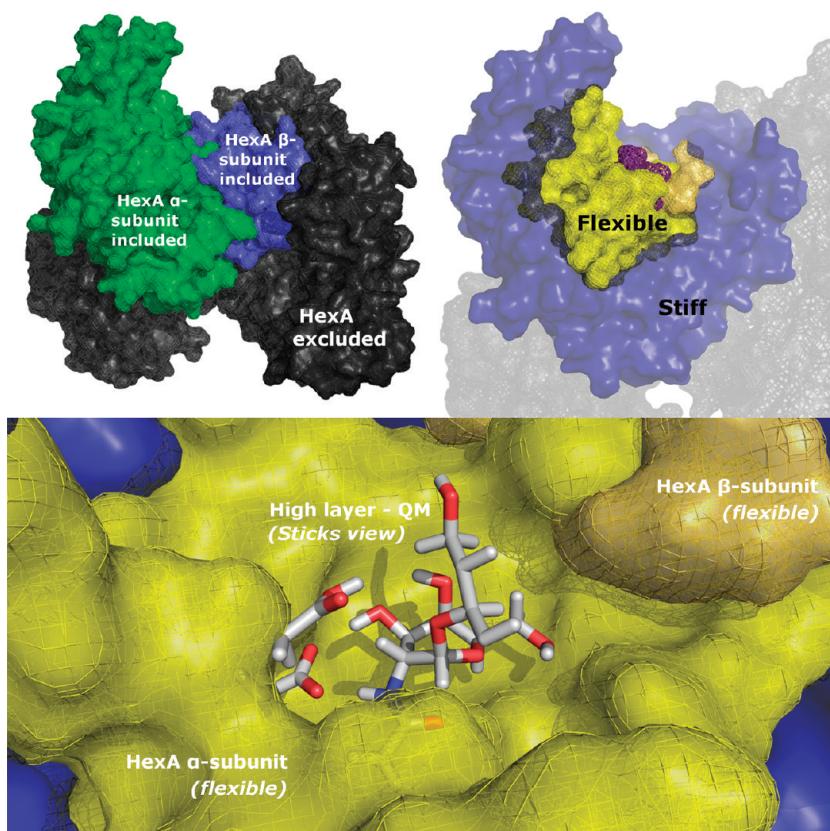
NGN was docked (not modeled) into the  $\alpha$  subunit active site of HexA using the GOLD software<sup>18</sup> with the Chemscore scoring function.<sup>19</sup> The active site was defined using as reference the catalytic  $\alpha$ Glu323 and a search radius of 10 Å. Several interactions identified by Lemieux et al.<sup>4</sup> were used as reference for the docking. We have defined five H-bond constraints and two distance constraints (shown in Figure 4). Table 1 in the Supporting Information summarizes the imposed constraints. Such interactions are believed to be essential for the proper fitting of the substrate into the active site pocket and are a prerequisite for the catalytic mechanism to be feasible. Taken together, these constraints give a very high reliability to the pose produced by the docking algorithm.

Before starting the QM/MM calculations, we performed a molecular dynamics simulation to relax the system using the Amber03 force field<sup>20</sup> for the enzyme and the Glycam04 force field<sup>21</sup> for the substrate. The system was solvated with explicit solvent with preequilibrated TIP3P<sup>22</sup> water molecules, inside a truncated octahedral box (minimum gap of 12 Å). Counterions were added after the water molecules. The MD simulations were performed using the Sander module of the Amber8<sup>23</sup> package through a three-stage minimization/relaxation: first only the water molecules and counterions were minimized. Afterward, the full system was minimized, and finally a small 20 ps MD simulation was carried out with the temperature increasing from 0 to 310 K in the NVT ensemble (Langevin thermostat<sup>24</sup>). Subsequently, we performed a 2 ns MD simulation with a time step of 2 fs. The SHAKE<sup>25</sup> algorithm was used to constraint all bond lengths involving hydrogen.

For the QM/MM<sup>26</sup> step, we have taken the final structure of the MD run, keeping all residues within a radius of 25 Å from the NGN substrate assuring that all important residues and interactions between  $\alpha$  and  $\beta$  subunits were included. A few residues



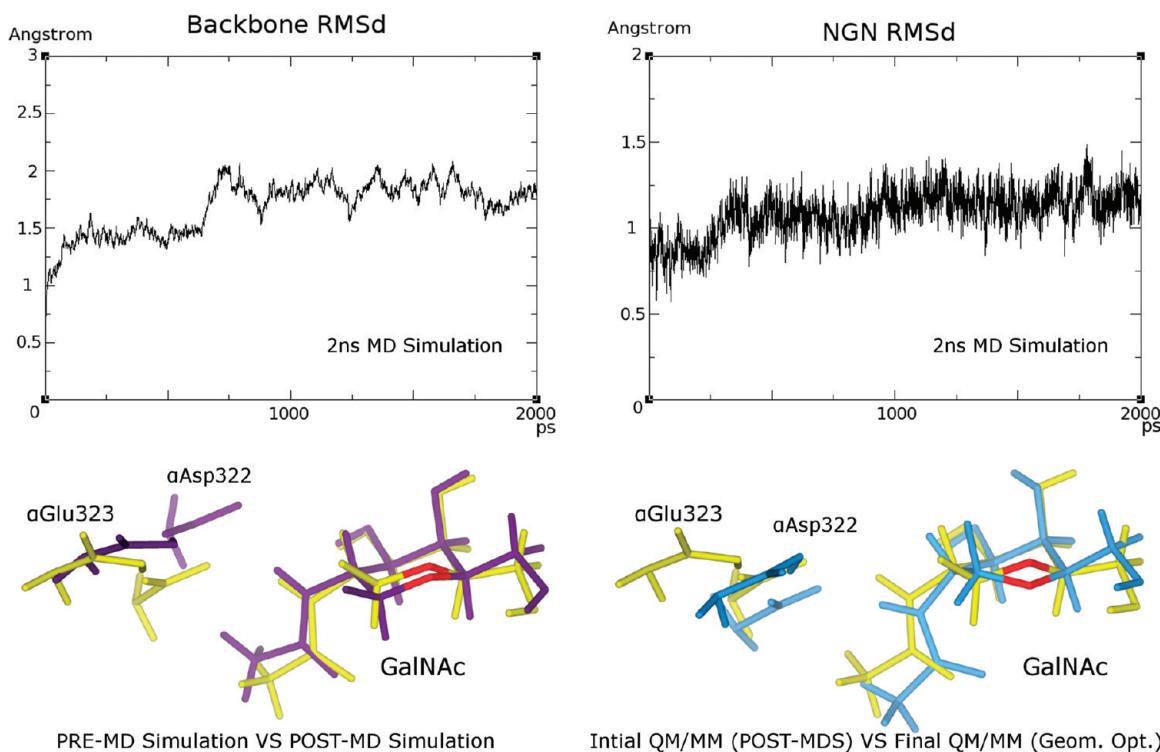
**Figure 4.** Knowledge-based hydrogen bond and distance constraints used to dock the substrate in the active site of HexA (marked as green and black dashed lines). The left and right pictures show two different orientations of the Michaelis complex. All enzymatic residues are from the  $\alpha$  subunit.



**Figure 5.** Top left: HexA dimer with the 25 Å radial cut around the  $\alpha$  active site colored green ( $\alpha$  subunit) and blue ( $\beta$  subunit). This was the region considered for the QM/MM calculations. The deleted HexA region is colored black. Top right: region that was allowed to move (15 Å shell) during the QM/MM geometry optimizations is colored yellow ( $\alpha$  subunit) and light-yellow ( $\beta$  subunit). The frozen region is colored blue. The dark shade is from the deleted region of HexA. The substrate is colored purple. Bottom: Detail of the active site. The region treated at the QM level is represented in sticks and colored by element. The remaining substrate was deleted for clarity.

beyond the 25 Å radius were included to avoid unnecessary gaps in the protein sequence. An outer shell of residues with a thickness of 10 Å was fixed, and only the inner 15 Å shell was allowed to move during the QM/MM geometry optimizations,

to avoid drifting through multiple minima unrelated to the reaction coordinate. The free and the fixed regions of the model are shown in Figure 5. It is important to mention that the NGN motif does not present the GalNAc boat conformation



**Figure 6.** Results of the MD simulations of the HexA:NGN complex. Top left: backbone RMSd of HexA. Top right: RMSd of NGN, calculated including all substrate atoms. Bottom left: GalNAc,  $\alpha$ Glu323, and  $\alpha$ Asp322, before and after the MD simulations (the remaining enzyme and substrate were deleted for clarity). The initial and final MD structures are colored purple and yellow. The sugar ring oxygen is colored red. The enzyme backbone was used to superimpose the two structures. It can be clearly seen that the substrate kept its overall position and orientation. Bottom right: initial and optimized QM/MM structures colored yellow and blue. The sugar ring oxygen is colored red.

postulated by Lemieux et al.<sup>4</sup> before the first the QM/MM geometry optimization. Only after such optimization is the boat conformation achieved.

Two-layer QM/MM ONIOM formalism<sup>27</sup> was used for our model with a total of 6512 atoms, 53 of those in the high layer. The geometries were optimized at the ONIOM(B3LYP/6-31G(d):Amber) level, and the energies were recalculated at the ONIOM(B3LYP/6-31++G(2d,2p):Amber)//ONIOM(B3LYP/6-31G(d):Amber) and ONIOM(M06-2X/6-311++G(2d,2p):Amber)//ONIOM(B3LYP/6-31G(d):Amber) levels. The latter were used to draw the final conclusions as M06-2X<sup>28</sup> is known to be very accurate for kinetics and B3LYP<sup>29</sup> is known to underestimate reaction barriers. The Coulombic interactions between the high and the low layers were treated with the electronic embedding scheme. This scheme includes the MM point charges in the QM Hamiltonian, and although much more CPU-consuming than the parent mechanical embedding scheme, the result has been known to compensate the computational effort.<sup>30</sup> The potential energy surface (PES) for the rate-limiting step of the reaction catalyzed by HexA was calculated using the Gaussian09 software<sup>31</sup> following two internal coordinates: the distance between the  $\alpha$ Glu323 acidic proton and the glycosidic oxygen of GalNAc (distance *a* in Figure 2) and the distance between the anomeric carbon and the glycosidic oxygen of GalNAc (distance *b* in Figure 2). The resolution of the PES was 0.10 Å for *a* and 0.18 Å for *b*. The reactants and products were subsequently freely optimized. Table 2 in the Supporting Information shows the step size of the linear transit scans used to build the PES.

## RESULTS AND DISCUSSION

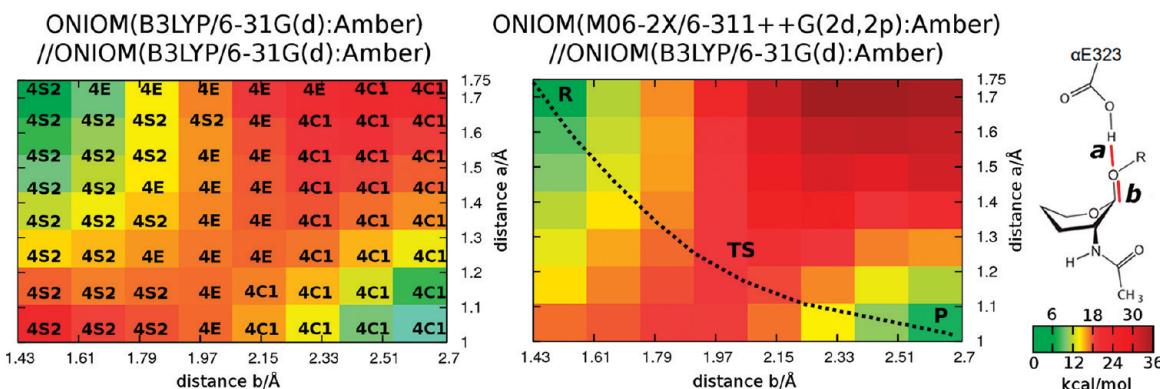
The HexA:inhibitor X-ray structure does not allow us to understand which is the conformation of GalNAc when it binds the active site. Therefore, we have modeled initially all substrate sugar rings in the chair conformation, expecting that the subsequent geometry optimizations and MD simulations would relax the conformation of the substrate into the most stable enzyme-bound state.

The chair conformation of GalNAc was kept during the docking stage. Subsequently, a 2 ns MD simulation with the full enzyme and substrate in explicit solvent was performed to relax the full system. There were no significant differences between the X-ray structure and the result of the MD run. The GalNAc ring kept the chair conformation during the whole run. The position and orientation of NGN in the enzyme was also kept (Figure 6).

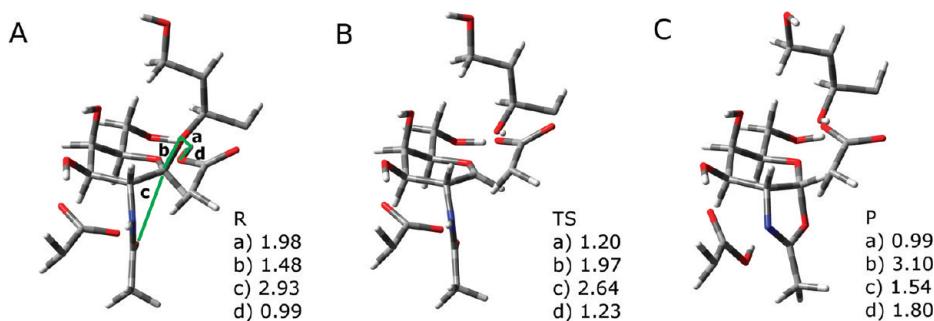
A shell of 25 Å (6512 atoms) was cut from the final structure of the MD simulation and used in the QM/MM study. The optimization of this system at the ONIOM(B3LYP/6-31G(d):Amber) level drove the substrate into a significant conformational rearrangement (variations on the protein were insignificant). In fact, the GalNAc ring changed from a chair to a boat conformation. Such change was probably due to the use of a more accurate QM/MM Hamiltonian. The difference between chair and boat should be small in the bound substrate, and small differences in the accuracy of the Hamiltonian can in these cases induce conformational changes.

At this stage, the enzyme:substrate built model has all the necessary requisites for the first step of the postulated mechanism to be studied, as GalNAc is in a distorted  $^4\text{B}_1$  boat conformation

### Potential Energy Surface for the First Step of the Reaction Mechanism



**Figure 7.** PES of the first, the rate-limiting step of the mechanism. The reactants (R), products (P), and the transition state (TS), as well as the minimum energy path between stationary points, are marked over the surface calculated at the higher level. In the surface calculated at the lower level is represented the conformation of the GalNAc sugar ring for each structure of the scan: E, envelope; S, skew boat; and C, chair.



**Figure 8.** (A) Reactant, (B) transition state, and (C) product for the rate-limiting step of HexA. The most relevant interatomic distances are also shown. Only the QM portion of the model is presented for clarity. All distances in Å.

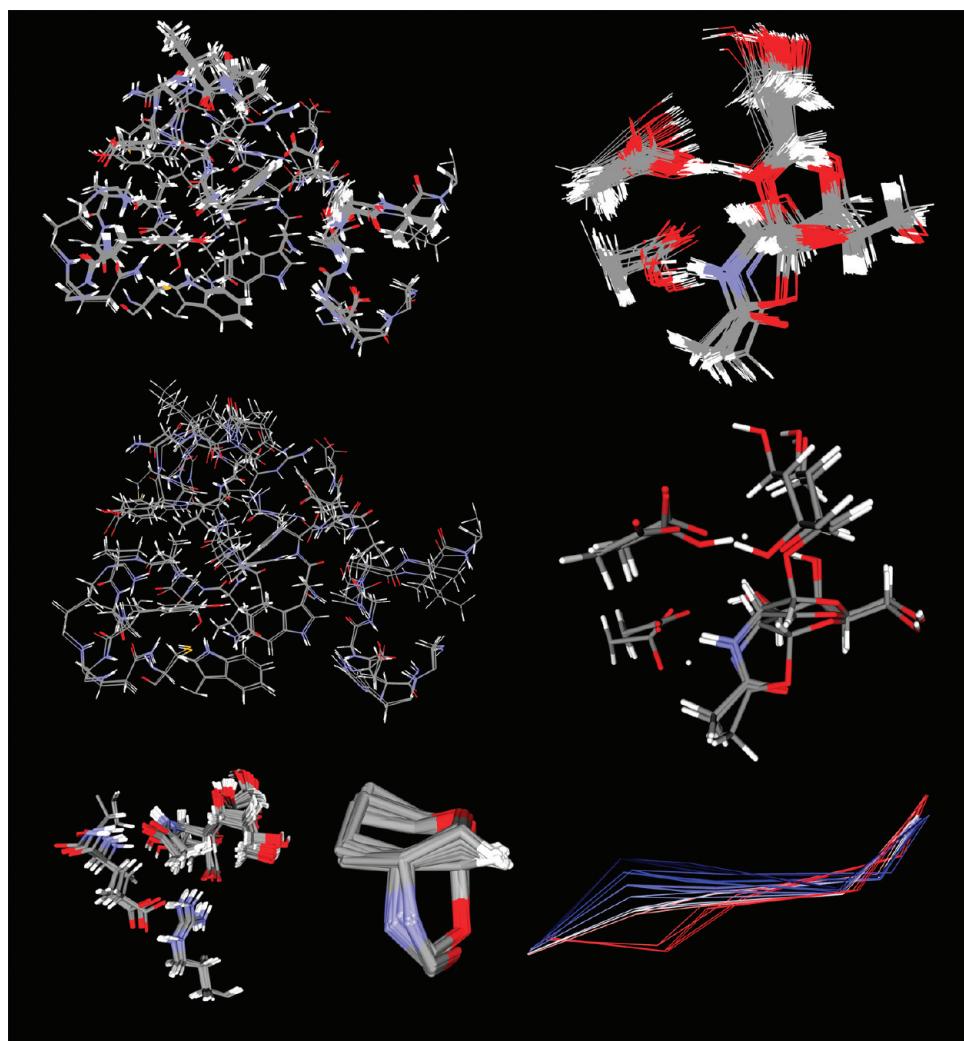
and very close to the catalytic residues  $\alpha$ Asp322 and  $\alpha$ Glu323. From here, two bidimensional PESs were drawn by calculating the respective energies at a set of specific points (114 in total) along the distances  $a$  and  $b$  shown in Figure 2. The specific distances are given in Table 2 in the Supporting Information. The geometry optimizations of all points in both PESs were obtained at the same theoretical level (see Methods), but the corresponding energies were calculated with two density functionals, B3LYP and M06-2X, which correspond to the two extremes of the spectrum of activation energies that are obtained using density functional theory on this system, with B3LYP at the lower border and M06-2X at the upper border.<sup>32</sup> Both density functionals describe the reaction with exactly the same mechanism, the difference lying only in a translation of the activation energies. In a preceding study we have compared the performance of density functionals in a small model system of HexA (which included only a reduced model of the substrate and the  $\alpha$ Asp322 and  $\alpha$ Glu323 catalytic residues), and it was found that (i) there was a clear positive correlation between the magnitude of the activation energy and the amount of HF exchange ( $HF_X$ ) and (ii) the difference in the electronic activation energy between B3LYP ( $HF_X = 20\%$ ) and M06-2X ( $HF_X = 54\%$ ) amounted to 10.5 kcal/mol. M06-2X resulted in the largest barrier of all functionals, and B3LYP resulted in the smallest. In the same study we have concluded that the larger barriers estimated with M06-2X were much closer to the values of high-level post-HF methods than the

smaller values of B3LYP. This means that the results of M06-2X are more accurate (despite its tendency for overestimating the barriers), and they were those chosen to discuss the results (the B3LYP results can be found in the Supporting Information for comparison).

Two PESs are shown in Figure 7. They correspond to the results obtained at the ONIOM(B3LYP/6-31G(d):Amber) level, used just for geometry optimization, and at the ONIOM-(M06-2X/6-311++G(2d,2p):Amber) level, used for final single-point energy calculation. We can notice the reactant (marked R in Figure 7), the product (marked P in Figure 7), and a first-order saddle point between R and P that corresponds to the transition state of the reaction (marked TS in Figure 7). The minimum energy pathway between the reactant and the product is also depicted in Figure 7.

The TS for the first step within these PESs can be unambiguously identified and is exactly the same irrespectively of the density functional (B3LYP or M06-2X) or the basis set (6-31G(d) or 6-311++G(2d,2p)) employed. We also note that the shapes of the PESs calculated with B3LYP (Figure 1 in the Supporting Information) and with M06-2X are very similar, even though the energetic barriers of the latter are larger, as expected.

Figure 8 shows the R, P, and TS taken from the PES (the reactant and product were further freely optimized). The TS presents a slightly distorted  $^4E$  conformation in the GalNAc sugar ring. At the TS the acidic  $\alpha$ Glu323 proton was halfway



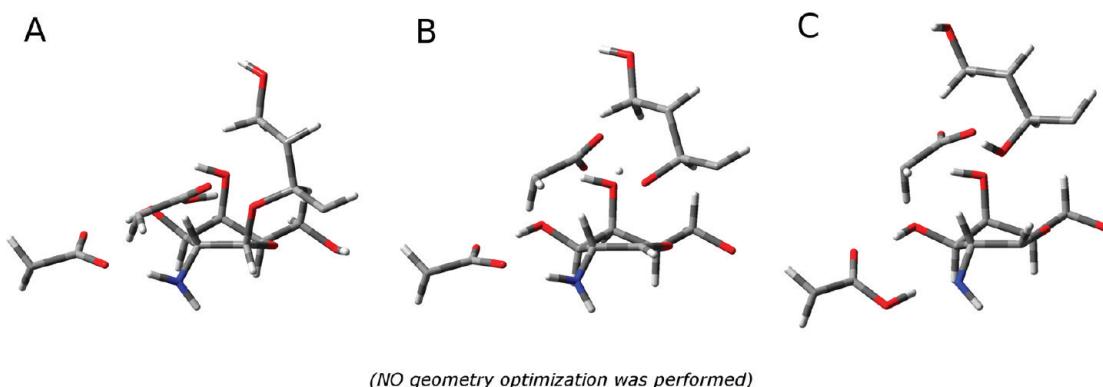
**Figure 9.** Top left: flexible region of the enzyme without the QM layer. All the structures of the PES were superimposed and represented. Top right: QM layer. All the structures of the PES were superimposed and represented. Center left: superposition of the R, TS, and P structures, representing only the flexible region of the enzyme without the QM layer. Center right: superimposition of the R, TS, and P structures representing only the QM layer. Bottom left: substrate without the QM layer,  $\alpha$ Arg424 and  $\alpha$ Glu362. All the structures of the PES were superimposed and represented. Bottom center: GalNAc sugar ring plus the N-acetyl group. All the structures of the PES were superimposed and represented. Bottom right: superposition of the GalNAc sugar ring of the structures along the reaction pathway. The representation goes from dark blue in the reactants to white at the transition state and to red in the products.

between the glutamate side chain ( $d_R = 0.99 \text{ \AA}$ ;  $d_{TS} = 1.23 \text{ \AA}$ ;  $d_P = 1.80 \text{ \AA}$ , with distance  $d$  given in Figure 8) and the glycosidic oxygen ( $a_R = 1.98 \text{ \AA}$ ;  $a_{TS} = 1.20 \text{ \AA}$ ;  $a_P = 0.99 \text{ \AA}$ , with distance  $a$  given in Figure 8). The glycosidic bond is significantly stretched at the transition state ( $b_R = 1.48 \text{ \AA}$ ;  $b_{TS} = 1.97 \text{ \AA}$ ;  $b_P = 3.10 \text{ \AA}$ , with distance  $b$  given in Figure 8). The bond between the N-acetyl oxygen and the anomeric carbon of ( $c_R = 2.93 \text{ \AA}$ ;  $c_{TS} = 2.64 \text{ \AA}$ ;  $c_P = 1.54 \text{ \AA}$ , with distance  $c$  given in Figure 8) is far from being formed at the TS, contrarily to the postulated mechanism, and only in the products we can see the intramolecular cyclization. The distances are also presented in Figure 8.

The superposition of the 114 structures used to calculate the PESs shows that the changes in geometry were mostly located in the QM layer. Glycosylases are known to be conformationally rigid enzymes. They are believed to act as electrostatic machines, lowering the barriers through the creation of a favorable active site electrostatic microenvironment. The geometric results obtained here are fully consistent with such concept. A detailed

analysis of the structures shows that the hydrogen bonds between NANA and the residues  $\alpha$ Arg424 and  $\alpha$ Glu396 were preserved all along the catalytic pathway. Those interactions, specific of the  $\alpha$  subunit, are determinant to stabilize the negative charge on this substrate. The structures also show that the products can be generated with minimal mobility and reorganization of the substrate (Figure 9).

An important change exists in the conformation of the GalNAc ring. During the breaking of the glycosidic bond the ring has moved from the reactants skew boat ( $^4S_2$ ) conformation to an envelope ( $^4E$ ) conformation at the transition state and, finally, to a chair ( $^4C_1$ ) conformation in the products. Such a conformational pathway has been proposed before. Our data gives now solid support for the earlier intuitive proposals. Figure 7 maps the conformation changes of the GalNAc ring over the PES obtained before. An interesting observation is that the transition from the  $^4E$  conformation to the  $^4C_1$  conformation is a necessary condition for the five-sided intramolecular ring



**Figure 10.** (A) Reactant, (B) transition state, and C) product for the rate-limiting step of HexA. Only the QM portion of the model is presented for clarity. They are the same reference structures but with the substitution of the N-acetyl's oxygen, carbons, and hydrogens by single hydrogen (suggestion: compare with Figure 8).

within GalNAc. The N-acetyl oxygen can only be close enough to the anomeric carbon to form a bond in the  ${}^4\text{C}_1$  conformation.

The activation energy (after a free geometry optimization of the reactants and products) corresponds to 22.5 kcal/mol. As M06-2X slightly overestimates this barrier<sup>32</sup> and ZPE corrections were not included (they would lower the barrier by about 3.3 kcal/mol<sup>32</sup>), this result can be considered in good agreement with the values found for other glycosidases (about 15 kcal/mol).<sup>33</sup> The reaction energy was calculated as -3.3 kcal/mol. The influence of the medium-range and long-range electrostatic environment created by the enzyme scaffold results in a decrease of the activation energies of -12.9 kcal/mol, having as a reference the values previously obtained with a reduced active site model.<sup>32</sup> This is an impressive catalytic effect, in particular considering that it is based only on medium- and long-range interactions without conformational rearrangements. The effect of such medium- and long-range interactions is even more significant in the reaction energy value. Comparing with the previous small active site model the difference stands as -19.0 kcal/mol. These variations cause an inversion in the thermodynamics of this step. In the very small model the step was endothermic, while within the enzyme the step is exothermic. We shall remember at this point that the objective of our previous work was not to model the enzymatic reaction but instead to study the intrinsic reactivity of the active site residues. In this context the differences in energy between a very small model and an enzyme are perfectly understandable and completely make sense.

Another aspect that becomes clear from the results is the transition state and intermediate stabilization gained by the formation of the intramolecular covalent bond in GalNAc. In fact, computer simulations allow for the separate observation of events that many times are extremely difficult to decouple experimentally. Here we can monitor the kinetics and thermodynamics of the reaction *without* the attack of the N-acetyl oxygen on the anomeric carbon. For that purpose we just replaced the carbons, oxygen, and hydrogens of the N-acetyl group in GalNAc by a proton, generating a GAINAc residue with an amine in position 2, incapable of binding the anomeric carbon (Figure 10).

We performed single-point calculations on these structures using ONIOM(M06-2X/6-311++G(2d,2p):Amber) level and no geometry optimization. The difference in energy in relation to the normal catalytic pathway gives us the contribution of the substrate-assisted catalysis for the overall enzyme kinetics. The

results show an activation energy of 27.5 kcal/mol (5.1 kcal/mol above the normal reaction pathway). Therefore, the substrate-assisted catalysis lowers the rate-limiting step, leading to a rate enhancement of 3–4 orders of magnitude. In the most common catalytic mechanism of glycosidases, the stabilization of the carbocation and the intermediate is made by a second catalytic Asp/Glu residue whose carboxylate binds covalently to the substrate C1, more favorably than in the substrate-assisted catalysis. The stabilization by the N-acetyl group implies the formation of a five-sided ring with small, but nevertheless significant, steric tension. Moreover, the N-acetyl group is less nucleophilic than the carboxylic side chain of Asp/Glu. The activation energy obtain for HexA is higher than the one obtained for enzymes that follow the general glycosylase mechanism. This was expected, making sense of the denomination of an alternative pathway for the substrate-assisted catalysis.

The use of the N-acetyl group to perform the catalysis constitutes an obvious case of adaptation of the catalytic machinery to the problems raised by a bulky substrate, which cannot use a second Asp/Glu in the catalysis, as the other glycosidases do when processing simpler oligosaccharides.

## CONCLUSIONS

The results constitute clear evidence of a substrate-assisted catalysis by Hex. Key features such as the sugar ring conformational pathway, which ranges from a skew boat to an envelope at the transition state and to a chair conformation in the products as well as the formation of the five-sided ring inside the substrate residue, were confirmed and detailed at an atomic level by the QM/MM study. The transition state stabilizing role of the N-acetyl group of the substrate was confirmed and measured as 5.1 kcal/mol. The need of this group seems obvious to us as a second catalytic Asp/Glu residue (as the ones present in other glycosidases, quite close to the position of  $\alpha$ Asp322) cannot reach the anomeric carbon and bind the substrate at the end of the first step. Even though,  $\alpha$ Asp322 still has a stabilization effect over the covalent intermediate, establishing a hydrogen bond between the aspartic acid oxygen and the positive amine hydrogen of the N-acetyl group. The results suggest that the evolutive pathway of Hex has been similar to many other glycosidases but with an extra capability to allow for the inclusion of a bulky substrate with limited access to the enzymatic catalytic residues. The particular specificity of Hex substrates (which cannot be processed by any

other enzyme) has resulted in a very unusual mechanism to deal with them.

## ■ ASSOCIATED CONTENT

**5 Supporting Information.** Constraints imposed for the docking of NGT in the HexA active site and the specific distances constrained in the linear transit scans; PES of the first, rate-limiting step of the mechanism at the ONIOM(B3LYP/6-311++G(2s,2p):Amber)//ONIOM(B3LYP/6-31G(d):Amber) level. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: mjramos@fc.up.pt.

## ■ ACKNOWLEDGMENT

This work has been financed by the program FEDER/COMPETE and by the Fundação para a Ciência e a Tecnologia (Project PTDC/QUI-QUI/102760/2008).

## ■ REFERENCES

- (1) Henrissat, B.; Davies, G. *Curr. Opin. Struct. Biol.* **1997**, *7*, 637–644.
- (2) (a) Sandhoff, K.; Christomanou, H. *Hum. Genet.* **1979**, *50*, 107–143.  
(b) Henrissat, B.; Bairach, A. *Biochem. J.* **1993**, *293*, 781–788.
- (3) Mahuran, D. J. *Biochim. Biophys. Acta—Mol. Basis Dis.* **1999**, *1455*, 105–138.
- (4) Lemieux, M. J.; Mark, B. L.; Cherney, M. M.; Withers, S. G.; Mahuran, D. J.; James, M. N. G. *J. Mol. Biol.* **2006**, *359*, 913–929.
- (5) Mark, B. L.; Vocadlo, D. J.; Knapp, S.; Triggs-Raine, B. L.; Withers, S. G.; James, M. N. G. *J. Biol. Chem.* **2001**, *276*, 10330–10337.
- (6) Mahuran, D. J. *Biochim. Biophys. Acta—Lipids Lipid Metab.* **1998**, *1393*, 1–18.
- (7) Werth, N.; Schuette, C. G.; Wilkening, G.; Lemm, T.; Sandhoff, K. *J. Biol. Chem.* **2001**, *276*, 12685–12690.
- (8) Cao, Z. M.; Petroulakis, E.; Salo, T.; TriggsRaine, B. *J. Biol. Chem.* **1997**, *272*, 14975–14982.
- (9) Leinekugel, P.; Michel, S.; Conzelmann, E.; Sandhoff, K. *Hum. Genet.* **1992**, *88*, 513–523.
- (10) Conzelmann, E.; Sandhoff, K. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 3979–3983.
- (11) Mark, B. L.; Mahuran, D. J.; Cherney, M. M.; Zhao, D. L.; Knapp, S.; James, M. N. G. *J. Mol. Biol.* **2003**, *327*, 1093–1109.
- (12) Hepbildikler, S. T.; Sandhoff, R.; Kolzer, M.; Proia, R. L.; Sandhoff, K. *J. Biol. Chem.* **2002**, *277*, 2562–2572.
- (13) Wright, C. S.; Li, S. C.; Rastinejad, F. *J. Mol. Biol.* **2000**, *304*, 411–422.
- (14) He, Y.; Macauley, M. S.; Stubbs, K. A.; Vocadlo, D. J.; Davies, G. *J. Am. Chem. Soc.* **2010**, *132*, 1807–+.
- (15) Bottoni, A.; Mischione, G. P.; Calvaresi, M. *Phys. Chem. Chem. Phys.* **2011**, *13*, 9568–9577.
- (16) *GaussView 4.0*; Gaussian, Inc.: Pittsburgh, PA, 2004.
- (17) Woods-Group. GLYCAM Web. Complex Carbohydrate Research Center. [http://glycam.crc.uga.edu/CCRC/biombuilder/biom\\_index.jsp](http://glycam.crc.uga.edu/CCRC/biombuilder/biom_index.jsp).
- (18) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727–748.
- (19) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. *Proteins: Struct., Funct., Genet.* **2003**, *52*, 609–623.
- (20) Yong duan, C. W.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (21) (a) Kirschner, K. N.; Woods, R. J. *J. Phys. Chem. A* **2001**, *105*, 4150–4155. (b) Kirschner, K. N.; Woods, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10541–10545. (c) Basma, M.; Sundara, S.; Calgan, D.; Vernali, T.; Woods, R. J. *J. Comput. Chem.* **2001**, *22*, 1125–1137.
- (22) Asensio, J. L.; Jimenezbarbero, J. *Biopolymers* **1995**, *35*, 55–73.
- (23) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, H. M.; Wang, B.; Pearlman, D. A.; et al. *Amber 8*; University of California: San Francisco, CA, 2004.
- (24) (a) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. *Biopolymers* **1992**, *32*, 523–535. (b) Izaguirre, J. A.; Catarello, D. P.; Wozniak, J. M.; Skeel, R. D. *J. Chem. Phys.* **2001**, *114*, 2090–2098.
- (25) Hammonds, K. D.; Ryckaert, J. P. *Comput. Phys. Commun.* **1991**, *62*, 336–351.
- (26) (a) Alberto, M. E.; Marino, T.; Ramos, M. J.; Russo, N. *J. Chem. Theory Comput.* **2010**, *6*, 2424–2433. (b) Himo, F. *Theor. Chem. Acc.* **2006**, *116*, 232–240. (c) Leopoldini, M.; Marino, T.; Michelini, M. D.; Rivalta, I.; Russo, N.; Sicilia, E.; Toscano, M. *Theor. Chem. Acc.* **2007**, *117*, 765–779. (d) Ramos, M. J.; Fernandes, P. A. *Acc. Chem. Res.* **2008**, *41*, 689–698.
- (27) (a) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170–1179. (b) Dapprich, S.; Komaromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct.—Theochem* **1999**, *461*, 1–21.
- (28) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
- (29) (a) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100. (b) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652. (c) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- (30) (a) Parandekar, P. V.; Hratchian, H. P.; Raghavachari, K. *J. Chem. Phys.* **2008**, *129*, 145101. (b) Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **2003**, *24*, 760–769.
- (31) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*, revision A.1; Gaussian, Inc.: Wallingford, CT, 2009.
- (32) Passos, O.; Fernandes, A. F.; Ramos, M. J. *Theor. Chem. Acc.* **2011**, *129*, 119–129.
- (33) Bras, N. F.; Fernandes, P. A.; Ramos, M. J. *J. Chem. Theory Comput.* **2010**, *6*, 421–433.