Theor Chem Acc (2011) 129:119–129
DOI 10.1007/s00214-011-0904-1

REGULAR ARTICLE

Theoretical insights into the catalytic mechanism of β-hexosaminidase

Óscar Passos · Pedro Alexandrino Fernandes · Maria João Ramos

Received: 24 September 2010 / Accepted: 5 February 2011 / Published online: 26 February 2011

Abstract The enzyme (β-Hex) has a specific role in the degradation of a particular type of glycolipid, the GM2 ganglioside. This specificity is also manifested by its unusual mechanism of substrate-assisted catalysis, which is considered to be an alternative pathway for the breakage of glycosidic bonds. We have studied its catalytic mechanism using a small model of the enzyme: substrate Michaelis complex with DFT and MP2/MP3 methods. The results reflect the intrinsic chemical reactivity of the active site, decoupled from the long-range enzyme electrostatic field. The mechanism supports, and adds further atomic detail, on the earlier mechanistic suggestions based on experimental data. Moreover, we also have compared the geometry and full potential energy surface obtained with nine different exchange–correlation functionals. It was surprising that the B3LYP activation energies were lower than the ones from some of the hybrid meta functionals (known by their excellent performance on kinetics) by as much as 10 kcal/mol and lower than the MP2 energies by more than 12 kcal/mol. It is known that B3LYP underestimates barriers but underestimations of this extent are unusual and surprising. The effect of the theoretical method on the geometry, usually supposed to be less significant, had in some cases a relevant influence in the mechanism. Therefore, a more thoughtful choice should be made when choosing a methodology for geometry optimization.

Keywords β-Hexosaminidase · Substrate-assisted catalysis · Mechanism · DFT · Benchmark

1 Introduction

The spontaneous hydrolysis of cellulose and starch has a half-life time of 5 million years, and the glycosidic bond between two glucoses is the most stable linkage found in natural biopolymers. Even though there is a class of enzymes (glycosidase) that can accomplish the break of this linkage with rate constants up to 1,000 s^{-1}, they catalyze the hydrolysis of polysaccharides into simple units as well as participate in the synthesis of new oligosaccharides. Glycosidases can accelerate a hydrolytic process up to 10^7-fold and belong to one of the most catalytically efficient group of enzymes [1].

The hydrolysis of glycosidic bonds can follow two mechanisms: one retaining the anomeric carbon configuration (α or β is conserved) and the other inverting this configuration (the configuration shifts from α to β or vice versa). Regardless the mechanism, the hydrolysis is believed to involve the formation of an oxocarbenium-ion-like transition state and is assisted by the carboxylate groups of two active-site residues (Asp or Glu). One of these residues is protonated and acts as a proton donor while the other, deprotonated, stabilizes the transition-state positive charge through nucleophilic addition to the oxocarbenium ion. The reaction occurs with inversion or retention configuration when these two residues are at a distance of about 10 ± 2 or 5.5 Å, respectively [2].

The formation of an oxocarbenium-ion features in most hydrolytic reactions. However, there are very interesting exceptions, such as the disease-related β-hexosaminidase, which develops its physiological route by a substrate-assisted

Electronic supplementary material The online version of this article (doi:10.1007/s00214-011-0904-1) contains supplementary material, which is available to authorized users.

Ó. Passos · P. A. Fernandes · M. J. Ramos (✉)
Requimte, Faculty of Sciences of Porto,
Rua do Campo Alegre S/N, 4169-007 Porto, Portugal
e-mail: mjramos@fc.up.pt

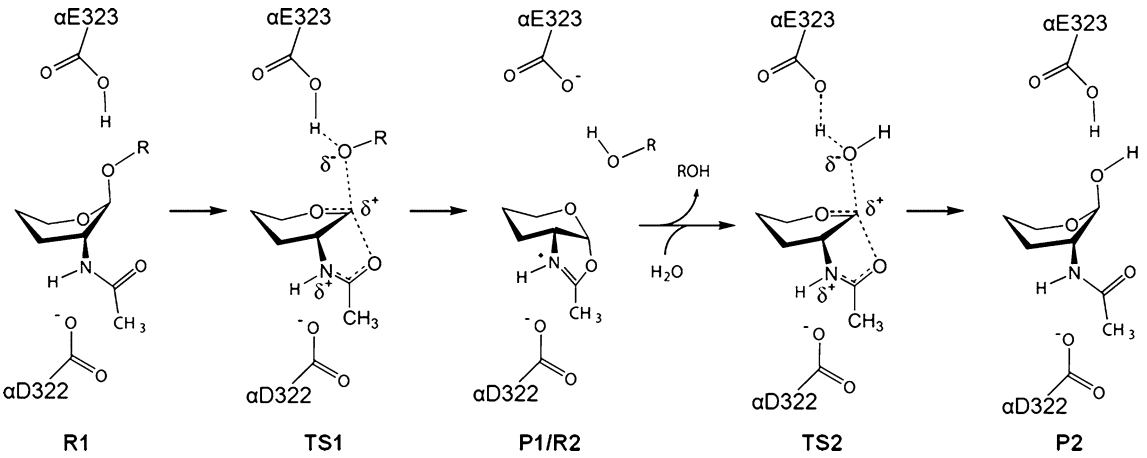

Fig. 1 Assisted catalysis mechanism proposed by Mark et al. [3] for β-hexosaminidase

catalytic mechanism, where it is believed that the *N*-acetyl of the substrate takes the place of the nucleophilic Asp/Glu, attacking the anomeric carbon and forming an oxazolinium ion intermediate as showed on Fig. 1 [3, 4].

The human β-hexosaminidase (β-Hex), a member of the family 20 hydrolases (EC 3.2.1.52) is a dimeric enzyme composed by two subunits, α and β, that share 60% of their amino acid sequence. There are three isoforms in which the subunit dimerizes, HexA ($\alpha\beta$) and HexB ($\beta\beta$), which are detectable at comparable amounts in normal human tissue, and HexS ($\alpha\alpha$) which is present in small amounts in tissues from patients suffering from Sandhoff disease [5, 6]. β-Hex participates in the degradation of glycoproteins, glycolipids, and proteoglycans through the removal of terminal β-glycosidically linked *N*-acetylglucosamine or *N*-acetylgalactosamine [7]. Its catalytic mechanism is poorly understood at the atomic level and is generally described through the paradigmatic mechanism of the glycoside hydrolase family, with the peculiarity of its substrate autocatalysis capability.

The aim of our work was to study and understand the atomic-level intrinsic reactivity of the β-Hex catalytic motif, assigning a catalytic role for each participants and eventually, giving a stronger theoretical background to the postulated mechanism [8]. For that purpose, we have kept the active-site model to a minimum, to separate clearly the intrinsic reactivity of the substrate and active site from the modulating interactions emanating from the enzyme scaffold [9]. Both steps of the reaction mechanism were considered at this stage. In addition to our primary objective, a large number of DFT functionals of different sophistication (M-GGA, H-GGA and HM-GGA) with different basis functions were tested in the calculations, together with high-level wave function methods (MP2 and MP3) to compare the influence of the functionals in the results of enzymatic studies. DFT results have a non-negligible dependence on the specific functional chosen for the study. DFT is widely used by many theoretical chemists working in enzymatic catalysis, due to its widespread applicability and accuracy, allowing for the prediction of important electronic properties with a much lower computational cost than post-Hartree–Fock methods [10]. Therefore, it is becoming increasingly important to understand and predict their influence on the results (both geometric and energetic) of catalytic mechanism studies.

There are five families of DFT functionals. Local (spin) density approximation (L[S]DA) is the first, the simplest, and depends only on the electronic density. It is not very adequate to describe chemical systems. LSDA explicitly includes the electron spin. The second uses the generalized gradient approximation (GGA), depending on the electronic density and on the gradient of the electronic density. GGA represent an evolution from LSDA but are still not accurate enough for a correct description of many chemical aspects. Meta GGA (M-GGA), the third one, was developed from GGA by including additional semi-local information beyond the first-order density gradient contained in the GGAs. M-GGAs depend also on the kinetic energy density; the fourth include hybrid GGA (H-GGA) and hybrid meta GGA (HM-GGA), including some percentage of HF exchange. H-GGA becomes a very popular choice in quantum chemistry and is now widely used; HM-GGA is a group of functionals that are derived from the M-GGA methods and are under active development. HM-GGA functionals represent an improvement over the previous formalisms, particularly in the determination of barrier heights and atomization energies [10]. Functionals of the last two families will be evaluated in this work. Nevertheless is important to mention an emerging new family of double-hybrid density functionals (DHDF). They are based on the existing hybrid functionals but replace a part of the semi-local GGA correlation by a second-order Møller–Plesset

type perturbation theory correction. Another important aspect that has been developed is inclusion of long-range van der Waals dispersion energy (DFT-D), which can be separately calculated (empirically) and added to the final DFT energy [11].

2 Methods

The model system used to simulate the reaction mechanism was built from the crystallographic structure of a human β-Hex (PDB ID: 2GK1, resolution of 2.8Å [4]). We deleted most of the enzyme and modeled the substrate from the substrate-analog inhibitor present in the crystallographic structure. The system included an *N*-acetylgalactosamine (GalNAc) molecule with a methyl group bound to the glycosidic oxygen and two molecules of ethanoic acid, one protonated and other deprotonated, modeling the side chains of the catalytic residues αE323 and αD322. The positions of the catalytic residues in our model were taken from the crystallographic structure. No restraints were introduced in the atomic positions during the geometry optimizations. These were performed at the B3LYP/6-31G(d) level and did not result in significant changes in their orientation. We have also checked that the truncation error of the basis set used for geometry optimization had no significant influence in the final energies. For that we have re-optimized all B3LYP/6-31G(d) structures with the B3LYP/6-311+G(d, p) level. The geometric differences were very small, and the energy differences between the B3LYP/6-311++G(2d, 2p)//B3LYP/6-311+G(d, p) and the B3LYP/6-311++G(2d, 2p)//B3LYP/6-31G(d) levels had a maximum (absolute) value of 0.8 kcal/mol and an mean unsigned difference of 0.45 kcal/mol. Further details can be found in support information. The system was thus completely decoupled of any enzymatic, stereochemical, or electrostatic influence beyond the active site; the chemical properties of the reaction were tested and analyzed without the influence of the natural environment provided by the enzyme, apart from the catalytic motif. Such strategy is perfect for the study of the intrinsic reactivity of the active site. Both reaction steps were considered—the substrate-assisted glycosidic bond cleavage with protonation by αGlu323 and the following water attack to the anomeric carbon (C1). The same general procedure was applied to their respective starting structures. Free geometry optimizations with vibrational frequency calculations were done in each transition-state structure along the mechanism. Intrinsic reaction coordinate calculations followed by further tighter geometry optimizations were performed to generate the correct reactants and products. In the second step of the mechanism, the methyl group released from the substrate (remember that this molecule modeled the leaving sugar and was a result of the previous glycosidic breakage and oxygen protonation) was replaced by a hydrogen, creating a water molecule in the vicinity of the anomeric carbon. The positions of the oxygen and native hydrogen in the water molecule were maintained from the first-step product. The second hydrogen was added along the O–CH_3 bond of methanol. This structure was used as reactant for the second step.

All geometry optimizations were performed in vacuum with the basis set 6-31G(d). Three different basis sets were used for the energy calculations of the optimized structures, namely, 6-31G(d), 6-31+G(d), and 6-311++G(2d, 2p). Recent studies suggest that the difference in the energy of activation and reaction calculated with a basis set more complete than 6-311++G(2d, 2p) is smaller than 0.5 kcal/mol for this kind of systems [12]. The computational effort involved when employing larger basis sets is not necessary for the purposes of this study. It is well known that activation barriers, as well as other properties, are highly dependent on the density functional, the ones with a higher fraction of HF exchange providing larger and more accurate barriers. This important aspect was investigated here, to understand the influence of the specific functional on the PES of this catalytic mechanism. The geometry and the energy of all the structures along the whole reaction mechanism were recalculated with 9 functionals (B3LYP, B97-2, B98, B1B95, PBE1PBE, BB1K, MPW1K, MPWB1K and M06-2X). The PES was also calculated with the wave function methods MP2 and MP3 (the last using the MP2 geometries). All the calculations were done with the Gaussian03 software package [13] except the ones using M06-2X functional that were done with Gaussian09 software package [14].

3 Results

There are two main objectives in this study; (1) understand the active-site intrinsic reactivity in the substrate-assisted mechanism of β-hexosaminidase; (2) comprehend the influence of the density functional in the energetic and geometric description of this kind of chemical reactions. We have considered the full reaction mechanism, which is accomplished in two steps. The first step is the breaking of the glycosidic bond and the nucleophilic attack of the anomeric carbon by an *N*-acetyl oxygen. The second step is the water attack to the anomeric carbon, with the subsequent breakage of the covalent bond formed in the first step.

The catalytic mechanism was studied using a wide range of density functionals. However, the structural and energetic results were quite similar among them all. Therefore, only the B3LYP mechanism will be discussed, for simplicity. The energies were calculated with three sets of

basis functions [6-31G(d), 6-31G+(d) and 6-311G++(2d, 2p)]. However, in this paragraph we will only discuss the more complete basis set results. Using, as a starting point, a 48 atoms model we optimized three structures for each step (reactant, transition state and product) in an overall of six structures corresponding to the full mechanism pathway (Fig. 2). It is crucial to understand that the comparison between these structures and those obtained at the active site of the enzyme is not pursued. The structural analysis will be focused on the chemical reactivity, which reveals important properties of the enzyme active site, but does not represent the whole catalytic machinery.

As proposed by Mark et al. [3, 4], the substrate binds to the enzyme with the sugar in the −1 subsite in a distorted sofa/boat conformation, placing the scissile bond in a pseudo-axial orientation (the optimized starting structure R1 has a 1S3 skew conformation which is in accordance to what is proposed). In the unroll of the reaction up to the transition state, this conformation would allow the C1, C2, C5, and O5 of the sugar in the −1 subsite to adopt a coplanar conformation necessary for an effective overlap between the O5 non-bonding lone pair electrons and the antibonding orbital at the deficient anomeric center of the postulated oxacarbenium-ion transition state. The DFT results have shown that C1 and the ring O describe an arc along the reaction pathway. As the glycosidic bond is broken and a new bond with the acetoamide oxygen is formed, arc moves from a distorced boat (1S3) to a chair (4C1) conformation in P1. This study thus provided a theoretical confirmation of the postulated conformational transitions along the catalytic mechanism of retaining glycosidases.

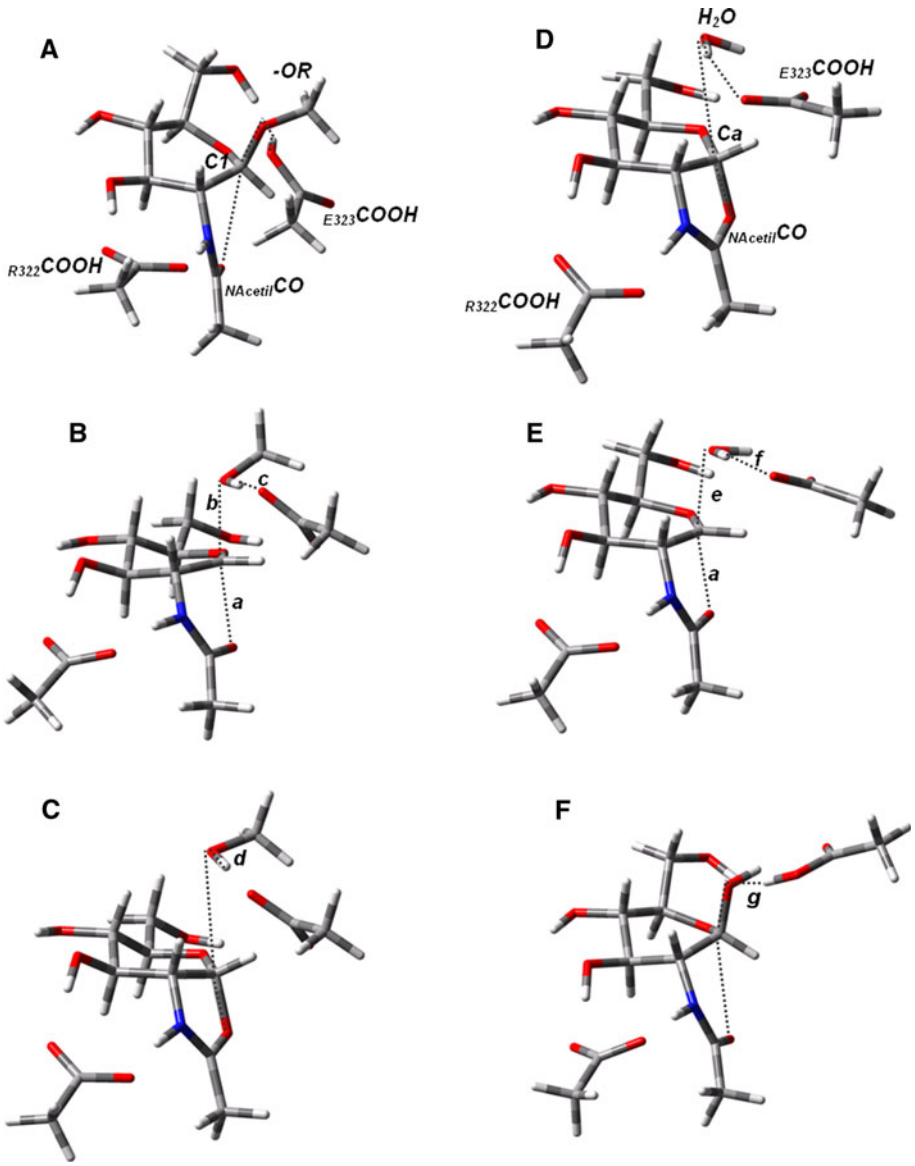

Fig. 2 The six stationary states along the mechanism pathway. Relevant distances are highlighted with dashed lines and their values are given in Table 1. **a** First-step reactant. **b** First-step transition state. **c** First-step product. **d** Second-step reactant. **e** Second-step transition state. **f** Second-step product

Table 1 Relevant distances along the catalytic cycle

Distance/Å	R1	TS1	P1	R2	TS2	P2
CO–C1 (a)	3.01	2.28	1.50	1.54	2.38	2.89
C1–OR (b)	1.44	2.21	3.09	–	–	–
COO–H (c)	0.99	1.45	1.63	–	–	–
E323COOH–O (d)	1.79	1.06	1.01	–	–	–
H2O–C1 (e)	–	–	–	3.26	2.26	1.45
HOH–OOC E323 (f)	–	–	–	1.89	1.80	1.00
HO–H (g)	–	–	–	0.99	0.99	1.73

The superscripts refer to the distances shown in Fig. 2. *C1* anomeric carbon, *CO* substrate amide carbonyl, *OR* leaving sugar, *COO* and *COOH* carboxylate groups of E322 or E323, *O* glycosidic oxygen

Figure 2 shows that the transition state has in fact a coplanar conformation, halfway along the arc described by C1. However, the TS apparent ^{4}E Envelope conformation is distorted; it is really a $^{4}H_3$ half-chair conformation. In Table 1, we can see that as the length of the glycosidic bond increases from the reactants of the first step (R1, 1.44 Å) to the first transition state (TS1, 2.21 Å), the acetoamide oxygen (CO) becomes nearer C1 (2.96 Å in R1 and 2.29 Å in TS1). The proton has already been transferred from E323 to the leaving sugar at this stage.

Upon the cleavage of the glycosidic bond, the sugar ring relaxes to the $^{4}C_1$ chair conformation (P1). This intermediate is characterized by the formation of an intramolecular covalent bond inside the substrate (Figs. 2c and 1d) between the CO group and the C1 carbon atom, with a bond length of 1.54 Å. According to the usual mechanism of several glycosidases, this bond is formed between a carboxylate of an Asp/Glu residue of the enzyme and C1. In both cases, the role of this bond is the stabilization of the positive charge formed on the C1 carbon atom during the glycosidic bond cleavage. The intrasubstrate linkage classifies this type of mechanism as substrate-assisted catalysis.

Hydrolysis of this intermediate leads to the formation of a product complex in a skew boat conformation and finally product release. In our model, this corresponds to the second step. The reactant for this step (R2) was built from P1 replacing the methanol for a water molecule (the methyl was replaced by a hydrogen along the carbon–oxygen bond, preserving the positions of the remaining atoms). It is believed that in this step, the C1 carbon atom traces the reverse arched path, as it breaks the bond with the oxazolinium ion and forms a covalent bond with a suitably positioned water molecule. Such expectations were confirmed by the theoretical results. As the water molecule approaches and reacts with C1, the sugar ring shifts from a $^{4}C_1$ back to a $^{1}S_3$ conformation. The second-step transition state (TS2) shares the TS1 $^{4}H_3$ conformation.

Another interesting aspect is the movement of residue E323. It is easy to see in Fig. 2 that the methyl terminus of this molecule is getting further away from residue D322. The fact that D322 does not participate directly in the reaction might be the reason for the almost inexistent displacement of this molecule relatively to the substrate. No position restraints were imposed and thus such movements were possible if they were favorable.

It is important to mention that the inclusion of D322 and E323 residues' side chains ($C\beta H_3COO^{-}$—D322) and $C\gamma H_3COOH$—E323, respectively) are essential in our model. In the enzyme, the importance of these catalytic residues seems to be obvious to the reaction mechanism, for transition-state stabilization by D322 (through a non-covalent ion–ion interaction) and acid/base general catalysis by E323. However, the importance of these residues to the sugar ring conformation was not anticipated beforehand. An analysis of the structure of the protein shows that the substrate must enter the active site already in a boat/skew conformation. The conformational rearrangement from chair to boat/skew in the biological environment might be assigned to the activator protein, the catalytic residues, or the surrounding residues of the active site, being unclear which one is the most relevant factor. It is important to notice that the necessary rearrangement, if carried out by the enzyme, would decrease the enzymatic efficiency.

The PES for this reaction was calculated at the B3LYP/6-311++G(2d, 2p)//B3LYP/6-31G(d). The first step has activation energy of 24.4 kcal/mol and a reaction energy of 13.5 kcal/mol. These energies are higher than those of the prototypical glycosylases, which use two Asp/Glu residues for catalysis (Ea 26.0 kcal/mol and ΔR 3.1 kcal/mol) [12], calculated with similar molecular models. Two factors contribute for the difference: the formation of a strained five sided ring in the sugar molecule, which increases the energy of the transition state and the products, and the nucleophilicity of CO, which is weaker than the one of the prototypical carboxylate oxygen.

Having the starting structure, R1, as a reference, the second step has an activation energy of 19.0 kcal/mol and a reaction energy of −0.4 kcal/mol. This is clearly the opposite profile of the first step. The reaction barrier is small (5.5 kcal/mol, considering only the second step), as the starting reactant is already high in energy, and the high penalty in the reaction energy is totally recovered in the second step.

To understand and quantify the two factors that dramatically raise the reaction energy in the first step (in relation to the prototypical glycosidases), we have created a new system in which we have disconnected the nitrogen atom from the ring. The valencies of both the ring carbon and the nitrogen atom were completed with hydrogens. The nucleophilicity of the carbonyl should not be greatly affected by this. This system is free from the steric strain

caused by the ring, and thus, the differences in activation and reaction energies can be attributed to the nucleophilic power of the amide carbonyl (in comparison with the more typical Asp/Glu carboxylate). The results (also at the B3LYP/6-311++G(2d, 2p)//B3LYP/6-31G(d) level) show that the reaction energy in this unstrained system amounts to less than 1 kcal/mol, compared with +13.5 kcal/mol for the same reaction with the natural substrate. Therefore, the very large increase in the reaction energy is mainly due to the formation of a strained five sided ring. It is difficult to understand the evolutive advantage of substrate autocatalysis (if any). Autocatalysis may just be a consequence of the chemical nature of the substrate. The *N*-acetyl group sterically precludes the attack of the anomeric carbon by a nucleophilic glutamate, as it happens with typical glycosidases. Substrate autocatalysis may be the only viable option to stabilize the developing charge in that center. The enzyme scaffold (or even the substrate:enzyme:activator protein complex) must give rise to an incremental stabilization to overcome this significant penalty and assure such vital reaction.

It is clear that the limiting step is the first one. The energy profile is represented in Fig. 3. However, different density functionals and methods might result in different energetic profiles, which was the reason why we have recalculated the PES with a significant set of density functionals and post-Hartree–Fock methods.

In the following paragraph, we will focus on the comparison between density functionals. Due to the size of the system (48 atoms, 190 electrons), it was impossible to perform high-level post-Hartree–Fock calculations with enough accuracy to serve as reliable benchmarking values. Therefore, we will focus on the comparison between density functionals, taking B3LYP as the reference. Such reference was chosen because B3LYP is by far the most widely used density functional, and not due to any preassumption about its accuracy. In fact, B3LYP is known to underestimate energy barriers and several recent meta and hybrid meta density functionals were shown to outperform B3LYP in that respect.

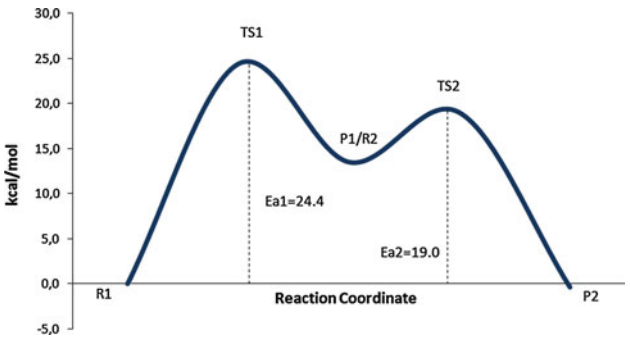

Fig. 3 PES for the studied reaction at the B3LYP/6-311++G(2d, 2p)//B3LYP/6-31G(d) level

The functionals were chosen from two different families (meta, hybrid and hybrid meta) in order to ensure a broad functional coverage. All the six structures along the reaction mechanism were re-optimized with each of the tested functionals. The differences in geometry were found to be minimal (except R2 optimized with MP2).

The results presented in Table 2 indicate the differences on the activation and reaction energies for both steps between B3LYP and the other functionals, using three different sets of basis functions.

The first important observation is that very large differences in activation energy were found between different functionals, the ones with a larger fraction of HF exchange giving higher energy barriers, and hybrid meta functionals giving larger barriers than the hybrid functionals. It is possible to see a correlation between the percentage of HF Exchange and the energetic barriers. To clarify this aspect, we have performed a series of single-point calculations over the B3LYP R1 and TS1 geometries, in order to analyze the relation between these two variables, comparing the first-step activation energy to the results of density functionals with different percentage of HF Exchange. For that we have chosen a larger number of functionals (seventeen) to form a group with a broader range of HF exchange percentage. The results plotted on Fig. 4 show that despite the existence of some outliners the relation between increased percentage of HF exchange and increased activation energy is clear. The linear fit is just illustrative, and there is no physical basis to expect a linear correlation. Moreover, many more factors are different between the functionals besides the percentage of HF exchange, and thus, the result cannot be attributed the HF exchange solely. Anyway, the correlation is evident. More data about these calculations can be found in support information.

B3LYP results in the lower of all the energy barriers, both for the first and for the second steps. The difference between the barriers of B3LYP and the hybrid meta functionals of the Truhlar group [15] in the first step amounts to 8–11 kcal/mol. Note that the latter are known for their accuracy in this kind of predictions. This observation emphasizes that the very popular B3LYP functional should be used with caution. Even though underestimations of 3–4 kcal/mol are known to occur, discrepancies up to 8 kcal/mol are beyond what we expected. It is somewhat frustrating that the system size does not allow for an estimation of a rigorous barrier, to realize whether the deviations are due to an underestimation of B3LYP, an overestimation of BB1K, MPWB1K and M06-2X or a combination of both.

In the second step, the discrepancies between the energy barriers are much smaller, but still significant (5–7 kcal/mol). Differences among density functionals are also much smaller in the reaction energies (average MUE of

Table 2 Activation and reaction energies for both reaction steps, calculated with nine density functionals and two post-Hartree–Fock methods

DFT method		First step						Second step					
		E_a (kcal/mol)			E_R (kcal/mol)			E_a (kcal/mol)			E_R (kcal/mol)		
Functionals	% HF	SB	MB	LB	SB	MB	LB	SB	MB	LB	SB	MB	LB
H-GGA													
B3LYP	20	28.2	25.9	24.4	17.9	15.0	13.5	20.0	19.7	19.0	0.4	−0.2	−0.4
B97-2	21	30.8	28.2	26.7	17.7	13.8	12.4	21.1	19.7	19.6	0.2	−1.4	−1.2
B98	22	31.1	28.8	27.4	16.5	13.0	11.5	21.2	20.1	20.2	−1.5	−2.6	−2.2
PBE1PBE	25	34.8	32.7	31.0	17.1	13.8	12.2	24.3	23.1	23.2	−2.1	−3.1	−2.6
MPW1K	43	32.3	29.8	28.3	17.8	13.9	12.5	22.4	20.9	20.9	−0.1	−1.7	−1.2
HM-GGA													
B1B95	25	37.2	34.8	33.1	20.1	16.3	14.6	26.6	25.0	25.1	0.7	−0.7	−0.2
BB1K	42	33.8	31.3	29.9	19.4	15.3	14.4	24.2	22.4	22.8	1.2	−0.6	0.3
MPWB1K	44	38.1	35.5	33.9	20.8	16.9	15.2	27.1	25.4	25.1	1.4	−0.1	0.3
M06-2X	54	39.5	37.2	35.4	20.8	18.3	15.7	28.5	28.0	26.8	3.7	3.7	3.0
Møller-Plesset method													
MP*n*													
MP2		37.2	33.9	33.1	19.7	17.4	16.9	30.8	28.5	29.8	6.4	4.1	7.5
MP3[a]		40.4	37.6	36.8	19.8	17.8	17.1	34.9	33.8	34.0	6.5	7.1	7.2

Three different basis sets were used. For the more complete basis set (LB), the energies were calculated using the geometries optimized with the smaller basis set, 6-31G(d), The geometries were re-optimized with every functional. All energies are related to the first-step reactant energy for each functional/method

[a] The structures used in the MP2 and MP3 energy calculations were optimized at the MP2 level

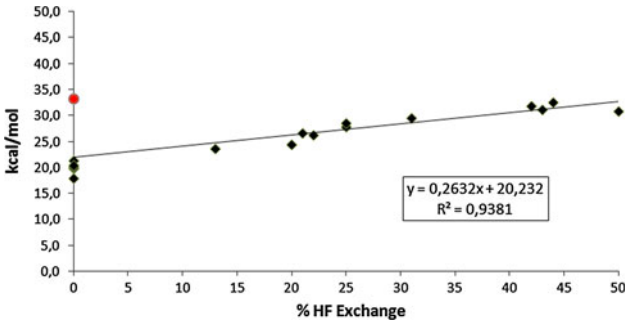

Fig. 4 First-step Ea versus percentages of HF Exchange from eighteen different DFT functionals. The structures used to calculate the activation energy were R1 and TS1 at the B3LYP/6-31G(d) level for all functionals. The single-point calculations were done with the more complete 6-311++G(2d, 2p) basis set. The *red mark* in the chart represents the VSXC result, as a severe outliner it was not used in the linear fitting

1.07 kcal/mol for the first and second step, taken together) clearly placing the tone of the problem in the energy barriers.

It was expected that different basis sets would result in different values for the activation and reaction energies. However, what was surprising to see was that the difference between the basis sets employed seems to be the same for all the functionals. The difference in energy between the larger basis set (LB) and the smaller basis set (SB) is in absolute average value 3.8 kcal/mol in activation energies and 4.6 kcal/mol in reaction energies. The surprise was that the average deviation of these values among the functionals was as small as 0.3 kcal/mol for the activation energies and 0.8 kcal/mol for the reaction energies. This behavior resembles that implicitly used in composite theoretical methods (e.g., G2, G3, etc.), where the additivity of basis set effects is assumed. Here, it is verified to an extent that may justify the use of a more time-consuming functional (such as MPWB1K) just for single-point free energy calculations rather than geometry optimization with the inclusion of polarization and diffuse functions for which much less time-consuming density functional would be suited (such as B3LYP). The calculation time for each functional is available in support information.

Before discussing the PES for the reaction with the several tested functionals, we will focus on their structural differences. As noted before, starting from the B3LYP optimized structures we performed a geometry re-optimization with all the functionals. The comparison between all of them would be tedious and thus we picked up only the cases where larger differences were expected. These correspond to the functionals with higher fraction of HF exchange (MPW1K, BB1K, MPWB1K and M06-2X), which gave rise to the largest energetic differences in relation to B3LYP (as shown in Table 2), together with MP2, a wave function–based method. No positional constraints were used in the geometry optimizations.

Table 3 Most important distances for the stationary points optimized with several methods and density functionals

Structure	First step						Second step					
	R1		TS1		P1		R2		TS2		P2	
Distances (Å)	CO–C1 (a)	C1–OR (b)	CO–C1	C1–OR	CO–C1	C1–OR	CO–C1	H_2O–C1	CO–C1	H_2O–C1	CO–C1	H_2O–C1
B3LYP	3.01	1.44	2.28	2.21	1.50	3.09	1.54	3.26	2.38	2.26	2.89	1.45
B97-2	3.02	1.43	2.28	2.20	1.48	3.13	1.52	3.32	2.36	2.26	2.90	1.44
B98	3.02	1.44	2.26	2.20	1.49	3.09	1.53	3.28	2.35	2.28	2.90	1.44
PBE1PBE	2.97	1.43	2.19	2.18	1.48	3.06	1.51	3.22	2.29	2.25	2.86	1.43
MPW1K	2.91	1.42	2.21	2.16	1.46	3.08	1.49	3.20	2.31	2.22	2.89	1.42
B1B95	2.93	1.43	2.20	2.19	1.48	3.07	1.51	3.10	2.33	2.20	2.86	1.43
BB1K	2.91	1.42	2.20	2.16	1.47	2.98	1.50	3.09	2.35	2.20	2.87	1.42
MPWB1K	2.90	1.42	2.19	2.16	1.46	2.96	1.49	3.08	2.35	2.19	2.86	1.42
M06-2X	2.90	1.43	2.17	2.22	1.46	3.02	1.48	3.10	2.41	2.26	3.04	1.43
MP2	2.86	1.44	2.19	2.18	1.49	3.24	1.45	3.40	2.42	2.21	2.88	1.44

The distances in the table are those represented in Fig. 2

Table 3 summarizes the most relevant interatomic distances and Table 4 the root mean square deviations (RMSd) in the atomic positions. In all stationary points, the sugar structure is strictly preserved has shown on Fig. 6. It is clear that the sugar ring conformation undergoes what has been previously suggested in the literature [3], skew [$^1S_3$, R1] → half-chair [$^4H_3$, TS1] → chair [$^4C_1$, P1, R2] → half-chair [$^4H_3$, TS2] → skew [$^1S_3$, R2], independently of the method or functional employed. Relevant distances are also preserved, such as the pseudo-glycosidic bond, the CO–C1 distance and the C1–H_2O distance. However, a closer look at Table 3 shows that B3LYP results in the higher distances among all DFT structures and MP2 results in the largest geometric differences from B3LYP. Non-covalent distances have a variation less than 0.26 Å, which is not very significant. Covalent bonds show variations up to 0.06 Å.

Table 4 RMSDs between the structures of the density functionals and MP2 taking B3LYP structures as the reference

Structure	First step			Second step		
RMSD (Å)	R1	TS1	P1	R2	TS2	P2
B97-2	0.38	0.68	1.09	0.13	0.25	0.28
B98	0.33	0.59	1.04	0.08	0.30	0.29
PBE1PBE	0.26	0.60	1.24	0.16	0.28	0.24
MPW1K	0.90	0.75	1.00	0.15	0.26	0.12
B1B95	0.90	0.25	1.07	0.19	0.30	0.18
BB1K	0.52	0.57	1.14	0.10	0.15	0.08
MPWB1K	0.52	0.32	1.19	0.05	0.13	0.02
M06-2X	0.38	0.29	0.88	0.30	0.38	0.66
MP2	1.16	0.42	1.22	0.60	0.21	0.53

The sugar rings of each pair of structures were superimposed

The higher (but still very small) RMSDs in Table 4 are mainly due to translations of the protein residues in relation to the substrate. This can be seen in Fig. 3d, e and f. The similarity between all the structures shows that there is no close relation between the RMSDs and the large energetic differences found within the density functionals (see below). Comparing the results shown in Tables 3, 4 and Fig. 6, we can see that almost identical geometries (e.g. the second-step species from BB1K and B3LYP) have significant energy differences. This is particularly evident in the first step calculated with B3LYP, BB1K, MPWB1K.

An important aspect though is the difference between B3LYP and post-Hartree–Fock methods in the second step. In the first step, the geometries are quite similar, while in the second step the situation is quite different. Again a closer look at Fig. 5d (MP2 structure) shows that D322 removes the *N*-acetyl nitrogen proton. The average length between this proton and the acid oxygen for the DFT methods is around 1.65 Å and in MP2 is 1.07 Å. If we use the B3LYP geometry to calculate the MP2 energy (which are essentially equivalent apart from the proton position), we can see that the transfer is responsible for a stabilization of 3.1 kcal/mol. With MP3, this stabilization corresponds to 5.0 kcal/mol energy. Because only the R2 structure is affected, the activation energy and reaction energy for the MP*n* levels present the same energetic variation.

4 Discussion

The direct comparison between our model and the enzymatic mechanism will not be pursued here. This discussion will only focus on the intrinsic reactivity of the active site of β-Hex. In this active-site model, the reaction mechanism

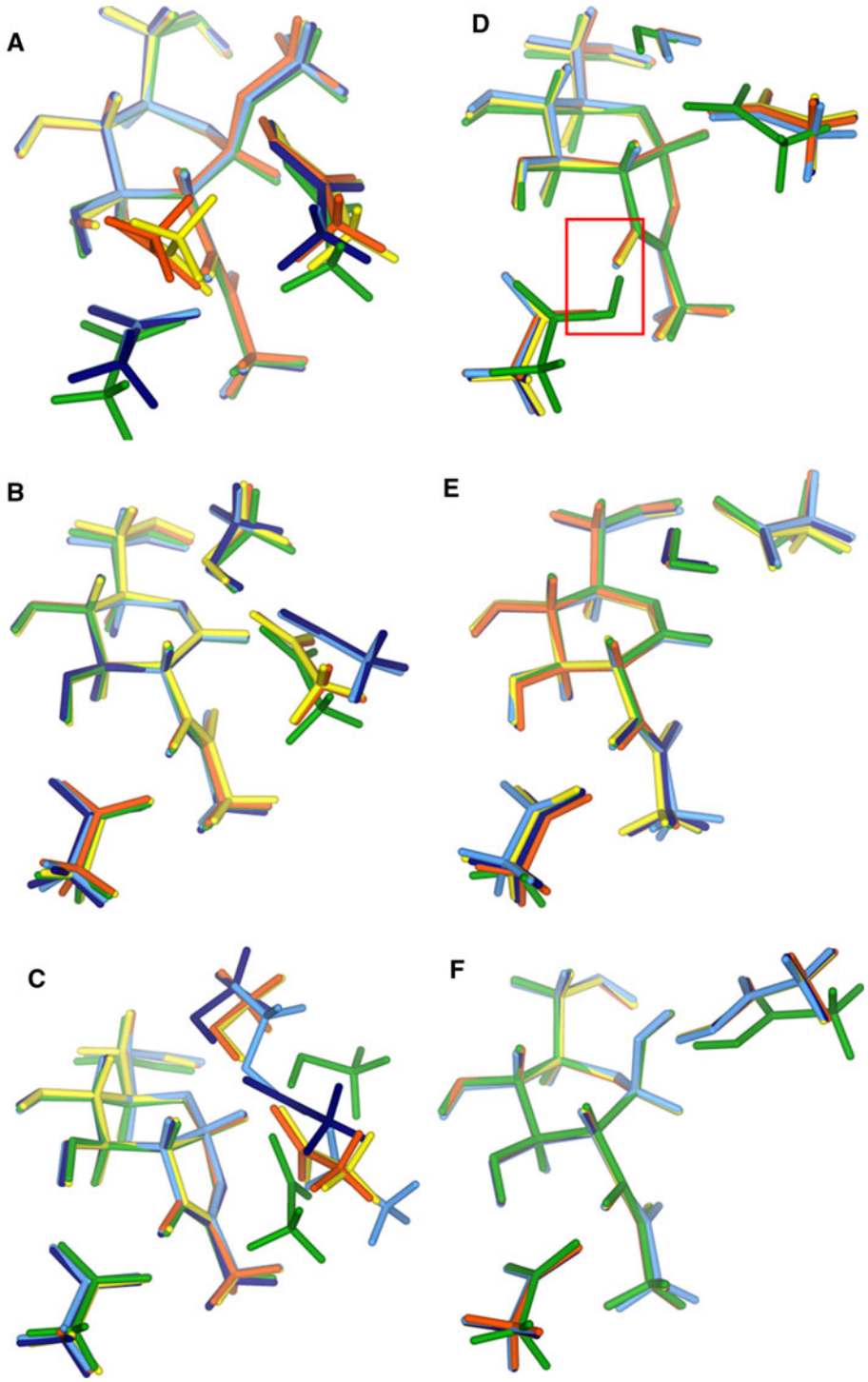

Fig. 5 The six stationary points along the PES calculated with B3LYP (*dark blue*), MPW1K (*light blue*), BB1K (*Orange*), MPWB1K (*Yellow*) and MP2 (*Green*) with 6-31G(d) basis set. **a** First-step reactant. **b** First-step transition state. **c** First-step product. **d** Second-step reactant. **e** Second-step transition state. **f** Second-step product. Only result from five functionals was represented for simplicity. The chosen five presented the most significant differences in relation to B3LYP

undergoes the postulated enzymatic mechanism proposed by Mark et al. [3]. The six conformations of the substrate ring mechanism are represented in Fig. 6. The sugar main ring shifts between skew and chair conformations. A five sided ring between the *N*-acetyl carboxide and the anomeric carbon is formed and stabilizes the intermediate (P1/R2) simultaneously with the cleavage of the pseudoglycosidic bond, both representing the first reaction step. In the second step, the breakage of the five sided ring and nucleophilic attack of the anomeric carbon by a water molecule will complete the reaction, resulting in a trimmed single sugar molecule. The five sided ring assumes the role of a second catalytic residue. The reaction proceeds via a pathway where the substrate stabilizes itself intramolecularly, even though the presence of a second molecule of acetic acid is crucial to maintain the boat conformation in

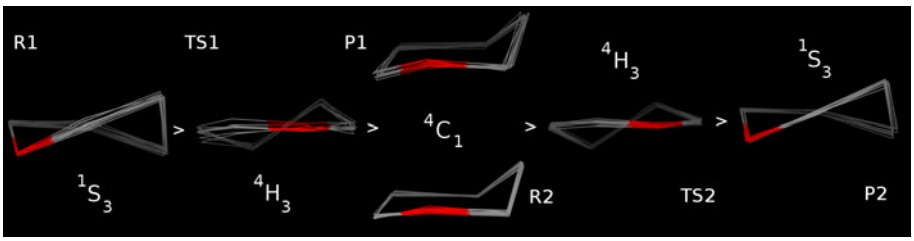

Fig. 6 The sugar ring conformations along the reaction coordinate for all optimized structures. Atoms were removed for clarity

the first-step reactant (data not shown), which may be essential for the substrate to enter the enzyme active site [3, 4].

Even though this catalytic mechanism and conformational changes have been consistently obtained with each of the density functionals tested here, significant quantitative differences were found. In fact, the computational chemist is currently confronted with a myriad of density functionals, with different merits and limitations. There is no consensus about the best functional, and different functionals can give different results in different systems. Due to their parametric nature, only general trends can be devised. However, the prediction of their accuracy for a specific system/property based on general trends is not the best way of choosing a density functional. Only benchmark studies will indicate which is the best functional to describe a specific chemical system/property. Moreover, the knowledge of the subtleties of the methods is very important for future developments. Benchmarking helps to build up "model chemistries".

In the description of these phenomena, B3LYP has been the most widely used density functional and is very easy to work. It is fast, stable and has a reasonable performance and well-known pitfalls. Despite all the good results that are associated with B3LYP, it is clear that it underestimates energy barriers. For the nine DFT hybrid functionals tested, B3LYP returned the lowest values. On the other hand, MPWB1K and M06-2X returned the highest values. Very large energetic differences were found between them (up to 11 kcal/mol). To rank the accuracy of the density functionals, we would need a high-level reference value, such as CCSD(T) with an extrapolation to a complete basis set, which is beyond the capacities of our own 800-core supercomputer (and of almost any supercomputer in the world with a system of this size!). It is still unclear whether the very large difference between B3LYP and more modern density functionals is due to an underestimation of B3LYP, an overestimation of the more modern functionals, or a combination of both (even though recent benchmarking has highlighted the adequacy of the more modern for kinetics).

Increasing amounts of HF exchange resulted in increasingly high activation energies. The highest energy result for DFT is close but still lower than the MP2 results.

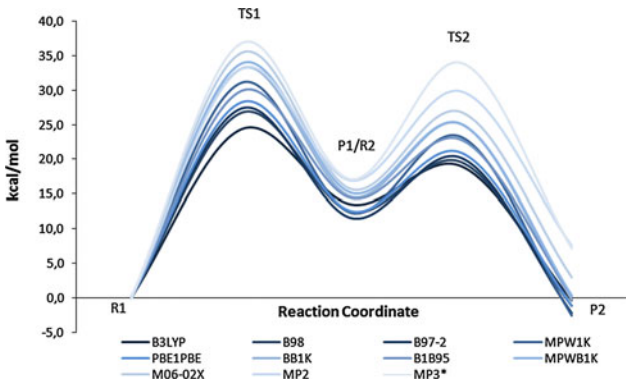

Fig. 7 PES for the reaction mechanism calculated with B3LYP, B97-2, B98, PBE1PBE, MPW1K, B1B95, BB1K, MPWB1K, M06-2X, MP2, and MP3 methods

The overall energetic results are presented on Fig. 7, in which it is possible to realize that, despite the translation of the energy of the stationary points, the use of different theoretical methods do not change the shape of the PES for the reaction.

5 Conclusion

The present study gathers strong evidences that support the catalytic mechanism suggested for this enzyme. It shows that the active site of β-Hex possesses the perfect characteristics to catalyze such a mechanism. Intrinsic activation and reaction energies were derived, and the catalytic role of the enzyme can be inferred through the difference between the active-site model and the observed kinetics. The ring strain and nucleophilic power were calculated and compared with the more standard reaction mechanism for these enzymes. The nucleophilic strength of the substrate seemed equivalent to a side chain carboxylate, but the ring strain of the substrate presented a large energy penalty for the reaction. In fact, the ring involved in the substrate autocatalysis mechanism penalizes the enzymatic efficiency by 13 kcal/mol [12], and is not clear why evolution has driven this enzyme through this particular structural way. Steric strain for a substrate conformation adequate for two-carboxylate catalysis or the existence of another rate-limiting step along the catalytic cycle are just a few hypotheses that merit further studies. The PES was studied through several

methodologies. Its shape was the same whichever method was used but very significant differences in the values of the activation barriers calculated with different density functionals were found, highlighting the importance of an appropriate choice at this level.

References

1. Wolfenden R, Lu XD, Young G (1998) Spontaneous hydrolysis of glycosides. J Am Chem Soc 120(27):6814–6815
2. Zechel DL, Withers SG (2000) Glycosidase mechanisms: anatomy of a finely tuned catalyst. Acc Chem Res 33(1):11–18
3. Mark BL et al (2001) Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase. J Biol Chem 276(13):10330–10337
4. Lemieux MJ et al (2006) Crystallographic structure of human beta-hexosaminidase A: interpretation of Tay-Sachs mutations and loss of G(M2) ganglioside hydrolysis. J Mol Biol 359(4):913–929
5. Proia RL, Soravia E (1987) Organization of the gene encoding the human beta-hexosaminidase alpha-chain. J Biol Chem 262(12):5677–5681
6. Korneluk RG et al (1986) Isolation of CDNA clones coding for the alpha-subunit of human beta-hexosaminidase—extensive homology between the alpha-subunits and beta-subunits and studies on Tay-Sachs disease. J Biol Chem 261(18):8407–8413
7. Sandhoff K, Wassle W (1971) Purification and characterisation of 2 forms of human N-acetyl-beta-D-hexosaminidase. Hoppe-Seylers Zeitschrift Fur Physiologische Chemie 352(8):1113–1119
8. Leopoldini M et al (2007) The role of quantum chemistry in the elucidation of the elementary mechanisms of catalytic processes: from atoms, to surfaces, to enzymes. Theor Chem Acc 117(5–6):765–779
9. Himo F (2006) Quantum chemical modeling of enzyme active sites and reaction mechanisms. Theor Chem Acc 116(1–3): 232–240
10. Sousa SF, Fernandes PA, Ramos MJ (2007) General performance of density functionals. J Phys Chem A 111(42):10439–10452
11. Schwabe T, Grimme S (2007) Double-hybrid density functionals with long-range dispersion corrections: higher accuracy and extended applicability. Phys Chem Chem Phys 9(26):3397–3406
12. Bras NF, Ramos MJ, Fernandes PA (2010) DFT studies on the beta-glycosidase catalytic mechanism: the deglycosylation step. J Mol Struct-Theochem 946(1–3):125–133
13. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE et al (2004) Gaussian 03, Revision C. 02. Gaussian, Wallingford CT
14. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE et al (2009) Gaussian 09, Revision A. 02. Gaussian, Wallingford CT
15. Zhao Y, Truhlar DG (2008) Density functionals with broad applicability in chemistry. Acc Chem Res 41(2):157–167