Theor Chem Account (2009) 122:283–296
DOI 10.1007/s00214-009-0507-2

REGULAR ARTICLE

Docking and molecular dynamics studies on the stereoselectivity in the enzymatic synthesis of carbohydrates

Natércia F. Brás · Pedro A. Fernandes · Maria J. Ramos

Received: 12 May 2008 / Accepted: 6 January 2009 / Published online: 22 January 2009

Abstract Glycosidases constitute a vast family of enzymes that catalyze the breaking and formation of glycosidic bonds. The synthesized oligosaccharides, being crucial to life, are involved in many biochemical processes, particularly in the pharmaceutical and food industries. The proposed catalytic mechanism of retaining glycoside hydrolases (glycosidases) occurs via a double displacement mechanism involving a covalent glycosyl enzyme intermediate. During the transglycosylation reactions, the control of the stereoselectivity for the formation of the new bond remains a complicated problem in the chemical synthesis of oligosaccharides. In this paper, docking and molecular dynamics methods were used to study the second step of the mechanism of transglycosylation in retaining glycosidases from six microorganisms with known stereoselectivity. Using the natural substrates as donor and acceptor molecules, we were able to corroborate and provide structural information about the active site, the trapped monosaccharide acceptor and the bound intermediates during the step that precedes transglycosylation, as well as identify and understand the commonly displayed stereoselectivity by these glycosidases in nature. The information obtained with this procedure helps to recognize, explain and predict the stereoselectivity of the sugars studied. These kind of procedures can be used to improve the efficiency of large-scale industrial synthesis of a specific sugar.

Keywords Galactosidases · Oligosaccharides · Transglycosylation reaction · Glycosyl enzyme intermediate · Stereoselectivity

1 Introduction

Oligosaccharides and their derivatives play a key role as therapeutic agents, as diagnostic tools and in cosmetic and food compounds [1, 2]. Galactooligosaccharides (GOS) are special glycosides containing galactose and glucose molecules. These sugars are considered to be functional food ingredients because they stimulate the growth of bifidobacteria and lactobacilli, combining prebiotic properties beneficial to the human health with physico-chemical properties beneficial in food processing [3, 4]. For that reason, they have been receiving increased attention from both research and industrial fields.

Significant progress has been made in increasing the industrial potential of these molecules over the past few years. For the enzymatic production of oligosaccharides, there are two strategies: the use of glycosyltransferases or the use of glycosidases [5, 6]. The employment of transferases showed an excellent activity and stereoselectivity. However, their substrates displayed limited availability and higher cost than those of glycosidases. Moreover, the latter are more stable, readily available, and both the substrates and the enzymes are relatively cheap. For these reasons, these enzymes are promising in large-scale industrial use for the synthesis of oligosaccharides, namely the efficient production of GOS [7].

Electronic supplementary material The online version of this article (doi:10.1007/s00214-009-0507-2) contains supplementary material, which is available to authorized users.

N. F. Brás · P. A. Fernandes · M. J. Ramos (✉)
REQUIMTE, Departamento de Química,
Faculdade de Ciências, Universidade do Porto,
Rua do Campo Alegre, 687,
4169-007 Porto, Portugal
e-mail: mjramos@fc.up.pt

These enzymes can be classified into two major groups: retaining and inverting glycosidases or glycoside hydrolases. They have different mechanisms of reaction and a different spatial arrangement of catalytic groups [8]. Retaining glycosidases maintain the initial conformation on the anomeric carbon. The catalytic mechanism of retaining glycosidases was proposed about 50 years ago to proceed via a double displacement mechanism involving a glycosylation and a deglycosylation step and is shown in Fig. 1 [9]. The catalytic groups are two carboxylic acids, which can be in the side chains of either Asp or Glu residues. In retaining glycosidases, these groups are around 5.5 Å apart. In the first step, one of the carboxyl groups functions as a general acid catalyst protonating the glycosidic oxygen, while the other acts as a nucleophile, forming a covalent glycosyl enzyme intermediate. The transition state is believed to have a mainly dissociative character, i.e., the breakage of the glycosidic bond happens before the attack by the nucleophile group. The second step of the reaction involves the attack of the monosaccharide-enzyme linkage by a water molecule simultaneously with the transfer of a proton from water to the proton donor, in a reverse manner of the first step. If an acceptor sugar attacks the active intermediate of enzyme-substrate, instead of a water molecule, a new glycosidic bond is formed, and in this case the reaction is called a transglycosylation [10]. However, yields for these transglycosylation reactions are typically low because the product itself is a substrate for the enzyme and undergoes hydrolysis.

These enzymatic catalysts are extremely important for the synthesis and modification of carbohydrates because most of the relevant reactions require high degrees of chemo, regio and stereo-selectivity [11]. The complex structure of oligosaccharides makes chemical synthesis difficult, namely the production of glycosides with a mixture of various linkages (i.e., formation of 1–2, 1–3, 1–4 and 1–6 bonds) and both anomers (α and β) [6]. In this regard, the control of the stereospecificity and the regiospecificity of bond formation remains a challenging problem in the chemical synthesis of oligosaccharides [1, 12]. A solution for this unsolved problem would be very important, as industrially there is only interest in the oligosaccharide target. Several studies were made in order to clarify the origin and underlying determinants for the specific stereoselectivity of several enzymes, but many questions still remain to be answered in this field [6, 7, 11, 13, 14].

In general, the stereoselectivity in the transglycosylation catalyzed by a given enzyme is highly correlated with the substrate specificity for the hydrolysis reaction [6]. Therefore, this information can be used to select enzymes that promote the synthesis of the target oligosaccharide. A strategy to control the stereoselectivity may be developed to promote the preferential formation of different linkages.

It is well known that the specific type of linkage between the constituent sugars of oligosaccharides depends not only on the nature of the enzyme, but also on the donor and acceptor involved in the transglycosylation reaction [5, 6, 14]. Although enzyme sources are important, stereoselectivity can be changed by the choice of the structure of the donor and acceptor [15]. Their nature can affect the stereoselectivity in the transglycosylation reactions and the yields of the products.

The search for new donors, acceptors, as well as new glycosidases having high transglycosylation activity, is essential for increasing the yield and improving the stereospecificity of the reaction. On the other hand, knowing

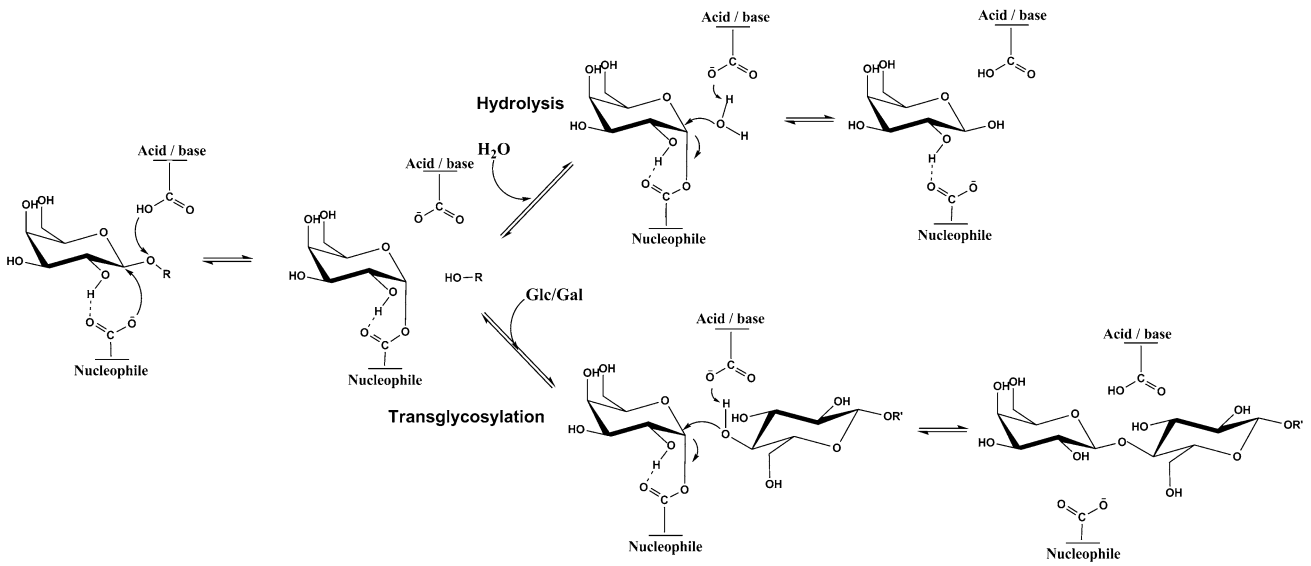

Fig. 1 The catalytic mechanism for the hydrolysis of the glycosidic bond, catalyzed by retaining glycosidases [9]

and understanding the behavior of natural substrates and acceptors for each enzyme during the glycosidic synthesis is crucial in order to clarify the most common mechanisms, and consequently the stereoselectivity displayed by these glycosidases in nature.

In many cases, the reason for the change in stereoselectivity is attributed to the structure of the acceptor, as particular constituent groups may lead to a greater affinity and proximity to the key residues present in the enzymatic center, which can control the conformation of the acceptor bound to the covalent glycosyl enzyme intermediate [16]. The large number of hydrogen bond interactions established between the acceptor and these amino acids can also influence the stereoselectivity of the linkage formed. The steric interaction between the acceptor substituent and the enzyme may therefore be a relevant factor. Zeng et al. have studied the influence of the nature of the acceptor in the stereoselectivity of the β-D-galactosidase from *B. circulans ATCC 31382* [16]. It was confirmed that its stereoselectivity is greatly influenced by the presence of the 2-acetamido group and the anomeric configuration of the aglycon in the acceptor. Therefore, the stereochemical environment between acceptor and the site of the glycosyl enzyme intermediate can explain the different stereoselectivity in the same enzyme.

To summarize, it is very important to choose carefully the method and enzyme for the synthesis, with high yields, of a particular oligosaccharide. For such reasons, we have studied retaining glycosidases from various microorganisms with known relative stereoselectivity for certain kind of acceptors. However, in all our studies, we always used the natural substrates, as donor and acceptor sugars, in order to understand the behavior of these carbohydrates within the corresponding enzymes.

Docking and molecular dynamics studies were performed, which allowed obtaining structural information of the binding site, in particular the identification of the atoms of the acceptor, which are closer to the catalytic carboxylic acids when the glycosyl enzyme intermediate is formed. This procedure also helps to understand, explain and predict the stereoselectivity obtained for the studied sugars, which may be useful to promote a highly regioselective syntheses of oligosaccharides. This information can now be used to improve the efficiency of large-scale industrial synthesis of a specific sugar.

2 Computational details

2.1 Molecular modeling

We studied six retaining glycosidases/galactosidases from different microorganisms, in order to evaluate the important hydrogen interactions established between the acceptor sugar and the glycosyl enzyme intermediate, as well as the specific stereoselectivity in the reaction of formation of glycosidic linkages. In the choice of these enzymes, among other factors, the knowledge of their stereoselectivities for some kind of sugars was specifically taken into account.

The X-ray crystallographic structures of all enzymes were obtained from the database protein databank [17]. Their PDB entries and resolutions are shown in Table 1.

All water molecules were deleted and hydrogen atoms were added using the Insight II software [18], taking into account all residues in their physiological protonation state. During the initial geometry optimization of the enzymes, all proteins were minimized in three stages, in order to release the bad contacts in the crystallographic structures. In the first stage, only the hydrogen atoms were minimized; in the second stage, the backbone was also minimized; in the third and final stage, the entire system was minimized. About 1,500 steps were used for each stage, with the first 500 steps performed using the steepest descent algorithm and the remaining steps carried out using conjugate gradient.

2.2 Molecular docking

A monosaccharide molecule (glucose or galactose depending on the particular case under study) was initially docked into the structure of the unligated protein, mimicking the glycosyl enzyme intermediate structure during the synthesis of the glycosidic bonds. In order to do that, we set up a single covalent link as implemented in the

Table 1 Pdb entries and their resolutions (Å) for the X-ray crystallographic structures of all the enzymes studied

Enzyme	Microorganism	Unbound enzyme	Covalent enzyme–substrate complex
β-Galactosidase	*Escherichia coli*	1DP0/1.70 [36]	1JZ2/2.10 [35]
β-Galactosidase	*Penicillium* sp.	1TG7/1.90 [37]	1XC6/2.10 [37]
β-Glucosidase	*Sulfolobus solfataricus*	1UWQ/2.02 [38]	1UWR/2.14 [38]
β-Glucanase	*Humicola insolens*	2A39/2.20 [39]	1OJK/1.50 [40]
Exo-1,3-β-Glucosidase	*Candida albicans*	1EQP/1.90 [41]	2PB1/1.90 [42]
β(1,3)-Glycosidase	*Thermus thermophilus*	1UG6/0.99 [42]	1E4I/2.00 [42]

docking program, which is able to dock covalently bound ligands. In the real process, the link atoms in both protein and the ligand should have a free valence available. However, in the docking algorithm, the link atom on the monosaccharide does not have a free valence, but a hydrogen atom instead, which is ignored in terms of its contribution to the fitness score. Using this tool, a covalent bond between the anomeric carbon of the sugar and the oxygen of the nucleophilic group was created. Furthermore, a distance constraint was introduced (1.80–2.50 Å) between the hydrogen atom of the hydroxyl group bound to the C_2 of the sugar molecule and the closest oxygen atom of the nucleophilic carboxylate group, because this hydrogen bridge is essential for the stabilization of the transition state (TS) structure and it is known to be highly conserved among glycoside hydrolases [19]. One other distance constraint (3.40–3.70 Å) was introduced between the conserved active-site tryptophan and the C3 atom of the docked sugar. This residue is important because it promotes the packing to the sugar ring during the catalytic mechanism, particularly through the formation of the glycosyl enzyme intermediate.

The docking procedure was made with GOLD [20], a program that predicts the binding modes of small molecules into protein binding sites. The program is based on a genetic algorithm that is used to place different ligand conformations in the protein binding site, recognized by a fitting points strategy. Two scoring functions are a posteriori available to rank the obtained solutions, i.e., GoldScore and ChemScore [21]. As the docking accuracy obtained with both scoring functions is similar, we used ChemScore as the scoring function. This is up to three times faster when compared with the other score function, which justifies the choice taken.

The best docking solutions were taken as starting structures for the subsequent minimization and second molecular docking procedures. In this second docking step, three independent docking processes were made, in which a second sugar molecule (glucose or galactose) was docked, defining a distance constraint (3.5–4.5 Å) between the anomeric carbon of the first sugar and the O3, O4 or O6 atoms of the acceptor sugar, respectively.

After analysis of all the solutions obtained, the best docking solutions were chosen as starting structures for the subsequent minimization and molecular dynamics studies.

Even though we could have used the X-ray structures of the enzyme–substrate covalent complexes (and possibly some modeling if what is connected to the enzyme is not exactly the substrate), instead of performing a docking calculation, we decided to model the whole process to check if the methodology leads to correct results. This will be very useful for the many cases in which a bound enzyme–substrate structure is not available.

2.3 Molecular dynamics

All geometry optimizations and molecular dynamics simulations were performed with the parameterization adopted in Amber 8 [22], using the Amber 1999 force field (parm99) for the proteins and the Glycam 2004 force field (Glycam-04 parameters) for the carbohydrates [23–25].

In these simulations, an explicit solvation model with pre-equilibrated TIP3P water molecules was used, filling a truncated octahedral box with a minimum 12 Å distance between the box faces and any atom of the protein [26]. Each structure was minimized in two stages. In the first stage, the protein was kept fixed and only the position of the water molecules and counter ions was minimized. In the second stage, the full system was minimized. Subsequently, using the Langevin temperature equilibration scheme [27, 28], a 500 ps molecular dynamics (MD) simulation at constant volume and with periodic boundaries was run starting from the optimized structures. After this, 2 ns of MD simulation was performed for each system in which Langevin dynamics was used (collision frequency of 1.0 ps^{-1}) to control the temperature [27, 28]. All simulations presented in this work were carried out using the sander module, implemented in the Amber8 simulations package, with the *Cornell* force field [29]. Bond lengths involving hydrogens were constrained using the SHAKE algorithm [30], and the equations of motion were integrated with a 2 fs time-step using the Verlet leapfrog algorithm. The nonbonded interactions were truncated with a 10 Å cutoff. The temperature of the systems was regulated by the Langevin thermostat to be maintained at 310.15 K [27, 28]. The only exceptions are the glycosidases from *Sulfolobus sulfataricus* and *Thermus thermophilus*, in which the optimal temperature values of the microbial niche occupied by these thermophiles organisms are 338.15 and 343.1 K, respectively [7, 31].

2.4 Calculation of binding free energies

The MM_PBSA script [32] implemented in Amber 8 [22] was used to calculate the binding free energies for all complexes. In each calculation, 50 snapshots of the complexes were extracted every 50 steps for the last 2,500 steps of the run. The internal energy (bond, angle and dihedral), the electrostatic and the van der Waals interactions were calculated using the Cornell force field [29] with no cutoff. The electrostatic solvation free energy was calculated by solving the Poisson–Boltzmann equation with the software Delphi v.4 [33, 34]. The nonpolar contribution to solvation free energy due to van der Waals interactions between the solute and the solvent and cavity formation was modeled as a term that is dependent on the solvent accessible surface area of the molecule. The value of the external dielectric

constant used was 80.0. The value of the internal dielectric constant was 40.0, because the active sites of all enzymes were significantly exposed to the solvent. To check the influence of the parameter on the values of $\Delta G_{\text{binding}}$, we also calculated $\Delta G_{\text{binding}}$ with an internal ε of 4, a value typically used in more hydrophobic, buried active sites. Almost all values (with only a single exception) were qualitatively similar (i.e., the ranking of the solutions was the same). However, the $\Delta G_{\text{binding}}$ were globally more negative (by circa 10 kcal/mol) with $\varepsilon = 4$.

3 Results and discussion

3.1 First docking procedure

The binding of carbohydrates in the step that precedes the transglycosylation reaction was simulated with docking and molecular dynamics methodologies. Initially, a structure for the glycosyl enzyme intermediate was built. For that purpose, a first docking step was performed in which a covalent linkage between the anomeric carbon of the sugar and the oxygen of the active site nucleophilic carboxylate group was made.

It is well known that the role of the hydrogen bridge between the nucleophile and the C_2–OH group of the ring in the transition state is very important to lower the activation barrier. Therefore, this interaction is conserved among retaining glycoside hydrolases. When there is no hydrogen bond, the activation energy rises to about 5.0 kcal mol^{-1}, which demonstrates the importance of this bridge in the catalytic mechanism of oligosaccharide synthesis [19]. For this reason the bridge was considered in the docking procedure, through a distance constrain between both atoms involved. Furthermore, in many glycosidases, the presence of several aromatic residues (Trp, Phe or Tyr) close to the active site provides the hydrophobic platform common to carbohydrate–protein interactions. These conserved amino acids show similar functions and display a position and orientation, which promotes the packing of the sugar rings during the catalytic mechanism. In order to obtain the correct position for binding, a distance constraint was also included between these aromatic residues and the docked monosaccharides.

With the purpose of validating this procedure, glycosidases with known stereoselectivity for some acceptor sugars were selected from the protein databank in both the unbound and bound forms.

After the calculations with the docking software, the results were compared with the ones from the crystallographic structures of complexes. The results are displayed in Table 2, which present the root mean square deviation (RMSD) between the crystallographic structures of each complex in the bound form and the results obtained with the docking software methodology. The RMSD values shown for the donor molecule were obtained by comparison with the carbon and oxygen atoms present in both sugars. In Fig. 2, we can see the superimposition of both structures for all the glycosidases studied.

According to the results above, it can be observed that the average values obtained for the proteins backbone (Cα) are minimal, which was expected due to the insignificant difference between the two crystallographic structures (unbound and bound with the sugar). For the monosaccharides docked in the active site, the values of RMSD are larger. In some cases, this is also expected because the sugar present in the crystallographic complex is not the same as the one that was docked. This occurred because it was very difficult to find the ideal cases for this study, i.e., glycosidases with (1) known relative stereoselectivity; (2) unbound crystallographic structures; and (3) crystallographic structures covalently bound with substrates as reaction intermediates. In order to solve this problem, some cases that contain non-covalent complexes with monosaccharides or disaccharides were chosen. In the first case, the result obtained with the docking procedure was directly compared, whereas in the disaccharides cases the sugars were modified into monosaccharides in order to compare with the docked solution. However, it should be noted that the non-existence of covalent linkage between the donor sugar and the enzyme, in some crystallographic structures, causes large values of RMSD because the position of the sugars is necessarily different.

Table 2 RMSD values (Å) between the crystallographic structure of the protein bound to the sugar and the docking results

Enzyme	Microorganism	Donor	C (α)/docking vs. PDB	Donor/docking vs. PDB
β-Galactosidase	*Escherichia coli*	α-Gal	0.27	1.03
β-Galactosidase	*Penicillium* sp.	α-Gal	0.16	2.15
β-Glucosidase	*Sulfolobus solfataricus*	α-Gal	0.21	2.31
β-Glucanase	*Humicola insolens*	α-Glc	0.06	1.74
Exo-1,3-β-glucosidase	*Candida albicans*	α-Glc	0.08	1.73
β(1,3)-glycosidase	*Thermus thermophilus*	α-Glc	0.21	1.68

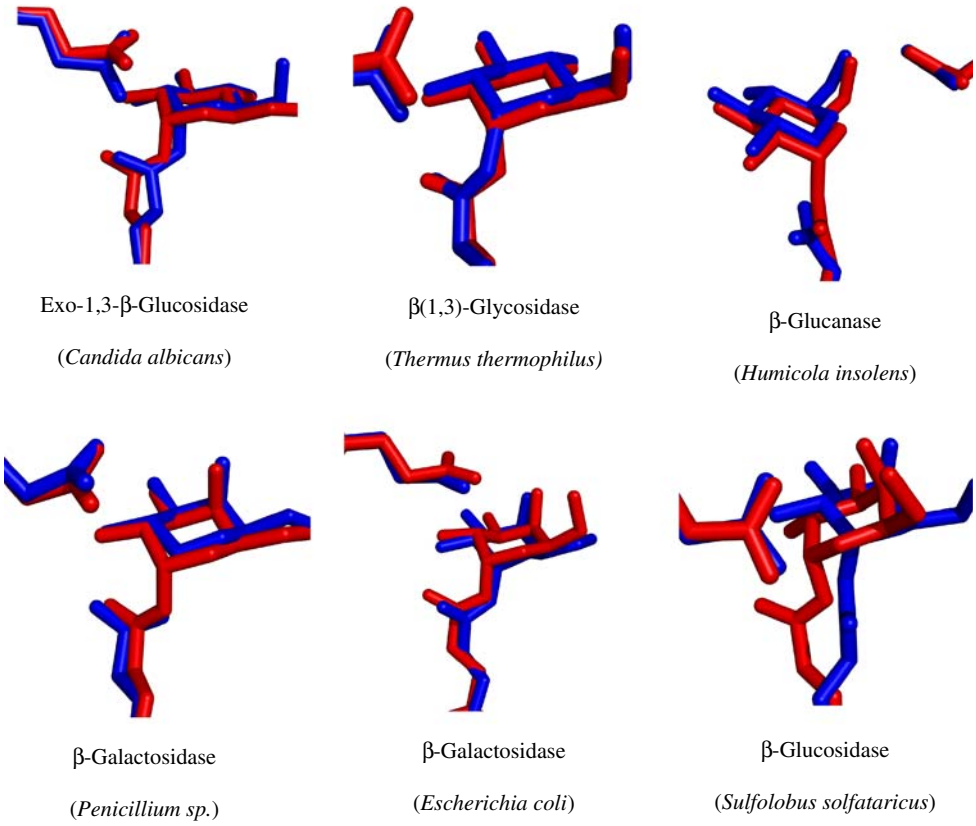

Fig. 2 The superimposition of both structures: that one obtained after the first docking process and the crystallographic file complexed with substrate analog

In the β-galactosidase (*Escherichia coli*), β-glucosidase (*Sulfolobus solfataricus*), Exo-1,3-β-glucosidase (*Candida albicans*) and β(1,3)-glycosidase (*Thermus thermophilus*) cases, the crystallographic files in the database were covalently complexed with substrate analogs, 2-*F*-α-D-glycosyl enzyme intermediate, mimicking the reaction intermediate structure. Therefore in these cases, a direct comparison between these two structures was made in order to calculate the values of RMSD. Furthermore, in the last case, a complex structure does not exist, but it is known that the structure of β-glycosidase from *Bacillus polymyxa* is similar to that one. After superimposition of both structures, it can be seen that the residues present in the active site did not differ significantly. Therefore, this structure was used in order to compare with the one obtained in the calculations. In the specific case of β-glucosidase (*Sulfolobus solfataricus*), the rotamer of glutamate nucleophilic in the unbound and bound crystallographic files is a little different. This fact can influence the binding pose by the donor sugar, and so the high value of RMSD (2.31 Å) is justified.

The structure of β-galactosidase from *Penicillium* sp. only exists non-covalently bound with the monosaccharide galactose. Therefore, the high value of RMSD (2.15 Å) can be understood, as the sugar is not bound to the nucleophile, even though the position of binding is similar. For the β-Glucanase (*Humicola insolens*) case, the crystallographic structure presents a non-covalent complex with the disaccharide Glu-β(1-4)-Glu. Additionally, in the β-glucanase the nucleophilic glutamate was mutated to a serine, so the value obtained for the RMSD, 1.74 Å, was expected. However, the position and orientation of the sugar in the docking solution is similar to the one present in the modeled crystallographic structure. In this way, one can assume the correct binding in the first docking.

In summary, the experimental structures of related complexes give us confidence in the adequacy of the docking results, even though a totally unambiguous quantitative RMSD evaluation cannot be made due to the inherent differences in the experimental and simulated systems.

3.2 Second docking procedure

After the first docking, a second step was taken, in which the acceptor sugar was docked three times independently, in the active site, in three positions induced by specific distance constraints (3.5–5.5 Å) between the anomeric carbon and the O3, O4 and O6 atoms of the acceptor sugar, respectively. The best docking solution for each constraint was used as a starting structure for the three subsequent molecular dynamics simulations. Analyzing the results obtained during these MD simulations, the average

distance between the C1 (anomeric carbon) and $C3_{Ac}$ (C3 of the acceptor sugar), $C4_{Ac}$ (C4 of the acceptor sugar) and $C6_{Ac}$ (C6 of the acceptor sugar) atoms and the average distance between each of the oxygens of the acid/base carboxylate and the proton of the 3, 4 and 6 hydroxyl groups bound to the acceptor sugar were calculated. The values of binding free energies for all complexes studied were also calculated and are shown in Table 3. Structural information on the active site, in particular the hydrogen interactions established between the acceptor molecule and the glycosyl enzyme intermediate, as well as, van der Waals contacts established between the sugar rings and the aliphatic side chains of neighbor aminoacids were analyzed. The average distances values were obtained after equilibration of MD simulations and are shown in the Tables SI-1, SI-2, SI-3, SI-4, SI-5 and SI-6 present in Supporting Information. All these distances give us an idea of which of the carbon atoms of the acceptor is in the best position to be attacked. In this way we try to see if there is a correlation between the position of the acceptor at the active site and the regiospecificity of the corresponding enzyme. Furthermore, the backbone RMSD values of both the whole proteins and the active site residues and for the acceptor group after equilibration of the system were also calculated and the average values to the last 1,000 ps are presented in Table 3. The active site residues considered included all the residues in direct contact with any atom of the substrate. In Supporting Information, the graphics SI-1, SI-2, SI-3, SI-4, SI-5 and SI-6 show the evolution of the RMSD values during all MD simulations.

3.3 Exo-1,3-β-glucosidase from *Candida albicans*

In this glycosidase, as in all others, when the second step of the docking was finished, three different poses of binding for the acceptor sugar were selected. One of the solutions has a preference for the position of O3 atom and in the other solutions the atoms directly positioned for the attack are O4 and O6.

The results obtained during each MD simulation for the average distance between the key atoms involved in the transglycosylation reaction are summarized in Table 3. We can see that in all cases the values of RMSD for the backbone, for the active site amino acids and the acceptor sugar are small, approximately 1 Å. In order to evaluate the relevant distances, we can observe that for solution 1, the hydroxyl oxygen atom of the acceptor nearest to the $C_{anomeric}$ of the donor is O3. Therefore, it can be assumed that this atom is in a better position to perform the attack, leading to the synthesis of 1–3 glycosidic linkages. Moreover, the HO–$C3_{Ac}$ proton is directed to the carboxylic group of the acid/base residue, which favors the transfer of the proton from one group to another, promoting the catalytic mechanism of synthesis. Solution 2 displayed an initial preference between atoms O4 and O6, but, during the MD simulation, the position of the acceptor sugar

Table 3 RMSD average values (Å) obtained for the backbone of all enzymes studied, for the amino acids of the active site and for the acceptor sugar

Enzyme	Solution/initial preference	RMSD C (α)/(Å)	RMSD residues of the active site/(Å)	RMSD acceptor/(Å)	Distance C1–$O3_{Ac}$/$O4_{Ac}$/$O6_{Ac}$	Distance COO^-–HO–$C3_{Ac}$/HO–$C4_{Ac}$/HO–$C6_{Ac}$	$\Delta G_{binding}$ (kcal/mol)
Exo-1,3-β-glycosidase from *Candida albicans*	1/O3	1.01	1.31	1.08	6.49/8.40/10.0	7.33/9.08/10.29	−25.14
	2/O4	0.97	1.32	0.97	7.19/4.50/4.62	2.90/2.45/5.73	−20.93
	3/O6	0.91	1.08	0.62	4.17/5.81/9.48	1.76/1.70/7.10	−22.05
β-Glucanase from *Humicola insolens*	1/O3	1.27	1.59	0.94	7.87/10.53/13.43	2.49/1.69/6.80	−16.75
	2/O6	1.11	2.09	0.92	12.95/11.69/9.46	8.26/5.90/1.80	−16.53
	3/O6	1.19	1.84	0.92	3.48/3.66/7.07	4.26/2.38/6.31	−19.02
β(1-3)-Glycosidase from *Thermus thermophilus*	1/O3	0.95	0.75	0.74	5.69/7.73/8.19	1.93/5.18/7.79	−18.17
	2/O4	1.02	0.66	1.04	14.94/16.08/18.58	12.81/14.22/17.62	−12.06
β-Galactosidase from *Escherichia coli*	1/O3	1.56	1.44	1.13	7.79/9.79/9.59	6.35/7.64/6.04	−20.36
	2/O4	1.59	1.71	0.42	4.07/4.36/7.75	1.72/1.75/7.16	−18.48
	3/O6	1.43	1.78	1.14	14.93/13.19/11.64	13.11/11.44/10.55	−16.18
β-Galactosidase from *Penicillium* sp.	1/O3	1.45	0.99	0.57	8.03/9.10/11.4	6.40/6.98/10.66	−14.87
	2/O4	1.40	1.32	0.99	8.06/6.20/7.59	10.23/7.80/5.99	−13.43
	3/O6	1.28	0.89	0.79	8.42/5.97/7.42	9.57/7.78/6.66	−17.97
β-Glycosidase from *Sulfolobus sulfaticuris*	1/O3	1.27	1.15	0.99	5.39/7.81/8.93	1.81/5.07/7.92	−21.47
	2/O4	1.28	0.88	0.67	5.51/3.85/6.97	1.91/1.87/6.74	−20.34
	3/O6	1.24	1.21	0.70	8.97/7.22/4.36	8.46/6.56/1.73	−23.12

Relevant average distances values were also analyzed during the last 1 ns of the MD simulations

changed directing atom O4 to the anomeric carbon of the donor sugar. The same change occurred in solution 3, but now towards atom O3. This fact is confirmed by the distances in Table 3, in which solution 2 shows a preference for position 4 (4.50 Å for $C_{anomeric}$–$O4_{Ac}$). Additionally, observing the second distance analyzed, the proton can be transferred both from the HO–$C3_{Ac}$ group and the HO–$C4_{Ac}$ group (2.90 or 2.45 Å). The third solution exhibited an initial preference for position 6 of the acceptor sugar, but, as previously discussed regarding this solution, the acceptor sugar changed its position in the active site. The small distances for $C_{anomeric}$–$O3_{Ac}$ (4.17 Å) and $COO_{acid/base}^{-}$–HO–$C3_{Ac}$ (1.76 Å) reveal the alteration that occurred in the stereoselectivity for the formation of the glycosidic linkage, i.e., the initial stereoselectivity, β(1–6), was changed to β(1–3), as seen in Fig. 3. These results are consistent with the results obtained experimentally, in which this enzyme catalyzes an efficient transglycosylation reaction with high yield of oligosaccharides, and β-1,3 linkages are specifically formed during this reaction of glycosidic synthesis, whereas the quantity of β-1,4 and β-1,6 glycosidic linkages formed is much lower [43].

The $\Delta G_{binding}$ values obtained for solution 1, 2 and 3 are −25.14, −20.93 and −22.05 kcal/mol, respectively. These results indicate that the interactions between the acceptor molecule and the covalent intermediate are not very different, even though they favor the formation of β-1,3 linkages, as expected.

Analyzing the main interactions established between the sugars (donor and acceptor) with the active site of this glycosidase, we can see that they occur via hydrogen bonds, particularly with the equatorial OH groups of the glucose monomers. Moreover, we can see also the existence of several van der Waals contacts that are promoted by the aliphatic side chains present at the interface, namely with Phe144 and Trp363. Table SI-1 present in Supporting Information summarizes the most important interactions that occur between the analyzed donor and the acceptor carbohydrates and the neighbor amino acids of the active site. We can see that for all solutions, the main hydrogen bonds were established with the amino acids Glu27, Asn146, Glu192, Tyr255 and one water molecule, differing only slightly in solution 1, in which the Asp145 residue establishes also one hydrogen bridge. Furthermore, the Phe144 and Trp363 amino acids allow several van der Waals contacts with the sugar rings via their aliphatic side chains.

3.3.1 β-*Glucanase from Humicola insolens*

In the study of β-glucanase, three different solutions for binding were obtained from the three sets of constraints imposed during the docking. One solution shows position 3 of the acceptor to be the closest to the $C_{anomeric}$ of the donor, while in the other solutions the anomeric atom points towards the $O6_{Ac}$ atom (O6 of the acceptor sugar). We can notice that there is not any docking solution with the $C_{anomeric}$ atom directed to the $O4_{Ac}$ atom (O4 of the acceptor sugar); however, solution 3 shows similar initial distances between the $C_{anomeric}$ atom and both $O4_{Ac}$/$O6_{Ac}$ atoms (we can see these distances in Fig. 4). This happens because the solutions obtained with this docking procedure are similar to those from the docking whose constraints involved the $O6_{Ac}$ atoms. On the other hand, both solutions directed for the $O6_{Ac}$ display a different configuration in the sugar ring of the acceptor.

Starting from these solutions, three MD simulations were run. The results obtained for the average distance

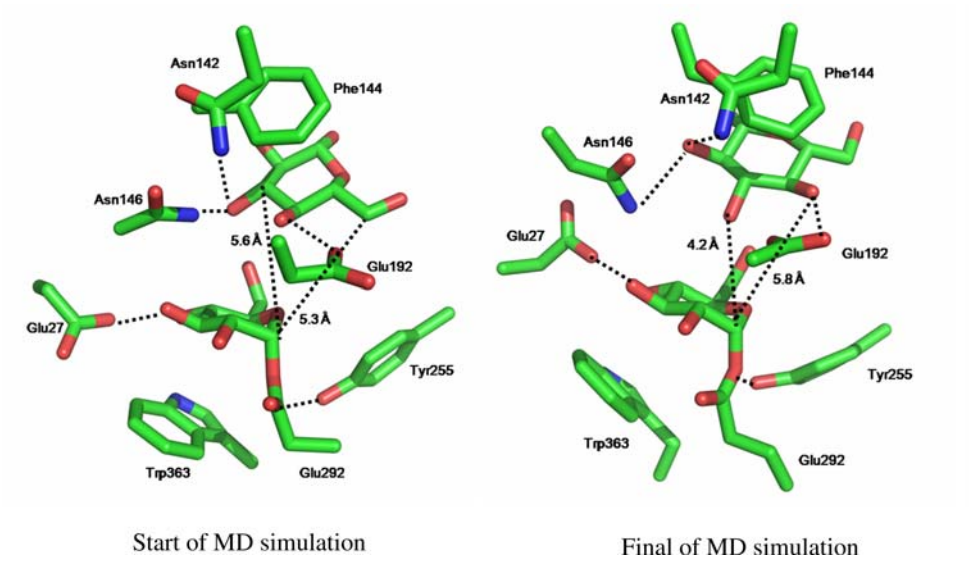

Fig. 3 The most important amino acids present in the active site of Exo-1,3-β-glucosidase (*Candida albicans*) as well as both carbohydrates at the start and end of MD simulation (solution 3). The distances between the $C_{anomeric}$ and some atoms of the acceptor are also shown

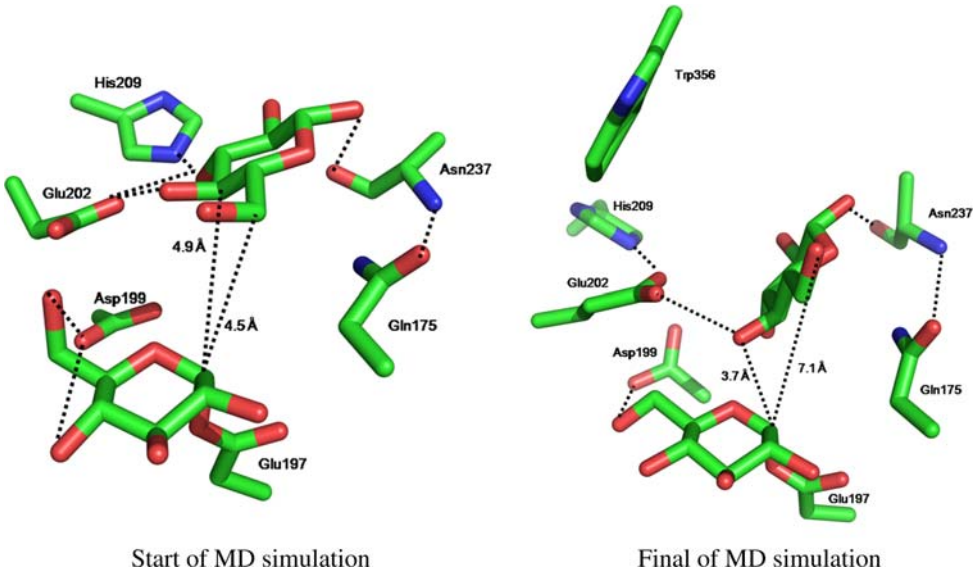

Fig. 4 Representation of the most important amino acids present in the active site of β-glucanase (*Humicola insolens*) and both carbohydrates at the start and end of the MD simulation (solution 3). The distances between atom $C_{anomeric}$ and some atoms of the acceptor are also shown

between the relevant atoms involved in the transglycosylation reaction are summarized in Table 3. As in the previous case, all values of RMSD for the protein backbone and the acceptor sugar are approximately 1 Å and this value is preserved after the equilibration process. The values of RMSD for the active site residues are slightly larger in this specific case, circa 2 Å, which is still a small and acceptable deviation. We can see that in the first solution, the initial preference to the position 3 of the acceptor is maintained during the MD simulation, so the oxygen atom of the acceptor with better position to promote the attack on the $C_{anomeric}$ atom is the $O3_{Ac}$ atom. However, analyzing the distance obtained between the hydroxyl hydrogen atoms of acceptor molecule and the OE1 atom of acid/base residue, we can observe that the smaller distance is obtained for the HO–$C4_{Ac}$. In the other two solutions, the initial preference is for the $O6_{Ac}$ atom. In the second solution, this preference is maintained during the MD simulation, while for solution 3 the position of the acceptor sugar was changed to another pose. This last solution displays the same initial preference for position 6, although the acceptor sugar changed its position in the active site during the MD simulations. In this case the modification in the stereoselectivity is visibly demonstrated by the smaller distances between atoms $C_{anomeric}$–$O4_{Ac}$ (3.66 Å) and atoms $COO^{-}_{acid/base}$–HO–$C4_{Ac}$ (2.38 Å).

It is well noticed that the first two solutions exhibit very high values to the key distances analyzed, revealing that the acceptor sugar is far from the catalytic residues. Therefore, only the closer distance values (by circa 5 Å) are relevant and conclusive to the promotion of the glycosidic linkages. Solution 3 shows the modification that occurred in the stereoselectivity for the formation of glycosidic linkage, i.e., the initial stereoselectivity, β(1-6), was changed to β(1-4) as seen in Fig. 4. The results obtained with experimental data show that the transglycosylation reactions catalyzed by *endo*-enzymes are always highly stereoselective. In that study, it is known that β-glucanase from *Humicola insolens* only forms β–1,4 linkages with whatever monosaccharide acceptors that are involved in the reaction. Furthermore, the Cel7B glycosynthase is a highly efficient catalyst for the synthesis of novel β(1-4)-oligo- and polysaccharides, which is in agreement with the results obtained for the wild-type enzyme [44, 45].

The smaller $\Delta G_{binding}$ value obtained for solution 3 (−19.02 kcal/mol) confirms that the complex is thermodynamically more stable and preferred, and emphasizes the preference of this enzyme for promoting the formation of β-1,4 glycosidic linkages.

Table SI-2 present in Supporting Information summarizes the most important interactions established between the donor and the acceptor carbohydrates and the neighbor residues of the active site. In this case, we can see that in all the structures the main residues that establish hydrogen bonds with the sugars are Asp173, Glu197, Asp199, Hie209, Glu202 and two water molecules. The only differences occur in solution 2 and 3, which also has the Arg108 and Asn237 residues, respectively, near the active site, close to the sugars. Additionally, the aliphatic side chain of Trp356 allows van der Waals contacts with the acceptor molecule.

3.3.2 β(1-3)-Glycosidase from Thermus thermophilus

A retaining glycosidase from the thermophilic species *Thermus thermophilus* has a cleaving preference for glycosidic linkages in the following order: β(1-3) > β(1-4) > β(1-6).

This enzyme also catalyzes the synthesis of all types of disaccharides by transglycosylation with the use of *o*-nitrophenyl β-D-Gal*p* (or Glc*p*) as a sugar donor. However, the β(1-3) disaccharides were obtained with higher yields than the other disaccharides [31].

In our study with this enzyme, two different docking solutions were obtained, which displayed atoms $O3_{Ac}$ and $O4_{Ac}$ of the acceptor close to the $C_{anomeric}$ atom of the donor. In this case, no solution directed to atom $O6_{Ac}$ was found because the solutions obtained during the docking process involving a constraint distance between atoms $C_{anomeric}$–$O6_{Ac}$ were similar to the ones obtained constraining the distance between atoms $C_{anomeric}$–$O4_{Ac}$. The overall results obtained for the average distance between the atoms involved in the transglycosylation reaction are summarized in Table 3. We can again observe the small values of RMSD for the backbone, for the residues of the active site and the acceptor in the two solutions studied.

Analyzing the distances between atoms $C_{anomeric}$ and $O_{acceptor}$ we can see that, for both solutions, the oxygen atom of the acceptor with better position to promote the attack for atom $C_{anomeric}$ is the $O3_{Ac}$ atom. It is noticed that in solution 1, the other distance analyzed is smaller for the HO–$C4_{Ac}$ atom; however during the MD simulation a conformational change occurs and in the end this distance is smaller for the HO–$C3_{Ac}$ atom. With the obtained results, it can be expected that the stereospecificity shown by this enzyme with the natural substrates is directed towards β(1-3). Therefore, these results are in agreement with those obtained experimentally, which demonstrates specificity in transglycosylation reactions by the β(1-3) disaccharides obtained with higher yields. In solution 2, the acceptor molecule is too far away from the active site, and consequently the results obtained with this solution are not relevant and not considered in the conclusions.

On the other hand, the smaller $\Delta G_{binding}$ value obtained for solution 1 (−18.17 kcal/mol) further confirms the preference of this enzyme to promote the formation of β-1,3 glycosidic linkages.

The carbohydrates interact with the active site of this enzyme mainly via hydrogen bonds, namely with the hydroxyl groups of the sugars, and also by some van der Waals contacts that are promoted by the aliphatic side chains present at the pocket site, namely with Trp385. Table SI-3 present in Supporting Information summarizes those main important interactions. We can see that for the two solutions, the residues that establish hydrogen bridges with the carbohydrates are Gln18, Glu164, Tyr284, Glu338 and Glu392. The Trp385 residue allows van der Waals contacts with the donor ring via its aliphatic side chains.

β-Galactooligosaccharides (β-GOS) are the major oligosaccharides that can be used in the food industry as dietary additives. The β-galactosidases and some β-glycosidases are responsible for their production. Here, three of these enzymes were studied, β-glycoside from the extremophilic *S. solfataricus* and nonthermostable β-galactosidase from *E. coli* and *Penicillium* sp. Subsequently, their production of different β-GOS was compared.

3.3.3 β-*Galactosidase from Escherichia coli*

β-galactosidase from *E. coli* is a retaining enzyme, which hydrolyzes the disaccharide lactose (with β-D-1-4 linkages) to galactose and glucose, as well as converting it into another disaccharide, allolactose (with β-D-1-6 linkages). This latter sugar is obtained with approximately 97% of yield during the transglycosylation reactions with lactose as substrate. In another study in which lactose was used as donor and *N*-acetyllactosamine as acceptor during the synthesis of β-GOS, *N*-acetylallolactosamine was detected as the main product obtained. However, Juers et al. [35] considered that this formation occurs via an internal return process in which that bond with galactose is created by the glucose whose β-D-1-4 linkage was cleaved. However, in the transglycosylation reaction with lactose as substrate and when glucose molecules were added into the reaction, other disaccharides were also formed [7].

In our study, we used a galactose molecule as the donor and a glucose molecule as the acceptor group. In the docking studies, three different solutions were obtained, displaying preference for atoms $O3_{Ac}$, $O4_{Ac}$ and $O6_{Ac}$. The results obtained for the average distance between the crucial atoms involved in the transglycosylation reaction are summarized in Table 3. The average values obtained for the RMSD of the backbone, for the amino acids present in the active site and for the acceptor are relatively small. In solution 1, this enzyme revealed an initial preference for atom $O3_{Ac}$.

During the MD simulation, the distance between atoms $C_{anomeric}$ and O_{Ac} is smaller for the O3 atom revealing that the first choice was maintained, whereas the other distance analyzed is smaller for the HO–$C6_{Ac}$ atom. In contrast, the second solution exhibits a preference for both atoms $O3_{Ac}$ and $O4_{Ac}$ and supports the formation of both glycosidic linkages. The small value obtained for the average RMSD of the acceptor molecule confirms that the initial configuration of this sugar is maintained during all the MD simulation, which indicates that the first choices in the stereoselectivity were preserved. Therefore, with these results, it can be expected that the configuration of acceptor in this solution promotes the formation of β(1-3) and β(1-4) glycosidic linkages. It can be seen that these two atoms are close together, so a minimal modification in orientation can modify the stereoselectivity in the glycosidic bond formed. For the third solution, the average values of distances between atoms $C_{anomeric}$ and O_{Ac} show

that all oxygen atoms of the acceptor are far from the $C_{anomeric}$ atom, whereas the O6 is the atom with smaller distances. On the other hand, the distance between the proton of one hydroxyl group of the acceptor and the carboxylic group of the acid/base residue is smaller for the hydroxyl hydrogen of $C6_{Ac}$ atom. However, solutions 1 and 3 show the acceptor molecule to be far away from the catalytic groups, which suggests that the results obtained with these two solutions are not relevant to the final conclusions. Taking into account only the results obtained with solution 3, these values suggest that during the transglycosylation reaction, the formation of both β-glucanase from *Humicola insolens*(1-3) and β(1-4) glycosidic linkages are possible. These results are somewhat different from those obtained experimentally. It is unclear at this stage what causes this discrepancy between theoretical predictions and experiments. Given that two of the three simulations are inconclusive, we believe that longer simulations with the acceptor molecule closer to the active site would be beneficial to shedding light on this issue.

Table SI-4 present in Supporting Information summarizes the most important interactions that occur between both carbohydrates and the neighbor amino acids close to the active site. Observing the hydrogen interactions established between the sugars (donor and acceptor) with the active site of this glycosidase, we can see that these bonds occurred with the Asn102, Asn460, Tyr503, Glu537 residues and water molecules. Solution 2 interacts also with Asp201, Glu461 and Lys517 residues. Moreover, we can also verify the existence of some van der Waals contacts that occur between the aliphatic side chains of Trp568 and Trp999 with the ring of donor and acceptor sugars, respectively.

3.3.4 *β-Galactosidase from Penicillium sp.*

The β-D-galactosidase was isolated from filamentous fungi having high transglycosylation activity. For that reason, Zinin et al. [15] studied the transglycosylation reaction occurring with 1-*O*-acetyl-β-D-galactopyranose (AcGal) and methyl β-D-galactopyranoside as substrates. The major product obtained in the first transglycosylation reaction was β-D-galactopyranosyl-(1-6)-1-*O*-acetyl-β-D-galactopyranose. In the second transglycosylation reaction studied, the main product was methyl β-D-galactopyranosyl-(1-6)-β-D-galactopyranoside, whereas the β-D-galactopyranosyl-(1-3)-β-D-galactopyranoside was also obtained with minor yield. Therefore, this enzyme in the presence of these substrates would have stereoselectivity by β-D-1-3 linkages and principally β-D-1-6 linkages.

As previously mentioned, in our study we always use the natural substrate for each enzyme, in order to simulate the most common process occurring in nature. For that reason, in this study we used a galactose molecule as the donor and glucose as the acceptor molecule, so a stereoselectivity different from these studies can be obtained.

After the docking studies, three different solutions were obtained. These display preference for atoms $O3_{Ac}$, $O4_{Ac}$ and $O6_{Ac}$, respectively. The results obtained for the average distance between the key atoms involved in the transglycosylation reaction are summarized in Table 3. Herein, one can again observe the small values of RMSD for the backbone, for the amino acids in the binding site and the acceptor in all solutions studied.

After calculating the average distance values between atoms $C_{anomeric}$ and O_{Ac}, it is noticeable that for the first solution, the oxygen atom of the acceptor positioned to promote the attack for atom $C_{anomeric}$ is the $O3_{Ac}$ atom. The results obtained for solution 2 show that the HO–$C6_{Ac}$ atoms are in a favorable position to promote the glycosidic linkage, whereas the oxygen atoms of the acceptor positioned to promote the attack for atom $C_{anomeric}$ are the $O4_{Ac}$ and $O6_{Ac}$ atoms. These data indicate a small modification in the configuration of the acceptor group during the MD simulation, and the initial preference, β(1-4), was changed to the formation of β(1-6) glycosidic bonds. Solution 3 shows that the average values of distances between atom $C_{anomeric}$ and the three oxygen atoms of the acceptor are large, but the $O4_{Ac}$ atom is the one that is nearest (5.97 Å). Moreover, the distance between the proton of one hydroxyl group of the acceptor and the carboxylic group of acid/base residue is smaller for hydroxyl hydrogen of $C6_{Ac}$ atom.

Our results suggest that this enzyme in the presence of the galactose and glucose molecules produces all the types of glycosidic linkages. These results can confirm that the stereoselectivity of transglycosylation reactions can be altered with the use of different molecules involved in the reaction. However, it is well noticed that all solutions displaying the acceptor sugar faraway from the active site of the glycosidase–galactose covalent intermediate, so these results can be inconclusive to those obtained experimentally.

Table SI-5 present in Supporting Information summarizes the most important interactions established between the donor and the acceptor carbohydrates with the neighbor residues of the active site. In this case, one can verify that for all solutions the main residues that establish hydrogen bonds with the carbohydrates were Asn140, Glu142, Asn199, Tyr261 and Glu299. Several van der Waals contacts were established between the sugar ring of the donor with the aliphatic side chain of Tyr343. Solution 3 interacts also with the Glu807 by a hydrogen bond and establishes some van der Waals contacts between the aliphatic side chain of Trp809 and the sugar ring of the acceptor molecule.

3.3.5 β-Glycosidase from Sulfolobus solfataricus

The oligosaccharide production with this thermostable glycosidase is higher at 65 °C. The behavior adopted by β-glycosidase from *S. solfataricus* in the production of oligosaccharides, namely its thermostability, the relationship between product yield and time required for maximum product formation was studied by Reuter et al. It was demonstrated that this enzyme with *N*-acetyllactosamine as an acceptor during the synthesis of β-GOS had specificity for β-D-(1-4), β-D-(1-6) and β-D-(1-3) linkages, whereas the disaccharide with higher yield was the β-D-(1-6) and the poorest yield was that of the β-D-(1-3) sugars [7].

In order to reproduce the most common process that occurs in the production of galactooligosaccharides, a galactose molecule was used as the donor and a glucose molecule as the acceptor. After our docking study, three different solutions were obtained directed to atoms $O3_{Ac}$, $O4_{Ac}$ and $O6_{Ac}$, respectively. The results obtained for the average distance between the key atoms involved in the transglycosylation reaction are summarized in Table 3. Analyzing the average values of the distances between atoms $C_{anomeric}$ and O_{Ac}, we can notice that in all solutions the initial preference was maintained during the MD simulations. All solutions displayed short key-distance values, that classified them as productive for the transglycosylation reaction. Therefore, this enzyme with galactose and glucose as the donor and acceptor, respectively, seems to be able to promote β(1-3), β(1-4) and β(1-6) glycosidic bonds. These results are in agreement with the others obtained by Reuter et al., which have used *N*-acetyllactosamine as an acceptor molecule and demonstrated that the transglycosylation reactions led to the formation of all types of glycosidic linkages, even though the β(1-6) glycosidic bonds were the most favored [7]. Analyzing the $\Delta G_{binding}$ values we can seen that solution 1 and 2 have close values, −21.47 and −20.34 kcal/mol, respectively, and solution 3 shows the most negative $\Delta G_{binding}$ value (−23.12 kcal/mol), in agreement with the preference for the formation of β(1-6) glycosidic bonds.

The interaction between proteins and carbohydrates can occur mainly via hydrogen bonds, namely with the equatorial OH groups of the monosaccharides. Furthermore, some van der Waals contacts promoted by the aliphatic side chains present in the protein with the rings of sugars can also occur. Table SI-6 summarizes the most important interactions that occurred between the analyzed donor and the acceptor carbohydrates with this enzyme. In this case, the solutions show that the main hydrogen bonds were established with the Gln20, Asn205, Glu206, Tyr320, Glu385, and Glu430 residues. Additionally, the Tyr423 amino acid allows van der Waals contacts with the ring of sugar donor via its aliphatic side chain.

4 Conclusions

The computational chemistry studies have generated detailed structural information on the second mechanistic step of transglycosylation reactions at the molecular and atomic levels. The attacks of a monosaccharide molecule at several glycosyl enzyme intermediates were studied.

The results obtained for the Exo-1,3-β-glucosidase from *Candida albicans* were consistent with the experimental data using the same substrates. As it was shown, the substrate of the active site of this enzyme promotes an efficient transglycosylation reaction in which the β-1,3 linkages are specifically formed. On the other hand, both sugars interact mainly via hydrogen bonds, with a central area in the active site that contains the amino acids Glu27, Asp145, Asn146, Glu192, Tyr255 and one water molecule. Moreover, the Phe144 and Trp363 residues allow several van der Waals contacts with the sugar rings by their aliphatic side chain, which results in great stabilization due to packing between both rings.

The solutions studied for the β-glucanase from *Humicola insolens* show that this enzyme displays a preference to form β(1-4) glycosidic linkages. The modifications that occurred in the position of the acceptor sugar during the MD simulations confirms that the initial stereoselectivity for β(1-6) glycosidic bonds was changed to β(1-4). Moreover, the most important amino acids present in the pocket site are the Arg108, Asp173, Glu197, Asp199, Glu202, His209 and Asn237, which establish hydrogen bonds with both sugars. These results are fully consistent with the available experimental data, which show that this enzyme is regioselective toward β(1-4) glycosidic linkages, whatever substrates are used.

When glucose is used as donor and acceptor molecules, glycosidase from the thermophilic species *Thermus thermophilus* has a preference for β(1-3) glycosidic linkages. Analyzing the most important hydrogen interactions that occurred with the donor and acceptor carbohydrates and the enzyme, we can verify that these were established with the Gln18, Glu164, Tyr284, Glu338 and Glu392 residues. These interactions happen mainly via hydrogen bonds, whereas some van der Waals contacts can occur by packing promoted by the aliphatic side chains of Trp385 with the donor ring. Experimental data show that the cleaving preference is β(1-3) > β(1-4) > β(1-6) and that higher yields of β(1-3) are obtained in transglycosylation reactions using *o*-nitrophenyl β-D-Gal*p* as the acceptor group. Taking into account these facts and our theoretical results, we can conclude that

in nature when glucose molecules can exist, this enzyme shows a stereoselectivity for β(1-3) disaccharides.

In the case of β-galactosidase from *Escherichia coli,* our results show that with another glucose molecule present in the solution, this enzyme promotes the formation of β(1-3) and β(1-4) glycosidic bonds. The most important amino acids present in the active site are Asn102, Asp201, Asn460, Glu461, Tyr503, Lys517 and Glu537. These stabilize both sugars during the second step of transglycosylation reaction via the hydrogen interactions established with them. It is important to emphasize the existence of van der Waals contacts that occur by the aliphatic side chains of Trp568 and Trp999 with the ring of donor and acceptor sugars, respectively.

Considering the results obtained for the β-galactosidase from *Penicillium* sp., we can conclude that when this galactosyl enzyme intermediate is attacked by a glucose molecule, it can promote the formation of all types of disaccharides, i.e., with β(1-3), β(1-4) and β(1-6) glycosidic linkages. Asn140, Glu142, Asn199, Glu200 and Glu299 were the main residues that established hydrogen bonds with both carbohydrates. On the other hand, van der Waals contacts were created between the sugar rings of donor and acceptor with the aliphatic side chains of Tyr343 and Trp809, respectively. Under the same conditions, β-glycosidase from *Sulfolobus solfataricus* also displays formation of all the three types of glycosidic linkages, although the β(1-6) glycosidic bonds are preferentially formed. In this case, the main hydrogen interactions were established with the Gln20, Asn205, Glu206, Tyr320, Glu385 and Glu430 residues. The side chain of the Tyr423 residue allows packing with the ring of sugar donor by van der Waals contacts.

These theoretical results confirm that the stereoselectivity of transglycosylation reactions can be altered with the use of different donor and acceptor molecules involved in the reaction. Furthermore, the data also demonstrates that the directional interactions that occur during the binding of carbohydrates with the enzymes are mainly established via hydrogen bonds. The presence of aromatic residues is also important, and we propose that these residues have a preponderant role in the reorientation of the carbohydrate pose and direct it to specific key amino acids responsible for the catalytic mechanism.

In summary, this procedure has shown reasonable success in identifying, explaining and predicting the stereoselectivity shown by the glycosidases during the formation of glycosidic linkages. This is very important to increase the efficiency of large-scale industrial productions of a particular carbohydrate, opening new doors to the *in silico* design of specific acceptors that promote the desired stereoselectivity and allowing the prediction of the adequate substrates, within a given set, which lead to the desired oligosaccharides.

Acknowledgments The authors would like to thank the Portuguese Science and Technology Foundation (FCT-MCTES) for financial support (scholarship SFRH/BD/31359/2006).

References

1. Perugino G, Trincone A, Rossi M, Moracci M (2004) Trends Biotechnol 22:31–37. doi:10.1016/j.tibtech.2003.10.008
2. Maugard T, Gaunt D, Legoy MD, Besson T (2003) Biotechnol Lett 25:623–629. doi:10.1023/A:1023060030558
3. Cruz-Guerrero AE, Gomez-Ruiz L, Viniegra-Gonzalez G, Barzana E, Garcia-Garibay M (2006) Biotechnol Bioeng 93:1123–1129. doi:10.1002/bit.20824
4. Holzapfel WH, Schillinger U (2002) Food Res Intern 35:109–116. doi:10.1016/S0963-9969(01)00171-5
5. Jakeman DL, Withers SG (2002) Trends Glycosci Glycotechnol 14(75):13–25
6. Ajisaka K, Yamamoto Y (2002) Trends Glycosci Glycotechnol 14(75):1–11
7. Reuter S, Nygaard AR, Zimmermann W (1999) Enzyme Microb Technol 25:509–516. doi:10.1016/S0141-0229(99)00074-5
8. Zechel DL, Withers SG (2000) Acc Chem Res 33:11–18. doi:10.1021/ar970172
9. Koshland DE (1953) Biol Rev Camb Philos Soc 28:416–436. doi:10.1111/j.1469-185X.1953.tb01386.x
10. Jahn M, Withers SG (2003) Biocatal Biotransformation 21:159–166. doi:10.1080/1024220310001614351
11. Crout DHG, Vic G (1998) Curr Opin Chem Biol 2:98–111. doi:10.1016/S1367-5931(98)80041-0
12. Mayer C, Jakeman DL, Mah M, Karjala G, Gal L, Warren RAJ, Withers SG (2001) Chem Biol 8:437–443. doi:10.1016/S1074-5521(01)00022-9
13. Gu QM (1999) J Environ Polym Degrad 7:1–7. doi:10.1023/A:1021833917049
14. Yoon JH, Ajisaka K (1996) Carbohydr Res 292:153–163
15. Zinin AI, Eneyskaya EV, Shabalin KA, Kulminskaya AA, Shishlyannikov SM, Neustroev KN (2002) Carbohydr Res 337:635–642. doi:10.1016/S0008-6215(02)00027-7
16. Zeng X, Murata T, Ajisaka K, Usui T (2000) Carbohydr Res 325:120–131. doi:10.1016/S0008-6215(99)00303-1
17. Bernstein FCKT, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) J Mol Biol 112(3):535–542. doi:10.1016/S0022-2836(77)80200-3
18. Accelrys (1993) InsightII v.2.3.0. San Diego, CA
19. Brás NF, Moura-Tamames SA, Fernandes PA, Ramos MJ (2008) J Comput Chem 29(15):2565–2574
20. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) J Mol Biol 267:727–748. doi:10.1006/jmbi.1996.0897
21. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Proteins Struct Funct Genet 52:609–623. doi:10.1002/prot.10465
22. Case DA, Cheatham TE, Darden TI, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods R (2005) J Comput Chem 26:1668
23. Kirschner KN, Woods RJ (2001) Proc Natl Acad Sci USA 98:10541–10545. doi:10.1073/pnas.191362798
24. Kirschner KN, Woods RJ (2001) J Phys Chem A 105:4150–4155. doi:10.1021/jp004413y

25. Basma M, Sundara S, Calgan D, Vernali T, Woods RJ (2001) J Comput Chem 22:1125–1137. doi:10.1002/jcc.1072
26. Asensio JL, Jimenez Barbero J (1995) Biopolymers 35:55–73. doi:10.1002/bip.360350107
27. Izaguirre JA, Catarello DP, Wozniak JM, Skeel RD (2001) J Chem Phys 114:2090–2098. doi:10.1063/1.1332996
28. Loncharich RJ, Brooks BR, Pastor RW (1992) Biopolymers 32:523–535. doi:10.1002/bip.360320508
29. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Am Chem Soc 117:5179–5197. doi:10.1021/ja00124a002
30. Hammonds KD, Ryckaert JP (1991) Comput Phys Commun 62:336–351. doi:10.1016/0010-4655(91)90105-T
31. Drone J, Feng HY, Tellier C, Hoffmann L, Tran V, Rabiller C, Dion M (2005) Eur J Org Chem 1977–1983. doi:10.1002/ejoc.200500014
32. Huo S, Massova I, Kollman PA (2002) J Comput Chem 23:15. doi:10.1002/jcc.1153
33. Rocchia W, Sridharan S, Nicholls A, Alexov E, Chiabrera A, Honig B (2002) J Comput Chem 23:128. doi:10.1002/jcc.1161
34. Rocchia W, Alexov E, Honig B (2001) J Phys Chem B 105:6507. doi:10.1021/jp010454y
35. Juers DH, Heightman TD, Vasella A, McCarter JD, Mackenzie L, Withers SG, Matthews BW (2001) Biochemistry 40:14781–14794. doi:10.1021/bi011727i
36. Juers DH, Jacobson RH, Wigley D, Zhang XJ, Huber RE, Tronrud DE, Matthews BW (2000) Protein Sci 9:1685–1699
37. Rojas AL, Nagem RAP, Neustroev KN, Arand M, Adamska M, Eneyskaya EV, Kulminskaya AA, Garratt RC, Golubev AM, Polikarpov I (2004) J Mol Biol 343:1281–1292. doi:10.1016/j.jmb.2004.09.012
38. Gloster TM, Roberts S, Ducros VMA, Perugino G, Rossi M, Hoos R, Moracci M, Vasella A, Davies GJ (2004) Biochemistry 43:6101–6109. doi:10.1021/bi049666m
39. Mackenzie LF, Sulzenbacher G, Divne C, Jones TA, Woldike HF, Schulein M, Withers SG, Davies GJ (1998) Biochem J 335:409–416
40. Ducros VMA, Tarling CA, Zechel DL, Brzozowski AM, Frandsen TP, von Ossowski I, Schulein M, Withers SG, Davies GJ (2003) Chem Biol 10:619–628. doi:10.1016/S1074-5521(03)00143-1
41. Cutfield JFSPA, Cutfield SM (2000) Protein Eng 13:735–738. doi:10.1093/protein/13.10.735
42. http://www.rcsb.org
43. Stubbs HJ, Brasch DJ, Emerson GW, Sullivan PA (1999) Eur J Biochem 263:889–895. doi:10.1046/j.1432-1327.1999.00581.x
44. Jahn M, Stoll D, Warren RAJ, Szabó L, Singh P, Gilbert HJ, Ducros VM-A, Davies GJ, Withers G (2003) Chem Commun 12:1327–1329. doi:10.1039/b302380j
45. Fort S, Boyer V, Greffe L, Davies GJ, Moroz O, Christiansen L, Schulein M, Cottaz S, Driguez H (2000) J Am Chem Soc 122:5429–5437. doi:10.1021/ja9936520